Persistency Model and Its Applications in Choice Modeling

Karthik Natarajan, Miao Song, Chung-Piaw Teo

Institutions: National University of Singapore, Massachusetts Institute of Technology

Published on: 01 Mar 2009 - Management Science (INFORMS)

Topics: Mixed logit, Discrete choice, Marginal distribution, Stochastic programming and Probability distribution

Persistency Model and Its Applications in Choice Modeling*

Karthik Natarajan[†] Miao Song[‡] Chung-Piaw Teo[§]

September 2006

Abstract

Given a discrete optimization problem $Z(\tilde{\boldsymbol{c}}) = \max\{\tilde{\boldsymbol{c}}'\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}\}$, with objective coefficients $\tilde{\boldsymbol{c}}$ chosen randomly from a distribution θ, we would like to evaluate the expected value $E_\theta(Z(\tilde{\boldsymbol{c}}))$ and the probability $P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k)$ where $x^*(\tilde{\boldsymbol{c}})$ is an optimal solution to $Z(\tilde{\boldsymbol{c}})$. We call this the persistency problem for a discrete optimization problem under uncertain objective, and $P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k)$, the persistence value of the variable x_i at k.

In general, this is a difficult problem to solve, even if θ is well specified. In this paper, we show that a subclass of this problem can be solved in polynomial time. In particular, we assume that θ belongs to the class of distributions Θ with given marginal distributions, or given marginal moments conditions. Under these models, we show that the persistency problem for $\theta^* \in \text{argmax}_{\theta\in\Theta} E_\theta[Z(\tilde{\boldsymbol{c}})]$ can be solved via a concave maximization problem.

The persistency model solved using this formulation can be used to obtain important qualitative insights to the behaviour of stochastic discrete optimization problems. We demonstrate how the approach can be used to obtain insights to problems in discrete choice modeling. Using a set of survey data from a transport choice modeling study, we calibrate the random utility model with choice probabilities obtained from the persistency model. Numerical results suggest that the persistency model is capable of obtaining estimates which perform as well, if not better, than classical methods such as logit and cross nested logit models. We can also use the persistency model to obtain choice probability estimates for more complex choice problems. We illustrate this on a stochastic knapsack problem, which is essentially a discrete choice problem under budget constraint. Numerical results again suggest that our model is able to obtain good estimates of the choice probabilities for this problem.

*Research supported in part by the Singapore-MIT Alliance.

[†]Department of Mathematics, National University of Singapore, Singapore 117543. Email: matkbn@nus.edu.sg

[‡]The Logistics Institute - Asia Pacific. Email: tlisongm@nus.edu.sg

[§]Department of Decision Sciences, NUS Business School, Singapore 117591. Email: bizteocp@nus.edu.sg. Part of the work was done when the author was at SKK Graduate School of Business, Sungkyunkwan University, Seoul, Korea.

1 Introduction

Consider a discrete optimization problem:

$$Z(\tilde{\boldsymbol{c}}) = \max\{\tilde{\boldsymbol{c}}'\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}\}, \tag{1}$$

where the feasible region $\mathcal{X}$ is given as:

$$\mathcal{X} = \left\{\boldsymbol{x} \;\middle|\; \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b},\ x_i \in \mathcal{X}_i \subseteq \mathcal{Z}^+\ \forall i \in \mathcal{N}\right\}, \tag{2}$$

The decision variables x_i are indexed in $i \in \mathcal{N} = \{1, 2, \ldots, n\}$ where the set $\mathcal{X}_i$ consists of non-negative integer values from α_i to β_i that x_i can take:

$$\mathcal{X}_i = \{\alpha_i, \alpha_i + 1, \ldots, \beta_i - 1, \beta_i\} \tag{3}$$

For a given distribution θ of the objective coefficients, the expected value of the discrete optimization problem in (1) can be expressed as:

$$E_\theta\Big(Z(\tilde{\boldsymbol{c}})\Big) = E_\theta\Big(\sum_{i\in\mathcal{N}} \tilde{c}_i x_i^*(\tilde{\boldsymbol{c}})\Big),$$

where $x_i^*(\tilde{\boldsymbol{c}})$ is an optimal value for the x_i decision variable for the objective $\tilde{\boldsymbol{c}}$. When $\tilde{\boldsymbol{c}}$ is random, $x_i^*(\tilde{\boldsymbol{c}})$ is a random variable. For ease of exposition, we will assume that the set of $\tilde{\boldsymbol{c}}$ such that $Z(\tilde{\boldsymbol{c}})$ has multiple optimal solutions has a support with measure zero. We can then rewrite the expected value as:

$$E_\theta\Big(Z(\tilde{\boldsymbol{c}})\Big) = \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{X}_i} k E_\theta\Big(\tilde{c}_i \;\Big|\; x_i^*(\tilde{\boldsymbol{c}}) = k\Big) P_\theta\Big(x_i^*(\tilde{\boldsymbol{c}}) = k\Big).$$

In this paper, we are interested in finding the value $E_\theta(Z(\tilde{\boldsymbol{c}}))$ and in particular $P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k)$ where the latter is called the **persistence value** of x_i at the value k[1]. We call this the persistency problem for discrete optimization under uncertain objective.

For many fixed choices of the distribution θ (e.g. $\tilde{c}_i$'s are independently and uniformly generated in [0,1]), there is by now a huge literature on finding approximations and bounds to the expected value of stochastic discrete optimization problem (cf. [13], [3], [8], [7]). However, finding precise persistence values for stochastic discrete optimization problems appears to be even harder, since we now need probabilistic information on the support of the optimal solutions. We propose instead an approach to compute $E_{\theta^*}(Z(\tilde{\boldsymbol{c}}))$ and $P_{\theta^*}(x_i^*(\tilde{\boldsymbol{c}}) = k)$, where

$$\theta^* \in \text{argmax}_{\theta\in\Theta} E_\theta\Big(Z(\tilde{\boldsymbol{c}})\Big),$$

[1] In an earlier work, Adams, Lassiter and Sherali studied the question of to what extent the solution to an LP relaxation can be used to fix the value of 0-1 discrete optimization problems. They termed this the persistency problem of 0-1 programming model. The motivation of their work, however is different from ours. See Adams, W.P., Lassiter, J.B., and Sherali, H.D., Persistency in 0-1 Polynomial Programming, Mathematics of Operations Research., Vol. 23, No. 2, 359-389, (1998).

and Θ is the set of joint distributions with a prescribed set of information on the marginal distributions. In this regard, θ^* can be viewed as an "optimistic" solution in the set Θ, as it is the distribution which attains the largest expected objective value. We guard against over-optimism by prescribing the entire marginal distributions (termed as Marginal Distribution Model, MDM), or a finite set of marginal moments (termed as Marginal Moments Model, MMM). To illustrate our interest in the persistency problem, we describe two applications next.

Application 1: Bounds For Stochastic Combinatorial Optimization Problems

As our model focuses on finding a "best case" probability distribution, we can provide bounds for various stochastic combinatorial optimization problems. Lyons, Pemantle and Peres [8], for instance, showed that whenever $\tilde{c}_i$ are exponential and independently distributed random variables with means μ_i, we have:

$$E\left(\min\left\{\sum_{i\in\mathcal{N}} \tilde{c}_i x_i : \boldsymbol{x} \in \mathcal{Q}\right\}\right) \;\geq\; \min\left\{\sum_{i\in\mathcal{N}} \mu_i x_i^2 : \boldsymbol{x} \in \mathcal{Q}\right\},$$

where $\mathcal{Q}$ denote the dominant of all $s-t$ paths in a graph G (refer to Lovasz [7] for the generalization to log-concave distributions). Using the results in this paper, we can obtain analogous bounds. Under the assumption that the $\tilde{c}_i$ are exponentially distributed, but without assuming independence, we obtain:

$$E\left(\min\left\{\sum_{i\in\mathcal{N}} \tilde{c}_i x_i : x \in \mathcal{Q}\right\}\right) \;\geq\; \min\left\{\sum_{i\in\mathcal{N}} \mu_i \Big(x_i + (1-x_i)\log(1-x_i)\Big) : \boldsymbol{x} \in \mathcal{Q}\right\},$$

$$E\left(\max\left\{\sum_{i\in\mathcal{N}} \tilde{c}_i x_i : x \in \mathcal{Q}\right\}\right) \;\leq\; \max\left\{\sum_{i\in\mathcal{N}} \mu_i \Big(x_i - x_i\log(x_i)\Big) : \boldsymbol{x} \in \mathcal{Q}\right\},$$

when $\mathcal{Q}$ denote a polytope with 0-1 vertices (the dominant of all $s-t$ path solutions being a special case). While our bound applies to general 0-1 problems, it is nevertheless weaker than the bound proposed in Lyons, Pemantle and Peres [8] since we drop the assumption of independence (see Figure 1). Our bounds have the advantage though that they can be shown to be tight, i.e., we can find the distribution, with suitable correlations built into the random variables, such that the expected optimal objective value approaches that obtained by our model.

Application 2: Choice Probability In Discrete Choice Models

Discrete choice models deal with problem of estimating the probability that a random customer will choose a particular product from a finite set of products. Assume that the utility function that a random customer attaches to product $i \in \mathcal{N}$ is expressed as

$$\tilde{U}_i = V_i + \tilde{\epsilon}_i,$$

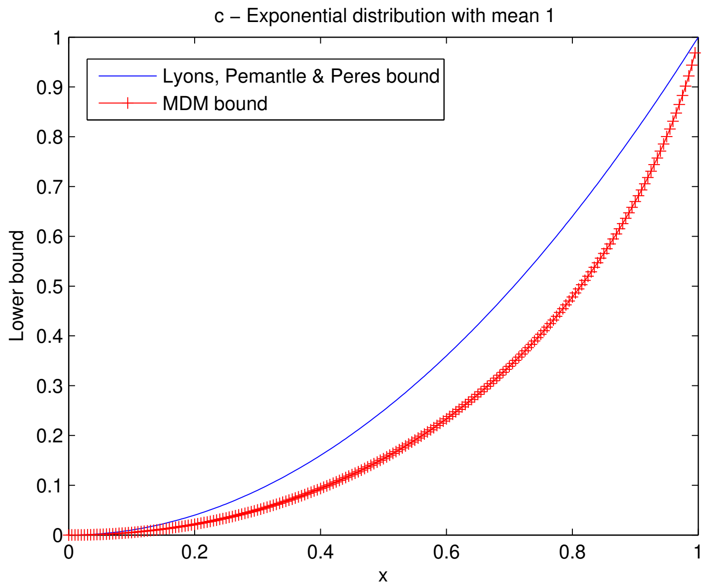

Figure 1: Comparison of Lyons, Pemantle and Peres [8] and MDM bound.

where V_i is the deterministic component that relates to the known attributes of the product and $\tilde{\epsilon}_i$ is the random error associated with the product due to factors not considered. Under the random utility maximization model, the probability that product i will be chosen is basically the persistence value $P(x_i^*(\tilde{\boldsymbol{c}}) = 1)$, where the problem of interest is:

$$Z(\tilde{\boldsymbol{c}}) = \max\left\{\tilde{\boldsymbol{c}}'\boldsymbol{x} : \sum_{i\in\mathcal{N}} x_i = 1,\ x_i \in \{0,1\}\ \forall i \in \mathcal{N}\right\},$$

with $\tilde{c}_i = V_i + \tilde{\epsilon}_i$. The classical logit model (see McFadden [9]) starts with the assumption that the error terms $\tilde{\epsilon}_i$'s are modeled by independent extreme value distributions, so that an elegant closed form solution for the choice probabilities can be obtained:

$$P(x_i^*(\tilde{\boldsymbol{c}}) = 1) = \frac{e^{V_i}}{\sum_{j\in\mathcal{N}} e^{V_j}}.$$

However, this approach has some drawbacks. For example, the formula implies the Independence of Irrelevant Alternatives (IIA) property wherein the relative ratio of the choice probabilities for two alternatives is independent of the remaining alternatives. This property is not always observed in practise wherein the entire choice set helps in determining the relative probabilities. The probit model, using correlated normal distributions, can be used to overcome this shortcoming, but at the added cost of finding choice probabilities through extensive simulation. Using the results in this paper, we can obtain simple new formulas for the choice probabilities.

Suppose the cumulative distribution function for the utility $\tilde{c}_i$ is $F_i(c)$, we obtain:

$$P_{\theta^*}(x_i^*(\tilde{\boldsymbol{c}}) = 1) = 1 - F_i(\lambda^*)$$

where λ^* is found by solving the equation:

$$\sum_{i \in \mathcal{N}} (1 - F_i(\lambda^*)) = 1.$$

Alternatively, suppose we know the mean V_i and variance σ_i^2 of the utility $\tilde{c}_i$ but the exact distribution function is unknown, we obtain:

$$P_{\theta^*}(x_i^*(\tilde{\boldsymbol{c}}) = 1) = \frac{1}{2}\left(1 + \frac{V_i - \lambda^*}{\sqrt{(V_i - \lambda^*)^2 + \sigma_i^2}}\right)$$

where λ^* is found by solving the equation:

$$\sum_{i \in \mathcal{N}} \frac{1}{2}\left(1 + \frac{V_i - \lambda^*}{\sqrt{(V_i - \lambda^*)^2 + \sigma_i^2}}\right) = 1.$$

It is easy to verify that the IIA property does not hold under our model due to the dependence on λ^*. Interestingly, the choice probabilities obtained under the persistence model and the logit model are strikingly similar for many numerical examples we have experimented on. Figure 2 compares the choice probabilities for a simple five product example using logit (closed form), probit (simulation) and the mean-variance MMM (root finding). The first example assumes that each product has a common variance of $\pi^2/6$, with mean:

$$V_1 = 1.2, V_2 = 1.5, V_3 = 1.8, V_4 = 2, V_5 = 2.3.$$

The choice probabilities obtained from the three models are strikingly similar, although they are obtained under very different assumptions. The product with the higher mean attribute will have a higher chance of being selected. The second example plots choice probabilities when the error terms related to the products are non identically distributed. We use the mean values as before but the variances are set to:

$$\sigma_1^2 = 4, \sigma_2^2 = 4, \sigma_3^2 = 3, \sigma_4^2 = 3, \sigma_5^2 = 0.1.$$

In this case, products with lower mean attributes may in fact have a higher chance of being selected. Surprisingly, product 5 with the highest mean attribute (2.3) and smallest variance (0.1) is now the product least likely to be selected. The model predicts accurately that the preferences for the products are in the order 4,3,2,1 and 5.

2 Problem Formulation

In this section, we discuss the model for the set of distributions Θ. The basic model assumes that the distribution function $F_i(c)$ is known for each objective coefficient $\tilde{c}_i$. No assumptions on

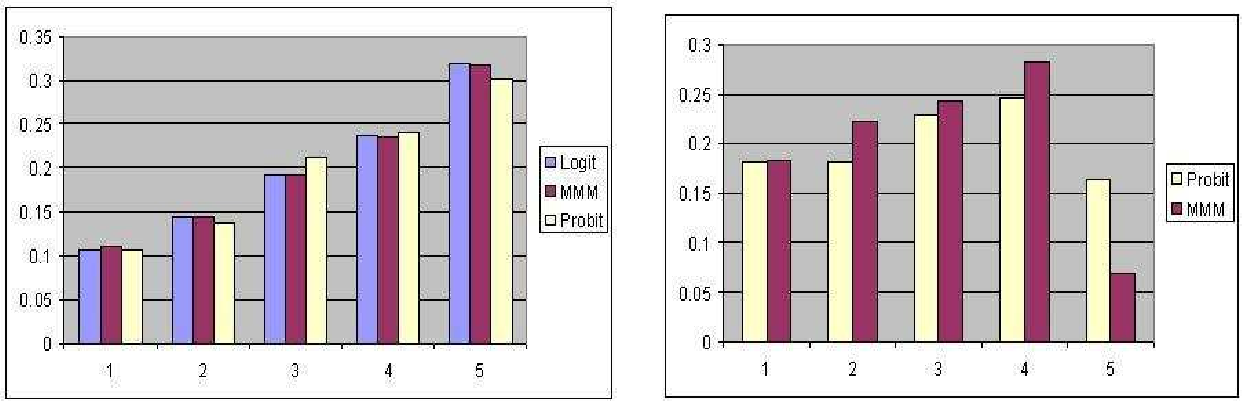

Figure 2: Choice Models Comparison: Constant Variance and Different Variance Case

independence of the marginal distributions is made. We term this as the Marginal Distribution Model (MDM). The earliest study of combinatorial optimization problems under this model was carried out by Meilijson and Nadas [10] and Nadas [11] for the stochastic project management problem. Their main result shows that the problem of finding the tight upper bound on the expected project duration can be obtained by solving a convex optimization problem. Weiss [15] generalized this result to network flow problems and reliability problems. Bertsimas, Natarajan and Teo [3], [4] considered 0-1 optimization problems under partial moment information on each objective coefficient and initiated the study of the persistency problem. In a similar spirit, the central problem that we focus on is:

$$Z^* = \max_{\theta \in \Theta} E_\theta \left(\max\{\tilde{\boldsymbol{c}}'\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}\} \right), \tag{4}$$

where $\mathcal{X}$ is the feasible region of a general discrete optimization problem, and we evaluate the persistence of variables under this model.

Since our interest lies in general discrete optimization problems, a formulation that is particularly useful is the full binary reformulation for general integer variables. We can represent any variable x_i in the set $\mathcal{X}_i$ as:

$$\begin{aligned} x_i &= \sum_{k \in \mathcal{X}_i} k y_{ik} \\ 1 &= \sum_{k \in \mathcal{X}_i} y_{ik} \\ y_{ik} &\in \{0,1\} \quad \forall k \in \mathcal{X}_i. \end{aligned}$$

Defining $\boldsymbol{y} = (y_{ik})_{k \in \mathcal{X}_i, i \in \mathcal{N}}$, the feasible region $\mathcal{X}$ for the discrete optimization problem can be transformed to $\mathcal{Y}$ using the full binary expansion:

$$\mathcal{Y} \quad = \quad \left\{ \boldsymbol{y} \;\middle|\; \boldsymbol{A}\left(\sum_{k\in\mathcal{X}_i} k y_{ik}\right)_{i\in\mathcal{N}} \leq \boldsymbol{b},\; \sum_{k\in\mathcal{X}_i} y_{ik} = 1\ \forall i \in \mathcal{N},\; y_{ik} \in \{0,1\}\ \forall k \in \mathcal{X}_i\ \forall i \in \mathcal{N} \right\}.$$

There is an unique one to one correspondence between the extreme points of $\mathcal{X}$ and $\mathcal{Y}$, that is $x_i = k$ if and only if $y_{ik} = 1$. Such binary reformulations have also been used in Sherali and Adams [12], though in a different context.

We expand upon the earlier work by Meilijson and Nadas [10], Weiss [15] and Bertsimas, Natarajan and Teo [4] in three ways.

(1) We generalize the persistency results to arbitrary discrete optimization problems. Particularly, the decision variables can assume more than two values which allows for more complicated problems. The key idea is to project the variables from the original space $\mathcal{X}$ into the higher dimensional space $\mathcal{Y}$, so that we can impose restrictions such that the expanded set of variables assumes only two possible values (0 or 1). For the formulation to be tight, however, we require an explicit characterization of the convex hull of the higher dimensional space $\mathcal{Y}$. The new facets identified in the projected polytope can now be interpreted as probabilistic inequalities associated with original problem.

(2) The work of Bertsimas, Natarajan and Teo [4] focuses on the case when the information given are marginal moments. Our work generalizes the approach to complete marginal distribution information. In this setting, our work can be viewed as a dual approach to the results of Meilijson and Nadas [10] in the 0-1 case. More importantly, instead of invoking duality results from infinite dimensional optimization, we obtain the results through a more direct constructive approach. We also indicate techniques to incorporate information on the shape of the distributions into our formulation.

(3) We investigate the use of the persistency model on a variety of new applications. For the discrete choice model, we perform both estimation and prediction using a real-life transportation data set. The predicted choice probabilities are comparable to logit and cross-nested logit models under weaker assumptions. In the case of choice modeling under budget constraints, the problem reduces to a stochastic knapsack problem. Through numerical simulations, we show that the persistence values obtained are close to the choice probabilities in this instance too.

3 Main Results

3.1 Marginal Distribution Model (MDM)

Assume that each objective coefficient $\tilde{c}_i$ in the discrete optimization problem is a continuously distributed random variable with marginal distribution function $F_i(c)$ and marginal density function

$f_i(c)$, i.e. $\tilde{c}_i \sim f_i(c)$. Let Θ denote the set of multivariate distributions θ for the objective coefficients with the given marginal distributions.

Define the following sets of non-negative variables[2]:

$$y_{ik}(c) = P\Big(x_i^*(\tilde{\boldsymbol{c}}) = k \;\Big|\; \tilde{c}_i = c\Big) \text{ and } y_{ik} = P\Big(x_i^*(\tilde{\boldsymbol{c}}) = k\Big) \quad \forall k \in \mathcal{X}_i \; \forall i \in \mathcal{N}.$$

Using conditional expectations, the objective function in (4) can be expressed as:

$$\begin{aligned}
E_\theta\Big(Z(\tilde{\boldsymbol{c}})\Big) &= E_\theta\Big(\sum_{i\in\mathcal{N}} \tilde{c}_i x_i^*(\tilde{\boldsymbol{c}})\Big) \\
&= \sum_{i\in\mathcal{N}} \int cE\Big(x_i^*(\tilde{\boldsymbol{c}}) \;\Big|\; \tilde{c}_i = c\Big) f_i(c)dc \\
&= \sum_{i\in\mathcal{N}} \sum_{k\in\mathcal{X}_i} \int cE\Big(x_i^*(\tilde{\boldsymbol{c}}) \;\Big|\; \tilde{c}_i = c, x_i^*(\tilde{\boldsymbol{c}}) = k\Big) P\Big(x_i^*(\tilde{\boldsymbol{c}}) = k \;\Big|\; \tilde{c}_i = c\Big) f_i(c)dc \\
&= \sum_{i\in\mathcal{N}} \sum_{k\in\mathcal{X}_i} \int kcP\Big(x_i^*(\tilde{\boldsymbol{c}}) = k \;\Big|\; \tilde{c}_i = c\Big) f_i(c)dc \\
&= \sum_{i\in\mathcal{N}} \sum_{k\in\mathcal{X}_i} k \int c y_{ik}(c) f_i(c)dc.
\end{aligned}$$

By the definition of the variables as conditional expectations, we have:

$$y_{ik} = \int y_{ik}(c) f_i(c)dc \quad \forall k \in \mathcal{X}_i \; \forall i \in \mathcal{N}.$$

Furthermore, for each realization c, the optimal value for x_i is one of the values in $\mathcal{X}_i$, implying that:

$$\sum_{k\in\mathcal{X}_i} y_{ik}(c) = 1 \quad \forall c \; \forall i \in \mathcal{N}.$$

Lastly, since the decision variables lie in $CH(\mathcal{Y})$ for all realizations, taking expectations we have:

$$\boldsymbol{y} = \Big(y_{ik}\Big)_{k\in\mathcal{X}_i, i\in\mathcal{N}} = \Big(P\Big(x_i^*(\tilde{\boldsymbol{c}}) = k\Big)\Big)_{k\in\mathcal{X}_i, i\in\mathcal{N}} \in CH(\mathcal{Y}).$$

It is clear from the previous discussion that we can obtain an upper bound on (4) by solving:

$$\begin{aligned}
Z^* \leq \max \quad & \sum_{i\in\mathcal{N}} \sum_{k\in\mathcal{X}_i} k \int c y_{ik}(c) f_i(c)dc \\
\text{s.t.} \quad & \int y_{ik}(c) f_i(c)dc = y_{ik} && \forall k \in \mathcal{X}_i \; \forall i \in \mathcal{N} \\
& \sum_{k\in\mathcal{X}_i} y_{ik}(c) = 1 && \forall c \; \forall i \in \mathcal{N} \\
& y_{ik}(c) \geq 0 && \forall c \; \forall k \in \mathcal{X}_i \; \forall i \in \mathcal{N} \\
& \boldsymbol{y} \in CH(\mathcal{Y}).
\end{aligned} \tag{5}$$

[2]We extend the definition of y_{ik} from Section 2 for simplicity of exposition.

For a given set of values of $\boldsymbol{y} \in CH(\mathcal{Y})$, the upper bound in (5) is separable across $i \in \mathcal{N}$. The ith subproblem that we need to solve is:

$$
\begin{aligned}
\max \quad & \sum_{k \in \mathcal{X}_i} k \int c y_{ik}(c) f_i(c) dc \\
\text{s.t.} \quad & \int y_{ik}(c) f_i(c) dc = y_{ik} && \forall k \in \mathcal{X}_i \\
& \sum_{k \in \mathcal{X}_i} y_{ik}(c) = 1 && \forall c \\
& y_{ik}(c) \geq 0 && \forall c \ \forall k \in \mathcal{X}_i.
\end{aligned}
$$

By relabeling the points in $\mathcal{X}_i$ if necessary, we may WLOG assume that $y_{ik} > 0$ for all $k \in \mathcal{X}_i$. The optimal values for the variables $y_{ik}(c)$ in the subproblem can be found using a greedy argument. Introducing the lagrange multipliers λ_{ik} for the first set of constraints, we obtain:

$$
\begin{aligned}
Z(\lambda) = \max \quad & \int \sum_{k \in \mathcal{X}_i} (ck - \lambda_{ik}) y_{ik}(c) f_i(c) dc + \sum_{k \in \mathcal{X}_i} \lambda_{ik} y_{ik} \\
\text{s.t.} \quad & \sum_{k \in \mathcal{X}_i} y_{ik}(c) = 1 \quad \forall c \\
& y_{ik}(c) \geq 0 \qquad \forall c \ \forall k \in \mathcal{X}_i.
\end{aligned}
$$

Thus, for a given value of c, we need to solve a linear program with a single budget constraint and non-negativity restrictions. Given a set of values in increasing order $\mathcal{X}_i = \{\alpha_i, \alpha_i + 1, \ldots, \beta_i\}$, the optimal solution is:

$$
y_{ik}(c) = \mathbb{I}(ck - \lambda_{ik} \geq cj - \lambda_{ij} \ \forall j \in \mathcal{X}_i)
$$

where $\mathbb{I}$ is the indicator function.

Let $\lambda^* = (\lambda^*_{ik})$ denote the optimal lagrange multipliers. In the solution to $Z(\lambda^*)$, our assumption that $y_{ik} > 0$ ensures that for all k, $y_{ik}(c) > 0$ for some c. Hence we must have

$$
y_{ik}(c) = \mathbb{I}(\lambda^*_{ik} - \lambda^*_{i,k-1} \leq c \leq \lambda^*_{i,k+1} - \lambda^*_{ik}) \quad \forall k \in \mathcal{X}_i,
$$

where $\lambda^*_{i,\alpha_i - 1} = -\infty$, $\lambda^*_{i,\beta_i + 1} = \infty$ (see Figure 3(a)). The Lagrange multipliers are chosen such that:

$$
y_{ik} = F_i(\lambda^*_{i,k+1} - \lambda^*_{ik}) - F_i(\lambda^*_{ik} - \lambda^*_{i,k-1}) \quad \forall k \in \mathcal{X}_i,
$$

with $\sum_{k \in \mathcal{X}_i} y_{ik} = 1$. Note that it is not possible for $\lambda^*_{ik} - \lambda^*_{i,k-1} \geq \lambda^*_{i,k+1} - \lambda^*_{ik}$ wherein $y_{ik}(c) = 0$ for all c and $y_{ik} = 0$ as indicated in Figure 3(b).

The values of the multipliers λ^*_{ik} that satisfy these set of equations are given as:

$$
\lambda^*_{ik} - \lambda^*_{i,k-1} = F_i^{-1}\left(\sum_{j \leq k-1} y_k \right).
$$

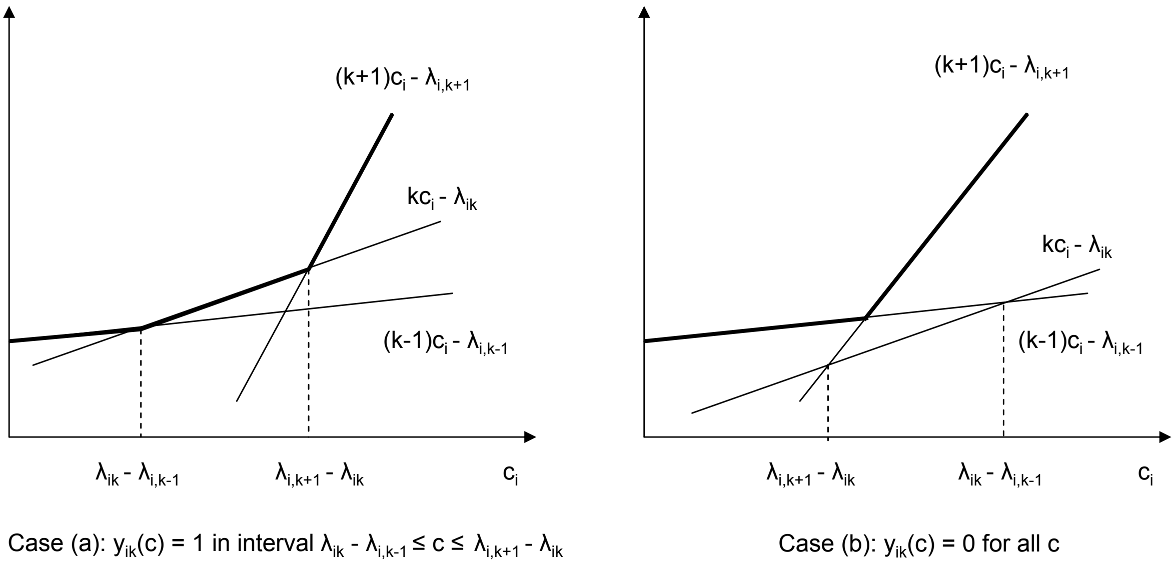

Figure 3: Optimal values for variables $y_{ik}(c)$.

Hence the upper bound on Z^* in (5) reduces to solving the problem:

$$\begin{aligned} Z^* \;\leq\; & \max \;\; \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{X}_i}\left(k\int_{F_i^{-1}(\sum_{j\leq k-1} y_{ij})}^{F_i^{-1}(\sum_{j\leq k} y_{ij})} cf_i(c)dc\right) \\ & \text{s.t.} \;\; \boldsymbol{y}\in CH(\mathcal{Y}). \end{aligned}$$

Using the substitution $t = F_i(c)$ and $dt = f_i(c)dc$, we can rewrite the upper bound as:

$$\begin{aligned} Z^* \;\leq\; & \max \;\; \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{X}_i}\left(k\int_{\sum_{j\leq k-1} y_{ij}}^{\sum_{j\leq k} y_{ij}} F_i^{-1}(t)dt\right) \\ & \text{s.t.} \;\; \boldsymbol{y}\in CH(\mathcal{Y}). \end{aligned}$$

In fact, this bound is tight under the marginal distribution model.

Theorem 1 *Under the marginal distribution model, Z^* and the persistence values are computed by solving the concave maximization problem:*

$$\begin{aligned} Z^* = max \;\; & \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{X}_i}\left(k\int_{\sum_{j\leq k-1} y_{ij}}^{\sum_{j\leq k} y_{ij}} F_i^{-1}(t)dt\right) \\ s.t. \;\; & \boldsymbol{y}\in CH(\mathcal{Y}). \end{aligned} \tag{6}$$

Proof. It is clear from the previous discussion that Formulation (6) provides an upper bound on Z^*. To show tightness, we construct an extremal distribution that attains the bound from the optimal $\boldsymbol{y}^*$ in Formulation (6). Expressing $\boldsymbol{y}^*$ as a convex combination of the extreme points of the polytope $\mathcal{Y}$ (denoted as $\mathcal{P}$), implies that there exist a set of numbers λ_p^* such that:

(i) $\lambda_p^* \geq 0$ for all $p \in \mathcal{P}$

(ii) $\sum_{p \in \mathcal{P}} \lambda_p^* = 1$

(iii) $y_{ik}^* = \sum_{p: y_{ik}[p]=1} \lambda_p^*$ for all $k \in \mathcal{X}_i,\ i \in \mathcal{N}$,

where $y_{ik}[p]$ denotes the value of the y_{ik} variable at the pth extreme point. We define the intervals:

$$I_{ik} = \left\{ c \;\middle|\; F_i^{-1}\left(\sum_{j \leq k-1} y_{ij}^*\right) \leq c \leq F_i^{-1}\left(\sum_{j \leq k} y_{ij}^*\right) \right\}. \tag{7}$$

We now generate the multivariate distribution θ^* as follows:

(a) Choose an extreme point p^* in $\mathcal{P}$ with probability λ_p^*

(b) For each $i \in \mathcal{N}$ with $y_{ik}[p^*] = 1$, generate $\tilde{c}_i \sim f_i(c)\mathbb{I}(c \in I_{ik})/y_{ik}^*$ (See Figure 4).

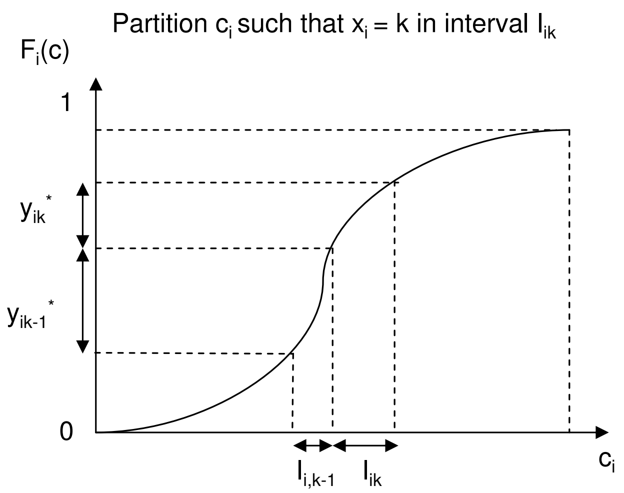

Figure 4: Constructing the extremal distribution for $\tilde{c}_i$.

Note that the cross dependency between the variables is not important here, and hence we can assume that for each fixed p, the distribution for $\tilde{c}_i$'s are generated independently. Under this construction, if $f_i'(c)$ denote the density function of $\tilde{c}_i$, then we have:

$$\begin{aligned} f_i'(c) &= \sum_{k \in \mathcal{X}_i} \sum_{p: y_{ik}[p]=1} \lambda_p^* \frac{f_i(c)\mathbb{I}(c \in I_{ik})}{y_{ik}^*} \\ &= \sum_{k \in \mathcal{X}_i} f_i(c)\mathbb{I}(c \in I_{ik}) \\ &= f_i(c). \end{aligned}$$

Thus the distribution generated in this way satisfies the marginal distribution condition. Furthermore, under $\tilde{\boldsymbol{c}}$, if we simply pick the pth solution with probability λ_p^*, instead of solving for $Z(\tilde{\boldsymbol{c}})$, we have

$$\begin{aligned} E_{\theta^*}\left(Z(\tilde{\boldsymbol{c}})\right) &\geq \sum_{p \in \mathcal{P}} \lambda_p^* \left(\sum_{i \in \mathcal{N}} \left(\sum_{k: y_{ik}[p]=1} \left(k \frac{\int c f_i(c)\mathbb{I}(c \in I_{ik}) dc}{y_{ik}^*} \right) \right) \right) \\ &= \sum_{i \in \mathcal{N}} \left(\sum_{k \in \mathcal{X}_i} \left(\sum_{p: y_{ik}[p]=1} \lambda_p^* \right) \left(k \frac{\int c f_i(c)\mathbb{I}(c \in I_{ik}) dc}{y_{ik}^*} \right) \right) \\ &= \sum_{i \in \mathcal{N}} \sum_{k \in \mathcal{X}_i} k \int c f_i(c)\mathbb{I}(c \in I_{ik}) dc \\ &= \sum_{i \in \mathcal{N}} \sum_{k \in \mathcal{X}_i} k \int_{\sum_{j \leq k-1} y_{ij}^*}^{\sum_{j \leq k} y_{ij}^*} F_i^{-1}(t) dt \end{aligned}$$

Since θ^* generates an expected optimal objective value that is greater than or equal to the optimal solution from Formulation (6) and satisfies the marginal distributions, it attains Z^*.
To check the concavity of the objective function, we define a new variable Y_{ik} by the affine transformation:

$$Y_{ik} = \sum_{j \leq k} y_{ij}.$$

The objective function can then be expressed as:

$$\sum_{i \in \mathcal{N}} \sum_{k \in \mathcal{X}_i} k \int_{Y_{i,k-1}}^{Y_{ik}} F_i^{-1}(t) dt.$$

The first derivative of the objective function with respect to Y_{ik} is then given as:

$$\frac{\partial \left(k \int_{Y_{i,k-1}}^{Y_{ik}} F_i^{-1}(t) dt + (k+1) \int_{Y_{ik}}^{Y_{i,k+1}} F_i^{-1}(t) dt \right)}{\partial Y_{ik}} = k F_i^{-1}(Y_{ik}) - (k+1) F_i^{-1}(Y_{ik}) = -F_i^{-1}(Y_{ik}).$$

This is a decreasing function in Y_{ik} since F_i^{-1} is the inverse cumulative distribution function. Hence the objective is concave in the Y_{ik} variables and therefore in the y_{ik} variables. ■

The result in Theorem 1 directly extends to minimization problems.

Corollary 1 *Under the marginal distribution model, we have:*

$$\min_{\theta \in \Theta} E_\theta \left(\min\{\tilde{\boldsymbol{c}}'\boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}\} \right) = \min \left\{ \sum_{i \in \mathcal{N}} \sum_{k \in \mathcal{X}_i} \left(k \int_{\sum_{j \geq k+1} y_{ij}}^{\sum_{j \geq k} y_{ij}} F_i^{-1}(t)dt \right) : \boldsymbol{y} \in CH(\mathcal{Y}) \right\}. \quad (8)$$

Theorem 1 indicates that the difficulty of solving the persistency problem for general discrete optimization problems is related to characterizing the convex hull of the extreme points in the binary reformulation. We compare this with a compact relaxation that uses the convex hull of the original $\mathcal{X}$ polytope:

$$\begin{aligned} Z_0^* = \max \quad & \sum_{i \in \mathcal{N}} \sum_{k \in \mathcal{X}_i} \left(k \int_{\sum_{j \leq k-1} y_{ij}}^{\sum_{j \leq k} y_{ij}} F_i^{-1}(t)dt \right) \\ \text{s.t.} \quad & \left(\sum_{k \in \mathcal{X}_i} k y_{ik} \right)_{i \in \mathcal{N}} \in \mathcal{CH}(\mathcal{X}). \end{aligned} \quad (9)$$

Clearly, $Z^* \leq Z_0^*$. Note that instead of the space $\mathcal{CH}(\mathcal{Y})$, which may not be readily available, we are now solving for the solution in $\mathcal{CH}(\mathcal{X})$. The latter is much easier to characterize for many problems. Hence Z_0^* can be evaluated if $\mathcal{CH}(\mathcal{X})$ is readily available. However, the relaxation is in general weak with the optimal values for each variable concentrated at the two extreme values α_i and β_i.

Proposition 1 *An upper bound on Z^* obtained from solving Formulation (9) is equivalent to:*

$$\begin{aligned} Z_0^* = max \quad & \sum_{i \in \mathcal{N}} \left(\alpha_i \int_0^{y_{i\alpha_i}} F_i^{-1}(t)dt + \beta_i \int_{1-y_{i\beta_i}}^1 F_i^{-1}(t)dt \right) \\ s.t. \quad & (\alpha_i y_{i\alpha_i} + \beta_i y_{i\beta_i})_{i \in \mathcal{N}} \in \mathcal{CH}(\mathcal{X}). \end{aligned} \quad (10)$$

Proof. Clearly, Formulation (10) is a restricted version of Formulation (9) at the two extreme values α_i and β_i for each $i \in \mathcal{N}$. To show tightness, consider an optimal solution to Formulation (9) denoted as y_{ik}^*. Now, define the variables for Formulation (10) as:

$$y_{i\alpha_i} = \sum_{k \in \mathcal{X}_i} \left(\frac{\beta_i - k}{\beta_i - \alpha_i} \right) y_{ik}^* \quad \text{and} \quad y_{i\beta_i} = \sum_{k \in \mathcal{X}_i} \left(\frac{k - \alpha_i}{\beta_i - \alpha_i} \right) y_{ik}^*.$$

This defines a feasible solution since:

$$y_{i\alpha_i} + y_{i\beta_i} = \sum_{k \in \mathcal{X}_i} y_{ik}^* = 1 \text{ and } \alpha_i y_{i\alpha_i} + \beta_i y_{i\beta_i} = \sum_{k \in \mathcal{X}_i} k y_{ik}^*.$$

Furthermore, we set

$$y_{i\alpha_i}(c) = \sum_{k \in \mathcal{X}_i} \left(\frac{\beta_i - k}{\beta_i - \alpha_i} \right) y_{ik}^*(c) \text{ and } y_{i\beta_i}(c) = \sum_{k \in \mathcal{X}_i} \left(\frac{k - \alpha_i}{\beta_i - \alpha_i} \right) y_{ik}^*(c),$$

where $y^*_{ik}(c) = \mathbb{I}(c \in I_{ik})$ in (7). The objective function to the restricted formulation can be expressed as:

$$
\begin{aligned}
\text{Objective function} &= \sum_{i\in\mathcal{N}} \left(\alpha_i \int c y_{i\alpha_i}(c) f_i(c) dc + \beta_i \int c y_{i\beta_i}(c) f_i(c) dc \right) \\
&= \sum_{i\in\mathcal{N}} \left(\alpha_i \sum_{k\in\mathcal{X}_i} \left(\frac{\beta_i - k}{\beta_i - \alpha_i} \right) \int c y^*_{ik}(c) dc + \beta_i \sum_{k\in\mathcal{X}_i} \left(\frac{k - \alpha_i}{\beta_i - \alpha_i} \right) \int c y^*_{ik}(c) f(c) dc \right) \\
&= \sum_{i\in\mathcal{N}} \sum_{k\in\mathcal{X}_i} k \int c y^*_{ik}(c) f(c) dc \\
&= \sum_{i\in\mathcal{N}} \sum_{k\in\mathcal{X}_i} \left(k \int_{\sum_{j\leq k-1} y^*_{ij}}^{\sum_{j\leq k} y^*_{ij}} F_i^{-1}(t) dt \right),
\end{aligned}
$$

which proves the desired result. ■

It is clear that the upper bound in Proposition 1 is tight for 0-1 optimization problems. This result can also be interpreted as the dual of the formulations obtained in Meilijson and Nadas [10] and Weiss [15]. Our constructive approach however appears to be more direct and elegant. A direct application of this result with F_i set to exponential distributions leads to the stochastic bounds for combinatorial optimization problems discussed in Application 1 in Section 1 of this paper.

We illustrate the key ideas developed thus far with a simple numerical example.

Example: Consider the two-dimensional integer polytope given as:

$$
\mathcal{X} := \left\{ (x_1, x_2) \in \Re^2 \;\middle|\; x_1 + x_2 \geq 1,\ -x_1 - x_2 \geq -3,\ x_1 - x_2 \geq -1,\ -x_1 + x_2 \geq -1 \right\}.
$$

The four extreme points of this polyhedron (see Figure 5) are $\{(0,1),\ (1,0),\ (1,2),\ (2,1)\}$ with $\mathcal{X}_1 = \{0,1,2\}$ and $\mathcal{X}_2 = \{0,1,2\}$. Assume that both $\tilde{c}_1$ and $\tilde{c}_2$ are independently and uniformly distributed on $[0,1]$. Note that under this model, the extreme points $\{(1,2),\ (2,1)\}$ have equal chances of attaining the optimal solution under the random objective function. We solve the case with only the marginal distribution conditions imposed on the problem.

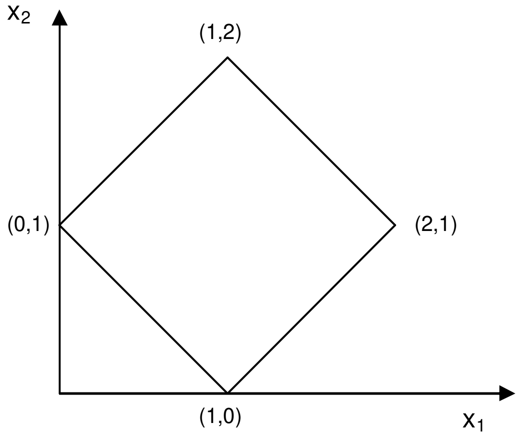

Figure 5: A diamond shaped polyhedron.

In this case, the upper bound in Proposition 1 reduces to

$$
\begin{aligned}
Z_0^* = \max \quad & 0.5\left(2 - y_{10}^2 - (y_{10}+y_{11})^2 + 2 - y_{20}^2 - (y_{20}+y_{21})^2\right) \\
\text{s.t.} \quad & (y_{11}+2y_{12}) + (y_{21}+2y_{22}) \geq 1 \\
& -(y_{11}+2y_{12}) - (y_{21}+2y_{22}) \geq -3 \\
& (y_{11}+2y_{12}) - (y_{21}+2y_{22}) \geq -1 \\
& -(y_{11}+2y_{12}) + (y_{21}+2y_{22}) \geq -1 \\
& y_{10}+y_{11}+y_{12} = 1 \\
& y_{20}+y_{21}+y_{22} = 1 \\
& y_{10}, y_{11}, y_{12}, y_{20}, y_{21}, y_{22} \geq 0
\end{aligned}
$$

This is a concave maximization problem over linear constraints. Solving this with CPLEX Version 9.1, OPL Studio Version 4.1 yields $Z_0^* = 1.875$ with persistence values:

$$(y_{10}, y_{11}, y_{12}, y_{20}, y_{21}, y_{22}) = (0.25, 0, 0.75, 0.25, 0, 0.75).$$

We can use the notion of cutting planes to strengthen the above formulation. One valid equality that is obtained from Figure 5 (or by looking at the $\mathcal{Y}$ polytope) which is not satisfied by the current solution is:

$$y_{10} + y_{12} + y_{20} + y_{22} = 1.$$

This inequality says that in the optimal solution, either $x_1 = 0$ or $x_2 = 0$. Solving the quadratic optimization problem with this added equality yields $Z^* = 1.75$ which is the tight bound. The corresponding optimal persistence values are:

$$(y_{10}, y_{11}, y_{12}, y_{20}, y_{21}, y_{22}) = (0, 0.5, 0.5, 0, 0.5, 0.5),$$

with the optimal solutions concentrated at $(x_1, x_2) = (1, 2)$ and $(x_1, x_2) = (2, 1)$ respectively. Interestingly, the persistence values obtained from our model is precisely the values under the independence model. ■

3.2 Marginal Moment Model (MMM)

We now relax the assumption on the knowledge of complete marginal distributions to the knowledge of a finite set of marginal moments. For simplicity, we focus on the case when only the mean μ_i and variance σ_i^2 of each objective coefficient $\tilde{c}_i$ is known with the support over the entire real line. For generalization to higher order moments, the reader is referred to Bertsimas, Natarajan and Teo [3], [4]. Let Θ denote the set of multivariate distributions θ for the objective coefficients such that they satisfy the given mean and variances for each $\tilde{c}_i$.

Under the marginal moment model with mean and variance information, Z^* is computed by solving:

$$
\begin{aligned}
Z^* = \max \quad & \sum_{i \in \mathcal{N}} \sum_{k \in \mathcal{X}_i} k w_{ik} \\
\text{s.t.} \quad & \sum_{k \in \mathcal{X}_i} z_{ik} = \mu_i^2 + \sigma_i^2 \quad && \forall i \in \mathcal{N} \\
& \sum_{k \in \mathcal{X}_i} w_{ik} = \mu_i && \forall i \in \mathcal{N} \\
& \sum_{k \in \mathcal{X}_i} y_{ik} = 1 && \forall i \in \mathcal{N} \\
& z_{ik} y_{ik} \geq w_{ik}^2 && \forall k \in \mathcal{X}_i \ \forall i \in \mathcal{N} \\
& \boldsymbol{y} \in \mathcal{CH}(\mathcal{Y}).
\end{aligned}
\tag{11}
$$

This formulation is obtained using argument similar to Section 3.1 and is skipped here. The variables can be interpreted as (scaled) conditional moments (refer to Bertsimas, Natarajan and Teo [4]):

$$
\begin{pmatrix} z_{ik} \\ w_{ik} \\ y_{ik} \end{pmatrix} = \begin{pmatrix} E_\theta\left(\tilde{c}_i^2 \,\middle|\, x_i^*(\tilde{\boldsymbol{c}}) = k\right) P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k) \\ E_\theta\left(\tilde{c}_i \,\middle|\, x_i^*(\tilde{\boldsymbol{c}}) = k\right) P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k) \\ P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k) \end{pmatrix},
$$

with the y_{ik} variables denoting the persistency values. The first three constraints model the margianl moment conditions. The constraints $z_{ik} y_{ik} \geq w_{ik}^2$ where $y_{ik} \in [0, 1]$ correspond to the moment feasibility conditions (Jensen's inequality) in this case.

Proposition 2 *Under the marginal moment model with mean and variance information, Z^* and the persistence values are computed by solving the concave maximization problem:*

$$\begin{aligned} Z^* = max \quad & \sum_{i \in \mathcal{N}} \left(\mu_i \sum_{k \in \mathcal{X}_i} k y_{ik} + \sigma_i \sqrt{\sum_{k \in \mathcal{X}_i} k^2 y_{ik} - \Big(\sum_{k \in \mathcal{X}_i} k y_{ik} \Big)^2} \right) \\ s.t. \quad & \boldsymbol{y} \in \mathcal{CH}(\mathcal{Y}). \end{aligned}$$

Proof. For given values of $\boldsymbol{y} \in \mathcal{CH}(\mathcal{Y})$, computing Z^* in Formulation (11) reduces to solving subproblems of the type:

$$\begin{aligned} \max_{\boldsymbol{z_i}, \boldsymbol{w_i}} \quad & \sum_{k \in \mathcal{X}_i} k w_{ik} \\ \text{s.t.} \quad & \sum_{k \in \mathcal{X}_i} z_{ik} = \mu_i^2 + \sigma_i^2 \\ & \sum_{k \in \mathcal{X}_i} w_{ik} = \mu_i \\ & z_{ik} y_{ik} \geq w_{ik}^2 \qquad \forall k \in \mathcal{X}_i. \end{aligned}$$

A relaxation to this subproblem is obtained from aggregating the last set of constraints[3]:

$$\begin{aligned} \max_{\boldsymbol{z_i}, \boldsymbol{w_i}} \quad & \sum_{k \in \mathcal{X}_i} k w_{ik} \\ \text{s.t.} \quad & \sum_{k \in \mathcal{X}_i} w_{ik} = \mu_i \\ & \sum_{k \in \mathcal{X}_i} \frac{w_{ik}^2}{y_{ik}} \leq \mu_i^2 + \sigma_i^2. \end{aligned}$$

This relaxation is tight since we can generate an optimal solution to the original subproblem by setting $z_{ik} = w_{ik}^2 / y_{ik} + \epsilon_{ik}$ with appropriate perturbations $\epsilon_{ik} \geq 0$ to ensure that the second moment constraint is met. Furthermore this does not change the objective which is independent of z_{ik}. Thus the ith subproblem reduces to maximizing a linear objective over a linear equality constraint and a convex quadratic constraint. Introducing multipliers λ_i and ν_i for the constraints, the Karush-Kuhn-Tucker feasibility and optimality conditions for this problem are:

$$\begin{aligned} &\text{(i)} \quad \sum_{k \in \mathcal{X}_i} w_{ik} = \mu_i \text{ and } \sum_{k \in \mathcal{X}_i} \frac{w_{ik}^2}{y_{ik}} \leq \mu_i^2 + \sigma_i^2 \\ &\text{(ii)} \quad -k + \lambda_i + 2\nu_i \frac{w_{ik}}{y_{ik}} = 0 \text{ for all } k \in \mathcal{X}_i \\ &\text{(iii)} \quad \nu_i \left(\mu_i^2 + \sigma_i^2 - \sum_{k \in \mathcal{X}_i} \frac{w_{ik}^2}{y_{ik}} \right) = 0 \\ &\text{(iv)} \quad \nu_i \geq 0. \end{aligned}$$

[3]We assume that all $y_{ik} > 0$. It is possible to extend this to allow for some of the $y_{ik} = 0$.

From condition (ii), we can assume that $\nu_i > 0$, else the problem is trivial with only a single value for $k \in \mathcal{X}_i$. Substituting $w_{ik} = (k - \lambda)y_{ik}/2\nu_i$ into the feasibility constraints yields the optimal values:

$$\nu_i = \frac{\sqrt{\sum_{k\in\mathcal{X}_i} k^2 y_{ik} - \left(\sum_{k\in\mathcal{X}_i} k y_{ik}\right)^2}}{2\sigma_i},$$

$$\lambda_i = \sum_{k\in\mathcal{X}_i} k y_{ik} - 2\nu_i \mu_i,$$

$$w_{ik} = \mu_i y_{ik} + \sigma_i y_{ik} \left(\frac{k - \sum_{k\in\mathcal{X}_i} k y_{ik}}{\sqrt{\sum_{k\in\mathcal{X}_i} k^2 y_{ik} - \left(\sum_{k\in\mathcal{X}_i} k y_{ik}\right)^2}} \right).$$

The corresponding optimal objective value for the ith subproblem is:

$$\mu_i \sum_{k\in\mathcal{X}_i} k y_{ik} + \sigma_i \sqrt{\sum_{k\in\mathcal{X}_i} k^2 y_{ik} - \Big(\sum_{k\in\mathcal{X}_i} k y_{ik}\Big)^2},$$

which proves the desired result. ■

The above proposition can be simplified further in the case of 0-1 optimization problem:

Corollary 2 *Given only mean and variance information, Z^* and the persistence values for 0-1 optimization problems are computed by solving the concave maximization problem:*

$$\begin{aligned} Z^* = max \quad & \sum_{i\in\mathcal{N}} \left(\mu_i x_i + \sigma_i \sqrt{x_i(1-x_i)} \right) \\ s.t. \quad & \boldsymbol{x} \in \mathcal{CH}(\mathcal{X}). \end{aligned}$$

A direct application with

$$\mathcal{X} = \left\{ \boldsymbol{x} \;\middle|\; \sum_{i\in\mathcal{N}} x_i = 1,\ x_i \in \{0,1\}\ \forall i \in \mathcal{N} \right\}$$

leads to the formula for choice probabilities for discrete choice models discussed in Application 2 in Section 1 of this paper.

We next provide a simple application of Corollary 2 for approximating a normal distribution.

Example: For a standard normal random variable, the distribution function is

$$P(\tilde{c} \le z) = \phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt,$$

for which no closed form expression exists but numerical estimates are easily available. Consider the following simple analytic approximation to the normal distribution. Let $\tilde{c}$ denote a random variable with mean μ and variance σ^2. Without loss of generality, let $\mu = 0$ and $\sigma^2 = 1$. Using Corollary 2, the persistency model for $Z(\tilde{c}) = \max(\tilde{c}, z)$ reduces to the following problem:

$$
\begin{aligned}
Z^* = \max \quad & \left(\sqrt{x_1(1-x_1)} + zx_2\right) \\
\text{s.t.} \quad & x_1 + x_2 = 1,\ x_1 \geq 0,\ x_2 \geq 0.
\end{aligned}
$$

We obtain:

$$
P_{\theta^*}(\tilde{c} \leq z) = P_{\theta^*}(x_2(\tilde{c}) = 1) = \frac{1}{2}\left(1 + \frac{z}{\sqrt{z^2+1}}\right).
$$

Figure 6 compares these values for $P(\tilde{c} \leq z)$ and $P_{\theta^*}(\tilde{c} \leq z)$ as a function of z. Clearly from the figure, these two values are observed to be in close agreement with an absolute error of at most 0.0321. This suggests that the extremal distribution obtained from the persistency model for a single variable is a good approximation to the normal distribution.

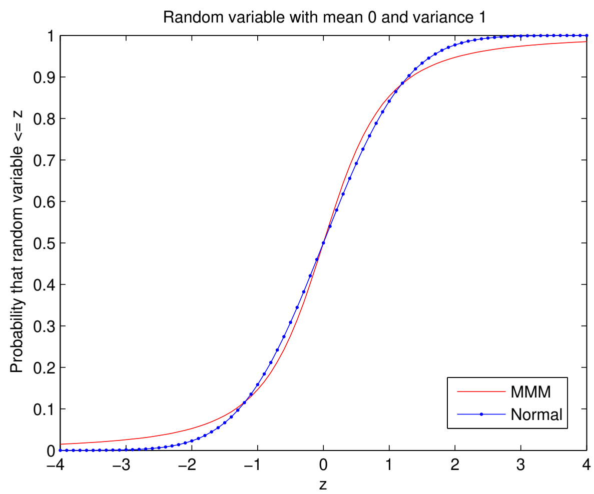

Figure 6: Comparison of $P(\tilde{c} \leq z)$ for normal distribution and MMM

3.3 Range and Unimodal Distribution

The approach described in this paper is flexible and can be enhanced in different ways. For instance, the Marginal Moment Model can be strengthened further if we know the range of support for each marginal distribution. For instance, other than the moments condition, if we know further that $\tilde{c}_i \in [\underline{c}_i, \overline{c}_i]$, then using the condition:

$$
(\overline{c}_i - \tilde{c}_i)(c_i - \underline{c}_i) \geq 0,
$$

we can capture the range conditions into the persistency model (11) through the addition of the following inequality:

$$
(\underline{c}_i + \overline{c}_i)w_{ik} \geq \underline{c}_i\overline{c}_i y_{ik} + z_{ik}.
$$

Other information concerning the distribution can also be introduced into the formulation. As an illustration, we describe next a simple method to tighten the formulation using information on the shape of the distribution.

The main results driving our enhanced reformulation is the following result due to Khintchine: A random variable $\tilde{X}$ has a unimodal distribution if and only if there exists a random variable $\tilde{Y}$ such that $\tilde{X} \sim \tilde{U}\tilde{Y}$ where $\tilde{U}$ is a uniform $[0,1]$ random variable independent of $\tilde{Y}$. Note that a distribution is unimodal if the density function $f(\cdot)$ is monotonically nondecreasing on $(-\infty, 0]$ and monotonically nonincreasing on $[0, \infty)$. Some examples of unimodal distributions are: the standard normal, exponential, Cauchy probability densities, and the class of stable distributions.

Assume that the objective coefficient $\tilde{c}_i$ is unimodal with mean μ_i and variance σ_i^2. We can use Khintchine's Theorem to refine our persistency model under the moments approach. Let $\tilde{c}_i = \tilde{u}_i\tilde{d}_i$ where $\tilde{u}_i$ is uniformly generated in [0,1] and independent of $\tilde{d}_i$. The first and second moments for $\tilde{d}_i$ is then given as:

$$E_\theta(\tilde{d}_i) = 2\mu_i \text{ and } E_\theta(\tilde{d}_i^2) = 3(\mu_i^2 + \sigma_i^2).$$

Furthermore, we have

$$\begin{aligned}
E_\theta\Big(\tilde{c}_i x_i(\tilde{\boldsymbol{c}})\Big) &= \int_0^1 E_\theta\Big(\tilde{u}_i\tilde{d}_i x_i(\tilde{\boldsymbol{c}}) \;\Big|\; \tilde{u}_i = u\Big)du \\
&= \int_0^1 uE_\theta\Big(\tilde{d}_i x_i(\tilde{\boldsymbol{c}}) \;\Big|\; \tilde{u}_i = u\Big)du \\
&= \int_0^1 u\sum_{k\in\mathcal{X}_i}\Big(kE_\theta\Big(\tilde{d}_i \;\Big|\; x_i(\tilde{\boldsymbol{c}}) = k, \tilde{u}_i = u\Big)P_\theta\Big(x_i(\tilde{\boldsymbol{c}}) = k \;\Big|\; \tilde{u}_i = u\Big)\Big)du.
\end{aligned}$$

Define the variables as (scaled) conditional moments:

$$\begin{pmatrix} z_{iku} \\ w_{iku} \\ y_{iku} \end{pmatrix} = \begin{pmatrix} E_\theta\Big(\tilde{d}_i^2 \;\Big|\; x_i^*(\tilde{\boldsymbol{c}}) = k, \tilde{u}_i = u\Big)P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k \;\Big|\; \tilde{u}_i = u) \\ E_\theta\Big(\tilde{d}_i \;\Big|\; x_i^*(\tilde{\boldsymbol{c}}) = k, \tilde{u}_i = u\Big)P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k \;\Big|\; \tilde{u}_i = u) \\ P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k \;\Big|\; \tilde{u}_i = u) \end{pmatrix}.$$

Since:

$$\int_0^1 P_\theta\Big(x_i(\tilde{\boldsymbol{c}}) = k \;\Big|\; \tilde{u}_i = u\Big)du = P_\theta\Big(x_i(\tilde{\boldsymbol{c}}) = k\Big),$$

we can reformulate the persistency model as shown below.

Proposition 3 *Under the marginal moment model given mean, variances and unimodal distribu-*

tions, Z^* and the persistence values are computed by solving the problem:

$$
\begin{aligned}
Z^* = max \qquad & \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{X}_i} k\int_0^1 uw_{iku}du \\
s.t. \qquad & \sum_{k\in\mathcal{X}_i} z_{iku} = 3(\mu_i^2+\sigma_i^2) && \forall u\in[0,1]\ \forall i\in\mathcal{N} \\
& \sum_{k\in\mathcal{X}_i} w_{iku} = 2\mu_i && \forall u\in[0,1]\ \forall i\in\mathcal{N} \\
& \sum_{k\in\mathcal{X}_i} y_{iku} = 1 && \forall u\in[0,1]\ \forall i\in\mathcal{N} \\
& z_{iku}y_{iku} \geq w_{iku}^2, && \forall u\in[0,1],\ \forall k\in\mathcal{X}_i\ \forall i\in\mathcal{N} \\
& \left(\int_0^1 y_{iku}du\right)_{k\in\mathcal{X}_i, i\in\mathcal{N}} \in \mathcal{CH}(\mathcal{Y}).
\end{aligned}
$$

One disadvantage of this model is that there are infinitely many variables, depending on $u \in [0,1]$. Numerically, this problem can be solved by discretizing the values of u in the range $[0,1]$. The increase in the number of variables, however, allows us to bring in characteristic of the shape of the marginal distribution into our formulation.

4 Applications: Discrete Choice Modeling

In this section, we test the performance of the persistency model on two canonical problems chosen from the literature: discrete choice models and integer knapsack problem (i.e. discrete choice under budget constraint). The former is selected to demonstrate estimation and prediction of choice probabilities for discrete choice models using commercial nonlinear solvers. The latter problem is selected to demonstrate how the projection to higher dimensional polytope can be executed for a NP-hard discrete optimization problem.

4.1 Estimation in Discrete Choice Models

Consider a discrete choice problem, where a set of alternatives $\mathcal{N} = \{1, 2, \ldots, n\}$ is considered. The utility that an individual $i \in \mathcal{I}$ assigns to alternative $j \in \mathcal{N}$ can be expressed as:

$$\tilde{U}_{ij} = V_{ij} + \tilde{\epsilon}_{ij}.$$

$V_{ij} = \beta'\mathbf{x}_{ij}$ is the systematic utility where $\mathbf{x}_{ij}$ is the vector of attributes characterizing the individual and the alternative and β are the parameters of the model that need to be estimated. The part of the utility that is assumed to be random is $\tilde{\epsilon}_{ij}$. Let P_{ij} denote the probability that alternative j is selected by individual i. Under the random utility maximizing model, the choice probability is evaluated as:

$$P_{ij} = P\left(\tilde{U}_{ij} \geq \tilde{U}_{ik}\ \forall k\in\mathcal{N}\right) = P\left(\beta'\mathbf{x}_{ij} + \tilde{\epsilon}_{ij} \geq \beta'\mathbf{x}_{ik} + \tilde{\epsilon}_{ik}\ \forall k\in\mathcal{N}\right).$$

The traditional approach to estimating the parameters β is to use a log-likelihood estimation technique (see Ben-Akiva and Leerman [2]). Let $y_{ij} = 1$ if alternative j is selected by individual i and 0 otherwise. The probability of each person in $\mathcal{I}$ choosing the alternative that they were actually observed to choose is:

$$\mathcal{L}(\beta) = \prod_{i \in \mathcal{I}} \prod_{j \in \mathcal{N}} (P_{ij})^{y_{ij}}.$$

The maximum log-likelihood estimator is obtained by solving the following optimization problem:

$$\max_{\beta} \mathcal{LL}(\beta) = \max_{\beta} \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{N}} y_{ij} \ln P_{ij}.$$

In this section, we compare three different models for the estimation procedure - the popular Multinomial Logit Model (MNL), the Cross-Nested Logit Model (CNL) and our proposed Marginal Moments Model (MMM).

(1) For MNL (see McFadden [9]), the error terms are assumed to be independent and identically distributed as extreme value distributions. The choice probabilities are given as

$$P_{ij} = \frac{e^{\beta' \mathbf{x}_{ij}}}{\sum_{k \in \mathcal{N}} e^{\beta' \mathbf{x}_{ik}}}.$$

The log-likelihood objective is known to be globally concave under this model implying that the optimal β can be estimated efficiently using various convex optimization solvers. A freely available package BIOGEME (developed by Bierlaire ([5]) was used for our computations. This is based on a sequential equality constrained quadratic programming method.

(2) For CNL (see Ben-Akiva and Bierlaire [1]), the choice set is partitioned into a set of nests denoted as $\mathcal{M}$. The parameter μ_m denotes the scale parameter for nest $m \in \mathcal{M}$ and α_{jm} denotes the cross nesting parameter for alternative j in nest m such that $\sum_{m \in \mathcal{M}} \alpha_{jm} = 1$. The choice probabilities are given as

$$P_{ij} = \sum_{m \in \mathcal{M}} \left(\frac{\left(\sum_{j' \in \mathcal{N}} \alpha_{j'm}^{\mu_m} e^{\mu_m \beta' \mathbf{x}_{ij'}} \right)^{\frac{1}{\mu_m}}}{\sum_{m' \in \mathcal{M}} \left(\sum_{j' \in \mathcal{N}} \alpha_{j'm'}^{\mu_{m'}} e^{\mu_{m'} \beta' \mathbf{x}_{ij'}} \right)} \times \frac{\alpha_{jm}^{\mu_m} e^{\mu_m \beta' \mathbf{x}_{ij}}}{\sum_{j' \in \mathcal{N}} \alpha_{j'm}^{\mu_m} e^{\mu_m \beta' \mathbf{x}_{ij'}}} \right).$$

The log-likelihood function is however not concave under this model with no guarantee of finding the global optimal solution. For our computations, we used BIOGEME to estimate the coefficients.

(3) For MMM, we assume that the error terms ϵ_{ij} has mean 0 and variance σ_{ij}^2. The random utility $\tilde{U}_{ij}$ is then distributed with mean $\beta' \mathbf{x}_{ij}$ and variance σ_{ij}^2. The variance term is used to

captures the variability due to factors not captured by the specified attributes of the model. The choice probabilities are given as

$$P_{ij} = \frac{1}{2}\left(1 + \frac{\beta'\mathbf{x}_{ij} - \lambda_i}{\sqrt{(\beta'\mathbf{x}_{ij} - \lambda_i)^2 + \sigma_{ij}^2}}\right)$$

where λ_i is found by solving the equation:

$$\sum_{i\in\mathcal{N}} \frac{1}{2}\left(1 + \frac{\beta'\mathbf{x}_{ij} - \lambda_i}{\sqrt{(\beta'\mathbf{x}_{ij} - \lambda_i)^2 + \sigma_{ij}^2}}\right) = 1.$$

The maximum log-likelihood estimation problem under MMM is:

$$\begin{aligned} \max \quad & \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{N}} y_{ij} \ln \frac{1}{2}\left(1 + \frac{\beta'\mathbf{x}_{ij} - \lambda_i}{\sqrt{(\beta'\mathbf{x}_{ij} - \lambda_i)^2 + \sigma_{ij}^2}}\right) \\ \text{s.t.} \quad & \sum_{j\in\mathcal{N}} \frac{1}{2}\left(1 + \frac{\beta'\mathbf{x}_{ij} - \lambda_i}{\sqrt{(\beta'\mathbf{x}_{ij} - \lambda_i)^2 + \sigma_{ij}^2}}\right) = 1 \qquad \forall i \in \mathcal{I}. \end{aligned} \tag{12}$$

The log-likelihood function is again not necessarily concave under this model. For our computations, we considered two versions of MMM. The first version assumed a fixed variance $\sigma_{ij}^2 = \pi^2/6$ for all the error terms as in logit. The second version assumed that the variances relate only to the alternatives, namely $\sigma_{ij}^2 = \sigma_j^2$. The variances are however not known and are estimated. We used LOQO to solve the estimation problem using an infeasible primal-dual interior point method applied to quadratic approximations to the original problem.

Example: We compare the performance of the three different discrete choice models, using a real-life discrete choice data set taken from Bierlaire et al. [6]. This is a transportation mode choice problem with three available alternatives: Train, Car and the Swiss Metro. The attributes that were modeled for each alternative are indicated in Table 1.

Table 1: Attributes and Alternatives.

	Alternative		
Attribute	Train	Swissmetro	Car
Age	B-Age		
Cost	B-Cost	B-Cost	B-Cost
Frequency	B-Freq	B-Freq	
Luggage			B-Luggage
Seats		B-Seats	
Time	B-Time	B-Time	B-Time
GA (Season Ticket)	B-GA	B-GA	
ASC (Alternative Specific Constant)		ASC-SM	ASC-Car

The data set consisted of a total of 10710 preferences (including revealed and stated preferences) for the three alternatives. The parameters for the model were estimated using half the data (N = 5355) and tested for accuracy using the other half of the data. For the MNL model, a total of nine parameters were estimated. For the CNL model, we estimated a total of fourteen parameters including μ-Exist, μ-Future, α-Exist-Car, α-Exist-SM and α-Exist-Train. These additional parameters correspond to the scale and cross nesting parameters for CNL. For the MMM, there were a total of nine parameters with two additional parameters in the version with different variances. Therein the two extra parameters were Variance-Train and Variance-SM with the variance for the Car alternative set to the default of $\pi^2/6$. The parameters estimated using the maximum log-likelihood formulation for the models are shown in Table 2. In terms of the log-likelihood objective it is clear that the MMM with different variances outperforms the MNL and CNL. This is achieved without capturing the rather complex structure of the CNL model.

Table 2: Estimation results using MNL, CNL and MMM.

Parameter	MNL	CNL	MMM (Fixed Variance)	MMM (Different Variance)
ASC-CAR	0.5942	0.3574	0.6826	0.1003
ASC-SM	0.5310	0.2746	0.5826	0.0642
B-Age	0.0639	0.0506	0.1276	0.1240
B-Cost	-0.0062	-0.0047	-0.0103	-0.0105
B-Freq	-0.0065	-0.0041	-0.0010	-0.0059
B-GA	1.8439	1.0261	2.3200	2.1808
B-Luggage	-0.1367	-0.1453	-0.1452	-0.2464
B-Seats	-0.1816	0.0013	-0.0186	-0.1312
B-Time	-0.0132	-0.0098	-0.0190	-0.0170
μ-Exist		2.5857		
μ-Future		1.0000		
α-Exist-Car		0.9665		
α-Exist-SM		0.0000		
α-Exist-Train		0.9548		
Variance-Car			1.6450	1.6450
Variance-Train			1.6450	0.1619
Variance-SM			1.6450	3.1465
Log-likelihood	-4242.05	-4117.23	-4146.43	-4065.58
Computational Time (seconds)	3	64	13.6	42.5
N	5355	5355	5355	5355

To test the quality of the estimations, we compare the predictions under the models with the actual choices made using the second half of the data. Tables 3-6 indicate the predicted probabilities classified by the actual choices made. All models obtain highly accurate predictions for the probability that a particular alternative is chosen. The error in terms of the aggregate deviation of actual and estimated percentages is slightly better under MMM with the different variances as compared to MNL and CNL.

Table 3: Prediction using MNL.

	Predicted Choices			
Actual Choices	Train	SM	Car	Total
Train	136.3	462.1	106.6	705
SM	410.3	1949.4	715.3	3075
Car	156.8	722.6	695.6	1575
Total	703.4	3134.1	1517.5	5355
Predicted Share	13.14%	58.53%	28.34%	
Actual Share	13.17%	57.42%	29.41%	
Error				2.207%

Table 4: Prediction using CNL.

	Predicted Choices			
Actual Choices	Train	SM	Car	Total
Train	160.4	442.0	102.5	705
SM	426.3	1920.0	728.7	3075
Car	94.7	772.6	707.8	1575
Total	681.4	3134.6	1539.0	5355
Predicted Share	12.72%	58.54%	28.74%	
Actual Share	13.17%	57.42%	29.41%	
Error				2.228%

Table 5: Prediction using MMM with fixed variances.

	Predicted Choices			
Actual Choices	Train	SM	Car	Total
Train	140.0	459.8	105.2	705
SM	398.5	1976.0	700.5	3075
Car	136.8	709.2	729.0	1575
Total	675.3	3145	1534.7	5355
Predicted Share	12.61%	58.73%	28.66%	
Actual Share	13.17%	57.42%	29.41%	
Error				2.614%

Table 6: Prediction using MMM with different variances.

	Predicted Choices			
Actual Choices	Train	SM	Car	Total
Train	175.4	427.2	102.4	705
SM	414.2	1933.6	727.2	3075
Car	86.3	770.6	718.1	1575
Total	675.9	3131.4	1547.7	5355
Predicted Share	12.62%	58.48%	28.90%	
Actual Share	13.17%	57.42%	29.41%	
Error				2.114%

Lastly, we use the systematic utility levels to do prediction for each sample. Suppose, we assume that the person chooses the alternative with the largest systematic utility level and compare this choice with the actual choice made. Table 7 provides the number of errors obtained using the predictions for each method.

Table 7: Sample Prediction Using Different Estimation Methods.

Models	MNL	CNL	MMM	MMM
			(Fixed Variance)	(Different Variance)
No. of Samples	5355	5355	5355	5355
No. of Errors	1858	1844	1826	1822

It is clear that once again the MMM outperforms the other two methods. Even when the variances for all alternatives are fixed, the performance of MMM is good. The results thus seem to suggest that it is possible to predict choice probabilities in discrete choice models under much weaker assumptions using the persistency results in this paper.

4.2 Discrete Choice Under Budget Constraint

Consider the discrete choice problem, where a set of items $\mathcal{N} = \{1, 2, \ldots, n\}$ is considered. Let $\tilde{c}_i$ denote the utility value accrued for each unit of item i, and a_i ($a_i > 0$) denote the cost of any item $i \in \mathcal{N}$. Assume also that the budget available to the consumer is b. Let x_i be the units of item i that the consumer needs to choose to maximize her utility values. This problem can be formulated as the following integer knapsack problem:

$$\begin{aligned} \max \quad & \sum_{i \in \mathcal{N}} \tilde{c}_i x_i \\ \text{s.t.} \quad & \sum_{i \in \mathcal{N}} a_i x_i \leq b \\ & x_i \in \{0, 1, ..., \lfloor b/a_i \rfloor\} \quad \forall i \in \mathcal{N}. \end{aligned} \tag{13}$$

In this section, we focus on the stochastic integer knapsack problem where the utility value $\tilde{c}_i$ is uncertain and investigate the persistency of each variable x_i under mean μ_i, variance σ_i^2 and range information $[\underline{c}_i, \overline{c}_i]$.

The persistence values under the marginal moment model is obtained by solving

$$\begin{aligned} Z^* = \max \quad & \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{X}_i} k w_{ik} \\ \text{s.t.} \quad & \sum_{k\in\mathcal{X}_i} z_{ik} = \mu_i^2 + \sigma_i^2 && \forall i \in \mathcal{N} \\ & \sum_{k\in\mathcal{X}_i} w_{ik} = \mu_i && \forall i \in \mathcal{N} \\ & \sum_{k\in\mathcal{X}_i} y_{ik} = 1 && \forall i \in \mathcal{N} \\ & z_{ik} y_{ik} \geq w_{ik}^2 && \forall k \in \mathcal{X}_i \ \forall i \in \mathcal{N} \\ & (\underline{c}_i + \overline{c}_i) w_{ik} \geq \underline{c}_i \overline{c}_i y_{ik} + z_{ik} && \forall k \in \mathcal{X}_i \ \forall i \in \mathcal{N} \\ & \boldsymbol{y} \in \mathcal{CH}(\mathcal{Y}). \end{aligned} \tag{14}$$

The region $\mathcal{Y}$ is given as:

$$\mathcal{Y} \quad = \quad \left\{ \boldsymbol{y} \;\middle|\; \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{X}_i} a_i k y_{ik} \leq b, \ \sum_{k\in\mathcal{X}_i} y_{ik} = 1 \ \forall i \in \mathcal{N}, \ y_{ik} \in \{0,1\} \ \forall k \in \mathcal{X}_i \ \forall i \in \mathcal{N} \right\}.$$

The difficulty in solving (14) lies in the last constraint which characterizes the convex hull of the set $\mathcal{Y}$. Therefore, we start with the linear programming relaxation of the constraint, that is

$$\boldsymbol{y} \in \mathcal{LP}(\mathcal{Y}) = \left\{ \boldsymbol{y} \;\middle|\; \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{X}_i} a_i k y_{ik} \leq b, \ \sum_{k\in\mathcal{X}_i} y_{ik} = 1 \ \forall i \in \mathcal{N}, \ y_{ik} \in [0,1] \ \forall k \in \mathcal{X}_i \ \forall i \in \mathcal{N} \right\}. \tag{15}$$

Applying the cutting plane approach, we can gradually cut off the solutions violating the feasibility constraints in $CH(\mathcal{Y})$ before getting the optimal solution.

Example: To illustrate the process, we use the Steinberg-Parks example problem [14] with the size of the knapsack to be 30. Table 8 shows the weight, mean and variance information. While the original problem assumes $\tilde{c}_i$ to be normally distributed, we set the support for $\tilde{c}_i$ to be $[\mu_i - 3\sigma_i, \mu_i + 3\sigma_i]$ so that the probability for a normally generated random variable to be in the support is 99.7%.

Table 8: The Steinberg-Parks numerical example.

i	1	2	3	4	5	6	7	8	9	10
a_i	5	7	11	9	8	4	12	10	3	6
μ_i	7	12	14	13	12	5	16	11	4	7
σ_i^2	15	20	15	10	8	20	8	15	20	25

Firstly, we solved (14) with the last constraint replaced with the linear programming relaxation. This problem is solved as a second order conic program[4]. The computational time was 1.051 seconds, and the approximation for the persistence values are displayed in Table 9.

Table 9: Approximation for the persistence values from the second order conic program.

	$P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k)$									
k	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0	0.8077	0.8437	1.0000	1.0000	0.9621	0.6923	1.0000	1.0000	0.6429	1.0000
1	0.0000	0.0000	0.0000	0.0000	0.0186	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0108	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0000		0.0000	0.0085	0.0000		0.0000	0.0000	0.0000
4	0.0000	0.1562				0.0000			0.0000	0.0000
5	0.0000					0.0000			0.0000	0.0000
6	0.1923					0.0000			0.0000	
7						0.3077			0.0000	
8									0.0000	
9									0.0000	
10									0.3571	

Exploiting the properties of the polytope $\mathcal{CH}(\mathcal{Y})$, we can add 255 more cuts to improve the solution. The computational time was 798.08 seconds and the persistence values are shown in Table 10.

Table 10: The persistence values for the Sternberg-Parks example under MMM.

	$P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k)$									
k	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0	0.8077	0.8437	1.0000	0.9213	0.9362	0.6923	1.0000	1.0000	0.6429	0.7159
1	0.0000	0.0000	0.0000	0.0787	0.0000	0.0000	0.0000	0.0000	0.0000	0.2841
2	0.0000	0.0772	0.0000	0.0000	0.0638	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	0.0787		0.0000	0.0000	0.0000		0.0000	0.0000	0.0000
4	0.0000	0.0003				0.0134			0.0000	0.0000
5	0.0000					0.0000			0.0000	0.0000
6	0.1923					0.2841			0.0000	
7						0.0102			0.0000	
8									0.0000	
9									0.0000	
10									0.3571	

[4]The experiments were run by a PC with a Intel Pentium 4 - M 1.72GHz CPU, 256 MB of RAM and Microsoft Windows XP Professional operation system. They are coded in MATLAB 6.5 and use SeDuMi 1.05R5 as the solver for convex optimization problems.

Some important observations can be made from this computational experiment:

- The approximation in Table 9 is very close to the persistence value in Table 10. The average difference is only 0.0316. Both models suggest that the variables x_3, x_4, x_5, x_7 and x_8 are very likely to be 0, and the variables x_1, x_2, x_6 and x_9 are more likely to take greater values, i.e., to be selected in the discrete choice problem.

- Although the whole cutting plane process took about 13 minutes, the largest conic program with all the cuts added was solved in 3.886 seconds. Therefore the bottleneck for the marginal moment model is to find an efficient way to generate the cuts.

To evaluate the persistence values obtained from the marginal moment model, we compared the results with simulation tests. Three distributions were used to generate the random coefficients of each item in the objective function: normal distribution (concentrated around the mean), uniform distribution (evenly spread out) and a two-point distribution (concentrated at extremes) with $\tilde{c}_i$ to be either $\mu_i - \sigma_i$ or $\mu_i + \sigma_i$. For each simulation, the $\tilde{c}_i$ were independently generated under the moment restrictions. A total of 10,000 cases were generated for each distribution. Figure 7 plots the persistence values obtained from MMM and the simulation results with corresponding distributions.

The horizontal axis shows the various values taken by the decision variables ($x_i = k$), whereas the vertical axis shows the persistence value $P_\theta(x_i^*(\tilde{\boldsymbol{c}}) = k)$. Eleven of the persistence values equal to 0 in both the MMM and the simulation, hence are not included in the figure. Generally, the persistency of the MMM agrees with the simulation. The average of the differences between the persistency from MMM and the simulations is merely 0.0298, and the MMM accurately predicts that x_3, x_4, x_5, x_7 and x_8 are very likely to be 0 while the rest of the variables usually take large values. Compared with the simulation results with any one of the distributions, 26 persistence values computed by the MMM have a difference less than 0.05, and there are only 8 persistence values with difference greater than 0.1. Even for these 8 values, the discrepancy is mainly caused by the difference in the simulation results using various distributions. Based on these results, we conclude that the MMM gives valuable insights for the stochastic knapsack problem.

Another interesting observation is that x_5 has much higher probability to be 0 than x_6 in both the MMM persistence values and the simulation results. However, if we consider the value/weight ratio using the expectation of $\tilde{c}_i$, the value/weight ratio of x_5 and x_6 are 1.50 and 1.25 respectively. Therefore, a naive decision maker would assign higher choice probability to x_5 than to x_6. It is indeed interesting that our MMM model, assuming only minimal conditions on the random parameters, is able to pick this anomaly out.

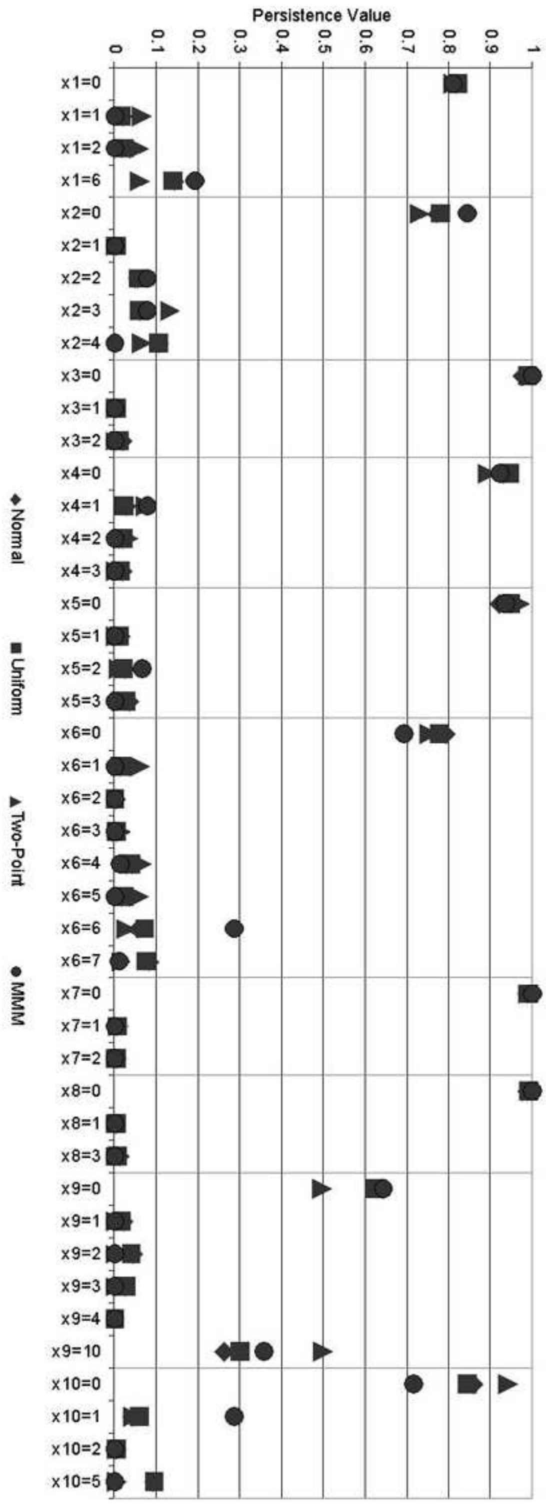

Figure 7: The persistence values of the MMM and the simulation results.

5 Conclusion

In this paper, we studied the problem of evaluating the probability distribution for the decision variable x_i of a general stochastic discrete optimization problem. Under a set of known marginal distributions or marginal moments for the objective coefficients, we propose a concave maximization technique to construct approximate solutions to the probability distribution of the optimal solution values. We show that the problem can be solved efficiently if we can characterize the convex hull of a projection of the original polytope to a higher dimension 0-1 polytope. For simpler problems, the above methodology can be implemented in an efficient manner. For instance, with only the first two moments, the persistency model for a 0-1 problem reduces to a simple concave maximization problem. For more general discrete optimization problem, we can often use the valid inequalities in the binary reformulation to generate cuts for the original problem. The computational results on discrete choice modeling problems with and without budget constraints provide encouraging evidence that the models can be used effectively to study discrete optimization problems under data uncertainty.

We believe that the approach opens up new research avenues for using structural properties of convex optimization models to study general discrete optimization problem under data uncertainty. In particular, is there a way to efficiently compute $CH(\mathcal{Y})$ for special classes of discrete optimization problems? The connection between the persistency model and the discrete choice model also opens up the possibility that simple convex optimization model can be used to calibrate empirical data obtained in more complicated choice decision problems. We leave these issues to future research.

Acknowledgment
We would like to thank Prof. Ben-Akiva for sharing with us the data on the transport choice model studied in Bierlaire et al. [6].

References

[1] Ben-Akiva, M., M. Bierlaire. 1999. *Discrete Choice Methods and their Applications to Short-Term Travel Decisions.* R. Hall (ed.), Handbook fo Transportation Science, Kluwer, 5-34.

[2] Ben-Akiva, M., S. Leerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, Massachusetts.

[3] Bertsimas, D., K. Natarajan and Chung Piaw Teo. 2004. Probabilistic combinatorial optimization: Moments, Semidefinite Programming and Asymptotic Bounds. *SIAM Journal of Optimization* **15**, No. 1, 185-209.

[4] Bertsimas, D., K. Natarajan and Chung Piaw Teo. 2006. Persistency in Discrete Optimization under Data Uncertainty. *Mathematical Programming* **108**, No. 2, 251-274.

[5] Bierlaire, M. 2003. BIOGEME: A Free Package For the Estimation of Discrete Choice Models. *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

[6] Bierlaire, M., K. Axhausen and G. Abay. 2001. The acceptance of modal innovation: The case of swissmetro. *Proceedings of the 1st Swiss Transportation Research Conference*, Ascona, Switzerland.

[7] Lovsz, L. 2001. Energy of convex sets, shortest paths, and resistance. *Journal of Combinatorial Theory A* **94**, 363-382.

[8] Lyon, R., R. Pemantle and Y Peres. 1999. Resistance bounds for first-passage percolation and maximum flow. *Journal of Combinatorial Theory A* **86**, 158-168.

[9] McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, ed., Frontiers in Econmetrics, Academic Press, New York, 105-142.

[10] Meilijson, I. and A. Nadas. 1979. Convex majorization with an application to the length of critical path. *Journal of Applied Probability* **16**, 671-677.

[11] Nadas, A. 1979. Probabilistic PERT. *IBM Journal of Research and Development* **23**, 339-347.

[12] Sherali, H. D., W. P. Adams. 1999. *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems.* Kluwer Academic Publishers.

[13] Steele, J. M. 1997. Probability Theory and Combinatorial Optimization. CBMS-NSF Regional Conference Series in Applied Mathematics **69**.

[14] Steinberg, E., M. S. Parks. 1979. A preference order dynamic program for a knapsack problem with stochastic rewards. *Journal of the Operational Resarch Society* **30**, 141-147.

[15] Weiss, G. 1986. Stochastic bounds on distributions of optimal value functions with applications to PERT, network flows and reliability. *Operations Research* **34**, 595-605.

[16] Vanderbei, R. 2006. LOQO User's Manual - Version 4.5.