Useful Bounds on the Expected Maximum of Correlated Normal Variables

Andrew M. Ross *

August 13, 2003

Abstract

We compute useful upper and lower bounds on the expected maximum of up to a few hundred correlated Normal variables with arbitrary means and variances. Two types of bounding processes are used: perfectly dependent Normal variables, and independent Normal variables, both with arbitrary mean values. The expected maximum for the perfectly dependent variables can be evaluated in closed form; for the independent variables, a single numerical integration is required. Higher moments are also available. We use mathematical programming to find parameters for the processes, so they will give bounds on the expected maximum, rather than approximations of unknown accuracy. Our original application is to the maximum number of people on-line simultaneously during the day in an infinite-server queue with a time-varying arrival rate. The upper and lower bounds are tighter than previous bounds, and in many of our examples are within 5 percent of each other.

Subject Classifications: Probability: bounds. Queues: Nonstationary.

1 Introduction

There are many cases where one wants an idea of the maximum load on a system over a period of time. Applications occur in structural engineering to withstand wind, wave, flood, or earthquake forces. Similar problems occur in surge suppression for electronic systems, and in designing power grids that should be able to handle the peak load. Maximum values are also important in applications other than load-determination. For example, critical paths in project scheduling can depend on how long the longest sequence of jobs takes until it is done. Circuit designs depend on how long it takes signals to propagate through a network of gates. Similarly, the lifetime of a system in reliability theory is related to the maximum of certain sums of component lifetimes. Also, factory capacity decisions depend on the maximum expected demand for a portfolio of products,

*Working Paper number 03W-004, Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA. amr5@lehigh.edu, www.lehigh.edu/∼amr5/

as do some inventory decisions. Our application comes from the Internet-access industry, where a company's bill is based in part on the maximum number of their customers that were on-line simultaneously during the day. In this application, we have (for example) 144 correlated Normal variables, each with a different mean and variance.

Extreme Value theory typically deals with order statistics (such as the maximum) as the size of the collection grows to infinity, and gives little indication of how many variables are needed before the limiting distribution of the maximum is a good approximation. It also typically assumes that the variables are IID (Independent and Identically Distributed), or that they form a stationary process. However, in many applications, the size of the collection is not very large, and the variables are what might be called DDD (Dependent and Differently Distributed). That is, the mean and variance might vary from one variable to another (often, changing with time).

In this paper, we focus on the expected value of the maximum, rather than on other moments or the whole distribution. This is because our original industrial problem was solely oriented toward the expected maximum, without any risk-aversion or other similar properties. After the literature survey in Section 2, we discuss in Section 3 the expected maximum of Perfectly Dependent and Differently Distributed (PDDD) Normal variables. Section 4 then treats the IDD (Independent and Differently Distributed) case. From there, Section 5 calculates upper and lower bounds on the DDD case for multivariate Normals. It uses an inequality due to Vitale (2000) for the form of the bounds, and uses mathematical programming to get the required data based on the covariance matrix. Section 6 then demonstrates the technique as applied to the expected maximum number of people in an $M_t/G/\infty$ queueing system at any time point.

We will use the relative distance between the upper and lower bounds as a measure of quality, but we realize that this is not an ideal measure in all cases. This is because it can be made arbitrarily small or large by adding a constant to all of the Normal variables, shifting them toward or away from zero. However, it is the natural measurement for our application. We note that our bounds do not work well for some applications. In particular, the expected maximum of finitely-sampled standard Brownian motion is difficult for our bounds to handle.

In many cases, the expected maximum can be computed fairly well using a simple simulation procedure. However, our technique using bounds can be useful when optimizing a system where the objective function value depends on the expected maximum. In our Internet-service application, we had the opportunity to shape the arrival rate, and wanted to find the optimal shape. Bounds like those presented below do not introduce stochastic noise into the objective function as simulation would do. This makes it easier to compute gradients.

Throughout this paper, we will use an overbar to indicate that a random variable or process has zero mean (such as $\bar{X}_i$), and a tilde to indicate that it has possibly non-zero mean (such as $\tilde{X}_i$). The mnemonic is that one has a "flat-line" mean function, while the other varies up and down, according to the time-of-day. We will use m_i for the mean of $\tilde{X}_i$, and always let $\bar{X}_i = \tilde{X}_i - m_i$.

Both variables will have variance σ^2_{Xi}, and covariances σ_{Xij}.

2 Previous Literature

Jensen's inequality gives us our first bound, since the "max" function is convex. This gives us an easily obtained but not very tight lower bound on the expected maximum:

$$\mathrm{E}\left[\max_i \tilde{X}_i\right] \geq \max_i \mathrm{E}\left[\tilde{X}_i\right] = \max_i m_i$$

Tippett (1925) gives tables for the expected value and variance of the maximum of IID Normals for $n = 2, 5, 10, 20, 60, 100, 200, 500, 1000$. He also gives tables for the CDF of the maximum. Most of the paper is concerned with the distribution of the range, though. Teichroew (1956) gives more detailed tables for $n = 2 \ldots 20$ for all Normal order statistics, along with their products. Again, this is for the IID case only. Clark and Williams (1958) consider the distribution of the order statistics for IID variables, but start by assuming that the CDF inverse is a polynomial. Thus, their method in this form is invalid for Normals. However, they extend it to require only differentiability. Bose and Gupta (1959) also consider the IID Normal case.

Owen and Steck (1962) considers the DID (Dependent but Identically Distributed) case, with standard Normals and all equal correlations. This is done by starting with $n+1$ IID Normals, and transforming them to n DID variables. They then consider multinomial distributions with equal cell probabilities.

Clark (1961) gives exact formulas for the first four moments of the maximum of two Normals in the DDD case. We summarize the formula for the expected maximum here: first, define

$$a \equiv \sqrt{\sigma^2_{X1} + \sigma^2_{X2} - 2\sigma_{X12}} \tag{1}$$

and

$$\alpha \equiv (m_1 - m_2)/a \tag{2}$$

Then

$$\mathrm{E}\left[\max(\tilde{X}_1, \tilde{X}_2)\right] = m_1 \cdot \Phi(\alpha) + m_2 \cdot \Phi(-\alpha) + a \cdot \phi(\alpha) \tag{3}$$

where ϕ and Φ are the Standard Normal density and cumulative distribution functions. He also provides a recursive approximation of the moments of the maximum for three, four, or more Normals, and shows some evidence that the approximation is fairly accurate. The approximation is to treat two of the variables first, and suppose that their maximum also has a Normal distribution, then combine that Normal with the third, etc. We will return to this approximation in Section 6.4. He points out that the completion time of a PERT network (Malcolm et al., 1959) can be represented as the maximum of all paths from start to finish, but that there are often too many paths to consider all of them explicitly. He then develops an approximation method not unlike Dijkstra's algorithm, where the time that each node occurs is updated based on its

predecessors. Kella (1986) gives the Laplace transform of the maximum for the DDD case with two Normal variables, and from it derives the first two moments in formulas that are equivalent to those from Clark.

David (1981) is a standard reference for order statistics, but does not include much on calculations for DDD variables. Leadbetter, Lindgren, and Rootzen (1983) has examples of the maximum of nonstationary processes, such as air pollution levels during the year (which increased during the winter, for the data set in question). Also considered is the maximum of a mixture of distributions, where the mixture is due to normal weather versus hurricanes. Coles (2001) considers several sets of nonstationary data, and explores when the stationarity or lack thereof becomes important. Unfortunately, a statistical approach to existing data sets will not help us answer "what-if" questions when designing systems.

Ross (2003) provides two results of interest. The first is an upper bound on $\mathrm{E}\left[\max \tilde{X}_i\right]$ in the DDD case, without requiring the variables to have Normal distributions. First, we have the bound, for all values of c,

$$\mathrm{E}\left[\max_i \tilde{X}_i\right] \leq c + \sum_{i=1}^{N} \int_c^{\infty} \Pr\left\{\tilde{X}_i > x\right\} dx \tag{4}$$

The optimal value of c (the one which produces the smallest upper bound using this system) is the c which satisfies

$$\sum_{i=1}^{N} \Pr\left\{\tilde{X}_i > c\right\} = 1 \tag{5}$$

Finding this optimal value of c requires a numerical rootfinding procedure. Note that these formulas do not include any covariance between random variables—they are valid for any covariances, but treat the variables independently. Once the value of c is determined, the bound may be evaluated by performing the integration. This bound was also discussed by Lai and Robinson (1976), but only in the DID case. We will refer to this bound as the LRR bound. In Appendix A, we specialize it to the Normal distribution, which allows us to avoid numerical integration.

Ross also provides a formula for the distribution of the maximum of standard Brownian motion on the interval $[0, T]$. The formula is equivalent to saying that the maximum has a Half-Normal distribution. This is one of the few closed-form results available for more than 20 or so correlated Normal random variables, but it goes to the other extreme of an uncountable set of Normals. It is also difficult to generalize the derivation to non-stationary Brownian motion.

Slepian (1962) introduced an inequality that has become a standard theorem in this field. It allows us to compare two zero-mean Gaussian processes and establish stochastic dominance of one maximum over the other. However, it requires that the variances stay the same from one process to the next. Adler (2003), on page 75, mentions that the Sudakov-Fernique inequality relaxes the

equal-variances condition of Slepian's inequality, but then loses the stochastic dominance condition. At that point, it can only guarantee that the expected values of the maxima are ordered. Still, it applies only to zero-mean variables.

Our main method for establishing upper and lower bounds on our DDD variables $\tilde{X}_i$ involves a theorem from Vitale (2000), which we state here in a slightly modified version. Let $\bar{W}_i$, $\bar{X}_i$, and $\bar{Y}_i$, for $i = 1 \ldots N$, be zero-mean DDD Normal random variables such that, for all i, j,

$$\mathrm{E}\left[(\bar{W}_i - \bar{W}_j)^2\right] \leq \mathrm{E}\left[(\bar{X}_i - \bar{X}_j)^2\right] \leq \mathrm{E}\left[(\bar{Y}_i - \bar{Y}_j)^2\right] \tag{6}$$

Then for arbitrary constants m_i,

$$\mathrm{E}\left[\max_i \bar{W}_i + m_i\right] \leq \mathrm{E}\left[\max_i \bar{X}_i + m_i\right] \leq \mathrm{E}\left[\max_i \bar{Y}_i + m_i\right] \tag{7}$$

We have chosen to also use m_i as the mean of $\tilde{X}_i$, so that we can write

$$\mathrm{E}\left[\max_i \tilde{W}_i\right] \leq \mathrm{E}\left[\max_i \tilde{X}_i\right] \leq \mathrm{E}\left[\max_i \tilde{Y}_i\right] \tag{8}$$

It is not too difficult to use simulation to estimate the expected value of the maximum. The multivariate Normal is simulated by using a vector of independent standard Normal values, and multiplying by the Cholesky decomposition of the covariance matrix, then adding the vector of means. This technique is summarized in Chapter 8.1.4 of Tong (1990) and was used in Ross (2001). However, as mentioned above, using simulation introduces noise into the results, which can make optimization more difficult.

A variety of papers have appeared that consider maximum values for queueing systems. Their techniques usually are particular to queueing systems, rather than applying to a wide class of processes (DDD Normal processes). Furthermore, they are typically confined to queueing systems with constant arrival rates, rather than allowing rates to vary with the time of day.

In the next two sections, we compute the expected maximum for two special covariance structures.

3 The Perfectly Dependent Case

It is the arbitrary structure of the covariance matrix that makes the expected maximum hard to compute. By imposing more structure on the covariances, we can obtain a process whose expected maximum is more amenable to computation. Our first simplification is the case when all the components of the process are perfectly correlated. That is, the correlations coefficients can only be $+1$ or -1. In the situation with identical distributions, perfect correlation makes all but the first variable redundant, so $\mathrm{E}\left[\max \tilde{X}_i\right] = m_1 = \ldots = m_N$. When the distributions are not the same, the situation is more complicated, but still relatively friendly.

Let $\tilde{V}_i$ be a set of DDD Normal variables, with all components being perfectly correlated. We call this the Perfectly-DDD, or PDDD, case. We may construct such a set by starting with a single standard Normal Z, and defining

$$\tilde{V}_i = s_i \cdot Z + m_i$$

where s_i may be any real number. This allows some correlations to be $+1$, while others are -1. We can then compute the expected maximum by conditioning on Z, then unconditioning. Note that $\mathrm{E}\left[\max \tilde{V}_i \,|Z = z\right]$ is a deterministic (and easily computed) function; we will call it $h(z)$. It is a convex, piecewise-linear function with at most one segment for each $\tilde{V}_i$ variable. In our application, below, it tends to have only a few segments, instead of one for each variable. If all the s_i are positive, $h(z)$ is increasing, and it is decreasing if all s_i are negative. In any case, the expected maximum is

$$\mathrm{E}\left[\max \tilde{V}_i\right] = \mathrm{E}\left[\mathrm{E}\left[\max \tilde{V}_i \,|Z\right]\right] = \int_{-\infty}^{\infty} h(z)\phi(z)dz$$

To compute the integral, we break it up into segments based on the breakpoints in $h(z)$. Let $z_1, z_2, \ldots$ be the breakpoints in increasing order, and suppose that variable $\tilde{V}_7$ is responsible for the value of $h(z)$ between z_2 and z_3. The integral for that portion is then

$$\int_{z_2}^{z_3} (s_7 \cdot z + m_7)\phi(z)dz \tag{9}$$

The first breakpoint is $-\infty$, and the last is $+\infty$. These integrals can be computed without numerical integration, using a technique similar to the one used in Appendix A. Thus, we have a rapid way to compute the expected maximum exactly in the PDDD case.

Before we go on, we will mention some interesting properties of the PDDD case. First, using $(-1)\cdot\vec{s}$ will give the same results as $\vec{s}$, because the distribution of Z is symmetric around zero. Next, the expected value doesn't change if we add a constant to each value of s_i. That is, if $\vec{1}$ is a vector of ones, using $\vec{s}+\delta\cdot\vec{1}$ gives the same expected maximum as using just $\vec{s}$, for any real δ, positive or negative. We can see this as follows:

$$\mathrm{E}\left[\max_i \left(\left(s_i + \delta\right) Z + m_i\right)\right] = \mathrm{E}\left[\delta Z + \max_i \left(s_i \cdot Z + m_i\right)\right] =$$

$$\delta\mathrm{E}\left[Z\right] + \mathrm{E}\left[\max_i \left(s_i \cdot Z + m_i\right)\right] = \mathrm{E}\left[\max_i V_i\right]$$

The last equality is because $\mathrm{E}\left[Z\right] = 0$. In contrast to the mean, the variance of the maximum changes with δ according to a convex quadratic function. The variance of the maximum is computed almost as easily as the expected value. Also, the gradient of the expected value with respect to $\vec{s}$ is not hard to compute analytically (see Appendix B), eliminating the need for finite-differences. We have found empirically that $\mathrm{E}\left[\max\right]$ is convex as a function of $\vec{s}$.

Next, we discuss a second way to restrict the covariance structure that makes the expected maximum computable. The random variables will be independent, but might have different distributions (the IDD case).

4 The Independent, Different Distributions Case

Suppose that we have a collection of independent random variables $\tilde{W}_i$ for $i = 1 \ldots N$; they may have different means and variances. The cumulative distribution function (CDF) of their maximum value is

$$\Pr\left\{\max_i \tilde{W}_i \leq w\right\} = \prod_{i=1}^{N} \Pr\left\{\tilde{W}_i \leq w\right\} \tag{10}$$

We could obtain the density by taking the derivative, using the product rule. However, we end up with a sum of products that takes roughly N times longer to evaluate than the CDF. Instead, we will use the CDF directly. It is well known that for any non-negative random variable R, the mean value of R may be computed using

$$\mathrm{E}\left[R\right] = \int_0^\infty \Pr\left\{R > r\right\} dr \tag{11}$$

A somewhat less common formula from David (1981), among other places, extends this to the case where the variable may take any value, positive or negative

$$\mathrm{E}\left[X\right] = \int_0^\infty (\Pr\left\{X > x\right\} - \Pr\left\{X < -x\right\}) dx \tag{12}$$

Combining Eqn. 10 with Eqn. 12, we get

$$\int_0^\infty \left(1 - \prod_{i=1}^{N} \Pr\left\{\tilde{W}_i \leq w\right\} - \prod_{i=1}^{N} \Pr\left\{\tilde{W}_i < -w\right\}\right) dw \tag{13}$$

Performing numerical integration gives us a relatively easy way to compute the expected maximum in the IDD case. Even though the integral has an infinite domain, the integrand approaches 0 very rapidly after a while, so not much is lost by stopping the integration then. In particular, we stopped integrating when the integrand underflowed using floating point arithmetic. That is, when the product terms come within roughly 10^{-16} of 1, then $1 - \prod$ evaluates to zero.

This formulation can easily accommodate some of the variables having a fixed value (a variance of zero). Such distributions can easily arise in applications, and in the bounding technique we will discuss below. Suppose that variable 10 has the largest mean of all the zero-variance distributions, and that $m_{10} \geq 0$. Then, we may start the numerical integration from m_{10} instead of from 0, because on the interval $[0, m_{10})$ we will have $\Pr\left\{\tilde{W}_{10} \leq w\right\} = 0 = \Pr\left\{\tilde{W}_{10} \leq -w\right\}$. Thus, on that interval, the integrand is exactly 1, which does

not require numerical integration. Furthermore, the $\prod \Pr\left\{\tilde{W}_i < -w\right\}$ term will be zero, and so may be left out. If the largest zero-variance distribution has a negative mean, the computation is still simplified somewhat, but not as much.

In the IDD case, we can compute higher moments using a formula similar to Eqn. 12:

$$\mathrm{E}\left[X^n\right] = \int_0^\infty nx^{n-1} \cdot \left(\Pr\left\{X > x\right\} + (-1)^n \Pr\left\{X < -x\right\}\right) dx \tag{14}$$

Since we are already evaluating the two CDF values just to get the expected value, computing this integral comes without much extra effort.

Having examined the expected value of the maximum in both the PDDD and IDD cases, we next turn to the DDD case with arbitrary covariance structures, and compute bounds on the expected maximum.

5 Bounds for the Correlated Case

Now that we have ways to calculate the expected maximum in the PDDD and IDD cases, we will compute bounds in the Normal DDD case by using the theorem from Vitale, in Eqns. 6 and 7. Note that, since the $\bar{X}_i$ have a mean of zero,

$$\mathrm{E}\left[(\bar{X}_i - \bar{X}_j)^2\right] = \sigma^2_{Xi} + \sigma^2_{Xj} - 2\sigma_{Xij} \equiv b_{ij} \tag{15}$$

Furthermore, this quantity is always non-negative, since it is the expectation of a squared quantity.

To compute a value for our lower bounds, we need to specify the variances and covariances of $\bar{W}_i$ so that:

1. we can actually compute $\mathrm{E}\left[\max_i \tilde{W}_i\right]$, and
2. Eqn. 6 is satisfied, and
3. $\mathrm{E}\left[\max_i \tilde{W}_i\right]$ is as large as we can reasonably make it.

The upper bound is similar, except we want to minimize rather than maximize. To be able to compute the expected maximum, we will first restrict ourselves to the PDDD case, in Section 5.1. Then, we will use the IDD case in Section 5.2.

5.1 Using PDDD Variables

To use PDDD variables to establish bounds on the expected maximum, we will keep the same mean values m_i for each variable. We must decide on the s_i values such that the conditions of Vitale's theorem are satisfied. The mathematical program for the lower bound is

$$\begin{array}{lll} (\text{maximize} & & \mathrm{E}\left[\max_i \tilde{V}_i\right]) \\ \text{s.t.} & \forall i,j: & (s_i - s_j)^2 \le b_{ij} \\ & \forall i: & s_i \text{ is unrestricted in sign} \end{array} \tag{16}$$

where b_{ij} comes from Eqn. 15. Our objective function is shown in parentheses because finding an optimal solution is not vital: any feasible solution establishes a bound. The inequalities for an upper bound are similar:

$$\begin{array}{lll} (\text{minimize} & & \mathrm{E}\left[\max_i \tilde{Y}_i\right]) \\ \text{s.t.} & \forall i,j: & (s_i - s_j)^2 \geq b_{ij} \\ & \forall i: & s_i \text{ is unrestricted in sign} \end{array} \tag{17}$$

Here, we are using $\tilde{Y}_i$ as the PDDD variables for the upper bound; hopefully this will not cause confusion with the IDD case, below.

These two programs have nonlinear (quadratic) constraints. For the lower bound, we can convert $(s_i - s_j)^2 \leq b_{ij}$ into two simultaneous linear inequalities:

$$(s_i - s_j) \leq \sqrt{b_{ij}} \qquad \text{and} \qquad -(s_i - s_j) \leq \sqrt{b_{ij}}$$

However, the feasible region for the upper bound program is not convex, so when we linearize the constraints we end up with a disjunctive condition: $(s_i - s_j)^2 \geq b_{ij}$ becomes

$$(s_i - s_j) \geq \sqrt{b_{ij}} \qquad \text{or} \qquad -(s_i - s_j) \geq \sqrt{b_{ij}}$$

This makes the problem much harder to solve for the upper bound than it is for the lower bound. However, we can at least find a starting feasible solution of the form $s_i = i \cdot \max \sqrt{b}$, though this is probably gives results far from the true expected value. Another initial feasible solution is $s_1 = 0$ and $s_j = \max_{i=1\ldots j-1}(s_i + \sqrt{b_{ij}})$. Either of these may be done with any ordering to the variables.

Because any feasible solution to the constraints gives a bound, it is not necessarily important to truly minimize or maximize the nonlinear objective function. We suggest starting with a linear objective function, whose weights are chosen heuristically. In the PDDD case, large expected maxima are obtained when two variables with large means have a large difference in their s_i values. For example, if s_4 is large and positive, and s_5 is large and negative, then regardless of the value of Z, at least one of the variables $\bar{V}_4$, $\bar{V}_5$ will be large and positive. From this reasoning, two obvious weight functions are $[-1, +1, -1, \ldots]$ and $[0, 0, \ldots, 0, +1, -1, 0, \ldots, 0]$, where the two nonzero components are near the maximum value of m_i. In some of our experiments, these two objective functions gave different values of $\vec{s}$, but the same $h(z)$ function and therefore the same final value of $\mathrm{E}[\max]$. In other experiments, the resulting $h(z)$ functions were only slightly different, and gave the same bounds to within 5 digits. When we used the true non-linear objective function on small problems, no change in the final value of $\mathrm{E}[\max]$ was obtained.

It is interesting to note that if $\vec{s}$ is a feasible solution to either of these programs (upper or lower bound), then $\vec{s} + \delta \cdot \vec{1}$ is also a feasible solution. Fortunately, as noted above, the objective function value is the same. The extra degree of freedom makes the feasible region unbounded in a way that can

confuse some optimization routines. To remove the problem, one may add a constraint like $s_1 = 0$ or $\sum s_i = 0$, since any feasible vector $\vec{s}$ may be converted to one of these standardized forms.

5.2 Using IDD Variables

We next present another way to obtain bounds on the DDD case, using IDD variables. Again, the mean values m_i are already determined, and we must decide on the variances such that the conditions of Vitale's theorem are satisfied. The covariances will be zero. To satisfy Eqn. 6, we need to find a feasible solution to the set of inequalities

$$\begin{array}{lll} (\text{maximize} & & \mathrm{E}\left[\max_i \tilde{W}_i\right]) \\ \text{s.t.} & \forall i,j: & \sigma^2_{Wi} + \sigma^2_{Wj} \leq b_{ij} \\ & \forall i: & \sigma^2_{Wi} \geq 0 \end{array} \tag{18}$$

where b_{ij} comes from Eqn. 15. Again, our objective function is shown in parentheses because finding an optimal solution is not vital: any feasible solution establishes a bound. The inequalities for an upper bound are similar:

$$\begin{array}{lll} (\text{minimize} & & \mathrm{E}\left[\max_i \tilde{Y}_i\right]) \\ \text{s.t.} & \forall i,j: & \sigma^2_{Yi} + \sigma^2_{Yj} \geq b_{ij} \\ & \forall i: & \sigma^2_{Yi} \geq 0 \end{array} \tag{19}$$

These problems have trivial feasible solutions ($\sigma^2_{Wi} = 0$ and $\sigma^2_{Yi} = \max b$), so our bounds always exist. Other simple solutions, better than the trivial ones, are $\sigma^2_{Wi} = 0.5 \min b$ and $\sigma^2_{Yi} = 0.5 \max b$. If all of the original covariances are non-negative, then we can get an upper bound without solving the program by simply ignoring the covariances, and letting $\sigma^2_{Yi} = \sigma^2_{Xi}$. A similar idea works for the lower bound if the original covariances are all non-positive. These simple bounds might not be near optimal, though.

We have the choice of using the standard deviations or the variances as our decision variables. If we use the standard deviations, we find that the objective function is empirically convex, but the constraints are nonlinear. In the lower-bound case, the feasible region is still convex, but not in the upper-bound case. If we use the variances, the objective function is not convex, but the constraints are linear, and the feasible region is convex for both the lower and upper bound problems. We have chosen to use the variances, because linear inequalities and a convex feasible region are much easier to deal with than the nonlinear constraints. Furthermore, the non-convexity of the objective is not as serious as it first appears. If we suppose that it is convex when using standard deviations, then its maxima occur at extreme points of the feasible region. If we then change to using variances, we see that the maxima will still be on the boundary, though possibly not at an extreme point. Similarly, any local minimum will be a global minimum when trying to find the smallest upper bound, whether we use the standard deviations or the variances.

Whether we use the standard deviations or variances as our decision variable, it takes N integrations to compute the gradient, whether done by taking the derivative symbolically, or by finite differences. Again, because any feasible solution gives us a bound, we suggest starting by using a linear objective function. If the bounds it gives are not satisfactory, then further effort can be put into solving the NLP, perhaps using the LP solution as a starting point. The program has one constraint for each pair of variables, which leads to a constraint matrix that is tall and thin, rather than short and wide. Thus, we have seen much faster solution times when solving the dual rather than the primal.

In using a linear program to find the bounds, we must choose an objective function. A few heuristics suggest themselves:

1. Emphasize just one variance
2. Emphasize the two or three variables with the highest means
3. Weight all variables using some function of their means

We have found that the final values of the bounds are not very different in our examples. For that reason, we have used the simple objective of equal weight on all variables.

We note briefly that our upper and lower bounds are equal in the 2-variable case, and therefore give the same answer that Eqn. 3 gives. This is because we may let $\sigma^2_{W1} = \sigma^2_{X1} + \sigma^2_{X2} - \sigma_{X12}$ and $\sigma^2_{W2} = 0$ to get the same value of a (in Eqn.1) as the $\tilde{X}_1, \tilde{X}_2$ variables give.

6 A Queueing Demonstration

In some sectors of the dial-up Internet access industry, one company will rent its modem banks to another company. The bill is based on the maximum number of simultaneous sessions seen in the system during the day. We will assume that the system has plenty of capacity, so that we can treat it as an $M_t/G/\infty$ system. This can be justified for modem banks that were built during the peak of the economic boom, but whose traffic has not risen to meet expectations. It is not uncommon to leave such a facility as it is, rather than canceling the phone lines, to avoid the trouble of rebuilding it later. Ross (2001) explored the optimal way to split the traffic between a company that uses this peak-based billing and one that bills on an hourly basis. Here, we use our upper and lower bounds on the expected maximum to get a good estimate of the expected peak of the day. We ignore the usual variation in average service durations throughout the day, though it is not difficult to include.

We will start by assuming that we know the arrival rate function, $\lambda(t)$, of a non-homogeneous Poisson process. Let $Q(t)$ be the number in the system at time t; it has a Poisson distribution with mean

$$m_\lambda(t) \equiv \int_{-\infty}^{t} \lambda(u) G^C(t-u) du \tag{20}$$

where $G^C(x) = \Pr\{S > x\}$ is the tail probability of the call duration distribution. For more information on this model, see Palm (1988), and Eick, Massey, and Whitt (1993ab). We use a subscript λ to denote the arrival rate that gives this particular mean. The covariance between time points is (adapted from Ross, 2003)

$$\mathrm{Cov}\left(Q(t), Q(t+\delta)\right) = \int_{-\infty}^{t} \lambda(u) \cdot G^C(t+\delta-u)du$$

If we let $\delta = 0$ we get $\mathrm{Var}\left(Q(t)\right)$. Since the number in the system at any time point is Poisson, the variance is equal to the mean: $\mathrm{Var}\left(Q(t)\right) = m_\lambda(t)$.

In the modem systems that we have dealt with, the average number of people on-line is typically large (300 or more), enough that a Normal approximation to the Poisson distribution is easily justified. This is what allows us to use the methods of Sections 4 and 5. At the points of very low usage around midnight, the Normal approximation might be more difficult to justify, but it is practically impossible for those time points to make a contribution to the maximum value, so we will ignore the goodness-of-fit issue.

While this describes the process in continuous time, it is difficult to analyze such a non-stationary process. Instead, we will sample the system at 10-minute intervals, and take the maximum. This actually matches the way it is done in industry. So, our problem is to find

$$\mathrm{E}\left[\max Q(00{:}00), Q(00{:}10), \ldots, Q(23{:}50)\right]$$

where our 144 samples have time-varying means and variances, and positive covariances. We will also also explore the effects of changing the 10-minute sample interval. Treating the $Q(t)$ variables as $\tilde{X}_i$, we want to find upper and lower bounds using the mathematical programming of Section 5.

Figure 1 shows the mean number in system $m_\lambda(t)$ for one particular sinusoidal arrival rate function, along with plus and minus two standard deviations (dotted curves), and a random sample. It also shows, off to the side, the value of $\mathrm{E}\left[\max\right]$, along with errorbars that indicate $\pm 2\sqrt{\mathrm{Var}\left(\max\right)}$, computed via simulation.

In Figure 2, we show the standard deviations of the original $\tilde{X}_i$ variables and the IDD $\bar{W}_i$ variables. The latter are from solving an LP with equal weights, though we show standard deviations instead of variances because the units are easier to interpret. The zigzags in the lower bound variables are common in our other examples, below. We have not graphed the upper-bound results, because they are practically indistinguishable from the original process–in particular, no zigzagging is seen.

Next, we generalize this arrival rate by allowing the height of the peak to vary, and explore changes to a few different system parameters.

6.1 A Class of Arrival Rate Functions

To investigate the behavior of the bounds in a more general situation, we will consider a class of sinusoidal arrival rate functions, similar to those used by

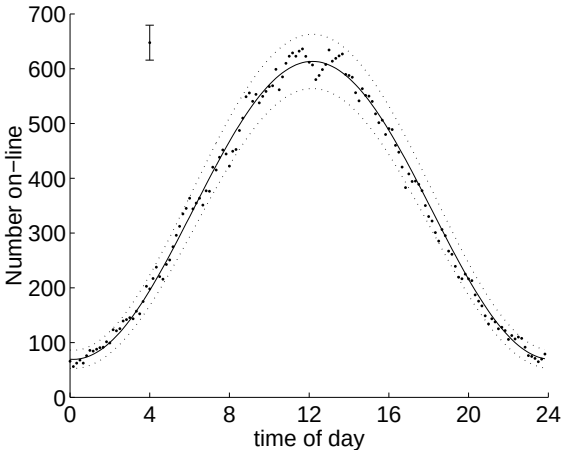

Figure 1: Mean, plus and minus two standard deviations, and a sample path

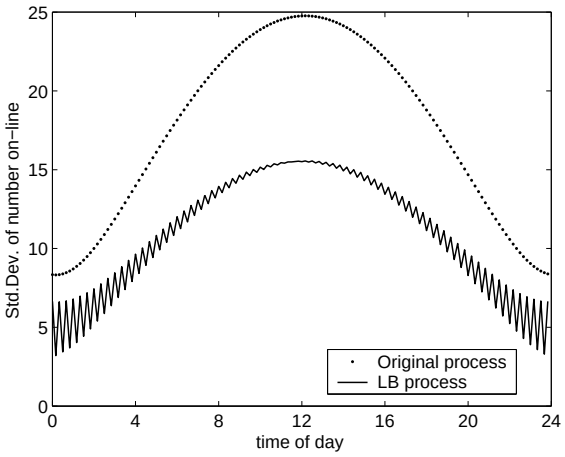

Figure 2: Standard deviations from the original and lower-bound processes

Green and Kolesar (1991):

$$\lambda(t) = \hat{\lambda} \cdot (1 - RA \cdot \cos(2\pi t/24))$$

where $\hat{\lambda}$ is the average arrival rate over the whole day, RA is the relative amplitude (between 0 and 1), and time t is measured in hours. We will consider the effects of three types of changes: the relative amplitude, the number of samples per day, and the service rate. The first relates to the scale of the process and what proportion of sample points are important in determining the maximum value (for higher relative amplitudes, fewer samples are near the maximum). The other two relate more to the dependence between sample points.

Our center case (illustrated in Fig. 1) is $\hat{\lambda} = 1024$ calls per hour, Exponential call durations with an average of 20 minutes per call ($\mu = 3$ calls per hour), and one sample every 10 minutes (144 per day). The relative amplitude is 0.8 unless otherwise stated. We have chosen Exponential durations not because we rely on any memoryless properties, but simply out of tradition. Other distributions (including heavy-tailed) would work just as well.

In Figure 3, we let the relative amplitude vary from 0 to 1 and plot the responses of our bounds and estimates. In the center is a line determined by simulation of the Gaussian process. No errorbars are shown on the data points from the simulation because they are small enough to be irrelevant. On either side are the bounds found using IDD variables and a linear objective function, putting equal weight on all variances. This gives us values for σ^2_{Wi} and σ^2_{Yi}, which are then used to calculate $\mathrm{E}\left[\max \tilde{W}_i\right]$ and $\mathrm{E}\left[\max \tilde{Y}_i\right]$ using the methods of Section 4. Immediately below the IDD lower bound is the PDDD lower bound. The lowest curve is from Jensen's inequality, and the uppermost curve is the LRR bound from Eqns 4 and 5. We see that the increase in the expected maximum is roughly linear for medium-large values of RA, where the peak arrival rate affects the expected maximum directly. However, for small values of RA, the expected maximum is not linear in RA. This is because the arrival rate peak is, heuristically speaking, washed out by the noise inherent in the system, and increasing RA a little bit does not increase the expected maximum much above the noise. A similar effect was seen in Ross (2001). The distance between the upper and lower bounds from the LP stays nearly unchanged as the relative amplitude increases. The relative distance decreases from 4.5 to 2.5 percent. For relative amplitudes very near 1, the number of people on-line at the lowest point might be so small that the Normal is no longer a valid approximation for the Poisson. However, these times have a negligible impact on the distribution of the maximum, so the approximation quality is not at all important. In fact, we could probably ignore any time point that was, say, 4 standard deviations below the Jensen's inequality lower bound, though we have not done so.

Conspicuously absent from Figure 3, and all subsequent figures, is any PDDD upper bound. This is because of the difficulty of finding good solutions to the mathematical program. The two feasible solutions given in Section 5.1 give values of the expected maximum that are far in excess of the LRR bounds. This held true even after considering various ways to order the variables. The

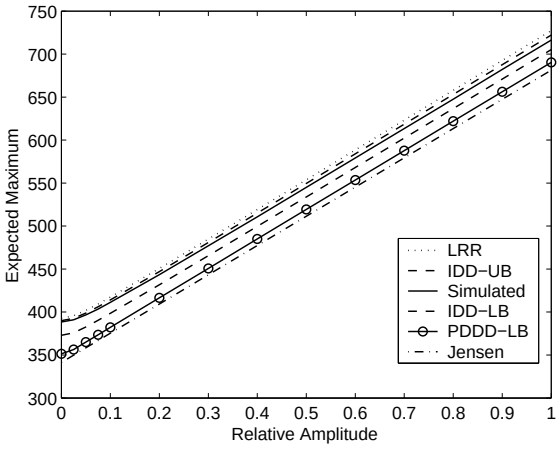

Figure 3: Bound behavior as the Relative Amplitude increases

orderings $1 \ldots 144$ and $144 \ldots 1$ produced essentially equal bounds. The orderings from largest to smallest mean (and vice versa) gave worse results (but equal to each other). Random orderings gave the worst results.

In Figure 4, we go back to $RA = 0.8$ but change the number of samples during the day. We might do this to get a better idea of the continuous-time maximum by sampling more often, or to see how a proposed change in the agreed-upon sampling interval would affect costs. As the samples become closer together, their correlations rise, and this makes the lower bounds from the LP not as tight. This is because the right-hand side values b_{ij} (from Eqn. 15) decrease as the covariances increase. The errorbars shown are at plus and minus 2 standard errors from the mean. The relative distance between the upper and lower bounds from the IDD processes is 0.01 percent for 60-minute samples, and increases to 6.5 percent for 2-minute samples.

In Figure 5, we go back to 144 samples per day, but change the average service duration. This is because different Internet service providers see different customer behavior. Here, we have taken special care to keep the mean values the same for different service durations. This is done by computing the damping coefficient as in Eick, Massey, and Whitt (1993a), and increasing the relative amplitude to compensate for the damping, so that the resulting mean-value curve has a relative amplitude of 0.8 in all cases. As the service durations get longer, the correlation between sample points goes up (people who were on-line at 1:10pm are more likely to still be on-line at 1:20pm, so the samples are less independent). Again, this affects the lower bounds from the mathematical programs through the right-hand side values. For a mean service of 120 minutes, the relative distance between the bounds is 4.9 percent, but it decreases to 0.48 percent when the mean service is 5 minutes.

Interestingly, the optimal solution for the upper bound LP is not very differ-

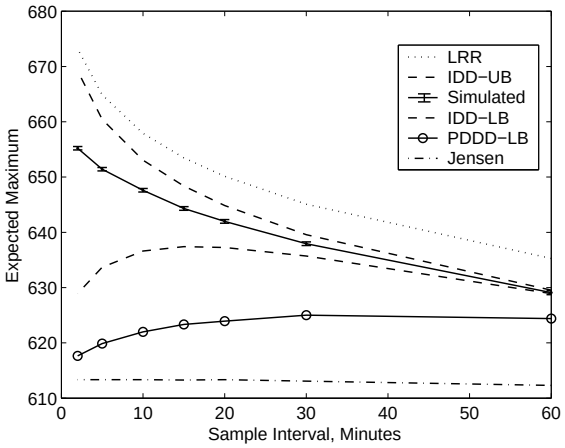

Figure 4: Bound behavior as the number of samples changes

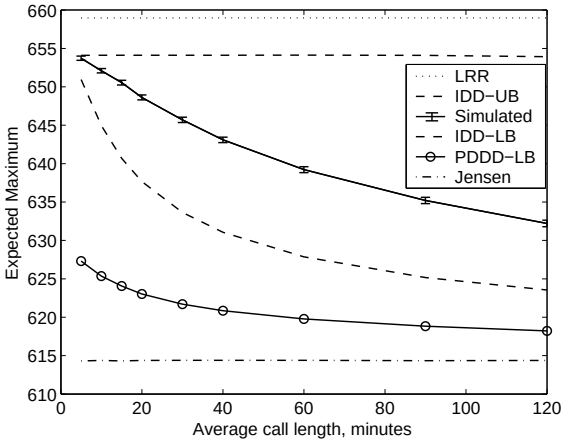

Figure 5: Bound behavior as the average service duration changes

ent than the original variances of the $\tilde{X}_i$, and so the value of the upper bound is practically the same as if we had chosen $\sigma^2_{Yi} = \sigma^2_{Xi}$ (as we mentioned in Section 5, since all covariances are positive) and not run the LP. However, we will see a case (below) where they are substantially different.

We mentioned that, in the IDD case, we can calculate higher moments of $\max \tilde{W}_i$ and $\max \tilde{Y}_i$ using Eqn. 14. However, there is no reason that they will be bounds on the higher moments of $\max \tilde{X}_i$. We found that, empirically,

$$\mathrm{Var}\left(\max \tilde{W}_i\right) \leq \mathrm{Var}\left(\max \tilde{Y}_i\right) \leq \mathrm{Var}\left(\max \tilde{X}_i\right)$$

in every case for the situations from Figures 3-5. In some sense, this is surprising—we might expect $\tilde{Y}_i$ to produce an upper bound on the variance rather than a lower bound, because it gives an upper bound on the expected value. However, since the variables $\tilde{Y}_i$ are IDD, where $\tilde{X}_i$ are DDD, we might expect that the positive correlations of the $\tilde{X}_i$ would increase the variance of the maximum, much as they would increase the variance if we were to take the sum.

6.2 Uncertainty in the Arrival Rate Function

In practical applications, we never know exactly what the non-homogeneous Poisson arrival rate $\lambda(t)$ is going to be. It is affected by weather, breaking news, and other unexpected events. The variation we see in arrival rates is much more than predicted by a Poisson process. For example, if we forecast an arrival rate of 100 calls for a particular hour next week, a Poisson model would say that we should see 100 plus or minus 10 calls (one standard deviation). However, from real data sets we see a standard deviation more on the scale of 20 or 30 calls. For this reason, we will model the arrival rate itself as being uncertain. See the monograph by Grandell (1997) for a general view of these types of models, which are sometimes called Cox processes, after Cox (1955).

There are many ways to model the uncertainty, but we will consider a very simple one. We will suppose that we know the shape of the arrival rate precisely, but its scale is subject to some forecast error. That is, for a particular known shape function $\ell(t)$, the arrival rate is

$$\Lambda(t) = S \cdot \ell(t)$$

where S is a random variable, and we have a prior distribution for it. We would typically take $\mathrm{E}[S] = 1$, but will leave it general for now. Let the prior CDF be $F_S(s)$. The multiplier is chosen once, just before the start of the day, rather than continuously changing as the day goes on. There is some evidence for this simple model being appropriate, as discussed in Thompson (1999), Henderson and Chen (2000), and Brown et al. (2002).

To compute the mean and variance at each sample point, and the covariance between sample points, we condition on the value of S and then uncondition:

$$m_{S\ell}(t) \equiv \mathrm{E}\left[\mathrm{E}\left[m(t)\,|S\right]\right] = \int_0^\infty \int_{-\infty}^t s \cdot \ell(u) G^C(t-u) du \; dF_S(s)$$

Here, we have left the inner integral starting at $-\infty$ even though the previous days have different values of S. This is not a problem as long as the average service duration is no more than a few hours, and the peak is not near the start of the day. Manipulating the integral further, we get

$$m_{S\ell}(t) = \int_0^\infty s \int_{-\infty}^t \ell(u) G^C(t-u) du \; dF_S(s) = \int_0^\infty s m(t) dF_S(s) =$$

$$m_\ell(t) \cdot \int_0^\infty s dF_S(s) = m_\ell(t) \cdot \mathrm{E}[S]$$

which is not surprising. It could also be seen by noting that the $M_t/G/\infty$ model is a linear system, as mentioned by Eick, Massey, and Whitt (1993a).

Now, $Q(t)$ is a mixture of Poisson distributions, instead of being a single Poisson, so we cannot get the second moments of $Q(t)$ as easily as before. To get the covariance between $Q(t)$ and $Q(t+\delta)$, we will use the conditional covariance formula, which in general terms is

$$\mathrm{Cov}(Y, Z) = \mathrm{E}[\mathrm{Cov}(Y, Z \,|X\,)] + \mathrm{Cov}(\mathrm{E}[Y \,|X\,], \mathrm{E}[Z \,|X\,])$$

In our terms, we have $\mathrm{Cov}(Q(t), Q(t+\delta)) =$

$$\begin{aligned}
\mathrm{E}[\mathrm{Cov}(Q(t), Q(t+\delta) \,|S\,)] &+ \mathrm{Cov}(\mathrm{E}[Q(t) \,|S\,], \mathrm{E}[Q(t+\delta) \,|S\,]) = \\
\mathrm{E}\left[\int_{-\infty}^t S \cdot \ell(t) \cdot G^C(u+\delta) du\right] &+ \mathrm{Cov}(S \cdot m_\ell(t), S \cdot m_\ell(t+\delta)) = \\
\mathrm{E}[S] \int_{-\infty}^t \ell(t) \cdot G^C(u+\delta) du &+ m_\ell(t) \cdot m_\ell(t+\delta) \cdot \mathrm{Var}(S)
\end{aligned}$$

To get the variance at any particular time, we let $\delta = 0$ to get $\mathrm{Var}(Q(t)) =$

$$\mathrm{E}[S] \int_{-\infty}^t \ell(t) \cdot G^C(u) du + m_\ell(t)^2 \cdot \mathrm{Var}(S) = \mathrm{E}[S] \, m_\ell(t) + m_\ell(t)^2 \cdot \mathrm{Var}(S)$$

Note that all of these formulas depend only on the first two moments of S, rather than the whole distribution. Thus, we do not need to decide exactly which type of prior (Normal, Gamma, etc.) to use. However, we assume that the final distribution of $Q(t)$ is still well approximated by a Normal distribution. This will be the case if the distribution has a strong central tendency (many Normals and Gammas). A case where it would not hold true would be a widely spread two-point discrete prior distribution that results in a bimodal final distribution of $Q(t)$. Another case would be an Exponential prior, which would give a Geometric final distribution.

In Figure 6, we start with our central case for $\ell(t)$, set $\mathrm{E}[S] = 1$, and let the coefficient of variation of the unknown scale S change from 0 to 30 percent. We have also added a curve showing the upper bound that results from letting $\sigma_{Yi}^2 = \sigma_{Xi}^2$, which we call "no covar" in the graph. This is the simple feasible

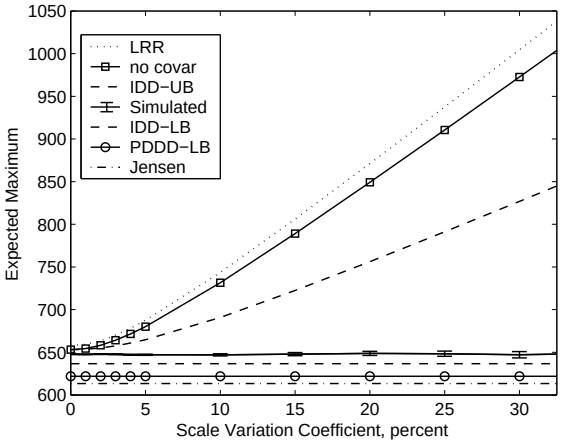

Figure 6: Bound behavior as the forecast uncertainty changes

solution to the IDD mathematical program in the case when all covariances are positive, suggested in Section 5. In the case where the arrival rate function was known exactly, this solution was similar to the solution of the LP for the IDD variables, but now in the uncertain-scale case it is substantially different. The relative distance between the upper and lower bounds from the LP in this case is not as good as in the previous examples: it is roughly equal to the coefficient of variation of the scale factor. That is, the bounds are within 4.5 percent of each other when the variation coefficient is 5 percent, and they are within 29.89 percent of each other when the variation coefficient is 30 percent. Nonetheless, they are still much closer together than the LRR bound and the Jensen's inequality bound.

Figure 7 is analogous to Figure 2: it shows the standard deviations from our two bounding processes, along with those from the original process. Now, the upper bound process is substantially different from the original. We still see the zigzagging in the lower bound process, but again not in the upper bounds.

Figure 8 shows the optimal solutions for the PDDD lower bound, for two different objective functions. The first uses the weights $[-1, +1, -1, \ldots]$, and the second uses $[0, 0, \ldots, 0, +1, -1, 0, \ldots, 0]$. For the first weighting function, the values of s_i zig-zag, alternating small and large. For the second, the values are essentially constant except for the two variables that had non-zero weights. We have adjusted the values of $\vec{s}$ in each case so that the smallest is zero; in this way, they may be thought of as standard deviations, and compared to Figure 7.

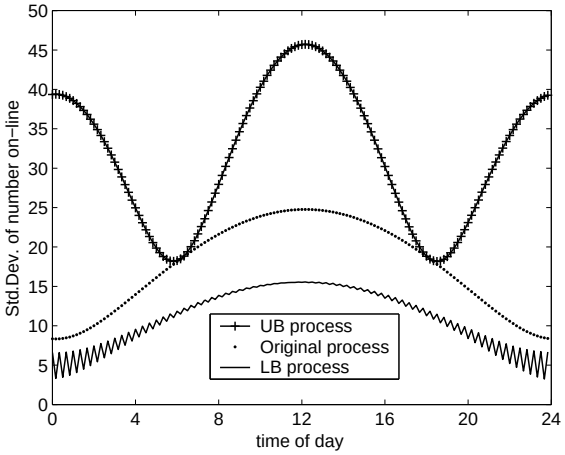

Figure 7: Standard deviations from the upper-bound, original, and lower-bound processes

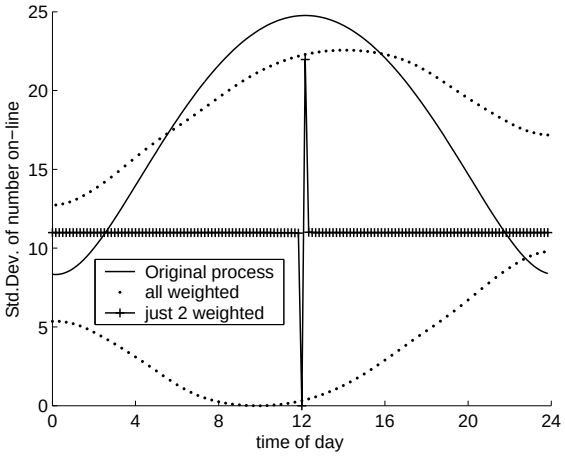

Figure 8: Standard deviations from two PDDD lower-bound processes

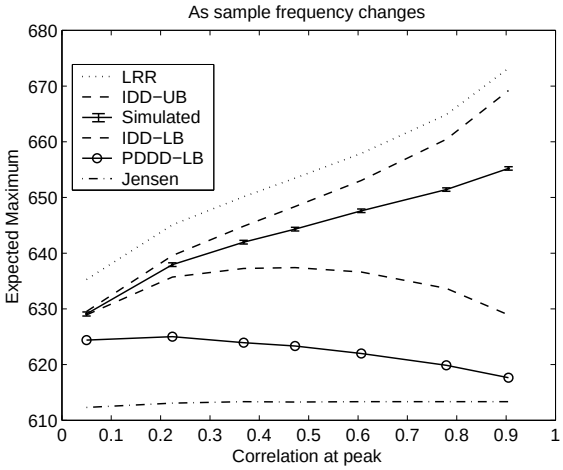

Figure 9: Bound behavior as the correlation changes due to sample frequency changes

6.3 Bound Quality and Pairwise Correlations

In Figures 4-6, we changed the system parameters in a way that affected the correlations between sample points. Now, we explore the effect of correlations more directly. Instead of graphing against the changing system parameters, as before, we will use the same data but look at the correlation between the two adjacent samples at the peak time of day. This is neither the highest nor the lowest correlation between adjacent points during the day—those occur during the lulls in the arrival rate (in the evening and the morning, respectively). Indeed, the correlation at peak is roughly the average of the entire day's adjacent correlations.

Figures 9-11 use this correlation at peak as the horizontal axis, and are analogous to Figures 4-6. We see that, as we anticipated from the earlier figures, the bounds move farther apart as the correlation changes, but there is little else that we can generalize.

6.4 Clark's Approximation

As mentioned in the literature survey, Clark (1961) proposed an approximation in the DDD case that uses the two-variable DDD results repeatedly. For example, starting with the 144 variables in our central case, we would pick two of them, create a new variable that is the maximum of the two, and assume that

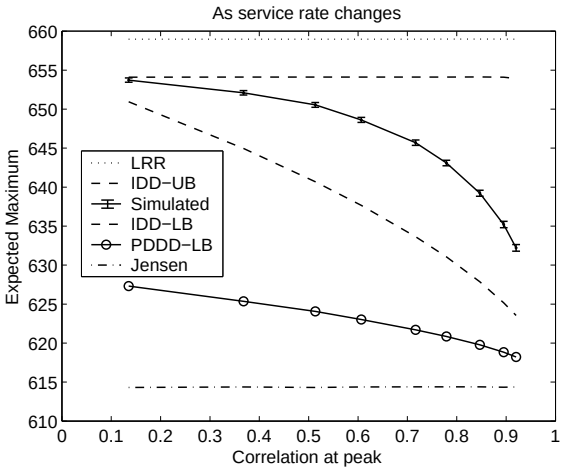

Figure 10: Bound behavior as the correlation changes due to service rate changes

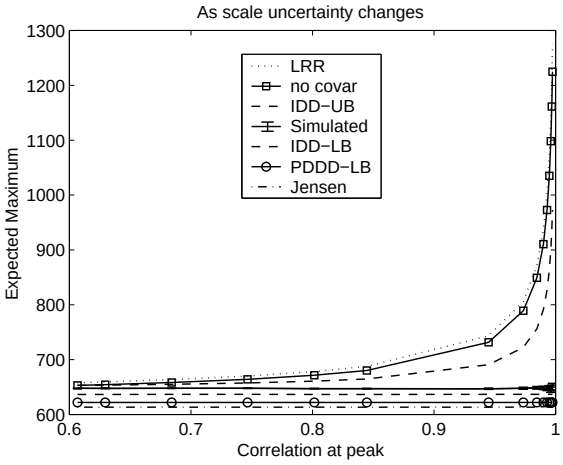

Figure 11: Bound behavior as the correlation changes due to forecast uncertainty changes

variable is normal. That is,

$$\begin{aligned} & \mathrm{E}\left[\max \tilde{X}_1, \ldots, \tilde{X}_{142}, \tilde{X}_{143}, \tilde{X}_{144}\right] \\ = & \mathrm{E}\left[\max \tilde{X}_1, \ldots, \tilde{X}_{142}, \max(\tilde{X}_{143}, \tilde{X}_{144})\right] \\ \approx & \mathrm{E}\left[\max \tilde{X}_1, \ldots, \tilde{X}_{142}, N_1\right] \end{aligned}$$

where N_1 is a two-moment Normal approximation to $\max(\tilde{X}_{143}, \tilde{X}_{144})$. The mean is computed via Eqn. 3; for the variance, see Clark (1961). New correlations are computed between N_1 and the other variables, and the procedure is repeated until two variables are left. The final expected value is then computed via Eqn. 3.

We have not seen in the literature a discussion of what order is best for the reduction. Several options are:

1. From 1 to N,
2. From N to 1,
3. From $\min m_i$ to $\max m_i$,
4. From $\max m_i$ to $\min m_i$, or
5. Random permutation.

One might also consider the variances along with the means when deciding the order, but we have not done so.

To evaluate the effects of the ordering, we have tried each of the above suggestions on some of our previous experiments. Figure 12 shows the results of the five orderings as we vary the average service rate (analogous to Figure 5). The $1 \ldots 144$ and $144 \ldots 1$ results are practically the same, and are closer to the results of the simulation than the $\min \ldots \max$, $\max \ldots \min$, and random-permutation results, which are themselves nearly indistinguishable.

However, we see different results for the various orderings in Figure 13, which varies the coefficient of variation in the forecast uncertainty (analogous to Figure 6). The $1 \ldots 144$ and $144 \ldots 1$ results are still very close. However, the $\min \ldots \max$ ordering seems very accurate compared to the simulation, while the $\max \ldots \min$ ordering is now worse, and the random ordering is the least accurate.

In all but one case here, we see that as the random variables become more correlated, the accuracy of the Clark approximation decreases.

7 Conclusions and Further Directions

We have demonstrated two ways to calculate lower bounds, and one way to compute upper bounds, that give tighter bounds than previously available results. While we have used surrogate objective functions, the results can only improve

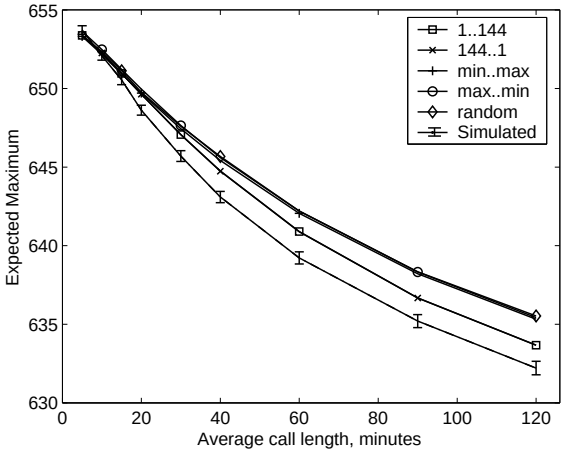

Figure 12: Clark approximations with various orderings as the average service rate changes

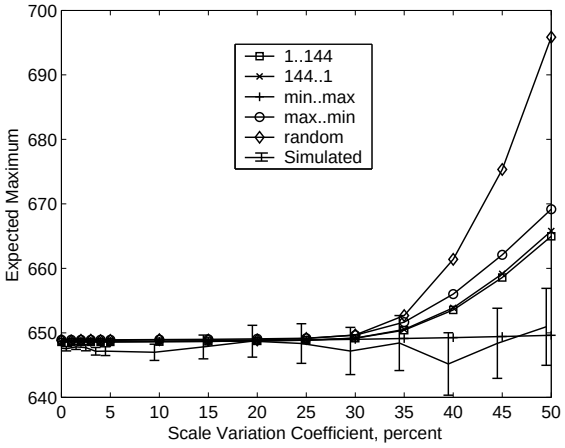

Figure 13: Clark approximations with various orderings as the forecast uncertainty changes

in the future by using the true non-linear objective function. Our results do not require choosing an ordering for the random variables, as Clark's approximation does. While the IDD lower bound was always better than the PDDD bound, there may be applications where the PDDD bound is superior. It does have the advantages of a closed-form way to evaluate the objective function and gradient.

Out of curiosity, it would be nice to have a proof (or counterexample) about the convexity of $\mathrm{E}[\max]$ in the standard deviations (IDD case) or s_i values (PDDD case). However, it would probably not dramatically affect the usefulness of the bounds. It would also be reassuring to have a more quantitative way to express the apparent fact that the bound values do not vary much when we change the weights in the LP.

It would also be nice to have an intuitive explanation of why, in the PDDD case, the expected value of the maximum is insensitive to adding $\delta \cdot \vec{1}$ to the value of $\vec{s}$. The current proof uses only elementary methods, but does not add probabilistic insight.

It might be possible to get better bounds with the currently IDD variables by allowing some $\bar{W}_i$ variables to be correlated. For example, we could make the covariance between $\bar{W}_1$ and $\bar{W}_2$ a decision variable, but have them be independent of all the other random variables. The same would apply for $\bar{W}_3$ and $\bar{W}_4$, etc. Then, instead of computing the expected maximum of (say) 144 IDD Normals, we would compute the expected maximum of 72 IDD Bivariate-Normals. This would require computing the CDFs for the bivariate Normals, which is possible but not as easy as it is for univariate Normals.

A The LRR Bound in the Normal case

The LRR bound in Eqns. 4 and 5 is valid for any distribution, but requires integration of the tail CDF function. Fortunately, for Normal random variables, we can manipulate the integral to get an expression that does not involve integration (other than the Normal CDF). We write $\Pr\left\{\tilde{X}_i > x\right\}$ as an integral, then reverse the order of integration of the resulting double integral. In the Normal case, we can then re-write the formula using the CDF and PDF of each variable.

First, we change the order of integration. Let $f_i(t)$ be the PDF for the random variable $\tilde{X}_i$. We have

$$\int_{x=c}^{\infty} \Pr\left\{\tilde{X}_i > x\right\} dx = \int_{t=c}^{\infty} t f_i(t) dt - c\Pr\left\{\tilde{X}_i > c\right\}$$

This is true for Normal, Gamma, and many other common distributions.

Next, assuming a Normal distribution, we get (after integrating by parts)

$$\sigma_{Xi}^2 \cdot f_i(c) + m_i \cdot \Pr\left\{\tilde{X}_i > c\right\} - c \cdot \Pr\left\{\tilde{X}_i > c\right\} =$$

$$\sigma_{Xi}^2 \cdot f_i(c) + (m_i - c) \cdot \Pr\left\{\tilde{X}_i > c\right\}$$

Using a Gamma distribution gives a similar formula. Overall, then, we have

$$\mathrm{E}\left[\max_i \tilde{X}_i\right] \leq c + \sum_{i=1}^{N} \left(\sigma_{Xi}^2 \cdot f_i(c) + (m_i - c) \cdot \Pr\left\{\tilde{X}_i > c\right\}\right)$$

B The Gradient in the PDDD Case

To optimize our bounds in the PDDD case, we need the gradient of the expected value $\mathrm{E}\left[\max \tilde{V}_i\right]$ with respect to the s_j values. Recall that the expected value in the PDDD case is a summation of integrals, each of the form in Eqn. 9. The breakpoints depend on the values of the s_j variables. Suppose we want to compute the partial derivative with respect to s_7. If variable 7 does not determine a segment of the piecewise-linear function $h(z)$, then the derivative is zero. Suppose instead that variable 7 determines the segment between breakpoints z_2 and z_3. We must then compute

$$\frac{\partial}{\partial s_7} \int_{z_2(s_7)}^{z_3(s_7)} (s_7 \cdot z + m_7)\phi(z)dz$$

We have noted the dependence of z_2 and z_3 on s_7. We use Leibnitz's rule for derivatives involving the integrand and integration limits, and we get a partial derivative of

$$\begin{array}{l} \int_{z_2(s_7)}^{z_3(s_7)} \frac{\partial}{\partial s_7}(s_7 \cdot z + m_7)\phi(z)dz \\ -(s_7 z_2 + m_7)\phi(z_2)\frac{\partial}{\partial s_7} z_2(s_7) \\ +(s_7 z_3 + m_7)\phi(z_3)\frac{\partial}{\partial s_7} z_3(s_7) \end{array} \tag{21}$$

Here, if z_2 appears alone, it is the current value. The integral component reduces to

$$\int_{z_2}^{z_3} z\phi(z)dz$$

which can be computed without numerical integration. We could compute the value of the other two components, but they will end up being entirely cancelled. This is because changing s_7 will also affect the previous and the next integrals, because their limits involve $z_2(s_7)$ and $z_3(s_7)$, respectively. When we apply Leibnitz's rule, we obtain terms that exactly cancel the last two terms in Eqn. 21. Overall, we have found

$$\frac{\partial}{\partial s_j}\mathrm{E}\left[\max \tilde{V}_i\right] = \begin{cases} \int_{z_k}^{z_{k+1}} z\phi(z)dz & \text{if } \tilde{V}_j \text{ determines segment } k \\ 0 & \text{if } \tilde{V}_j \text{ does not determine a segment} \end{cases}$$

There are two types of border cases that still result in the above value for the gradient. The first is for the variables that define the leftmost and rightmost segments of $h(z)$ (the leftmost segment extends to $-\infty$, the rightmost to $+\infty$). The second case is if three line segments meet at a point. For example, consider $\vec{s} = [-1, 0, 1]$ and $\vec{m} = [0, 0, 0]$. Variable 1 defines the leftmost segment, variable 3 the rightmost, and variable 2 does not appear in the piecewise-linear representation of $h(z)$.

Acknowledgments

We would like to thank Brian Borchers for his suggestion on speeding up the LP. We are grateful to Yuval Nov for providing useful bibliographic references. We would also like to thank colleagues for their comments, and an anonymous colleague who introduced us to the term "DDD".

References

Adler, R. J., and J. Taylor. 2003. *Random fields and their geometry*. Birkhaüser, in preparation.

Bose, R., and S. S. Gupta. 1959, Dec. Moments of order statistics from a normal population. *Biometrika* 46 (3/4): 433–440.

Brown, L. D., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. H. Zhao. 2002. Statistical analysis of a telephone call center: A queueing-science perspective. working paper.

Clark, C. E. 1961, Mar-Apr. The greatest of a finite set of random variables. *Operations Research* 9 (2): 145–162.

Clark, C. E., and G. T. Williams. 1958, Sep. Distributions of the members of an ordered sample. *Annals of Mathematical Statistics* 29 (3): 862–870.

Coles, S. G. 2001. *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer.

Cox, D. R. 1955. Some statistical methods connected with series of events. *J. Royal Statistical Society B* 17:129–164.

David, H. A. 1981. *Order statistics, 2nd ed.* Wiley.

Eick, S. G., W. A. Massey, and W. Whitt. 1993a, February. $M_t/G/\infty$ queues with sinusoidal arrival rates. *Management Science* 39 (2): 241–252.

Eick, S. G., W. A. Massey, and W. Whitt. 1993b, July-August. The physics of the $M_t/G/\infty$ queue. *Operations Research* 41 (4): 731–742.

Grandell, J. 1997. *Mixed poisson processes*. Number 77 in Monographs on Statistics and Applied Probability. Chapman and Hall.

Green, L. V., P. J. Kolesar, and A. Svoronos. 1991, May–June. Some effects of nonstationarity on multiserver Markovian queueing systems. *Operations Research* 39 (3): 502–511.

Henderson, S. G., and B. P. Chen. 2000. Two issues in setting call centre staffing levels. working paper.

Kella, O. 1986. On the distribution of the maximum of bivariate normal random variables with general means and variances. *Communications in Statistics–Theory and Methods* 15 (11): 3265–3276.

Lai, T., and H. Robbins. 1976, Feb. Maximally dependent random variables. *Proceedings of the National Academy of the Sciences USA* 73 (2): 286–288.

Leadbetter, M., G. Lindgren, and H. Rootzen. 1983. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer.

Malcolm, D., J. Roseboom, C. Clark, and W. Fazar. 1959. Application of a technique for research and development program evaluation. *Operations Research* 7:646–669.

Owen, D., and G. Steck. 1962, Dec. Moments of order statistics from the equicorrelated multivariate normal distribution. *Annals of Mathematical Statistics* 33 (4): 1286–1291.

Palm, C. 1988. *Intensity variations in telephone traffic (translation of 1943 article in Ericcson Technics, 44, 1-189)*. North-Holland, Amsterdam.

Ross, A. M. 2001, Aug. *Queueing systems with daily cycles and stochastic demand with uncertain parameters*. Ph. D. thesis, University of California, Berkeley.

Ross, S. M. 2003. *Introduction to probability models, 8th edition*. Academic Press.

Slepian, D. 1962. The one-sided barrier problem for Gaussian processes. *Bell System Technical Journal* 41:463–501.

Teichroew, D. 1956, Jun. Tables of expected values of order statistics and products of order statistics for samples of size twenty or less from the normal distribution. *Annals of Mathematical Statistics* 27 (2): 410–426.

Thompson, G. M. 1999. Setting staffing levels in pure service environments when the true mean daily customer arrival rate is a normal random variate. working paper.

Tippett, L. 1925, Dec. On the extreme individuals and the range of samples taken from a normal population. *Biometrika* 17 (3/4): 364–387.

Tong, Y. L. 1990. *The multivariate normal distribution*. Springer.

Vitale, R. A. 2000. Some comparisons for Gaussian processes. *Proceedings of the American Math. Society* 128 (10): 3043–3046.