7-2015

Least squares approximation to stochastic optimization problems

Zhichao ZHENG
Singapore Management University, DANIELZHENG@smu.edu.sg

Karthik NATARAJAN
Singapore University of Technology and Design

Chung-Piaw TEO
National University of Singapore

Least Squares Approximation to Stochastic Optimization Problems

Zhichao Zheng* Karthik Natarajan† Chung-Piaw Teo‡

July 14, 2013

Abstract

This paper is motivated by the following question: How to construct good approximation for the distribution of the solution value to linear optimization problem when the objective function is random? More generally, we consider any mixed zero-one linear optimization problem, and develop an approach to approximate the distribution of its optimal value when the random objective coefficients follow a multivariate normal distribution. Linking our model to the classical Stein's Identity, we show that the least squares normal approximation of the random optimal value can be computed by solving the persistency problem, first introduced by Bertsimas et al. (2006). We further extend our method to construct a least squares quadratic estimator to improve the accuracy of the approximation, in particular, to capture the skewness of the objective. We use this approach to construct good estimators for (a) the fill rate of an inventory system in a finite horizon; (b) the waiting time distribution of the nth customer in a G/G/1 system when the arrival rate equals the service rate; and (c) the project completion time distribution.

Key words: distribution approximation; persistency; Stein's Identity; project management; fill rate; G/G/1 queue; transient solution

1 Introduction

Many problems in the emerging area of predictive analytics can be cast as a likelihood inferencing problem for a stochastic system, where the focus is on finding a way to predict the

*Lee Kong Chian School of Business, Singapore Management University, Singapore. Email: danielzheng@smu.edu.sg

†Department of Engineering Systems and Design, Singapore University of Technology and Design, Singapore. Email: natarajan_karthik@sutd.edu.sg

‡Department of Decision Sciences, NUS Business School, National University of Singapore, Singapore. Email: bizteocp@nus.edu.sg

outcome/behaviour of a related stochastic optimization problem, given additional side information on a set of predictor variables. Take for instance the customer choice model prevalent in many revenue optimization problem: for given product attributes and choice bundle, the consumer optimizes a utility function to obtain the choice decision, and the challenge in this predictive analytical task is to estimate the choice probability of a random customer. Another well-known example is the queue inference engine (QIE) problem (cf. Larson (1990)): given the service durations of customers in a busy period (e.g., captured in the ATM machine transaction data), one would like to predict the queueing behaviour in the system during this busy period. These problems have received a considerable amount of attention in literature.

The project completion time prediction is an early example of predictive analytics in the OR literature. Given a set of activities, precedence relationship and (random) duration of each activity, we would like to estimate the time it takes to complete all the activities in the project. This problem is often represented by a directed acyclic graph (DAG). In this paper, we adopt the conventional activity-on-arc representation of the project network, where arcs represent activities and nodes represent the milestones that indicate the starting or ending of the activities. The length of an arc is the duration of the activity represented by that arc. The project completion time is simply the longest path in this network. If all the activities have deterministic durations, finding the project completion time is as easy as solving a linear programming (LP) problem[1]. However, when the activity durations are stochastic, the analysis of the random project completion time becomes nontrivial.

It has long been the interest of both researchers and practitioners to estimate the distribution of the project completion time. Over the past few decades, various methods have been proposed to approximate this distribution (cf. Dodin (1985), Cox (1995), etc.). Unfortunately, to the best of our knowledge, most of the existing approaches are derived using ad hoc heuristics or work on specific problem instances. In this paper, we partially address this issue under the assumption that the activity durations follow a multivariate normal distribution, and construct a normal distribution approximation for the random project completion time that is optimal under the L^2-norm. In fact, our method applies to any general random mixed 0-1 LP problem under objective uncertainty:

$$Z\left(\tilde{\boldsymbol{c}}\right) := \max_{\boldsymbol{x}\in\mathcal{P}} \sum_{j=1}^{n} \tilde{c}_j x_j, \tag{1}$$

where $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \ldots, \tilde{c}_n)^T$ is the random coefficient vector following a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, denoted as $\tilde{\boldsymbol{c}} \sim N(\boldsymbol{\mu}, \Sigma)$, and $\mathcal{P}$ is the

[1]In fact, for the deterministic case, this problem can be solved in a more efficient way, which is dynamic programming, in effort proportional to the number of arcs in the DAG.

domain of the feasible solutions (assumed to be bounded) defined by

$$\mathcal{P} := \{\boldsymbol{x} \in \mathbb{R}^n \ : \ \boldsymbol{a}_i^T \boldsymbol{x} = b_i, \forall i = 1, \ldots, m, \ x_j \in \{0,1\}, \forall j \in \mathcal{B} \subseteq \{1, \ldots, n\}, \ \boldsymbol{x} \geq \boldsymbol{0}\}.$$

In the project management problem, $\mathcal{P}$ characterizes the incidence vector of paths in the project network, and $\tilde{c}_j$ is the random duration of activity j. We assume that $\mathcal{P}$ is nonempty and bounded so that $\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right]$ is finite. Throughout this paper, we use bold face letters to denote column vectors. We use $\sigma_{j,k}$, $j,k = 1, \ldots, n$, to denote the covariance between $\tilde{c}_j$ and $\tilde{c}_k$, i.e., $(j,k)-$term of the covariance matrix Σ. We also use σ_j^2, $j = 1, \ldots, n$, to denote the variance of $\tilde{c}_j$, i.e., the jth diagonal term of Σ.

There is by now a huge literature on finding the distribution of $Z\left(\tilde{\boldsymbol{c}}\right)$ for various combinatorial optimization problems, including minimum assignment, spanning tree, and traveling salesman problem (cf. Aldous & Steele (2003)). These problems are notoriously hard, and often only partial results (e.g., asymptotic results with independent and identically distributed (i.i.d.) random variables) are known. Finding the exact distribution for the general mixed 0-1 LP problem appears to be almost impossible.

In this paper, we develop a generic approach to construct good approximation to the distribution of $Z\left(\tilde{\boldsymbol{c}}\right)$. Our approach has interesting applications in various other domains:

- In classical inventory theory, the fill rate of an order-up-to inventory system in K periods is given by
$$\frac{\min\left\{\tilde{D}_1, Q\right\} + \ldots + \min\left\{\tilde{D}_K, Q\right\}}{\tilde{D}_1 + \ldots + \tilde{D}_K}.$$
where Q is the order-up-to level, and $\tilde{D}_i$ is the random demand in period i. Suppose that $\tilde{D}_i$'s are i.i.d. Finding the distribution, or even the expectation of the fill rate performance, is a challenging problem. The fill rate is often approximated by $\mathbf{E}[\min\{\tilde{D}_1, Q\}]/\mathbf{E}[\tilde{D}_1]$, but it is well-known that this is a weak lower bound, especially if the number of periods, K, is small. We construct a new estimator for the fill rate in this paper, using our least squares linear estimator for the random function $\min\{\tilde{D}_i, Q\}$.

- In a single server queue when the arrival rate equals the service rate, it is well-known that the system is not stable. However, in many service system such as a clinic, the number of customers served is usually moderate in each day. In these systems, the waiting time of the nth customer may be an important performance measure in the system. In this paper, we develop an approach to approximate such waiting time distribution, utilizing classical results in random walks and results from Spitzer (1956) and the arcsine law.

- For the project management problem, under the Critical Path Method (CPM), which is commonly used in practice, the random project completion time is estimated by replacing $\tilde{c}_j$ with its expected value μ_j, i.e., $Z(\boldsymbol{\mu})$ is used to approximate the project completion time. In the classical Program Evaluation and Review Technique (PERT), this is taken one step further where the distribution of the project completion time is approximated by $\sum_{j=1}^{n} \beta_j \tilde{c}_j = \sum_{j=1}^{n} \beta_j(\tilde{c}_j - \mu_j) + Z(\boldsymbol{\mu})$, with

$$\beta_j = \begin{cases} 1, \text{ if arc } j \text{ is on the longest path when solving Z}(\boldsymbol{\mu}), \\ 0, \text{ otherwise.} \end{cases}$$

  To simplify the exposition, here we impose the conventional assumption that there is a unique optimal solution when we compute $Z(\boldsymbol{\mu})$. Due to the simplicity of the approach, PERT has gained a lot of popularity, and the random project networks are sometimes also called PERT networks. However, simply using the distribution of one critical path to approximate the distribution of the project completion time suffers from severe estimation errors. In particular, PERT has been widely criticized for significant underestimation of the mean project completion time and overestimation of the variability of the project completion time. We improve on these deficiencies of the PERT method in this paper.

The above leads us to a natural estimation problem:

$$\text{(P)} \qquad \min_{\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n} \ \mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha - \sum_{j=1}^{n} \beta_j(\tilde{c}_j - \mu_j)\right)^2\right],$$

where the challenge is to solve for the *least squares normal approximation* (or the best normal approximation in L^2-norm) to the random optimal objective value, as an affine function of the individual normally distributed random coefficients. We also refer to this as the *least squares linear estimator*. In this paper, we use these two terms interchangeably. We explicitly obtain the solution to this optimization problem, and link it to the *persistency* problem.

Bertsimas et al. (2006) introduced the notion of the persistency of a binary decision variable in Problem (1) as the probability that the variable is active (i.e., takes value of 1) in an optimal solution to Problem (1). We generalize this concept to include continuous variables as follows:

Definition 1 *The persistency of the decision variable x_j in Problem (1) is defined as $\mathbf{E}[x_j(\tilde{\boldsymbol{c}})]$, where $x_j(\tilde{\boldsymbol{c}})$ denotes an optimal value of x_j as a function of the random vector $\tilde{\boldsymbol{c}}$. If x_j is a binary variable, $\mathbf{E}[x_j(\tilde{\boldsymbol{c}})] = \mathbf{P}(x_j(\tilde{\boldsymbol{c}}) = 1)$.*

Remark 1 *When $\tilde{\boldsymbol{c}}$ is continuous and spans the whole space of $\mathbb{R}^n$, the support of $\tilde{\boldsymbol{c}}$ over which Problem (1) has multiple optimal solutions has measure zero and $\boldsymbol{x}(\tilde{\boldsymbol{c}})$ is unique almost surely*[2]*. In other situations, if there exist multiple optimal solutions over a support of strictly positive measure, $\boldsymbol{x}(\tilde{\boldsymbol{c}})$ is defined to be an optimal solution randomly selected from the set of optimal solutions at $\tilde{\boldsymbol{c}}$.*

In this paper, we assume that $\tilde{\boldsymbol{c}}$ is non-degenerate, i.e., the covariance matrix Σ is symmetric positive definite (denoted as $\Sigma \succ 0$). Together with the normality assumption, we are sure that $\boldsymbol{x}(\tilde{\boldsymbol{c}})$ is unique almost surely. The notion of persistency generalizes "criticality index" in project networks and "choice probability" in discrete choice models (cf. Bertsimas et al. (2006), Natarajan et al. (2009), Mishra et al. (2012)). By persistency problem, we refer to the problem of estimating the persistency values.

One critical drawback of the estimated distribution from solving Problem (P) is that it is restricted to be normal, which is symmetric about the mean. However, in most circumstances, $Z(\tilde{\boldsymbol{c}})$ is skewed. PERT also suffers from a similar issue. To strengthen the approximation, we propose to extend the estimator to include higher order terms on $\tilde{\boldsymbol{c}}$. In particular, we also find a quadratic estimator, $Q(\tilde{\boldsymbol{c}})$, to the distribution of $Z(\tilde{\boldsymbol{c}})$ of the following form:

$$Q(\tilde{\boldsymbol{c}}) = \alpha + \sum_{j=1}^{n} \beta_j(\tilde{c}_j - \mu_j) + \sum_{j_1=1}^{n} \sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2}),$$

where α, β_j and γ_{j_1,j_2} are adjustable parameters. Interestingly, the *least squares quadratic estimator* is also closely related to the persistency problem, and shares some common components with the least squares linear estimator.

Through this paper, we use the project management problem as the main example to illustrate the our results and related concepts with extensive numerical analysis. Applications in other areas are presented separately with relatively less numerical analysis.

Outline of this paper: In the next section, we review the related literature. In Section 3, we build our least squares linear approximation with applications in fill rate and queue performance estimation. The extension to least squares quadratic estimation is developed in Section 4 with an illustration using project management problems. In Section 5, we briefly review some ways for persistency estimation, and present the results from our computational studies. Finally, we provide some concluding remarks and future research directions in Section 6.

[2]Note that the feasible region of Problem (1) is a bounded polytope, so it has multiple optimal solutions only when $\tilde{\boldsymbol{c}}$ realized to be a normal vector of a facet of the polytope. Since the number of facets is finite for a give polytope, the probability measure over all the normal vectors is zero. For example, consider a polytope in $\mathbb{R}^2$, for any polytope, its normal vectors are just lines in $\mathbb{R}^2$. If $\tilde{\boldsymbol{c}}$ is continuous and spans the whole space of $\mathbb{R}^2$, the probability measure over all these lines is zero, since the number of these lines is finite.

2 Literature Review

Our problem of interest has a long history, and it is related to the classical "distribution problem of stochastic linear programming" literature (cf. Ewbank et al. (1974), Prekopa (1966) and the references therein). The distribution of the optimal value is often approximated by numerical methods such as the Cartesian integration method (cf. Bereanu (1963)). These methods have been studied under the general framework when the uncertain parameters may appear in the objective, constraint matrix, or the right hand side of the LP problem. However, the total number of random variables are very limited due to the numerical methods employed. In the case of project management, finding the distribution of completion time in a PERT network is still an active area of research with a rich literature (cf. Yao & Chu (2007) and the references therein). Most of the work in this area has been focused on using some graphical approaches to reduce the size of the graph and to reduce the complexity of estimating the distribution of the project completion time (e.g., Dodin (1985)). Another line of research tries to find a good normal approximation to the project completion time distribution using Central Limit Theorem and moment estimation methods (e.g., Cox (1995)). We solve this problem and show that the best normal approximation to the completion time distribution, under L^2-norm, can be obtained by solving the related persistency problem introduced by Bertsimas et al. (2006), and further studied in Natarajan et al. (2009).

Brown et al. (1997) brought up the issue of persistence and persistent modeling in optimization through a series of case studies. Although the idea of persistence conveyed in that paper is very broad and different from the persistency defined above, these two concepts are closely related through the issue of data uncertainty and robust optimization. The authors point out that from the perspective of persistence, robust optimization seeks a baseline solution that will persist as best as possible with a number of alternate forecast revisions. On the other hand, persistency describes the degree of persistence of each individual decision variable in an optimization problem with data uncertainty. Indeed, we can further generalize Definition 1 to the persistency of a feasible solution, i.e., the probability that this feasible solution is optimal. However, this is beyond the scope of the current paper.

Over the past few years, a substream of research in the field of persistency estimation has yielded a series of semidefinite programming (SDP) models based on the connection between the moment cone and the semidefinite cone. A common feature of these models is that they only assume the knowledge of moment information of the uncertainty rather than the exact form of the distribution. Hence, they are also referred as distributionally robust stochastic programming (DRSP) models.

Bertsimas et al. (2006) introduced arguably the first computational approach to approxi-

mate the persistency by solving a class of SDPs called Marginal Moment Model (MMM) under the assumption that the random vector $\tilde{\boldsymbol{c}}$ is described only through the marginal moments of each $\tilde{c}_j$ and all the decision variables in Problem (1) are binary. Natarajan et al. (2009) extended MMM to general mixed-integer LP problems, but their model formulation is based on the characterization of the convex hull of the binary reformulation which is typically difficult to derive. Lasserre (2010) studied the class of parametric polynomial optimization problems, which includes the mixed 0-1 linear programming problem as a special case. The author described the uncertainty using a combination of joint probability measure on the parameters and optimal solutions and marginal probability measures on the parameters. A hierarchy of semidefinite relaxations was proposed to solve the problem. However, the size of the semidefinite relaxation grows rapidly which makes solving the higher order semidefinite relaxations numerically challenging. Mishra et al. (2012) presented a SDP model named Cross Moment Model (CMM) for $\tilde{\boldsymbol{c}}$ described by both the marginal and cross moments. The formulation of CMM is based on the extreme point enumeration of Problem (1). Hence, the size of CMM becomes exponential for general LP problems. Inspired by a recent application of conic optimization on mixed 0-1 LP problems due to Burer (2009), Natarajan et al. (2011) developed a parsimonious but NP-hard conic optimization model to estimate the persistency of a general mixed 0-1 LP problem when $\tilde{\boldsymbol{c}}$ is described by both the marginal and cross moments. They referred to their model as Completely Positive Cross Moment Model (CPCMM). In this paper, we mainly exploit this model to estimate the persistency values. We will review it in more details in Section 5.

A recent paper by Agrawal et al. (2012) investigated the loss incurred by ignoring correlations in a DRSP model and proposed a new concept called price of correlations (POC). They showed that POC is bounded from above for a certain class of cost functions, suggesting that the intuitive approach of assuming independent distributions may actually work well for these problems. However, independence conditions can be extremely difficult to capture as well. One of the negative results is given by Hagstrom (1988), who showed that computing the expected value of the longest path in a directed acyclic graph is $\#\mathcal{P}$-complete when the arc lengths are restricted to taking two possible values and independent of each other. Perhaps a DRSP model with correlation conditions is more tractable. On the other hand, Agrawal et al. (2012) also show that for some cost functions, POC can be particularly large, indicating the need of DRSP models to capture correlations. Fortunately, CPCMM partially fills this gap, which in turns further strengthens our approximation method.

In the literature of project management, there is only limited sensitivity analysis with correlated activity times. For example, Banerjee & Paul (2008) showed that in the case of a project network with multivariate normal activity completion times and a covariance matrix characterized by only nonnegative terms, the completion times of activities are positively correlated. To

the best of our knowledge, none of the previous studies address the issues of correlated activities for the project management problem when approximating the distributions of the project completion times. Our research contributes to fill this gap by assuming a general non-degenerate multivariate normal distribution for the activity times when constructing the approximating distributions.

The key contributions of this paper are summarized next:

- We systematically study the distribution approximation problem under the least squares framework and take into account correlations among the random coefficients.
- Linking our problems to Stein's Identity, we explicitly derive the expressions of both the least squares linear and quadratic approximations.
- We provide a new perspective to the distribution approximation problem by transforming it into the related persistency estimation problem, for which there exist many well-established results to provide good estimates.
- In the context of project management problem, we show that knowing the criticality indices of arcs is the key to estimate the variability in the project completion time.
- By comparing against existing methods through extensive numerical studies, we demonstrate the superiority of bringing persistency into the distribution approximation problem.

3 Least Squares Linear Estimator for the Distribution

As discussed in the introduction, our main idea is to approximate the distribution of $Z(\tilde{\boldsymbol{c}})$ by a normal distribution, $W(\tilde{\boldsymbol{c}})$, with the following form:

$$W(\tilde{\boldsymbol{c}}) = \alpha + \sum_{j=1}^{n} \beta_j (\tilde{c}_j - \mu_j), \tag{2}$$

where α and β_j's are adjustable parameters. Note that the linear estimator in Equation (2) is also a normal distribution. The objective is to choose α and β_j's such that the expected squared deviation between $W(\tilde{\boldsymbol{c}})$ and $Z(\tilde{\boldsymbol{c}})$ is minimized. In particular, we aim to solve:

$$\text{(P)} \quad \min_{\alpha \in \mathbb{R}, \boldsymbol{\beta} \in \mathbb{R}^n} \mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha - \sum_{j=1}^{n} \beta_j (\tilde{c}_j - \mu_j)\right)^2\right],$$

i.e., we want to find the least squares normal approximation to the distribution of $Z(\tilde{\boldsymbol{c}})$. It turns out that the solution to Problem (P) under the normality assumption of $\tilde{\boldsymbol{c}}$ is related to

the concept of persistency in a straightforward manner as shown in the following theorem.

Theorem 1 *When $\tilde{\boldsymbol{c}} \sim N(\boldsymbol{\mu}, \Sigma)$ and $\Sigma \succ 0$, the unique solution to Problem (P) is*

$$\alpha^* = \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right], \quad \beta_k^* = \mathbf{E}\left[x_k(\tilde{\boldsymbol{c}})\right], k = 1, \ldots, n.$$

The proof of Theorem 1 utilizes the following classical covariance identity due to Stein, and its proof is enclosed in Appendix A for completeness.

Lemma 1 *[Stein's Identity] Let the random vector $\tilde{\boldsymbol{c}} = (\tilde{c}_1, \ldots, \tilde{c}_n)^T$ be multivariate normally distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. For any function $h(c_1, \ldots, c_n) : \mathbb{R}^n \to \mathbb{R}$ such that $\partial h(c_1, \ldots, c_n)/\partial c_j$ exists almost everywhere and $\mathbf{E}[|\partial h(\tilde{\boldsymbol{c}})/\partial c_j|] < \infty$, $\forall j = 1, \ldots, n$, denote $\nabla h(\tilde{\boldsymbol{c}}) = (\partial h(\tilde{\boldsymbol{c}})/\partial c_1, \ldots, \partial h(\tilde{\boldsymbol{c}})/\partial c_n)^T$. Then $Cov(\tilde{\boldsymbol{c}}, h(\tilde{\boldsymbol{c}})) = \Sigma \mathbf{E}[\nabla h(\tilde{\boldsymbol{c}})]$. Specifically,*

$$Cov\left(\tilde{c}_k, h(\tilde{c}_1, \ldots, \tilde{c}_n)\right) = \sum_{j=1}^{n} Cov\left(\tilde{c}_k, \tilde{c}_j\right) \mathbf{E}\left[\frac{\partial}{\partial c_j} h(\tilde{c}_1, \ldots, \tilde{c}_n)\right], \forall k = 1, \ldots, n.$$

Proof of Theorem 1. It is obvious that Problem (P) is convex. Then the necessary and sufficient optimality conditions of Problem (P) are

$$\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right) - \alpha^* - \sum_{j=1}^{n} \beta_j^* \left(\tilde{c}_j - \mu_j\right)\right] = 0, \text{ and}$$

$$\mathbf{E}\left[\left(Z\left(\tilde{\boldsymbol{c}}\right) - \alpha^* - \sum_{j=1}^{n} \beta_j^* \left(\tilde{c}_j - \mu_j\right)\right)\left(\tilde{c}_k - \mu_k\right)\right] = 0, \forall k = 1, \ldots, n.$$

Hence, an optimal solution to (P), $(\alpha^*, \boldsymbol{\beta}^*)$ should satisfy

$$\alpha^* = \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right], \text{ and}$$

$$\mathbf{E}\left[\left(Z\left(\tilde{\boldsymbol{c}}\right) - \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right] - \sum_{j=1}^{n} \beta_j^* \left(\tilde{c}_j - \mu_j\right)\right)\left(\tilde{c}_k - \mu_k\right)\right] = 0, \forall k = 1, \ldots, n.$$

Rearranging the second set of conditions, we get

$$Cov\left(\tilde{c}_k, Z\left(\tilde{\boldsymbol{c}}\right)\right) = \sum_{j=1}^{n} \beta_j^* \sigma_{j,k}, \forall k = 1, \ldots, n. \tag{3}$$

The optimal objective value $Z(\tilde{\boldsymbol{c}})$ satisfies the conditions in Stein's identity since $\partial Z(\tilde{c})/\partial c_k = x_k(c)$ almost everywhere. By applying Stein's Identity on $\tilde{\boldsymbol{c}}$ and $Z(\tilde{\boldsymbol{c}})$, we have

$$Cov\left(\tilde{c}_k, Z(\tilde{\boldsymbol{c}})\right) = \sum_{j=1}^{n} \sigma_{j,k} \mathbf{E}\left[\frac{\partial Z(\tilde{\boldsymbol{c}})}{\partial c_j}\right], \forall k = 1, \ldots, n.$$

Observe that $\forall j = 1, \ldots, n$,

$$\begin{aligned}
\mathbf{E}\left[\frac{\partial Z(\tilde{\boldsymbol{c}})}{\partial c_j}\right] &= \mathbf{E}\left[\frac{\partial}{\partial c_j}\left(\sum_{k=1}^{n} \tilde{c}_k x_k(\tilde{\boldsymbol{c}})\right)\right] \\
&= \mathbf{E}\left[\sum_{k=1}^{n} \tilde{c}_k \frac{\partial x_k(\tilde{\boldsymbol{c}})}{\partial c_j} + x_j(\tilde{\boldsymbol{c}})\right] \\
&= \mathbf{E}\left[x_j(\tilde{\boldsymbol{c}})\right].
\end{aligned}$$

The last equality follows from our assumptions on $\tilde{\boldsymbol{c}}$, i.e., normal and non-degenerate, so that for all $j, k = 1, \ldots, n$, $\partial x_k(\boldsymbol{c})/\partial c_j$ exists almost everywhere and equals to zero whenever it exists[3]. Thus, we get $\beta_j^* = \mathbf{E}\left[x_j(\tilde{\boldsymbol{c}})\right], j = 1, \ldots, n$ as one solution to Equation (3), which is also unique since Σ is positive definite. Thus, the proof is complete. ■

With Theorem 1, the problem of finding the least squares normal approximation to the distribution of $Z(\tilde{\boldsymbol{c}})$ is transformed into computing the persistency in Problem (1) as well as estimating $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. From these results, we know that the mean of estimated distribution $W(\tilde{\boldsymbol{c}})$ is the same as the mean of $Z(\tilde{\boldsymbol{c}})$. However, the variance of $W(\tilde{\boldsymbol{c}})$ is governed by the persistency values, and it is not necessarily equal to the variance of $Z(\tilde{\boldsymbol{c}})$. Indeed, the variance of $W(\tilde{\boldsymbol{c}})$ is a lower bound of the variance of $Z(\tilde{\boldsymbol{c}})$, i.e.,

$$\begin{aligned}
Var\left(W\left(\tilde{\boldsymbol{c}}\right)\right) &= Var\left(\sum_{j=1}^{n} \mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right] \tilde{c}_j\right) \\
&= \left(\mathbf{E}\left[\boldsymbol{x}\left(\tilde{\boldsymbol{c}}\right)\right]\right)^T \Sigma \left(\mathbf{E}\left[\boldsymbol{x}\left(\tilde{\boldsymbol{c}}\right)\right]\right) \\
&\leq Var\left(\sum_{j=1}^{n} x_j\left(\tilde{\boldsymbol{c}}\right) \tilde{c}_j\right) \\
&= Var\left(Z\left(\tilde{\boldsymbol{c}}\right)\right).
\end{aligned}$$

[3]Note that $\partial x_k(\tilde{\boldsymbol{c}})/\partial c_j$ is not defined when there are multiple optimal solutions to Problem (1), but in other situations, $x_k(\tilde{\boldsymbol{c}})$ does not change with a small perturbation of c_j. Please refer to the footnote in Remark 1 for the detailed discussion on the probability measure over the set of $\tilde{\boldsymbol{c}}$ that leads to multiple optimal solutions. Precisely, we should write the derivation process in integral form, i.e., expressing all the expectations in integral form. Then it will be clear that $\partial x_k(\tilde{\boldsymbol{c}})/\partial c_j$ can only be integrated over the support of $\tilde{\boldsymbol{c}}$ where it is defined, and hence only zero values remain in the integration expression for $\mathbf{E}[\tilde{c}_k \partial x_k(\tilde{\boldsymbol{c}})/\partial c_j]$.

The inequality above is due to Cacoullos (1982), where equality holds if and only if $\mathbf{E}[x_j(\tilde{\boldsymbol{c}})]$ is constant for every $j = 1, \ldots, n$. Note that although Cacoullos' inequality,

$$Var\left(g\left(\tilde{\boldsymbol{c}}\right)\right) \geq \left(\mathbf{E}\left[\nabla g\left(\tilde{\boldsymbol{c}}\right)\right]\right)^T \Sigma \left(\mathbf{E}\left[\nabla g\left(\tilde{\boldsymbol{c}}\right)\right]\right),$$

holds for any absolutely continuous real-valued function $g\left(\tilde{\boldsymbol{c}}\right)$ with finite variance, we still need those properties of $Z\left(\tilde{\boldsymbol{c}}\right)$ and $\mathbf{E}\left[\boldsymbol{x}\left(\tilde{\boldsymbol{c}}\right)\right]$ as used in the proof of Theorem 1 to derive the above result. Though a lower bound, the variance of the least squares linear estimator is significantly closer to the true variance than those estimated from existing distribution approximation methods. We will illustrate this point using examples in Section 5.1.

Remark 2 *Empirically, instead of using the observed persistency values to estimate the values for $\boldsymbol{\beta}$, we can also use $Cov(\tilde{c}_j, Z(\tilde{\boldsymbol{c}}))/\sigma_j^2$ to estimate β_j when $\tilde{c}_j$'s are independent of each other (cf. Equation (3)). This is exactly the formula used in linear regression. One such example is estimating the beta coefficient of a risky asset under the capital asset pricing model (CAPM) in finance. This approach comes in handy when only $Z(\tilde{\boldsymbol{c}})$ is observed but not the optimal choices made, as is the case in linear regression.*

3.1 Application: Fill Rate

Consider a finite horizon order-up-to inventory system, where demand the $\tilde{D}_i$'s are i.i.d. with mean μ and standard deviation σ. Let Q denote the order up to level. We are interested to approximate the fill rate over K periods, given by

$$\frac{\min\left\{\tilde{D}_1, Q\right\} + \ldots + \min\left\{\tilde{D}_K, Q\right\}}{\tilde{D}_1 + \ldots + \tilde{D}_K}.$$

Chen et al. (2003) established that

$$\mathbf{E}\left[\frac{\min\left\{\tilde{D}_1, Q\right\}}{\tilde{D}_1}\right] \geq \mathbf{E}\left[\frac{\sum_{i=1}^K \min\left\{\tilde{D}_i, Q\right\}}{\sum_{i=1}^K \tilde{D}_i,}\right] \geq \frac{\mathbf{E}\left[\min\left\{\tilde{D}_1, Q\right\}\right]}{\mathbf{E}\left[\tilde{D}_1\right]}, \forall K = 1, 2, \ldots.$$

Thomas (2005) argued that the distribution of the fill rate measurement affects the stocking decision, and hence the choice of the planning horizon in the fill rate measurement is an important consideration in the design of the fill rate target. Define $\theta(Q) = \mathbf{P}(\tilde{D}_1 < Q)$. Assume $\tilde{D}_i$'s are normally distributed. Let $\mathcal{L}(x) = \mathbf{E}[\max(Z - x, 0)]$ denote the standardized normal

loss function, where Z is the standard normal random variable. Then

$$\begin{aligned}\frac{\sum_{i=1}^{K}\min\left\{\tilde{D}_i,Q\right\}}{\sum_{i=1}^{K}\tilde{D}_i} &\approx \frac{\sum_{i=1}^{K}\left\{\theta\left(Q\right)\tilde{D}_i-\theta\left(Q\right)\mu+Q+\mathbf{E}\left[\min\left\{\tilde{D}_i-Q,0\right\}\right]\right\}}{\sum_{i=1}^{K}\tilde{D}_i}\\ &= \frac{\sum_{i=1}^{K}\left\{\theta\left(Q\right)D_i+Q-\theta\left(Q\right)\mu-\sigma\mathbf{E}\left[\max\left\{Z-\frac{\mu-Q}{\sigma},0\right\}\right]\right\}}{\sum_{i=1}^{K}\tilde{D}_i}\\ &= \theta\left(Q\right)+\frac{K\left[Q-\theta\left(Q\right)\mu-\sigma\mathcal{L}\left(\frac{\mu-Q}{\sigma}\right)\right]}{\sum_{i=1}^{K}\tilde{D}_i}.\end{aligned}$$

Example 1 *Consider the case when the i.i.d. demand D_i is normally distributed with mean $\mu = 10$ and standard deviation $\sigma = 3$.*

The mean fill rates obtained from simulation and the linear estimator developed above are summarized in Table 1.

Q	$K=2$		$K=10$		$K=20$	
	Simulation	Estimation	Simulation	Estimation	Simulation	Estimation
6	0.6116	0.6137	0.5918	0.5919	0.5895	0.5895
7	0.6997	0.7025	0.6797	0.6798	0.6773	0.6774
8	0.7780	0.7814	0.7592	0.7593	0.7569	0.7570
9	0.8441	0.8479	0.8278	0.8279	0.8258	0.8258
10	0.8969	0.9006	0.8837	0.8838	0.8820	0.8821
11	0.9360	0.9393	0.9264	0.9264	0.9251	0.9251
12	0.9630	0.9657	0.9565	0.9566	0.9556	0.9556
13	0.9802	0.9821	0.9762	0.9762	0.9756	0.9756
14	0.9902	0.9915	0.9880	0.9880	0.9876	0.9876

Table 1: Comparison between simulated and estimated finite-horizon fill rates

The mean of the estimator obtained using the persistency approach is surprisingly close to the simulated fill rate performance, for all values of Q. Moreover, the effect of the review periods is more visible for small Q. For instance, when $Q = 6$, the mean fill rate is around 61% when $K = 2$, but drops to 59% when K is around 10 to 20.

Note that

$$\begin{aligned}\mathbf{E}\left[\frac{\sum_{i=1}^{K}\min\left\{\tilde{D}_i,Q\right\}}{\sum_{i=1}^{K}\tilde{D}_i}\right] &\approx \mathbf{E}\left[\theta\left(Q\right)+\frac{K\left(Q-\theta\left(Q\right)\mu-\sigma\mathcal{L}\left(\frac{\mu-Q}{\sigma}\right)\right)}{\sum_{i=1}^{K}\tilde{D}_i}\right]\\ &= \theta\left(Q\right)+\left[Q-\theta\left(Q\right)\mu-\sigma\mathcal{L}\left(\frac{\mu-Q}{\sigma}\right)\right]\mathbf{E}\left[\frac{K}{\sum_{i=1}^{K}\tilde{D}_i}\right].\end{aligned}$$

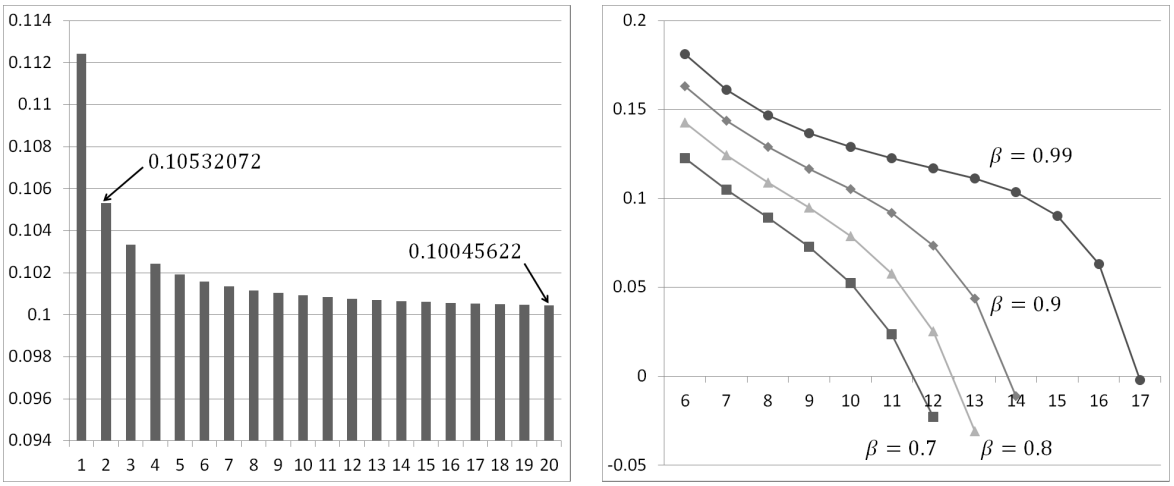
Hence to meet a target expected fill rate of β, the order up to level Q and the number of review periods K, approximately satisfy the relationship

$$\theta(Q) + \left[Q - \theta(Q)\mu - \sigma\mathcal{L}\left(\frac{\mu - Q}{\sigma}\right)\right]\mathbf{E}\left[\frac{K}{\sum_{i=1}^{K}\tilde{D}_i}\right] = \beta,$$

i.e.,

$$K\mathbf{E}\left[\frac{1}{\sum_{i=1}^{K}\tilde{D}_i}\right] = \frac{\beta - \theta(Q)}{Q - \theta(Q)\mu - \sigma\mathcal{L}\left(\frac{\mu-Q}{\sigma}\right)}.$$

For Example 1, we can numerically integrate the function on both sides of the equation. Figure 1 shows how the left-hand side (LHS) and right-hand side (RHS) vary as a function of K, Q and β, respectively. For $K = 2$ (with corresponding LHS value of 0.1053), and $\beta = 0.7$, the plots indicate that we only need Q to be around 7 to attain the fill rate of 70%. However, for the target of $\beta = 0.99$, we need Q to be around 14.



(a) (b)

Figure 1: LHS as a function of K (a) versus RHS as a function of Q and β (b)

3.2 Application: Maximum Partial Sum and G/G/1 Queue

Here we discuss another application of our result on an important statistical problem: approximating the distribution of the maximum partial sum of normal random variables. The problem is critical in many areas of application, including hydrology and testing for a change-point (cf. Hurst (1951), James et al. (1987), Conniffe & Spencer (2000)). It also arises in the transient analysis of single server queue. Combining our result with some classical results in probability theory, we present a closed-form expression for the least squares normal approximation of the

maximum partial sum of normal random variables, from which many interesting applications follow.

Suppose that $\tilde{c}_j$'s ($j = 1, \ldots, n$) are i.i.d. normal random variables with zero mean and finite standard deviation σ. Let $S_0 = 0$ and

$$\tilde{S}_k = \tilde{c}_1 + \cdots + \tilde{c}_k,\ k = 1, \ldots, n.$$

The problem is to estimate the distribution of $\tilde{S}_{max} := \max_{k \in \{0,\ldots,n\}} \tilde{S}_k$, i.e., the maximum partial sum of $\tilde{c}_j$'s (or the maximum value of the random walk from $\tilde{c}_j$'s). Note that

$$\begin{aligned} S_{max} = \max_{\sum_{k=0}^{n} y_k = 1, \boldsymbol{y} \geq 0} \sum_{k=0}^{n} S_k y_k &= \max_{\sum_{k=0}^{n} y_k = 1, \boldsymbol{y} \geq 0} \sum_{k=1}^{n} \left(\sum_{j=1}^{k} \tilde{c}_j \right) y_k \qquad (4) \\ &= \max_{\sum_{k=0}^{n} y_k = 1, \boldsymbol{y} \geq 0} \sum_{j=1}^{n} \left(\sum_{k=j}^{n} y_k \right) \tilde{c}_j \qquad (5) \end{aligned}$$

Applying Theorem 1, we get the following expression of the least squares normal approximation to $\tilde{S}_{max}$:

$$\mathbf{E}\left[\tilde{S}_{max}\right] + \sum_{j=1}^{n} \left(\sum_{k=j}^{n} \mathbf{E}\left[y_k\left(\tilde{\boldsymbol{c}}\right)\right] \right) \tilde{c}_j, \qquad (6)$$

where $\mathbf{E}[y_k(\tilde{\boldsymbol{c}})]$ is the persistency in Problem (5), i.e., the probability that the partial sum attains its maximum value at step k. The classical finite arcsine law (cf. Andersen (1953)) states that this probability does not depend on the distribution of $\tilde{c}_j$ provided that $\tilde{c}_j$ is symmetric around the mean 0:

$$\mathbf{E}\left[y_k(\tilde{\boldsymbol{c}})\right] = \left(\begin{array}{c} 2k \\ k \end{array} \right) \left(\begin{array}{c} 2n-2k \\ n-k \end{array} \right) \frac{1}{2^{2n}},\ k = 1, \ldots, n.$$

Observe that the variance of our approximation in Equation (6) is solely determined by the second term through persistency. Hence, we get the following closed-form lower bound to the variance of the maximum partial sum:

$$\frac{\sigma^2}{2^{4n}} \sum_{j=1}^{n} \left[\sum_{k=j}^{n} \left(\begin{array}{c} 2k \\ k \end{array} \right) \left(\begin{array}{c} 2n-2k \\ n-k \end{array} \right) \right]^2.$$

The above result expands the current literature by providing a different way to estimate the variance of the maximum partial sum of i.i.d. normal random variables. Note that there exists various methods in literature to compute or estimate $\mathbf{E}[\tilde{S}_{max}]$, with which we get a complete characterization of the least squares normal approximation as shown in Equation (6).

For instance, when $\mu = 0$, i.e., $\tilde{c}_i$'s are independent normal random variables with zero means, Spitzer (1956) showed that

$$\mathbf{E}\left[\tilde{S}_{max}\right] = \sum_{k=1}^{n} \frac{1}{k} \mathbf{E}\left[\tilde{S}_k^+\right] = \mathbf{E}\left[\tilde{c}_1^+\right] \sum_{k=1}^{n} \frac{1}{\sqrt{k}},$$

where $\tilde{S}_k^+ = \max\{0, \tilde{S}_k\}$, and $\tilde{c}_1^+ = \max\{0, \tilde{c}_1\}$.

We use the results established above to develop an estimator for the waiting time distribution of the nth customer in a G/G/1 queue, where the arrival rate equals the service rate. Let $\tilde{c}_{n-i+1} = \tilde{T}_{i-1} - \tilde{A}_i$, where $\tilde{T}_{i-1}$ is the service time duration of the $(i-1)$th customer, and $\tilde{A}_i$ the inter-arrival time between the arrivals of the $(i-1)$th and ith customer. In this way, the waiting time of the nth customer in the $G/G/1$ queue is known to be given by the maximum partial sum $\tilde{S}_{max}$. We have shown that

$$\mathbf{E}\left[\tilde{S}_{max}\right] + \sum_{j=1}^{n} \left(\sum_{k=j}^{n} \mathbf{E}\left[y_k\left(\tilde{\boldsymbol{c}}\right)\right] \right) \tilde{c}_j$$

is the optimal least squares linear estimator for the waiting time distribution of the nth customer. Note that this estimation could return a negative value since $\tilde{c}_j$'s are normally distributed. It is obvious that such estimate is wrong. Thus, we modify the estimator a bit by truncating the negative estimates. We look at the numerical performance of our modified linear estimator using the following example.

Example 2 *Suppose that $\tilde{c}_i$ is normally distributed with mean 0 and standard deviation 2. Approximate the waiting time distribution of the 12th customer.*

For this example, Spitzer's Identity gives rise to a mean waiting time of 4.4771 units. Figure 2 shows the performance of the linear estimator (denoted by Predict(ALL) since we are using all the $\tilde{c}_i$ to predict the waiting time), vis-à-viz the waiting time obtained from simulation, across 100 sample paths.

Interestingly, the linear estimators Predict(ALL) performs consistently well across all sample paths. The average square deviation is around 2.7305 squared units.

In practice, we are also interested to evaluate the waiting time given additional information on the queueing system. For instance, in some applications, we want to evaluate the average waiting time distribution of all the customers, given the additional information that, say

$$\mathcal{S} := \left\{\tilde{c}_1 + \tilde{c}_2 \ldots + \tilde{c}_n > 0\right\}.$$

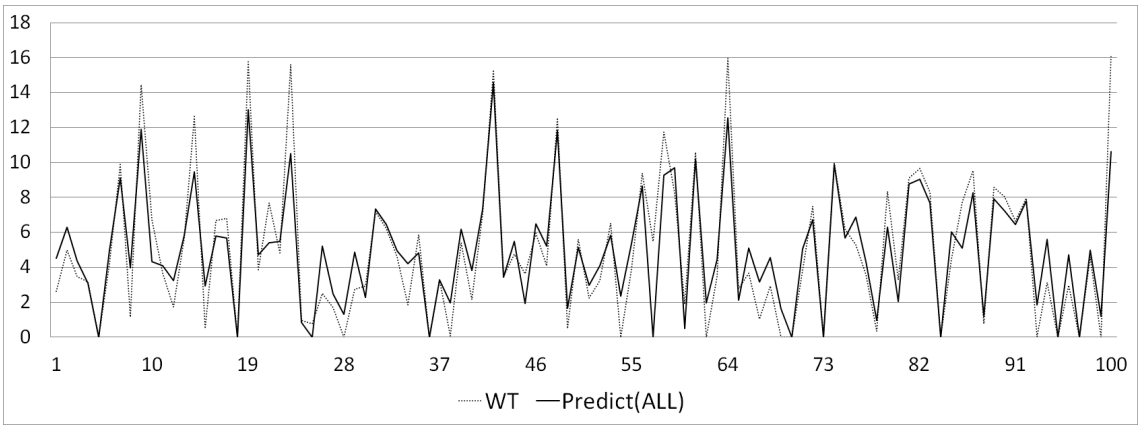


Figure 2: Estimations (Predict(ALL)) versus simulated (WT) waiting times of the 12th customer across 100 sample paths

This is the case in many overloaded systems, or during a busy period. Note that

$$\mathbf{E}\left[\tilde{c}_i \middle| \sum_{i=1}^{n} \tilde{c}_i > 0\right] = \frac{1}{n}\mathbf{E}\left[\sum_{i=1}^{n} \tilde{c}_i \middle| \sum_{i=1}^{n} \tilde{c}_i > 0\right] = \frac{\sigma}{n}\sqrt{\frac{2n}{\pi}}.$$

We use the linear estimator constructed above to approximate the expected waiting time of the nth customer, which yields

$$\begin{aligned}\mathbf{E}\left[\mathbf{E}\left[\tilde{S}_{max}\right] + \sum_{j=1}^{n}\left(\sum_{k=j}^{n}\mathbf{E}\left[y_k\left(\tilde{\boldsymbol{c}}\right)\right]\right)\tilde{c}_j \middle| \mathcal{S}\right] &= \mathbf{E}\left[\tilde{S}_{max}\right] + \sum_{j=1}^{n}\left(\sum_{k=j}^{n}\mathbf{E}\left[y_k\left(\tilde{\boldsymbol{c}}\right)\right]\right)\mathbf{E}\left[\tilde{c}_j \middle| \mathcal{S}\right] \\ &= \mathbf{E}\left[\tilde{S}_{max}\right] + \sum_{j=1}^{n}\left(\sum_{k=j}^{n}\mathbf{E}\left[y_k\left(\tilde{\boldsymbol{c}}\right)\right]\right)\frac{\sigma}{n}\sqrt{\frac{2n}{\pi}}.\end{aligned}$$

For Example 2, the above approximation gives rise to an estimate of 7.2410 units. This estimate is surprisingly accurate, as compared to the simulation result, 7.2419 units, from 10^6 sample paths.

4 Least Squares Quadratic Estimator for the Distribution

In the previous section, we show how to approximate the distribution of $Z(\tilde{\boldsymbol{c}})$ using a linear estimator $W(\tilde{\boldsymbol{c}})$. By "linear", we mean that $W(\tilde{\boldsymbol{c}})$ is linear in $\tilde{\boldsymbol{c}}$. As discussed in the introduction, to address the problem of skewness in $Z(\tilde{\boldsymbol{c}})$, we propose to extend our estimator to incorporate

higher order terms on $\tilde{\boldsymbol{c}}$. The estimator we consider is denoted as $Q(\tilde{\boldsymbol{c}})$ with the following form:

$$Q(\tilde{\boldsymbol{c}}) = \alpha + \sum_{j=1}^{n} \beta_j(\tilde{c}_j - \mu_j) + \sum_{j_1=1}^{n} \sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2}),$$

where α, β_j's and γ_{j_1,j_2}'s are adjustable parameters. Then the least squares quadratic estimation problem can be formulated as:

$$\text{(Q)} \qquad \min_{\alpha\in\mathbb{R},\boldsymbol{\beta}\in\mathbb{R}^n,\Gamma\in\mathbb{R}^{n\times n}} \mathbf{E}\left[\Big(Z(\tilde{\boldsymbol{c}}) - \alpha - \textstyle\sum_{j=1}^{n} \beta_j(\tilde{c}_j - \mu_j) - \sum_{j_1=1}^{n} \sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})\Big)^2\right],$$

where the matrix Γ is defined in such a way that makes our notation compact, $\Gamma_{j_1,j_2} \doteq (1/2)\gamma_{j_1,j_2}$, for $1 \le j_1 < j_2 \le n$, $\Gamma_{j_1,j_2} \doteq (1/2)\gamma_{j_2,j_1}$, for $1 \le j_2 < j_1 \le n$, and $\Gamma_{j_1,j_2} \doteq \gamma_{j_1,j_2}$, for $j_1 = j_2 = 1, \ldots, n$.

Following a similar approach as in Section 3, we can also derive the solution to Problem (Q). Interestingly, adding the quadratic term does not affect the solution of $\boldsymbol{\beta}$, which are still the persistency values, as presented in the following theorem. Notation-wise, we use "$\bullet$" to denote the inner product of two matrices.

Theorem 2 *When $\tilde{\boldsymbol{c}} \sim N(\boldsymbol{\mu}, \Sigma)$, a solution $(\alpha^*, \boldsymbol{\beta}^*, \Gamma^*)$ to Problem (Q) can be characterized as follows:*

$$\alpha^* = \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] - \Sigma \bullet \Gamma^*, \quad \beta_k^* = \mathbf{E}\left[x_k(\tilde{\boldsymbol{c}})\right], \, k = 1, \ldots, n,$$

and Γ^* is symmetric and satisfies the following system of $(n^2+n)/2$ linear equations:

$$\sum_{j_1=1}^{n} \sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}^* \left(\sigma_{j_1,k_1}\sigma_{j_2,k_2} + \sigma_{j_1,k_2}\sigma_{k_1,j_2}\right) = \sum_{j=1}^{n} \left(\mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] - \mu_{k_1}\mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right]\right)\sigma_{j,k_2}, \; \forall 1 \le k_1 \le k_2 \le n.$$

The proof of Theorem 2 is similar to that of Theorem 1. Hence, we omit it here but refer the readers to Appendix B for the details.

From Theorem 2, the problem of finding the least squares quadratic estimator for the distribution of $Z(\tilde{\boldsymbol{c}})$ is again transformed into a persistency problem, i.e., estimating $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$, $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$, and $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. The additional requirement to estimate $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$, i.e., the interaction between random coefficients and the optimal solution, can be interpreted as the increased difficulty of adding the quadratic terms in the estimation. However, we shall see in Section 5.1 that $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$ can be obtained as a by-product when we estimate the persistency using semidefinite programming methods.

In general, Γ^* may not be unique due to the correlation structures. However, when $\tilde{c}_j$'s are uncorrelated and not degenerate, we do have a simple and unique solution.

Corollary 1 *When $\tilde{c}_j$'s are uncorrelated and each follows a normal distribution with $\sigma_j^2 > 0$, there is a unique solution to Problem (Q) as follows:*

$$\begin{aligned}
\alpha^* &= \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] - \Sigma \bullet \Gamma^*, \\
\beta_k^* &= \mathbf{E}\left[x_k(\tilde{\boldsymbol{c}})\right], \, k = 1, \ldots, n, \\
\gamma_{k_1,k_2}^* &= \frac{\mathbf{E}\left[\tilde{c}_{k_1} x_{k_2}\left(\tilde{\boldsymbol{c}}\right)\right] - \mu_{k_1}\mathbf{E}\left[x_{k_2}\left(\tilde{\boldsymbol{c}}\right)\right]}{\sigma_{k_1}^2}, \, \forall 1 \leq k_1 < k_2 \leq n, \\
\gamma_{k,k}^* &= \frac{\mathbf{E}\left[\tilde{c}_k x_k\left(\tilde{\boldsymbol{c}}\right)\right] - \mu_k\mathbf{E}\left[x_k\left(\tilde{\boldsymbol{c}}\right)\right]}{2\sigma_k^2}, \, \forall k = 1, \ldots, n.
\end{aligned}$$

It would be interesting to know whether the least quadratic estimation is convex in $\tilde{\boldsymbol{c}}$. Unfortunately, Hertog et al. (2002) observed that the least squares quadratic approximation of a multivariate convex function in a finite set of points is not necessarily convex even though it is convex for a univariate convex function. Similarly for our problem, we cannot guarantee that the least quadratic estimation is convex. It is however possible to enforce convexity through imposing a semidefinite constraint on Γ, but the resulting problem will not exhibit a nice and explicit characterization of the solution as the unconstrained version.

4.1 Application: Project Management

We would like to know how accurate the least squares linear approximation can be and how the least squares quadratic approximation can improve the estimation accuracy. Using the exact persistency values, we rule out the impact of errors from estimating persistency values, which might either increase or decrease the accuracy of our least squares estimators and complicate the analysis. By "exact", we mean the persistency values are directly computed from simulation (i.e., sample estimates of $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$, $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$ and $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$) rather than some persistency estimation models.

The key performance indicator we consider is the expected square deviation (ESD), which is also the objective function we try to minimize in obtaining our least square approximations. Unfortunately, almost all the approximating distributions derived using previous methods do not reside in the same probability space as $Z(\tilde{\boldsymbol{c}})$, which makes it impossible to compute the squared deviation from $Z(\tilde{\boldsymbol{c}})$. This problem arises since the traditional approaches solely focus on the distribution (like tail probabilities, etc.) but overlook the approximation error between the approximated completion time and the true completion time under a specific realization of the random activity durations. For example, Cox (1995) assumed the project completion time

to be normally distributed at first, and then tried to estimate the moments of the completion time. Hence, we have to resort to other measures to compare the performance of different approximation methods including descriptive statistics, like mean, standard deviation, and skewness. In addition, we also employ the following measure to quantify the distance between two distributions:

$$\text{Square Norm Distance}(F,G) = \text{SND}(F,G) := \int_0^1 \left[F^{-1}(y) - G^{-1}(y)\right]^2 dy$$

where F and G are the cumulative distribution functions of two distributions.

Example 3 *The project network consists of four nodes and five arcs as shown in Figure 3. All activities are independent and normally distributed with mean and variance both equal to one.*

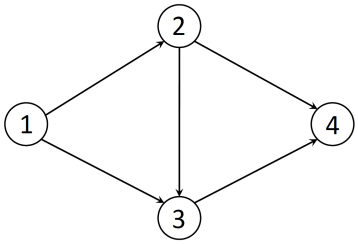


Figure 3: The Project Network Example

The network in Example 3 is the "Wheatstone bridge" network from Lindsey (1972) and later regarded as the "forbidden graph" by Dodin (1985) since it is the basic evidence of graph irreducibility. Ord (1991) summarized the results for this graph documented in literature with normally distributed activity durations, and also provided the results from his discrete approximation method with a parameter k indicating the number of discrete points used to approximate the normal distribution. Indeed, the approximated distributions obtained by Ord (1991) should be a discrete distribution. However, we extend his theory in computing the square norm distance by assuming the final approximated distribution follows a normal distribution with the moments derived from his original procedure. All these results are presented in Table 2, where T denotes the project completion time, and $\sigma(T)$ denotes its standard deviation, and $sk(T)$ denotes its skewness. "Error on $\sigma(T)$" is computed as the absolute relative error against the simulation result. The new result from our method is also presented in Table 2 under "LSN" and "LSQ", where "LSN" stands for "Least Squares Normal" and "LSQ" stands for "Least Squares Quadratic". We conducted 10^6 simulation runs to estimate the persistency values.

Approximation Method		$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	SND
10^6 simulation		3.516	1.39	-	0.28	-	-
Numerical integration		3.483	1.47	5.76%	0	-	0.017
Ord (1991)	$k=2$	3.261	0.70	49.64%	0	-	0.543
	$k=3$	3.485	1.04	25.18%	0	-	0.128
	$k=4$	3.525	1.08	22.32%	0	-	0.101
	$k=5$	3.582	1.15	17.27%	0	-	0.068
	$k=6$	3.594	1.15	17.27%	0	-	0.069
Cox (1995)		3.639	1.69	21.58%	0	-	0.116
PERT		3.000	1.73	24.46%	0	0.973	0.395
LSN		3.515	1.27	8.63%	0	0.311	0.021
LSQ		3.515	1.36	2.16%	0.47	0.078	0.005

Table 2: Estimation results for Example 3 with simulated parameters for least squares approximating distributions

From Table 2, we can see that except the numerical integration approach and our quadratic estimator, the least squares linear estimation gives the best estimate for the standard deviation, in terms of absolute relative error. Regardless of the high accuracy, the integration approach would be too tedious to be applicable for even medium-size networks. This suggests that using persistency could be a promising way to estimate the variability in the project completion time. Recall in our approximation model, the variance is solely determined by the persistency values (i.e., β_j's in Equation(2)). Adding the quadratic terms not only helps capture the right direction of skewness, but more interestingly, it significantly improves the estimation on variance. The added variability comes from the quadratic components of the estimator, as the linear term in the least squares quadratic estimation shares the same coefficients as the least squares linear estimator, i.e., persistency. Overall, the least squares linear approximation is remarkably effective with extremely low ESD and SND, and the least squares quadratic approximation even pushes the SND below the numerical integration approach. Figure 4 plots the density and cumulative distribution functions of PERT and our least squares estimations together with the simulation results. It is obvious from the plots that both least square estimators fit closely with simulation results. With the right skewness direction, the cumulative distribution function of the quadratic estimator almost overlaps with that of simulation.

For the example problems we studied above, the skewness in the optimum distribution is not very strong. In order to better demonstrate the impact of the quadratic estimator, we study a simple problem discussed by Zhan et al. (2005) in the next example.

Example 4 *Approximate the distribution of the maximum of two normal random variables,* $N(0, 0.5^2)$ *and* $N(1, 3^2)$*. In this case, the persistency values can be accurately obtained from integration.*

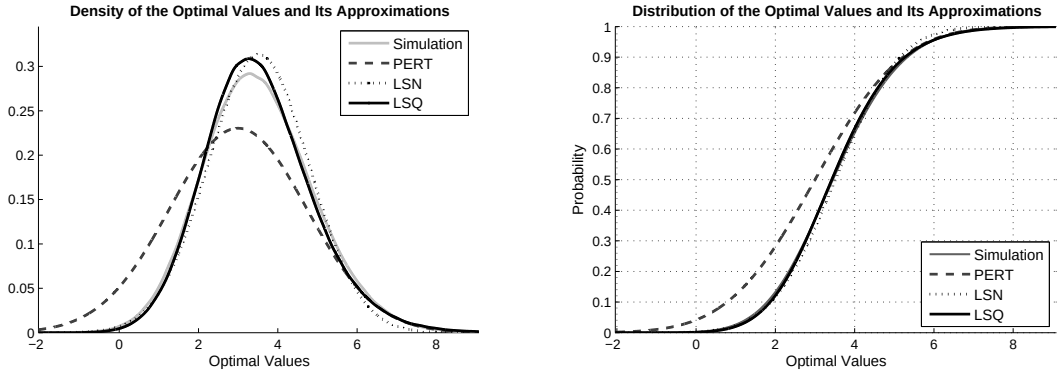


Figure 4: Distributions for Example 3

The results are plotted in Figure 5, and the improvement from the quadratic estimator is obvious. We can conclude that the advantage of adding quadratic terms is larger if the true distribution is suspect to be very skewed.

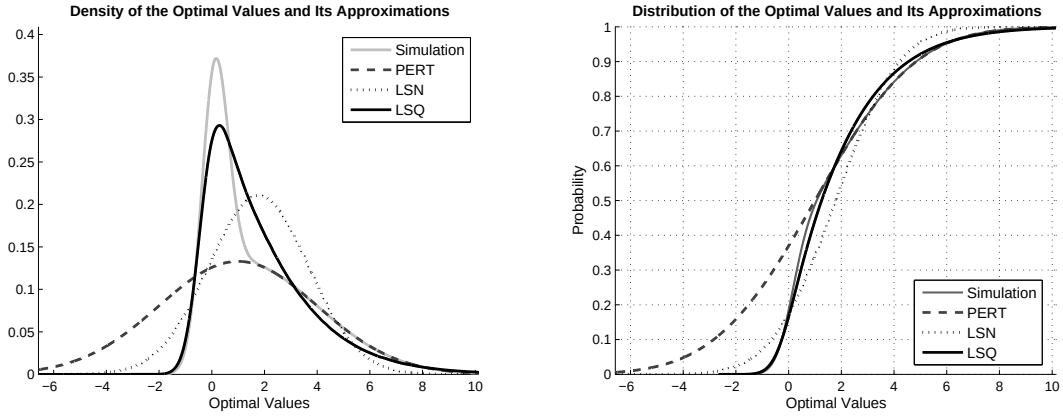


Figure 5: Distributions for Example 4

5 Approximating Persistency Values

In practical applications, if simulation can be easily conducted, we can directly obtain an empirical distribution for the random optimization problem without resorting to other distribution approximation techniques. However, if the deterministic problem is $\mathcal{NP}$-hard or very complicated, then simulation may require tremendous effort or resources to achieve satisfactory results. In that case, we might benefit from choosing appropriate ways to approximate the distribution. Under the least squares approximation framework, our results tell that the task

is transformed to estimate persistency values and some related parameters. Hence, the success of our approximation methods hinge on the accuracy of the estimation on these parameters. In what follows, we will briefly review some estimation methods that we adopt in the numerical analysis.

Note that we are not bound to use just one model to estimate all the necessary parameters for the approximating distributions, $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$, $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$, and $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. Indeed, we can choose any methods deemed appropriate for each parameter. From this point of view, there is a huge literature we can make use of to estimate these parameters, especially for $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. What we show next is only one possible approach. To avoid the criticism of speculation, we only choose some basic and generic estimation methods without sophisticated modifications to tailor to our test problems. Hence, we leave plenty of room for users to improve the approximation accuracy for specific applications and better demonstrate the power of our least squares approximations.

In the literature, the problem of estimating the expected objective value of a stochastic optimization problems has been studied for a long time. In case of the project management problem, the search for the expected project completion time started half century ago (cf. Fulkerson (1962)) and is still an active research topic (cf. Yao & Chu (2007)). In this section, two naïve methods are used. For small networks, we use the classical estimation method proposed by Clark (1961), which is the building block of most modern distribution approximation methods, especially for the project management problem. However, we only use the original estimation methods from Clark (1961) to estimate $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ without considering any further extensions and refinements. For larger networks, implementing Clark's methods may require some programming effort, so we simply use PERT to give a rough estimate of $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$, since it is a popular tool in practice.

On the other hand, although the concept of persistency has only been brought into the optimization area since Bertsimas et al. (2006), it has long been studied under different guises such as the criticality index in the project management area. The majority of the research work on estimating criticality has been focusing on developing heuristics algorithms based on the topological properties of the project networks, and the uncertainty is usually treated by discretization and/or stochastic dominance considerations (cf. Dodin (1984), Dodin & Elmaghraby (1985), etc.). More advanced method combines the strength of different approaches to obtain some hybrid method. For example, Bowman (1995) utilized the geometric properties of the networks to reduce the computational requirement of simulation. The common limitation of these methods is the lack of consideration of correlations among different activity completion times. Besides these specific estimation methods for the project management method, there is a series of generic conic programming based models for persistency estimation as reviewed before (cf. Natarajan et al. (2009), Mishra et al. (2012), Natarajan et al. (2011), Kong et al.

(2013) etc.). By "generic" we mean that these methods work on any optimization problems and do not exploit any specific problem structure like the network flow in the project management problem. In what follows, we will review in more details the most recent progress on the persistency estimation, i.e., CPCMM, mainly contributed by Natarajan et al. (2011). We will make use of this model in the numerical studies.

Natarajan et al. (2011) consider the following stochastic optimization problem:

$$Z_P := \sup_{\tilde{\boldsymbol{c}} \sim (\boldsymbol{\mu}, \Sigma)^+} \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right],$$

where $\tilde{\boldsymbol{c}} \sim (\boldsymbol{\mu}, \Sigma)^+$ means that the set of distributions of the random coefficient vector $\tilde{\boldsymbol{c}}$ (assumed to be nonempty) is defined by the nonnegative support $\mathbb{R}^n_+$, finite mean vector $\boldsymbol{\mu}$ and finite covariance matrix Σ, i.e., $\tilde{\boldsymbol{c}} \in \{\tilde{\boldsymbol{X}} : \mathbf{E}[\tilde{\boldsymbol{X}}] = \boldsymbol{\mu}, \mathbf{E}[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T] = \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T, \mathbf{P}(\tilde{\boldsymbol{X}} \geq \mathbf{0}) = 1\}$. They proved that Z_P can be solved as the following convex conic optimization problem:

$$\begin{array}{rl} Z_C = \max & \sum_{j=1}^n Y_{j,j} \\ \text{s.t.} & \boldsymbol{a}_i^T X \boldsymbol{a}_i - 2b_i \boldsymbol{a}_i^T \boldsymbol{x} + b_i^2 = 0, \, \forall i = 1, \ldots, m \\ & X_{j,j} = x_j, \, \forall j \in \mathcal{B} \\ & \begin{pmatrix} 1 & \boldsymbol{\mu}^T & \boldsymbol{x}^T \\ \boldsymbol{\mu} & \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T & Y^T \\ \boldsymbol{x} & Y & X \end{pmatrix} \succeq_{cp} 0 \end{array}$$

(i.e., $Z_P = Z_C$) where the decision variables are $\boldsymbol{x} \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times n}$, and $Y \in \mathbb{R}^{n \times n}$. For a matrix $A \in \mathbb{R}^{n \times n}$, $A \succeq_{cp} 0$ means that A lies in the cone of completely positive matrices of dimension n defined as

$$\mathcal{CP}_n := \left\{ A \in \mathbb{R}^{n \times n} \,|\, \exists V \in \mathbb{R}^{n \times k}_+, \text{ such that } A = VV^T \right\}.$$

The linear program over the convex cone of the completely positive matrices is called a completely positive program (CPP), and Z_C is a typical CPP. That is why this model is called Completely Positive Cross Moment Model. Furthermore, they extended CPCMM by relaxing the nonnegative support assumption on $\tilde{\boldsymbol{c}}$, which makes CPCMM suitable for our case, because the support of a multivariate normal distribution is the whole Euclidean space. The only change needed is to modify the conic constraint. For ease of exposition, we still keep the basic CPCMM formulation for the following illustration. The support extension can be uniformly applied through modifying the conic constraint. A key reason that we choose this model is its ability to capture correlations among random coefficients.

In the formulation of Z_C, the variables $\boldsymbol{x}$, Y and X attempt to encode the information $x_j = \mathbf{E}[x_j(\tilde{\boldsymbol{c}})]$, $Y_{i,j} = \mathbf{E}[\tilde{c}_j x_i(\tilde{\boldsymbol{c}})]$ and $X_{i,j} = \mathbf{E}[x_i(\tilde{\boldsymbol{c}}) x_j(\tilde{\boldsymbol{c}})]$ under the worst case distribution.

Thus, through solving Z_C, the optimal objective value gives the value of $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$, and the optimal value of $\boldsymbol{x}$ is simply the persistency, also under the worst case distribution. In addition, the "by-product" of solving CPCMM, Y, gives necessary information to construct the quadratic estimator, which is obtained without any additional effort.

However, a key drawback of CPCMM is that it ignores the distributional information. Hence, when $\tilde{\boldsymbol{c}}$ is normally distributed, CPCMM only gives an upper bound on $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ and an estimate of the persistency and $\mathbf{E}[\tilde{c}_j x_i(\tilde{\boldsymbol{c}})]$. A direct cure to this problem is to add some ellipsoidal constraints on the probability mass of $\tilde{\boldsymbol{c}}$ that are known for multivariate normal random variables, so that CPCMM can be gradually refined to incorporate the distributional information. For an illustration of this technique, please refer to Natarajan et al. (2010). In this paper, however, we do not implement this method, because the persistency estimates from CPCMM are good enough for most examples we will discuss later and we want to keep the focus of this paper on distribution approximation rather than persistency estimation.

Another issue with CPCMM is that it is $\mathcal{NP}$-hard to solve despite the fact that the completely positive cone is closed, convex and pointed. Fortunately, there are various hierarchies of tractable approximations for the completely positive cone, e.g., Bomze et al. (2000), Parrilo (2000) and Klerk et al. (2002) etc. In the following computational study, we use a simple SDP approximation of the completely positive constraint, i.e., $A \succeq_{cp} 0$ is relaxed to $A \succeq 0$ and $A \geq 0$, where $A \succeq 0$ means that A is positive semidefinite. Such relaxation is also called doubly nonnegative relaxation.

Despite all these numerical inaccuracies, we show that our approximation methods are still practically attractive due to the use of persistency in the approximation and the flexibility of our methods.

5.1 Computational Study

Consider Example 3 again, and we will construct our least squares approximating distributions using estimated persistency values. As discussed above, we implement the estimation scheme from Clark (1961) to estimate the mean project completion time, i.e., the parameter α in our models. For persistency estimates, we solve the SDP relaxation of CPCMM reviewed in Section 5. The results are summarized in Table 3, where we add a lower case letter "e" after "LSN" and "LSQ" to indicate the results from estimated persistency parameters.

From the table, we can see that when estimated parameters are used instead of the exact ones, the distributions constructed from our least squares method still perform very well. For the least squares linear approximation, the estimated variance only deteriorates a little bit, which highlights the accuracy of persistency estimates from CPCMM and the power of using persistency in distribution approximation. Although the estimation error on $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$ has

Approximation Method		$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	SND
10^6 simulation		3.516	1.39	-	0.28	-	-
Numerical integration		3.483	1.47	5.76%	0	-	0.017
Ord (1991)	$k=2$	3.261	0.70	49.64%	0	-	0.543
	$k=3$	3.485	1.04	25.18%	0	-	0.128
	$k=4$	3.525	1.08	22.32%	0	-	0.101
	$k=5$	3.582	1.15	17.27%	0	-	0.068
	$k=6$	3.594	1.15	17.27%	0	-	0.069
Cox (1995)		3.639	1.69	21.58%	0	-	0.116
PERT		3.000	1.73	24.46%	0	0.973	0.395
LSNe		3.518	1.26	8.80%	0	0.311	0.022
LSQe		3.519	1.44	3.76%	0.60	0.124	0.014

Table 3: Estimation results for Example 3 with estimated parameters for least squares approximating distributions

some impact on the least squares quadratic approximation, it still improves the performance from the least squares linear approximation. In particular, the variability estimate still outperforms the numerical integration approach, and the SND is below the numerical integration approach and much better than any other existing methods.

To further justify the performance of our models, we also test our least square estimations on a series of random project networks of larger sizes, and compare the results with PERT. In this case, we use PERT to estimate $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$, which will be used in calculating the optimal parameter α in our least squares quadratic estimator. We drop the comparison on SND in this example, since all the distributions here allow the computation of ESD from the true distribution of $Z(\tilde{\boldsymbol{c}})$.

Example 5 *Approximate the completion time distributions of the random projects generated by the following algorithm:*

Random Project Network Generation Algorithm

Step 1. Randomly set the number of nodes (m') in the project network.

Step 2. Construct a zero adjacency matrix. Go through every matrix entry in the upper triangle (above the diagonal), and replace 0 by 1 if an independent realization of a uniform random variable $U(0,1)$ is greater than s, where $s \in [0,1]$. s can be used to control the density of the graph. More precisely, after this step, the random network will have an expected number of arcs $\mathbf{E}[n'] = s \cdot m'(m'-1)/2$, and each node will have $s(m'-1)$ expected number of neighbours. We randomly set s from 0.2 to 0.8 in our experiments.

Step 3. Remove all the isolated nodes in the network.

Step 4. Create an initial node s. For each node i without incoming arcs, add an arc $s \to i$.

Step 5. Create a terminal node t. For each node i without outgoing arcs, add an arc $i \to t$. After this step, the structure of the network is fixed. Denote the final number of nodes as m and the final number of arcs as n.

Step 6. For arc i, generate the random arc length with mean μ_i uniformly drawn between 1 and 10, and standard deviations uniformly drawn from 0 to $0.7\mu_i$.

Step 7. Randomly generate a correlation matrix for the activities.

The results for ten random networks are presented in Table 4[4]. The sample size for all the simulations is 2×10^4. From Table 4, it is clear that our findings observed in small example network carry on to larger networks, and both least squares approximations demonstrate consistent superior performance. It is worthwhile to mention that the quadratic estimator consistently provides very accurate estimation of the variability in project completion time. For quite a few cases, the estimation errors are less than 1%.

	$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD
	$m = 20$, $n = 29$					$m = 22$, $n = 41$				
SIMU	4.1392	0.2220	-	0.0605	-	3.5374	0.1942	-	0.3421	-
PERT	4.0185	0.2725	22.72%	0	0.0445	3.3992	0.2713	39.70%	0	0.0547
LSNe	4.0185	0.1968	11.34%	0	0.0229	3.3992	0.1572	19.05%	0	0.0296
LSQe	4.0197	0.2262	1.91%	0.5085	0.0183	3.3997	0.2040	5.06%	0.9225	0.0238
	$m = 10$, $n = 28$					$m = 22$, $n = 35$				
SIMU	4.2635	0.3561	-	0.1281	-	3.8964	0.1553	-	0.3447	-
PERT	4.1912	0.3986	11.93%	0	0.0171	3.7742	0.2102	35.35%	0	0.0378
LSNe	4.1912	0.3416	4.09%	0	0.0101	3.7742	0.1228	20.93%	0	0.0230
LSQe	4.1928	0.3540	0.60%	0.5213	0.0100	3.7744	0.1713	10.28%	0.7489	0.0207
	$m = 24$, $n = 44$					$m = 12$, $n = 37$				
SIMU	3.5860	0.2115	-	0.6954	-	5.0994	0.2843	-	0.1939	-
PERT	3.5031	0.2807	32.72%	0	0.0296	5.0289	0.3311	16.46%		0.0212
LSNe	3.5031	0.1847	12.63%	0	0.0176	5.0289	0.2664	6.30%		0.0130
LSQe	3.5034	0.2241	6.00%	1.2997	0.0136	5.0304	0.2932	3.13%	0.6096	0.0108
	$m = 8$, $n = 16$					$m = 9$, $n = 20$				
SIMU	3.5774	0.3298	-	0.2027	-	4.2892	0.1644	-	0.5189	-
PERT	3.5350	0.3682	11.65%	0	0.0105	4.2159	0.2082	26.66%	0	0.0196
LSNe	3.5350	0.3203	2.87%	0	0.0062	4.2159	0.1474	10.34%	0	0.0112
LSQe	3.5352	0.3317	0.59%	0.5817	0.0045	4.2174	0.1769	7.64%	1.0683	0.0078
	$m = 25$, $n = 42$					$m = 11$, $n = 33$				
SIMU	3.5276	0.2756	-	0.5382	-	5.1726	0.2911	-	0.0965	-
PERT	3.4938	0.3234	17.35%	0	0.0110	5.0934	0.3156	8.43%	0	0.0164
LSNe	3.4938	0.2531	8.19%	0	0.0073	5.0934	0.2712	6.85%	0	0.0105
LSQe	3.4925	0.2767	0.40%	1.2969	0.0071	5.0958	0.2913	0.08%	0.4413	0.0091

Table 4: Estimation results for random project networks in Example 5

[4]For this example, we have constructed and analyzed more than one hundred random networks, and the results share the same pattern throughout the experiment. Hence, we only show ten instances as a demonstration.

6 Conclusion

In this paper, we show that the distribution approximation problem under least squares framework and normality assumption can be transformed into the related persistency problem. Various applications and computational experiments are presented to demonstrate the advantages of our approximation method, especially the benefits of introducing persistency into the distribution approximation problem. Better estimation on persistency values is then becoming critical and hence worth more exploration, especially under the normality assumption.

The results in this paper can be developed further in several ways. In particular, with the knowledge on the distribution of the optimal value, we can now conduct more in-depth risk analysis or parameter calibration for the underlying stochastic mixed zero-one linear optimization problem. We leave these and other related issues for future research.

Appendix A. Proof of Lemma 1

The proof is consolidated from Stein (1972), Stein (1981) and Liu (1994).

The first result is the univariate version of Stein's Identity (cf. Stein (1972) and Stein (1981)).

Let $\tilde{c}$ follow a standard normal distribution, $N(0,1)$, and $\phi(c)$ denote the standard normal density with the derivative satisfying $\phi'(c) = -c\phi(c)$. For any function $h : \mathbb{R} \to \mathbb{R}$ such that h' exists almost everywhere and $\mathbf{E}[|h'(\tilde{c})|] < \infty$,

$$\begin{aligned}
\mathbf{E}\left[h'(\tilde{c})\right] &= \int_{-\infty}^{\infty} h'(c)\phi(c)\,dc \\
&= \int_{0}^{\infty} h'(c)\left[\int_{c}^{\infty} z\phi(z)\,dz\right]dc + \int_{-\infty}^{0} h'(c)\left[\int_{-\infty}^{c} -z\phi(z)\,dz\right]dc \\
&= \int_{0}^{\infty} z\phi(z)\left[\int_{0}^{z} h'(c)dc\right]dz - \int_{-\infty}^{0} z\phi(z)\left[\int_{z}^{0} h'(c)dc\right]dz \\
&= \left(\int_{0}^{\infty} + \int_{-\infty}^{0}\right)\left[z\phi(z)\left[h(z) - h(0)\right]\right]dz \\
&= \int_{-\infty}^{\infty} z\phi(z)\,h(z)dz \\
&= \mathbf{E}\left[\tilde{c}h(\tilde{c})\right],
\end{aligned}$$

where the third equality is justified by Fubini's Theorem. Note that since $\mathbf{E}[\tilde{c}] = 0$ and $Var(\tilde{c}) = 1$, the equality proved above is essentially

$$Cov\left(\tilde{c}, h(\tilde{c})\right) = Var(\tilde{c})\mathbf{E}\left[h'(\tilde{c})\right]. \tag{7}$$

Next, we present the generalization of the result to the multivariate case (cf. Stein (1981) and Liu (1994)).

Let $\tilde{\boldsymbol{z}} = (\tilde{z}_1, \ldots, \tilde{z}_n)^T$, where $\tilde{z}_j$'s are independent and identically distributed standard normal random variables. From Equation (7) it follows that for any function $\hat{h} : \mathbb{R}^n \to \mathbb{R}$ satisfying the same conditions as h in the Theorem,

$$\mathbf{E}\left[\tilde{z}_1\hat{h}(\tilde{\boldsymbol{z}})\middle|(\tilde{z}_2, \ldots, \tilde{z}_n)\right] = \mathbf{E}\left[\frac{\partial\hat{h}(\tilde{\boldsymbol{z}})}{\partial z_1}\middle|(\tilde{z}_2, \ldots, \tilde{z}_n)\right].$$

Taking the expectation of both sides, we get

$$\mathbf{E}\left[\tilde{z}_1\hat{h}(\tilde{\boldsymbol{z}})\right] = \mathbf{E}\left[\frac{\partial\hat{h}(\tilde{\boldsymbol{z}})}{\partial z_1}\right].$$

Using a similar argument for the remaining random variables, we can show that

$$Cov\left(\tilde{\boldsymbol{z}}, \hat{h}\left(\tilde{\boldsymbol{z}}\right)\right) = \mathbf{E}\left[\nabla \hat{h}\left(\tilde{\boldsymbol{z}}\right)\right].$$

Note that the random vector $\tilde{\boldsymbol{c}}$ can be written as $\tilde{\boldsymbol{c}} = \Sigma^{1/2}\tilde{\boldsymbol{z}} + \boldsymbol{\mu}$. Consider $\hat{h}\left(\tilde{\boldsymbol{z}}\right) = h\left(\Sigma^{1/2}\tilde{\boldsymbol{z}} + \boldsymbol{\mu}\right)$, then $\nabla \hat{h}\left(\tilde{\boldsymbol{z}}\right) = \Sigma^{1/2}\nabla h\left(\tilde{\boldsymbol{c}}\right)$. Hence,

$$Cov\left(\tilde{\boldsymbol{c}}, h\left(\tilde{\boldsymbol{c}}\right)\right) = Cov\left(\Sigma^{1/2}\tilde{\boldsymbol{z}}, \hat{h}\left(\tilde{\boldsymbol{z}}\right)\right) = \Sigma^{1/2}\mathbf{E}\left[\nabla \hat{h}\left(\tilde{\boldsymbol{z}}\right)\right] = \Sigma \mathbf{E}\left[\nabla h\left(\tilde{\boldsymbol{c}}\right)\right].$$

Appendix B. Proof of Theorem 2

Since Problem (Q) is convex, its necessary and sufficient optimality conditions are

$$\mathbf{E}\left[Z(\tilde{\boldsymbol{c}}) - \alpha^* - \sum_{j=1}^{n} \beta_j^*(\tilde{c}_j - \mu_j) - \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}^*(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})\right] = 0,$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^* - \sum_{j=1}^{n} \beta_j^*(\tilde{c}_j - \mu_j) - \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}^*(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})\right)(\tilde{c}_k - \mu_k)\right] = 0,$$
$$\forall k = 1, \ldots, n, \text{ and}$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^* - \sum_{j=1}^{n} \beta_j^*(\tilde{c}_j - \mu_j) - \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}^*(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})\right)(\tilde{c}_{k_1} - \mu_{k_1})(\tilde{c}_{k_2} - \mu_{k_2})\right] = 0,$$
$$\forall 1 \le k_1 \le k_2 \le n.$$

Hence, an optimal solution $(\alpha^*, \boldsymbol{\beta}^*, \Gamma^*)$ should satisfy

$$\alpha^* = \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] - \Sigma \bullet \Gamma^*,$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^* - \sum_{j=1}^{n} \beta_j^*(\tilde{c}_j - \mu_j)\right)(\tilde{c}_k - \mu_k)\right]$$
$$-\mathbf{E}\left[\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}^*(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})(\tilde{c}_k - \mu_k)\right] = 0, \forall k = 1, \ldots, n, \text{ and}$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^* - \sum_{j=1}^{n} \beta_j^*(\tilde{c}_j - \mu_j)\right)(\tilde{c}_{k_1} - \mu_{k_1})(\tilde{c}_{k_2} - \mu_{k_2})\right]$$
$$-\mathbf{E}\left[\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n} \gamma_{j_1,j_2}^*(\tilde{c}_{j_1} - \mu_{j_1})(\tilde{c}_{j_2} - \mu_{j_2})(\tilde{c}_{k_1} - \mu_{k_1})(\tilde{c}_{k_2} - \mu_{k_2})\right] = 0, \forall 1 \le k_1 \le k_2 \le n.$$

From Isserlis' Theorem, if random variable $(\tilde{z}_1, \ldots, \tilde{z}_n)$ follows a zero mean multivariate normal distribution, then

$$\mathbf{E}\left[\prod_{i=1}^{n}\tilde{z}_i\right] = \begin{cases} 0, & \text{if } n \text{ is odd,} \\ \sum\prod \mathbf{E}\left[\tilde{z}_i\tilde{z}_j\right], & \text{if } n \text{ is even,} \end{cases}$$

where $\sum\prod$ means summing over all distinct ways of partitioning $(\tilde{z}_1, \ldots, \tilde{z}_n)$ into pairs (cf. Isserlis (1918)). In particular, when $n = 3, 4$,

$$\mathbf{E}\left[\tilde{z}_1\tilde{z}_2\tilde{z}_3\right] = 0, \text{ and}$$

$$\mathbf{E}\left[\tilde{z}_1\tilde{z}_2\tilde{z}_3\tilde{z}_4\right] = \mathbf{E}\left[\tilde{z}_1\tilde{z}_2\right]\mathbf{E}\left[\tilde{z}_3\tilde{z}_4\right] + \mathbf{E}\left[\tilde{z}_1\tilde{z}_3\right]\mathbf{E}\left[\tilde{z}_2\tilde{z}_4\right] + \mathbf{E}\left[\tilde{z}_1\tilde{z}_4\right]\mathbf{E}\left[\tilde{z}_2\tilde{z}_3\right].$$

Applying Isserlis' Theorem, we can reduce the optimality conditions into

$$\alpha^* = \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] - \Sigma \bullet \Gamma^*,$$

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^* - \sum_{j=1}^{n}\beta_j^*(\tilde{c}_j - \mu_j)\right)(\tilde{c}_k - \mu_k)\right] = 0, \ \forall k = 1, \ldots, n, \tag{8}$$

and

$$\begin{aligned} &\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^*\right)(\tilde{c}_{k_1} - \mu_{k_1})(\tilde{c}_{k_2} - \mu_{k_2})\right] \\ &\quad - \sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\gamma_{j_1,j_2}^*\left(\sigma_{j_1,j_2}\sigma_{k_1,k_2} + \sigma_{j_1,k_1}\sigma_{j_2,k_2} + \sigma_{j_1,k_2}\sigma_{k_1,j_2}\right) = 0, \ \forall 1 \le k_1 \le k_2 \le n. \end{aligned} \tag{9}$$

Further simplifying Equation (8), we get

$$\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})(\tilde{c}_k - \mu_k)\right] = \sum_{j=1}^{n}\beta_j^*\sigma_{j,k}, \ \forall k = 1, \ldots, n.$$

Since $\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})(\tilde{c}_k - \mu_k)\right] = Cov\left(\tilde{c}_k, Z(\tilde{\boldsymbol{c}})\right)$, we arrive at the same conditions as Equation (3) in Theorem 1. Therefore, following the same argument, we have $\beta_k^* = \mathbf{E}\left[x_k(\tilde{\boldsymbol{c}})\right]$, $k = 1, \ldots, n$, which is unique if Σ is positive definite.

Consider a part of the first term in Equation (9),

$$\begin{aligned} \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})(\tilde{c}_{k_1} - \mu_{k_1})(\tilde{c}_{k_2} - \mu_{k_2})\right] &= \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}\tilde{c}_{k_2}\right] - \mu_{k_1}\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_2}\right] \\ &\quad -\mu_{k_2}\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}\right] + \mu_{k_1}\mu_{k_2}\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] \\ &= \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}\tilde{c}_{k_2}\right] - \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}\right]\mu_{k_2} \\ &\quad -\mu_{k_1}\left(\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_2}\right] - \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right]\mu_{k_2}\right) \\ &= Cov\left(Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}, \tilde{c}_{k_2}\right) - \mu_{k_1}Cov\left(Z(\tilde{\boldsymbol{c}}), \tilde{c}_{k_2}\right). \end{aligned}$$

It is straightforward to apply Stein's Identity on $Cov\left(Z(\tilde{\boldsymbol{c}}), \tilde{c}_{k_2}\right)$ as we have done before, i.e.,

$$Cov\left(Z(\tilde{\boldsymbol{c}}), \tilde{c}_{k_2}\right) = \sum_{j=1}^{n} \mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right] \sigma_{j,k_2}.$$

For the other term, $Cov\left(Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}, \tilde{c}_{k_2}\right)$, we can also use Stein's Identity,

$$\begin{aligned} Cov\left(Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}, \tilde{c}_{k_2}\right) &= \sum_{j=1}^{n} \mathbf{E}\left[\frac{\partial Z(\tilde{\boldsymbol{c}})\tilde{c}_{k_1}}{\partial c_j}\right] Cov\left(\tilde{c}_j, \tilde{c}_{k_2}\right) \\ &= \sum_{j=1}^{n} \mathbf{E}\left[\tilde{c}_{k_1}\frac{\partial Z(\tilde{\boldsymbol{c}})}{\partial c_j} + Z(\tilde{\boldsymbol{c}})\frac{\partial \tilde{c}_{k_1}}{\partial c_j}\right] \sigma_{j,k_2} \\ &= \sum_{j=1}^{n} \mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] \sigma_{j,k_2} + \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] \sigma_{k_1,k_2}, \end{aligned}$$

where the last equality follows from the same argument as in the proof of Theorem 1. Therefore,

$$\begin{aligned} \mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}}) - \alpha^*\right)\left(\tilde{c}_{k_1} - \mu_{k_1}\right)\left(\tilde{c}_{k_2} - \mu_{k_2}\right)\right] &= \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\left(\tilde{c}_{k_1} - \mu_{k_1}\right)\left(\tilde{c}_{k_2} - \mu_{k_2}\right)\right] - \alpha^*\sigma_{k_1,k_2} \\ &= \sum_{j=1}^{n} \mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] \sigma_{j,k_2} + \mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] \sigma_{k_1,k_2} \\ &\quad -\mu_{k_1} \sum_{j=1}^{n} \mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right] \sigma_{j,k_2} \\ &\quad -\left(\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right] - \Sigma \bullet \Gamma^*\right) \sigma_{k_1,k_2} \\ &= \sum_{j=1}^{n} \left(\mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] - \mu_{k_1}\mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right]\right) \sigma_{j,k_2} \\ &\quad +\sigma_{k_1,k_2}\Sigma \bullet \Gamma^*. \end{aligned}$$

Substituting this into Equation (9), we get a system of $(n^2+n)/2$ linear equations on Γ^*,

$$\begin{aligned} &\sum_{j=1}^{n} \left(\mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] - \mu_{k_1}\mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right]\right) \sigma_{j,k_2} + \sigma_{k_1,k_2}\Sigma \bullet \Gamma^* \\ &\quad - \sum_{j_1=1}^{n} \sum_{j_2=j_1}^{n} \gamma^*_{j_1,j_2}\left(\sigma_{j_1,j_2}\sigma_{k_1,k_2} + \sigma_{j_1,k_1}\sigma_{j_2,k_2} + \sigma_{j_1,k_2}\sigma_{k_1,j_2}\right) = 0, \; \forall 1 \le k_1 \le k_2 \le n, \end{aligned}$$

which reduces to

$$\begin{aligned} &\sum_{j_1=1}^{n} \sum_{j_2=j_1}^{n} \gamma^*_{j_1,j_2}\left(\sigma_{j_1,k_1}\sigma_{j_2,k_2} + \sigma_{j_1,k_2}\sigma_{k_1,j_2}\right) \\ &\quad = \sum_{j=1}^{n} \left(\mathbf{E}\left[\tilde{c}_{k_1} x_j\left(\tilde{\boldsymbol{c}}\right)\right] - \mu_{k_1}\mathbf{E}\left[x_j\left(\tilde{\boldsymbol{c}}\right)\right]\right) \sigma_{j,k_2}, \; \forall 1 \le k_1 \le k_2 \le n. \end{aligned}$$

Thus, we complete the proof.

References

Adcock, C. J. (2007) *Extensions of Stein's Lemma for the skew-normal distribution*, Communications in Statistics–Theory and Methods, **36**, pp. 1661–1671.

Agrawal, S., Y. Ding, A. Saberi, Y. Ye (2012) *Price of correlations in stochastic optimization*, Operations Research, **60**, pp. 150–162.

Aldous, D., M. Steele (2003) *The objective method: Probabilistic combinatorial optimization and local weak convergence*, in Probability on Discrete Structures, H. Kesten (ed), Springer, Berlin, **110**, pp. 1–72.

Andersen, E. P. (1953) *On the fluctuations of sums of random variables*, Mathematica Scandinavica, **1**, pp. 263–285.

Banerjee, A., Paul, A. (2008) *On path correlation and PERT bias*, European Journal of Operational Research, **189**, pp. 1208–1216.

Bereanu, B. (1963) *On stochastic linear programming. I: Distribution problems: A single random variable*, Romanian Journal of Pure and Applied Mathematics, **8**, pp. 683–697.

Bertsimas, D., K. Natarajan, C. P. Teo (2006) *Persistence in discrete optimization under data uncertainty*, Mathematical Programming, **108**, pp. 251–274.

Bomze, I. M., M. Dür, E. D. Klerk, C. Roos, A. J. Quist, T. Terlaky, (2000) *On copositive programming and standard quadratic optimization problems*, Journal of Global Optimization, **18**, pp. 301–320.

Bowman, R. A. (1995) *Efficient estimation of arc criticalities in stochastic activity networks*, Management Science, **41**, pp. 58–67.

Brown, G. G., R. F. Dell, R. K. Wood (1997) *Optimization and persistence*, Interfaces, **27**, pp. 15–37.

Burer, S. (2009) *On the copositive representation of binary and continuous nonconvex quadratic programs*, Mathematical Programming, **120**, pp. 479–495.

Cacoullos, T. (1982) *On upper and lower bounds for the variance of a function of a random variable*, The Annals of Probability, **10**, pp. 799–809.

Chen, J., D. K. Lin, D. J. Thomas (2003) *On the item fill rate for a finite horizon*, Operations Research Letters, **31**, pp. 199–123.

Clark, E. C. (1961) *The greatest of a finite set of random variables*, Operations Research, **9**, pp. 145–162.

Conniffe, D., J. E. Spencer (2000) *Approximating the distribution of the maximum partial sum of normal deviates*, Journal of Statistical Planning and Inference, **88**, pp. 19–27.

Cox, M. A. (1995) *Simple normal approximation to the completion time distribution for a PERT network*, International Journal of Project Management, **13**, pp. 265–270.

Dodin, B. M. (1984) *Determining the K most critical paths in PERT networks*, Operations Research, **32**, pp. 859–877.

Dodin, B. M. (1985) *Bounding the project completion time distribution in PERT networks*, Operations Research, **33**, pp. 862–881.

Dodin, B. M., S. E. Elmaghraby (1985) *Approximating the criticality indices of the activities in Pert networks*, Management Science, **31**, pp. 207–223.

Ewbank, J. B., B. L. Foote, H. J. Kumin (1974), *A method for the solution of the distribution problem of stochastic linear programming*, SIAM Journal on Applied Mathematics, **26**, pp. 225–238.

Fulkerson, D. R. (1962) *Expected critical path lengths in PERT networks*, Operations Research, **10**, pp. 808–817.

Hagstrom, J. N. (1988) *Computational complexity of PERT problems*, Networks, **18**, pp. 139–147.

Hertog, D. den, E. de Klerk, J. Roos (2002) *On convex quadratic approximation*, Statistica Neerlandica, **56**, pp. 376–385.

Hurst, H. E. (1951) *Long term storage capacity of reservoirs*, Transactions of the American Society of Civil Engineers, **56**, pp. 376–385.

James, B., K. L. James, D. Siegmund (1987) *Test for a change-point*, Biometrika, **74**, pp. 71–83.

Isserlis, L. (1918) *On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables*, Biometrika, **12**, pp. 134–139.

Klerk, E. de, D. V. Pasechnik (2002) *Approximation of the stability number of a graph via copositive programming*, SIAM Journal on Optimization, **12**, pp. 875–892.

Kleindorfer, G. B. (1971) *Bounding distributions for a stochastic acyclic network*, Operations Research, **19**, pp. 1586–1601.

Kong, Q., C. Y. Lee, C. P. Teo, Z. Zheng (2013) *Scheduling arrivals to a stochastic service delivery system using copositive cones*, forthcoming in Operations Research.

Larson, R.C. (1990) *The Queue Inference Engine: Deducing Queue Statistics from Transactional Data*, Management Science, Vol. **36**, No. 5, pp. 586-601

Lasserre, J. B. (2010) *A "joint+marginal" approach to parametric polynomial optimization*, SIAM Journal of Optimization, **20**, pp. 1995–2022.

Liu, J. S. (1994) *Siegel's formula via Stein's identities*, Statistics & Probability Letters, **21**, pp. 247–251.

Lindsey, J. H. (1972) *An estimate of expected critical-path length in PERT networks*, Operations Research, **20**, pp. 800–812.

Mishra, V. K., K. Natarajan, H. Tao, C. P. Teo (2012) *Choice Prediction with Semi-definite Optimization when utilities are correlated*, IEEE Automatic Control, **57**, pp. 2450–2463.

Natarajan, K., M. Sim, J. Uichanco (2010) *Tractable robust expected utility and risk models for portfolio optimization*, Mathematical Finance, **20**, pp. 695–731.

Natarajan, K., M. Song, C. P. Teo (2009) *Persistency model and its applications in choice modeling*, Management Science, **55**, pp. 453–469.

Natarajan, K., C. P. Teo, Z. Zheng (2011) *Mixed zero-one linear programs under objective uncertainty: a completely positive representation*, Operations Resaerch, **59**, pp. 713–728.

Ord, J. K. (1991) *A simple approximation to the completion time distribution for a PERT network*, The Journal of the Operational Research Society, **42**, pp. 1011–1017.

Parrilo, P. A. (2000) *Structured semidefinite programs and semi-algebraic geometry methods in robustness and optimization*, Ph.D. Dissertation, California Institute of Technology.

Prekopa, A. (1966) *On the probability distribution of the optimum of a random linear program*, SIAM Journal on Control and Optimization, **4**, pp. 211–222.

Siegel, A. F. (1993) *A surprising covariance involving the minimum of multivariate normal variables*, Journal of the American Statistical Association, **88**, pp. 77–80.

Spitzer, F. (1956) *A combinatorial lemma and Its application to probability theory*, Transactions of the American Mathematical Society, **82**, pp. 323–339.

Stein, C. M. (1972) *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, **2**, pp. 583–602.

Stein, C. M. (1981) *Estimation of the mean of a multivariate normal distribution*, The Annals of Statistics, **9**, pp. 1135–1151.

Thomas, D. J. (2005) *Measuring item fill-rate performance in a finite horizon*, Manufacturing & Service Operations Management, **7**, pp. 74–80.

Yao, M. J., W. M. Chu (2007) *A new approximation algorithm for obtaining the probability distribution function for project completion time*, Computers & Mathematics with Applications, **54**, pp. 282–295.

Zhan, Y., A. J. Strojwas, X. Li, L. T. Pileggi, D. Newmark, m. Sharma (2005) *Correlation-aware statistical timing analysis with non-Gaussian delay distributions*, Proceedings of the 2005 Design Automation Conference, pp. 77–82.