11-2016

Least squares approximation to the distribution of project completion times with Gaussian uncertainty

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

Karthik Natarajan
Singapore University of Technology and Design

Chung-Piaw TEO
National University of Singapore

DOI: https://doi.org/10.1287/opre.2016.1528

Submitted to *Operations Research*
manuscript OPRE-2013-07-401

Least Squares Approximation to the Distribution of Project Completion Times with Gaussian Uncertainty

Zhichao Zheng
Lee Kong Chian School of Business, Singapore Management University, Singapore 178899, danielzheng@smu.edu.sg

Karthik Natarajan
Engineering Systems and Design, Singapore University of Technology and Design, Singapore 487372, natarajan_karthik@sutd.edu.sg

Chung-Piaw Teo
Department of Decision Sciences, NUS Business School, National University of Singapore, Singapore 119245, bizteocp@nus.edu.sg

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 random objective coefficients follow a multivariate normal distribution? Using 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. Computational studies show that the new approach provides more accurate estimates of the distributions of project completion times compared to existing methods.

Key words: distribution approximation, persistency, Stein's Identity, project management, statistical timing analysis

History: This paper was first submitted on July 14, 2013 and has been with the authors for 3 years for 4 revisions.

Published in Operations Research, November 2016, Volume 64 (6), Pages 1406-1421.
http://dx.doi.org/10.1287/opre.2016.1528

1. Introduction

One of the fundamental problems in project management is to identify the project completion time when the activity durations are random. This problem is often represented by a directed acyclic graph (DAG)[1]. Over the past few decades, various methods have been proposed to approximate the distribution of project completion time (cf. Dodin (1985), Cox (1995), Yao & Chu (2007), etc.). Unfortunately, to the best of our knowledge, most of the existing approaches are derived using ad hoc heuristics or exploit specific problem features.

In this paper, we address the estimation of the project completion time, under the assumption that the activity durations follow a multivariate normal distribution, and construct an estimator for the random project completion time that is optimal among the family of normal distributions under the L^2-norm. In fact, our method applies to any general random mixed 0-1 LP problem under objective uncertainty:

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

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 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 $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ 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 $Z(\tilde{\boldsymbol{c}})$ for specific 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 intractable.

Back to the project management problem, under the Critical Path Method (CPM), which is often used by the project management community, 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

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

Here we impose the conventional assumption that $Z(\boldsymbol{\mu})$ has unique optimal solution. 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 (or longest path) 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. This 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 $\alpha \in \mathbb{R}$, and $\boldsymbol{\beta} \in \mathbb{R}^n$ are determined by minimizing the expected squared deviation of the linear approximation from the true distribution. Problem (P) is the central question addressed by this paper. We 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*. 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 Lebesgue measurable with a non-vanishing density, 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 $\Sigma \succ 0$). Under 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.

Outline of this paper: In the next section, we review the related literature. In Section 3, we build our least squares linear approximation. The extension to least squares quadratic estimation is developed in Section 4. In Section 5, we briefly review some ways to estimate persistency values, and present the results from our computational studies. Finally, we provide some concluding remarks and future research directions in Section 6.

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 approximate 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 5.3.

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 studies on 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 the project and 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.

There is another version of the stochastic longest path problem from the engineering domain, which has gained much research interest in recent years. The problem arises in analyzing signal delay in the digital circuit design, and it is known as "Statistical Timing Analysis" (STA). For a recent review on STA, please refer to Blaauw et al. (2008). The problem that STA tries to address is to estimate the delay time distribution of sending a signal through a digital circuit with various gates and connecting routes. There are many sources of uncertainty in delay times, and the delay of the signals passing through a gate is usually modeled as a Gaussian random variable. The whole problem can be cast into a stochastic longest path problem in a DAG, just like the project management problem. Correlations in STA come from various reasons, e.g., sharing of common physical parameters, or proximity in physical locations (a.k.a., spatial correlations), etc. There is a growing literature in STA to model these correlations, e.g., Tsukiyama et al. (2001), Agrawal et al. (2003), Le et al. (2004), Chang & Sapatnekar (2005), Zhan et al. (2005), Zhang et al. (2005), Li et al. (2007), Khandewal & Srivastava (2007), Tang et al. (2012), etc. Commonly used methods for STA are based on CPM, PERT and approximation methods developed by Clark (1961). They are very similar to the approach adopted by Cox (1995) from project management literature (cf. Blaauw et al. (2008)). Clark (1961) studied the moment estimation problem for the maximum of a finite set of random variables following a multivariate normal distribution. The method is iterative in the sense that $\max\{\tilde{c}_1, \tilde{c}_2, \tilde{c}_3\} \cong \max\{\max\{\tilde{c}_1, \tilde{c}_2\}, \tilde{c}_3\}$, where $\max\{\tilde{c}_1, \tilde{c}_2\}$ is assumed to be normal. Clark (1961) presented a set of analytical expressions to compute the moment estimates by taking into account of correlations among the set of random variables.

More recently, several researchers in this field have begun to look beyond the simple normality assumption and tried to model the delay time at the gate as a quadratic function of normal

random variables (cf. Zhang et al. (2005), Zhan et al. (2005), Khandewal & Srivastava (2007), etc.). The key reason to extend to quadratic models is to capture the skewness in the delay time distribution in the project management problem. Our approach is different from these works which try to estimate the coefficients of the quadratic function through moment matching equalities and topologically go through every node in the network applying approximation techniques on "sum" and "max" operations. On the other hand, by linking the problem to Stein's Identity, we can explicitly characterize the expression of the least squares quadratic estimator. Besides the obvious theoretical elegance, our method is indeed more accurate computationally, which we will illustrate using examples in Section 5.

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) has also 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:

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

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*

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

The proof of Theorem 1 utilizes the following classical covariance identity due to Stein, and its proof is enclosed in the online electronic companion to this paper 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,*

$$\int_a^b \frac{\partial h\left(c_1, ..., c_n\right)}{\partial c_j} dc_j = h\left(c_1, ...c_{j-1}, b, c_{j+1}, c_n\right) - h\left(c_1, ...c_{j-1}, a, c_{j+1}, c_n\right), \forall a, b \in \mathbb{R},$$

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

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

Note that the function $Z(\tilde{\boldsymbol{c}})$ is the finite maximum of linear functions and hence convex and differentiable almost everywhere. Therefore, $Z(\tilde{\boldsymbol{c}})$ satisfies the conditions for Stein's identity. 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. Q.E.D.

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.,

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

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 can be significantly closer to the true variance than those estimated from existing distribution approximation methods. We will illustrate this using examples in Section 5.

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.

REMARK 3. The normality assumption plays an important role in deriving the simple closed-form expression for $\boldsymbol{\beta}^*$. Stein's identity transforms the complicated function $\mathbf{E}[\tilde{c}_j Z(\tilde{\boldsymbol{c}})]$ into one that involves only $Z(\tilde{\boldsymbol{c}})$, in particular $\mathbf{E}[\partial Z(\tilde{\boldsymbol{c}})/\partial c_j] = \mathbf{E}[x_j(\tilde{\boldsymbol{c}})]$. When the normality assumption does not hold, we can still derive a closed-form expression for $\boldsymbol{\beta}^*$ using the well-known size-bias distribution under a mild assumption that $\tilde{\boldsymbol{c}}$ is nonnegative. This assumption is generally satisfied for activity completion times in project management problems, and product prices in consumer choice problems, etc. For $j \in \{1, \ldots, n\}$, we say that $\tilde{\boldsymbol{c}}^j = (\tilde{c}_1^j, \ldots, \tilde{c}_n^j)^T$ has the $\tilde{\boldsymbol{c}}$-size biased distribution in the jth coordinate if $\mathbf{E}[\tilde{c}_j f(\tilde{\boldsymbol{c}})] = \mu_j \mathbf{E}[f(\tilde{\boldsymbol{c}}^j)]$ for all functions $f(\cdot)$ for which the expectations exist. The distribution function of $\tilde{\boldsymbol{c}}^j$ can be characterized as $c_j dF(\boldsymbol{c})/\mu_j$, where $F(\boldsymbol{c})$ is the joint distribution function of $\tilde{\boldsymbol{c}}$ (cf. Goldstein & Rinott (1996), Ross (2011)). Then it is straightforward to derive from Equation (3) that $\boldsymbol{\beta}^* = \Sigma^{-1}(\boldsymbol{\mu} \circ \boldsymbol{\Delta})$, where $\circ$ denotes Hadamard product, and $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_n)^T = (\mathbf{E}[Z(\tilde{\boldsymbol{c}}^1)] - \mathbf{E}[Z(\tilde{\boldsymbol{c}})], \ldots, \mathbf{E}[Z(\tilde{\boldsymbol{c}}^n)] - \mathbf{E}[Z(\tilde{\boldsymbol{c}})])^T$.

Another way to look at the normality assumption is using Gram-Charlier Type A (GCA) series. In general, the GCA series expands any probability density function (pdf), say $\psi(\boldsymbol{x})$, in terms of the differentiations of the Gaussian density $\phi(\boldsymbol{x})$ as a reference pdf. There are various derivations of multivariate GCA series, and a recent paper by Bhaveshkumar (2015) provides a simple representation in terms of the Hermite polynomials as follows:

$$\psi(\boldsymbol{x}) = \phi(\boldsymbol{x}) \left[1 + \sum_{i=3}^{\infty} \boldsymbol{\alpha}_i^T \mathbf{H}_i(\boldsymbol{x}; \mathbf{0}, \Sigma) \right],$$

where $\mathbf{H}_i(\boldsymbol{x};\mathbf{0},\Sigma)$'s are multivariate Hermite polynomials defined by Holmquist (1996), and $\boldsymbol{\alpha}_i$'s are coefficient vectors involving higher order cumulants of $\psi(\boldsymbol{x})$. This expansion assumes that the first and second order cumulants are the same for $\psi(\boldsymbol{x})$ and $\phi(\boldsymbol{x})$. If the higher order cumulants of $\psi(\boldsymbol{x})$ are near to zero, then the error of approximating $\psi(\boldsymbol{x})$ using the Gaussian density $\phi(\boldsymbol{x})$ is small. Consequently, the error of using Stein's identity will be small.

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:

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

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:*

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

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 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 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. In all our numerical studies, we choose the least-norm solution (i.e., the solution with minimum Euclidean norm) for Γ^* whenever it is not unique to reduce the impact of estimation errors[4]. The least-norm solutions perform consistently well in all our numerical tests. Nevertheless, 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:*

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

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 positive semidefinite constraint on Γ, but the resulting problem will not exhibit a nice and explicit characterization of the solution as the unconstrained version. From our numerical experiments, we find that when the persistency values can be accurately estimated, the added positive semidefinite constraint has very little impact on the accuracy of the approximating distributions, i.e., the optimality gaps in term of the expected squared deviation (the objective function) are consistently less than 1% from the cases without the positive semidefinite constraint on Γ. However, we do observe some issues when the persistency parameters are not accurate, which can result in distorted approximating distributions and even cause Problem (Q) to be non-convex numerically. Interestingly, allowing the quadratic estimator to be non-convex help mitigate this problem significantly through the analytical solution. The resulting approximating distributions could still perform very well as we show in Section 5.3.

Before presenting the numerical analysis on the performance of our approximation methods, we would like to illustrate the impact of having quadratic terms in the approximation scheme through the following simple example.

EXAMPLE 1. Let $Z(\tilde{\boldsymbol{c}}) = \max\{\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_n\}$ be the maximum of n i.i.d. standard normal random variables. Then $Z(\tilde{\boldsymbol{c}})$ can be computed through the following linear optimization problem:

$$\begin{aligned} Z\left(\tilde{\boldsymbol{c}}\right) = \max \ & \sum_{j=1}^{n} \tilde{c}_j x_j \\ s.t. \ & \sum_{j=1}^{n} x_j = 1 \\ & \boldsymbol{x} \geq 0. \end{aligned}$$

This problem can be regarded as a project management problem with n parallel activities.

By symmetry, we have $\mathbf{E}[x_j(\tilde{\boldsymbol{c}})] = 1/n$, for all $j = 1, 2, \ldots, n$. Then we can immediately obtain our least squares linear approximation of $Z(\tilde{\boldsymbol{c}})$ as follows:

$$W\left(\tilde{\boldsymbol{c}}\right) = \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right] + \frac{1}{n}\sum_{j=1}^{n} \tilde{c}_j.$$

It is obvious that $W(\tilde{\boldsymbol{c}})$ is normally distributed with mean $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ and variance $1/n$.

To construct our least squares quadratic approximation, we need to find $\mathbf{E}[\tilde{c}_i x_j(\tilde{\boldsymbol{c}})]$, for all $i,j = 1,2,\ldots,n$, which can be computed by using the law of total expectation and the property of symmetry from the i.i.d. condition. When $i=j$, we have

$$\mathbf{E}\left[\tilde{c}_i x_i\left(\tilde{\boldsymbol{c}}\right)\right] = \mathbf{E}\left[\tilde{c}_i | \tilde{c}_i = Z\left(\tilde{\boldsymbol{c}}\right)\right] \mathbf{P}\left(\tilde{c}_i = Z\left(\tilde{\boldsymbol{c}}\right)\right) = \frac{1}{n}\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right], \forall i = 1,2,\ldots,n.$$

When $i \neq j$, by symmetry, $\mathbf{E}[\tilde{c}_i | \tilde{c}_j = Z(\tilde{\boldsymbol{c}})]$ must equal to each other for different i. Note that there are $(n-1)$ such terms. Then from the the law of total expectation, we have

$$\begin{aligned}\sum_{k=1}^{n}\mathbf{E}\left[\tilde{c}_k | \tilde{c}_j = Z\left(\tilde{\boldsymbol{c}}\right)\right] &= \mathbf{E}\left[\tilde{c}_j | \tilde{c}_j = Z\left(\tilde{\boldsymbol{c}}\right)\right] + \sum_{k\neq j}\mathbf{E}\left[\tilde{c}_k | \tilde{c}_j = Z\left(\tilde{\boldsymbol{c}}\right)\right] \\ &= \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right] + (n-1)\,\mathbf{E}\left[\tilde{c}_i | \tilde{c}_j = Z\left(\tilde{\boldsymbol{c}}\right)\right], \forall i,j = 1,2,\ldots,n,\ i \neq j.\end{aligned}$$

It is easy to see that $\sum_{k=1}^{n}\mathbf{E}[\tilde{c}_k | \tilde{c}_j = Z(\tilde{\boldsymbol{c}})] = 0$ using similar argument, so together with the above equations, we get $\mathbf{E}[\tilde{c}_i | \tilde{c}_j = Z(\tilde{\boldsymbol{c}})] = -\mathbf{E}[Z(\tilde{\boldsymbol{c}})]/(n-1)$, and consequently,

$$\mathbf{E}\left[\tilde{c}_i x_j\left(\tilde{\boldsymbol{c}}\right)\right] = \mathbf{E}\left[\tilde{c}_i | \tilde{c}_j = Z\left(\tilde{\boldsymbol{c}}\right)\right] \mathbf{P}\left(\tilde{c}_j = Z\left(\tilde{\boldsymbol{c}}\right)\right) = -\frac{\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right]}{n(n-1)}, \forall i,j = 1,2,\ldots,n,\ i \neq j.$$

Using Corollary 1, we obtain the quadratic estimator of $Z(\tilde{\boldsymbol{c}})$ as follows:

$$\begin{aligned}Q\left(\tilde{\boldsymbol{c}}\right) &= \mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right] - \sum_{j=1}^{n}\frac{\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right]}{2n} + \frac{1}{n}\sum_{j=1}^{n}\tilde{c}_j + \sum_{j=1}^{n}\frac{\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right]}{2n}\tilde{c}_j^2 - \sum_{j_1=1}^{n-1}\sum_{j_2=j_1+1}^{n}\frac{\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right]}{n(n-1)}\tilde{c}_{j_1}\tilde{c}_{j_2} \\ &= \frac{\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right]}{2} + \frac{1}{n}\sum_{j=1}^{n}\tilde{c}_j + \frac{\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right]}{2n}\sum_{j=1}^{n}\tilde{c}_j^2 - \frac{\mathbf{E}\left[Z\left(\tilde{\boldsymbol{c}}\right)\right]}{n(n-1)}\sum_{j_1=1}^{n-1}\sum_{j_2=j_1+1}^{n}\tilde{c}_{j_1}\tilde{c}_{j_2}.\end{aligned}$$

In the above expression,both the second and forth terms are symmetrically distributed around zero, and the skewness is brought into the approximation only through the third term, $\sum_{j=1}^{n}\tilde{c}_j^2$, which follows a chi-squared distribution with n degrees of freedom.

It is interesting to observe that both linear and quadratic estimators require the calculation of $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ only. One can complete the approximations using various results in literature that compute exactly or estimate the expectation of the maximum of finite i.i.d. normal random variables, e.g., Royston (1982).

5. Computational Study

We divide the computational analysis in this section into two main parts. In the first part, we assess the quality of our linear and quadratic estimations by assuming the availability of exact persistency values. In the second part, we build approximating distributions based on the estimated persistency values discussed in the previous section. For both parts, we mainly focus on the application in project management problems, and the performance of our estimators are gauged against various existing methods. All the computational studies in this section are conducted in the MATLAB environment with the YALMIP interface (cf. Löfberg (2004)). Linear programming problems are solved using CPLEX solver and semidefinite programming problems (used in estimating persistency values) are solved using MOSEK solver.

5.1. Performance Measures

We evaluate the performance of the estimators obtained in this paper, against previous methods, using the measure of expected square deviation (ESD), which is also the objective function we try to minimize in obtaining our least square approximations. Unfortunately, some of 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 expected squared deviation from $Z(\tilde{\boldsymbol{c}})$[5]. 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:

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

where F and G are the cumulative distribution functions of two distributions. Note that this is the square of Wasserstein distance of order 2 (cf. Villani (2009), Mallows (1972)).

5.2. Computational Analysis With Exact Persistency Values

The purposes of this analysis are two fold. First, 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. Second, the results from such analysis serve as a benchmark for the following subsection where we discuss the approximations based on estimated persistency values.

By "exact", we mean the persistency values are computed from extensive 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 $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. For small project management problems, extensive simulation is possible because their deterministic versions are solvable in polynomial time. We use 10^6 samples in our simulation to obtain the sample estimates of persistency values.

Example 2. The project network consists of four nodes and five arcs as shown in Figure 1. All activities are independent and normally distributed with mean and variance both equal to one.

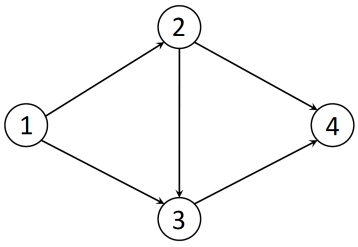

Figure 1 The project network example

The network in Example 2 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 1, 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 1 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 1 Estimation results for Example 2 with simulated parameters for least squares approximating distributions

From Table 1, 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 2 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.

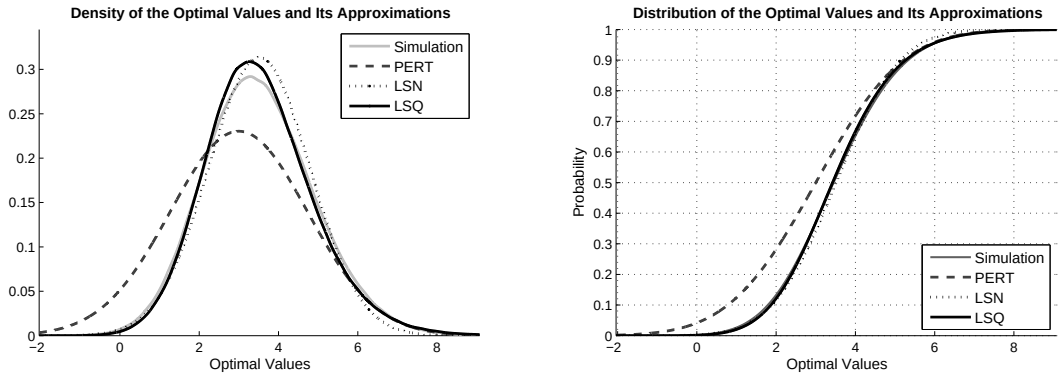

Figure 2 Distributions for Example 2

Note that the numerical values of activity completion times and project completion time can take negative values with a large probability in Example 2. It should be pointed out that these values may not have a practical interpretation in the settings of a project management problem. Nevertheless, the problem is still a valid longest path problem in an acyclic graph, which fits into the framework of Problem (1) we consider in this paper.

For Example 2, 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 3. Approximate the distribution of the maximum of two independent 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.

The results are plotted in Figure 3, 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.

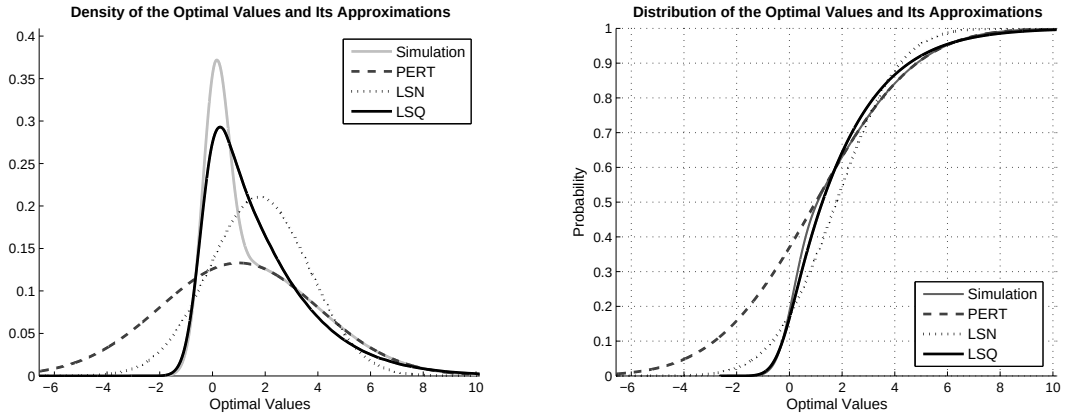

Figure 3 Distributions for Example 3

5.3. Approximating Persistency Values

The performance of our approximation methods hinge on the accuracy of the estimation on the persistency values. In what follows, we will briefly review some estimation methods that we can adopt in the numerical analysis.

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:

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

(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.

5.4. Computational Analysis with Estimated Persistency Values

Consider Example 2 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 5.3. The results are summarized in Table 2, where we add a lower case letter "e" after "LSN" and "LSQ" to indicate the results from estimated persistency parameters.

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 2 Estimation results for Example 2 with estimated parameters for least squares approximating distributions

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 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 4. 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 $0.7\mu_i$[6].

Step 7. Randomly generate a correlation matrix for the activities using the MATLAB function $gallery('randcorr', n)$.

The results for ten random networks are presented in Table 3[7]. The sample size for all the simulations is 2×10^4. From Table 3, 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%.

The machine used to perform all the computation is an Acer personal computer, Veriton X6620G with Intel® Core™ i7-3770 CPU @ 3.40GHz, RAM 16GB, Microsoft Windows 7 Enterprise. The computational times used by different methods for the ten instances reported in Table 3 are summarized in Table 4 (rounded to seconds). Our approach has a clear advantage over simulation in terms of computational effort, because it only requires to solve two optimization problems (one for estimating $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$ and the other for estimating $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$ and $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$) and constructing the approximating distributions is done through analytical equations.

Before we end this section, we present a final example on STA considering correlations between delays. The example is taken from Tsukiyama et al. (2001). As we shall see later, the input values for this example are quite extreme. However, our least squares approximation method can still provide accurate approximations to the delay time distributions, which also demonstrates the robustness of the new approach.

Example 5. Consider the digital circuit and its network representation as shown in Figure 4. All the delay times (i.e., arc lengths) follow normal distributions. Arc $(1,3)$ and $(1,4)$ have distributions of $N(20, 1.4)$, while Arc $(2,3)$, $(2,4)$, $(3,5)$, and $(4,5)$ follow $N(10, 0.7)$. Arc $(1,2)$ is normally distributed with mean μ_{12} and standard deviation σ_{12}. Correlation exists only between Arc $(3,5)$ and Arc $(4,5)$, and their correlation coefficient is denoted as ρ.

For this example, we directly use the mean delay estimates from Tsukiyama et al. (2001) as the value of α for our least squares approximations. Tsukiyama et al. (2001) studied four scenarios of different input parameter values, i.e., dμ_{12}, σ_{12}, and ρ. We report our approximation results under different scenarios in Table 5.

	$\mathbf{E}[T]$	$\sigma(T)$	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
	$m=26$, $n=42$					$m=11$, $n=33$				
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

Instance	2×10^4 Simulation	LSNe	LSQe
$m = 20$, $n = 29$	5533	7	8
$m = 22$, $n = 41$	5551	30	41
$m = 10$, $n = 28$	5502	7	8
$m = 22$, $n = 35$	5620	18	23
$m = 24$, $n = 44$	5334	40	51
$m = 12$, $n = 37$	5446	19	24
$m = 8$, $n = 16$	5434	2	3
$m = 9$, $n = 20$	5440	2	3
$m = 26$, $n = 42$	5572	34	43
$m = 11$, $n = 33$	5434	11	14

Table 4 CPU time (in seconds) used for computation in Example 4

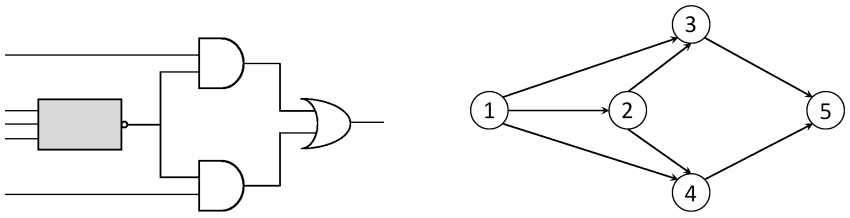

Figure 4 The digital circuit and its network representation in Example 5

As mentioned above, the numerical values are extreme in this example. The coefficients of variation for all the delay times are larger than 14. Furthermore, with the normality assumption on delay times, PERT is expected to provide excellent approximation results, and the impact of correlations is minimal and hence difficult to capture. From Table 5, we find that despite the extreme inputs, our least squares quadratic approximation still gives the best estimate on the variance of circuit delay. Overall, both least squares approximations perform much better than PERT in terms of lower ESD and SND. Although the numerical errors cause the least squares quadratic approxi-

Approximation Method	$\mathbf{E}[T]$	$\sigma(T)$	Error on $\sigma(T)$	$sk(T)$	ESD	SND
$\mu_{12}=10$, $\sigma_{12}=0.7$, $\rho=0.122$						
10^6 simulation	31.32	1.08	-	0.26	-	-
PERT	30.00	1.21	12.28%	0	3.220	1.769
LSNe	31.34	0.86	20.74%	0	0.432	0.054
LSQe	31.34	1.18	8.93%	0.62	0.162	0.018
$\mu_{12}=20$, $\sigma_{12}=1.4$, $\rho=0.800$						
10^6 simulation	40.43	1.66	-	0.01	-	-
PERT	40.00	1.71	3.43%	0	0.585	0.191
LSNe	40.56	1.60	3.43%	0	0.124	0.020
LSQe	40.56	1.66	0.06%	0.26	0.073	0.026
$\mu_{12}=40$, $\sigma_{12}=2.8$, $\rho=0.941$						
10^6 simulation	60.41	2.94	-	0.00	-	-
PERT	60.00	2.97	0.96%	0	0.520	0.171
LSNe	60.57	2.91	0.96%	0	0.118	0.024
LSQe	60.57	2.94	0.07%	0.17	0.062	0.038
$\mu_{12}=80$, $\sigma_{12}=5.6$, $\rho=0.980$						
10^6 simulation	100.40	5.67	-	0.00	-	-
PERT	100.00	5.69	0.27%	0	0.500	0.158
LSNe	100.60	5.64	0.56%	0	0.132	0.041
LSQe	100.60	5.66	0.26%	0.18	0.122	0.097

Table 5 Estimation results for Example 5

mation to have a higher SND than the least squares linear approximation, the order of ESD is still consistent with our optimization scheme and the least squares quadratic approximation provides the lowest ESD.

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. 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

$$\alpha^*=\mathbf{E}\left[Z(\tilde{\boldsymbol{c}})\right]-\Sigma\bullet\Gamma^*,$$
$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}})-\alpha^*-\sum_{j=1}^{n}\beta_j^*(\tilde{c}_j-\mu_j)\right)(\tilde{c}_k-\mu_k)\right]=0,\ \forall k=1,\ldots,n, \tag{4}$$

and

$$\mathbf{E}\left[\left(Z(\tilde{\boldsymbol{c}})-\alpha^*\right)(\tilde{c}_{k_1}-\mu_{k_1})(\tilde{c}_{k_2}-\mu_{k_2})\right]$$
$$-\sum_{j_1=1}^{n}\sum_{j_2=j_1}^{n}\gamma_{j_1,j_2}^*\left(\sigma_{j_1,j_2}\sigma_{k_1,k_2}+\sigma_{j_1,k_1}\sigma_{j_2,k_2}+\sigma_{j_1,k_2}\sigma_{k_1,j_2}\right)=0,\ \forall 1\le k_1\le k_2\le n. \tag{5}$$

Further simplifying Equation (4), 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 (5),

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

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 (5), 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.

Endnotes

[1]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.

[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 given 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.

[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 $\mathbf{E}[\tilde{c}_k \partial x_k(\tilde{\boldsymbol{c}})/\partial c_j]$.

[4]Theoretically, any Γ^* obtained from solving the system of linear equations shown in Theorem 2 should produce the same expected squared deviation for the quadratic estimator. However, this is not true numerically, because we do not solve the quadratic estimation problem exactly in our numerical studies, i.e., the persistency values are all estimated, including $\mathbf{E}[\boldsymbol{x}(\tilde{\boldsymbol{c}})]$, $\mathbf{E}[\tilde{\boldsymbol{c}}\boldsymbol{x}(\tilde{\boldsymbol{c}})^T]$, and $\mathbf{E}[Z(\tilde{\boldsymbol{c}})]$. Values of Γ^* are then determined through solving the system of linear equations involving these estimated persistency values. Therefore, estimation errors are unavoidable, and intuitively, the errors will be exaggerated by choosing Γ^* with larger norms. We have tested various randomized solutions when Γ^* is not unique, and found that other solutions generally give worse approximations than the least-norm solution in all performance measures that we concern.

[5]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.

[6]Note that the random samples thus generated may not represent real life project management problems, where the project activity durations are usually positive. Since the project activity durations are assumed to follow normal distributions–which is consistent with many studies in the literature–it is impossible to completely rule out negative realizations in our simulation. We would like to emphasize that these randomly generated instances are still valid longest path problems, and the results show that our approach is able to find better approximations to the distributions of the longest paths. Nevertheless, one can always scale the means and variances so that the simulated samples contain fewer negative realizations, which can better represent project management problems.

[7]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.

Acknowledgments

The authors would like to thank Professor Aharon Ben-Tal, Professor Amir Beck, the Associate Editor, and other anonymous reviewers for their valuable comments and suggestions on improving this manuscript.

References

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

Agrawal S, Blaauw D, Zolotov V (2003) Statistical timing analysis for intra-die process variations with spatial correlations. *Proceedings of the 2003 International Conference on Computer Aided Design.* 900–907.

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

Andersen EP (1953) On the fluctuations of sums of random variables. *Mathematica Scandinavica.* 1: 263–285.

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

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

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

Bhaveshkumar CD (2015) Multivariate Generalized Gram-Charlier series using only calculus of several variables, instead tensor calculus. *Working paper.*

Blaauw D, Chopra K, Srivastava A, Scheffer L (2008) Statistical timing analysis: From basic principles to state of the art. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.* 27: 589–607.

Bomze IM, Dür M, Klerk ED, Roos C, Quist AJ, Terlaky T, (2000) On copositive programming and standard quadratic optimization problems. *Journal of Global Optimization.* 18: 301–320.

Bowman RA (1995) Efficient estimation of arc criticalities in stochastic activity networks. *Management Science.* 41: 58–67.

Brown GG, Dell RF, Wood RK (1997) Optimization and persistence. *Interfaces.* 27: 15–37.

Burer S (2009) On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming.* 120: 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: 799–809.

Chang H, Sapatnekar SS (2005) Statistical timing analysis under spatial correlations. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems.* 24: 1467–1482.

Chen J, Lin DK, Thomas DJ (2003) On the single item fill rate for a finite horizon. *Operations Research Letters.* 31: 199–123.

Clark EC (1961) The greatest of a finite set of random variables. *Operations Research.* 9: 145–162.

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

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

Dodin BM (1984) Determining the K most critical paths in PERT networks. *Operations Research.* 32: 859–877.

Dodin BM (1985) Bounding the project completion time distribution in PERT networks. *Operations Research.* 33: 862–881.

Dodin BM, Elmaghraby SE (1985) Approximating the criticality indices of the activities in Pert networks. *Management Science.* 31: 207–223.

Ewbank JB, Foote BL, Kumin HJ (1974), A method for the solution of the distribution problem of stochastic linear programming. *SIAM Journal on Applied Mathematics.* 26: 225–238.

Fulkerson DR (1962) Expected critical path lengths in PERT networks. *Operations Research.* 10: 808–817.

Goldstein L, Rinott Y (1996) Multivariate normal approximations by Steins method and size bias couplings. *Journal of Applied Probability.* 31: 1–17.

Hagstrom JN (1988) Computational complexity of PERT problems. *Networks.* 18: 139–147.

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

Holmquist B (1996) The d-variate vector Hermite polynomial of order k. *Linear Algebra and its Applications.* 237/238: 155–190.

Hurst HE (1951) Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers.* 56: 376–385.

James B, James KL, Siegmund D (1987) Test for a change-point. *Biometrika.* 74: 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: 134–139.

Khandewal V. Srivastava A (2007) A quadratic modeling-based framework for accurate statistical timing analysis considering correlations. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems.* 15: 206–215.

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

Kong Q, Lee CY, Teo CP, Zheng Z (2013) Scheduling arrivals to a stochastic service delivery system using copositive cones. *Operations Research.* 61: 711–726.

Larson RC (1990) The queue inference engine: Deducing queue statistics from transactional data. *Management Science.* 36: 586–601.

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

Le J, Li X, Pileggi LT (2004) STAC: Statistical timing analysis with correlation. *Proceedings of the 2004 Design Automation Conference.* 83–88.

Li H, Koh CK, Balakrishnan V, Chen Y (2007) Statistical timing analysis considering spatial correlations. *Proceedings of the 2007 International Symposium on Quality Electronic Design.* 102–107.

Lindsey JH (1972) An estimate of expected critical-path length in PERT networks. *Operations Research.* 20: 800–812.

Liu JS (1994) Siegel's formula via Stein's identities. *Statistics & Probability Letters.* 21: 247–251.

Löfberg J (2004) YALMIP: A Toolbox for Modeling and Optimization in MATLAB. *In Proceedings of the CACSD Conference, Taipei, Taiwan.* Available online at: http://control.ee.ethz.ch/∼joloef/yalmip.php.

Mallows CL (1972) A note on asymptotic joint normality. *Annals of Mathematical Statistics.* 43: 508–515.

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

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

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

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

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

Parrilo PA (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: 211–222.

Ross N (2011) Fundamentals of Steins method. *Probability Surveys.* 8: 210–293.

Royston JP (1982) Algorithm AS 177 - Expected normal order statistics (exact and approximate). *Journal of the Royal Statistical Society.* 31: 161–165.

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

Stein CM (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: 583–602.

Stein CM (1981) Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics.* 9: 1135–1151.

Tang Q, Zjajo A, Meijs N-van-der (2012) Transistor-level gate model based statistical timing analysis considering correlations. *Proceedings of the 2012 Design, Automation & Test in Europe Conference & Exhibition.* pp. 917–922.

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

Tsukiyama S, Tanaka M, Fukui B (2001) A statistical static timing analysis considering correlations between delays. *Proceedings of the 2001 Asia and South Pacific Design Automation Conference.* pp. 353–358.

Villani C (2009) Optimal Transport: Old and New. *Grundlehren der Mathematischen Wissenschaften.* (Springer, Berlin).

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

Zhan Y, Strojwas AJ, Li X, Pileggi LT, Newmark D, Sharma M (2005) Correlation-aware statistical timing analysis with non-Gaussian delay distributions. *Proceedings of the 2005 Design Automation Conference.* 77–82.

Zhang L, Chen W, Hu Y, Gubner JA, Chen CCP (2005) Correlation-preserved non-Gaussian statistical timing analysis with quadratic timing model. *Proceedings of the 2005 Design Automation Conference.* 83–88.

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 differentiable function $h : \mathbb{R} \to \mathbb{R}$ such that $\mathbf{E}[|h'(\tilde{c})|] < \infty$,

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

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

$$Cov\left(\tilde{c}, h\left(\tilde{c}\right)\right) = Var(\tilde{c})\mathbf{E}\left[h'\left(\tilde{c}\right)\right]. \tag{EC.1}$$

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

Let $\tilde{\boldsymbol{z}} = \left(\tilde{z}_1, \ldots, \tilde{z}_n\right)^T$, where $\tilde{z}_j$'s are independent and identically distributed standard normal random variables. From Equation (EC.1) it follows that for any function $\hat{h} : \mathbb{R}^n \to \mathbb{R}$ satisfying the same conditions as h in the Theorem,

$$\mathbf{E}\left[\tilde{z}_j \hat{h}\left(\tilde{\boldsymbol{z}}\right) \middle| \left(\tilde{z}_1, \ldots, \tilde{z}_{j-1}, \tilde{z}_{j+1}, \ldots, \tilde{z}_n\right)\right] = \mathbf{E}\left[\frac{\partial \hat{h}\left(\tilde{\boldsymbol{z}}\right)}{\partial z_j} \middle| \left(\tilde{z}_2, \ldots, \tilde{z}_{j-1}, \tilde{z}_{j+1}, \ldots, \tilde{z}_n\right)\right], \forall j = 1, \ldots, n.$$

Taking the expectation of both sides, we get

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

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,

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

References

Liu JS (1994) Siegel's formula via Stein's identities. *Statistics & Probability Letters.* 21: 247–251.

Stein CM (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: 583–602.

Stein CM (1981) Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics.* 9: 1135–1151.