Research Notes for Chapter 6*

As discussed in the research notes for Chapter 1, the development of stochastic sequencing models was slow until the late 1970s. Nonetheless, the basic stochastic counterpart models covered in this chapter were discovered earlier. Theorems 6.1 and 6.2 are due to Rothkopf (1966). Theorem 6.3 is due to Crabill and Maxwell (1969), who were also the first to note that minimizing the maximal expected tardiness is not identical to minimizing the expected maximal tardiness. Theorem 6.4 is due to Hodgson (1977). Corollary 4.1 was actually published earlier than the theorem itself, by Banerjee (1965). We are not aware of earlier sources of Theorems 6.5, 6.6, 6.7 and 6.8. Theorem 6.6 could be used in a branch and bound application for finding the sequence that minimizes the expected maximal penalty, but research to test whether it is indeed effective for that purpose is lacking. A basic linear-association model was proposed by Trietsch (2005), and, because it addresses projects, we discuss it in the research notes of Chapter 18. Essentially, the results on linear association given in this chapter, Chapter 11 and Appendix A were developed to avoid the ubiquitous stochastic independence assumption, and yet maintain tractability. Characterizing the cases for which traditional results based on stochastic independence can be extended to linearly-associated distributions is an open research problem. But Theorem 6.8 is one example where Theorem 6.7 suffices for this purpose. In Chapter 7 we present another example, Algorithm 7.1 (see also Theorem 7.1), where Theorem 6.7 can be used to expand the conditions of a result from stochastic independence to linear association. Extending the analysis to more elaborate associations is another area that requires further research.

Stored samples are central to our computations for stochastic models. They also help us avoid the stochastic independence assumption. Furthermore, in subsequent chapters, stored samples are vital to our calculations of safety time. It is therefore important to discuss the theoretical underpinnings of this approach. But first, we lay to rest a common myth, according to which the use of a stored sample—often associated with simulation—is in some sense less precise than the use of analytic models. From a practical point of view, the opposite is true. It is the typical analytic model that is more divorced from reality. Key to any analytic model is the use of given parameters. But where do these parameters actually come from? At best, such parameters are estimated from practical samples (which, under a stored-sample approach, could have been used directly), and typically the models also involve fitting theoretical distributions to the data collected. When fitting distributions, goodness-of-fit tests apply, and if the tests are passed, we can say that the fit cannot be rejected. We can never say that the fit is correct, however. So there are two sources of potential error in this approach: estimating parameters and selecting a distribution. Again, if a real-life sample is given, our approach

---

would not require estimating anything from it: we could often simply use it as our stored sample. As a result, we would avoid both the estimation errors and the errors in fitting theoretical distributions. An additional advantage, even if we use simulated samples (which, admittedly, are not necessarily better representations of reality), is that we don't need strong assumptions, such as stochastic independence. Thus, we achieve more realistic representations, and we do not require restrictions designed mainly to achieve tractability in analytic models.

Analytic models have their value, especially in studying the behavior of stochastic systems in a stylized framework that allows seeing fundamental issues more clearly. Indeed, we utilize such models in this text, and we do so for that very purpose. But they should not always be touted as a better approach for practical problems than using sample data. There is a caveat, however. Using a larger stored sample implies longer computation times. As a rule of thumb, we can estimate the relative computation time as roughly equal to r times that of the deterministic counterpart, where r is the number of realizations in the sample. So, if we can estimate distributions well and then use them efficiently without relying on simulation, we may be able to solve larger problems than we could with a stored sample. For instance, if computation time is proportional to 2^n, we might expect to be able to solve for 10 fewer jobs if we use a sample of 1000 realizations.

We now turn to the historical development of stored-sample analysis. For this purpose we assume the sample is actually simulated: when the sample is given from real data, the issues we now discuss become moot. Essentially, using a sample is a very intuitive concept and has probably been reinvented repeatedly, making it difficult to identify the first source of this idea. For instance, in Chapter 15 we discuss job shop simulations that date back to the 1960s, and they essentially involve the use of stored samples. Early applications of computerized simulation, however, did not address optimization. Instead, they involved comparison of few alternatives (e.g., sequences) presented by the user. For this purpose, they relied on the statistical sampling approach combined with artificial generation of samples. In the early days, fast access memory was a severe bottleneck, and in many cases, the sample was regenerated for each new test. To make such comparisons more precise, it is possible and recommended to regenerate the exact same sample for each alternative solution by using the same seed for generating random numbers. Furthermore, it is possible to calculate the mean and the variance of a sample during the process of generating it without keeping the data in memory or storage. When memory is at a premium and storage is expensive, this facility is very attractive. The same approach can also be used in sample-based optimization; i.e., it is not necessary to actually store the sample. For instance, Gurkan et al. (1994) recommend this approach. Today, it is much easier to keep a sufficiently large stored sample in fast access memory, so the idea of storing a sample and then using it repeatedly is technically more attractive. This is especially true if each stored number is the result of a simulation from a complex distribution. For example, in an early experiment, we used a very large stored sample with data generated by an Excel distribution function. In that case, it took 75 seconds to generate a sample and write it in a spreadsheet; but afterwards it took only 0.2 seconds to optimize the precise schedule for it. There were no sequencing decisions involved, or the optimization might have taken much longer, but this case demonstrates that using a stored sample is a very quick operation relative to generating it—that is, regenerating the same sample repeatedly would be highly wasteful.

Regardless of whether we store the sample or regenerate it, sample-based optimization starts with an assumption (implicit in the main body of our text) that processing times are *stationary*. Processing times are stationary if they do not depend on jobs' start times; i.e., they are independent of the sequence or the schedule. The processing times of two jobs may be statistically correlated and yet stationary. For optimization, we utilize this stationarity to justify using the same sample under different schedules. By contrast, if processing times were a function of their start time (the schedule), or the sequence, this approach would require modification. For instance, in problems involving highway travel time, the stationarity assumption does not hold because travel time tends to be longer during rush hour. In such a case, our use of stored samples requires modification. Such modification may involve transformations of the distribution as described by Trietsch and Quiroga (2009) in a different context. We also use this approach for stochastic crashing in Chapter 18.

Two types of optimization are relevant. One is associated with optimizing nonlinear functions; e.g., precise scheduling decisions, where the decision variable is continuous and the function has a continuous derivative, but stochastic noise is also present, necessitating the clever use of samples. Shapiro and Homem-de-Mello (1998) discuss such a case with a multivariate normal distribution. For such cases, it makes sense to use small samples when we are far from the optimum and larger samples as we approach the final solution. Indeed, these authors recommend that approach. A related subject is to find the optimal setting of a complex process; e.g., in chemical engineering. In this case, the underlying function—say yield—is not known theoretically but must be estimated empirically by trial and error. If so, the problem calls for a statistical experimental design approach, also known as *response surface optimization* (RS—see Box and Draper, 1987). RS can also be useful when few discrete decisions are involved; e.g., whether to use ingredient *X* or ingredient *Y*. It is not likely to be useful for sequencing decisions involving many jobs, however, because each possible sequence would become a special case requiring estimation. Related papers discuss the use of stored samples in the context of integer programming—which conceptually can be used for sequencing decisions (see Appendix C); e.g., see Kleywegt et al. (2001) and Verweij et al. (2003). Nonetheless, for the purpose of making sequencing decisions, there is less evidence that it is safe to start the search with a small sample and increase the sample size as we approach the final solution. It may be a useful heuristic to do so, but this subject requires further research. The state of the art in our context is to use neighborhood searches for the optimal sequence.

A related approach is pursued by Healy and Schruben (1991). They generate and store a sample and then optimize for each repetition separately and select the solution that is correct most frequently. This measurement—the frequency at which a sequence is optimal—is also known as the *optimality index*, a term coined by Dodin (1996) in a sequencing context. The use of optimality indices may look attractive, but Portougal and Trietsch (1998) cautioned that it is not a robust criterion. They demonstrated that maximizing the optimality index may favor the selection of schedules whose distributions have a low mode but high variance and high mean. Such distributions tend to be superior very often—that is why they have high optimality indices—but when they fail, the failure is worse than it could have been. The practical conclusion is that if we need to resort to heuristics, we are likely to be better off using the deterministic counterpart sequence as a

basis for scheduling than using a sequence with a higher optimality index. This choice is guaranteed to be easier computationally and likely to be at least as good once we take stochastic variation into account.

Portougal and Trietsch (1998) is also the source of the result cited in Section 6.6 that if $Y \geq_{st} X$ and Y and X are independent, then $\Pr\{Y \geq X\} \geq 0.5$. (This result can be extended to the linearly-associated case by invoking Theorem 6.7.) Therefore, when comparing two stochastically-ordered distributions, the one that is stochastically smaller will have a higher optimality index. This result is highly intuitive but the proof is not immediate. On the subject of stochastic dominance, we should mention that several stronger forms of stochastic dominance are often mentioned in the research literature. Perhaps the most important one is the strongest possible dominance, where one variable dominates the other with probability 1 (*w.p.*1), or *almost surely*. If Y is larger than X almost surely we can write $Y \geq_{as} X$. Because dominance *w.p.*1 implies stochastic dominance, every result that can be proved for distributions with stochastic dominance also applies for stochastic dominance *w.p.*1. Suppose that two random variables Y and X are independent. If $Y \geq_{as} X$ then the cdf of Y must reach 1 before the cdf of X can exceed 0. For this reason, independent random variables with this strong dominance between them are also described as having *non-overlapping* distributions. But it is easy to show that this dominance does not require non-overlapping distributions when the variables are correlated (Ross 1996). For example, if Z is a nonnegative random variable and $Y = X + Z$, then $Y \geq_{as} X$ but the two can have overlapping distributions nonetheless. In fact, the assumption of independence is so ubiquitous that sometimes results that are stated for non-overlapping distributions could actually be proven for regular dominance *w.p.*1.

Emerging Research Areas

A useful research area, suggested by our discussion above, is cataloguing classical results that assume independence and classifying them according to whether they can or cannot be generalized to linearly-associated distributions. That includes, for example, distinguishing between results that require dominance with probability one and those that actually require non-overlapping distributions. In general, doing that is a task that requires studying in detail a very large number of publications. Whereas some cases would be almost immediate to analyze, others may require careful study. In Appendix A, among other things, we discuss a case where Theorem 6.7 cannot be used because it involves due dates that are not subject to the common bias. Hence, it is clear that not every result obtained under the independence assumption can be extended. It would be especially useful to devise general rules that can help in making such decisions.

We now discuss an important open area of future research that combines stochastic analysis with conventional mathematical programming techniques, such as branch and bound (B&B) and dynamic programming (DP). We address specifically the stochastic T- and T_w-problems, but part of the challenge is to identify additional models that might be addressed that way. As a rule, we can apply B&B, DP or Integer Programming to practically any stochastic problem by sample-based optimization. For instance, Gutjahr et al. (1999) apply B&B within a sample-based optimization framework for the T-problem. Similarly, some of the references discussed before involve sample-based optimization by various mathematical programming approaches. We also remark that stochastic

programming with recourse typically utilizes a set of scenarios, which we might as well call a sample. Our aim here, however, is to show the applicability of the analytical approach for some distributions *without* a stored sample, and thus achieve more efficient computation. To begin, we give a streamlined proof of a slight generalization of Theorem 6.8. The generalization allows agreeable weights and we address any two jobs. That is, we actually prove a generalized version of Theorem 2.8 (see Exercise 2.8*g*). Theorem 6.8 is essentially a corollary of our new result.

Theorem RN6.1 In the T_w-problem, let the processing times, p_j, of all jobs be linearly associated, and let jobs 1 and 2 satisfy $p_1 \leq_{st} p_2$, $d_1 \leq d_2$ and $w_1 \geq w_2$, then job 1 precedes job 2 in an optimal sequence.

Proof.

»» Again, we prove for independent processing times and then invoke Theorem 6.7 to cover linear association. In Figure RN6.1, the expected tardiness of a job is depicted as a tail to the right of its due date, above the distribution that applies to it and below the upper horizontal line of 1. The relevant distributions are either F_k if job k is scheduled first (k = 1, 2), or F_{1+2} if job k is scheduled second. These three distributions also reflect any preceding jobs that have already been scheduled, or any jobs scheduled between jobs 1 and 2. As the figure shows, job 1 is stochastically smaller and has a lower due date, per the condition of the theorem. Let $T_{F,d}$ denote the area of the tail above distribution F (where F = 1, 2 or 1+2) to the right of due date d (where d = 1, 2). $T_{F,d}$ measures an expected tardiness; for instance, $T_{1+2,1}$ is the expected tardiness of job 1 if it is sequenced second and is thus subject to the completion time distribution F_{1+2}. We start with the sequence 1-2, assuming the two jobs are adjacent. By an API, the tardiness cost of job 1 increases by $w_1(T_{1+2,1} - T_{1,1}) \geq w_1(T_{1+2,2} - T_{1,2})$, whereas the tardiness cost of job 2 decreases by $w_2(T_{1+2,2} - T_{2,2}) \leq w_2(T_{1+2,2} - T_{1,2})$. But, because $w_2 \leq w_1$, $w_2(T_{1+2,2} - T_{1,2}) \leq w_1(T_{1+2,2} - T_{1,2})$, so the gain is bounded from above by a lower bound of the loss and the change cannot decrease and may increase the total weighted tardiness. Now allow additional jobs (which need not be stochastically ordered) between jobs 1 and 2. If we interchange the two jobs, all these intermediary jobs follow a stochastically larger job so their expected tardiness cannot decrease. Hence, such jobs cannot provide incentive to perform the interchange either. ««

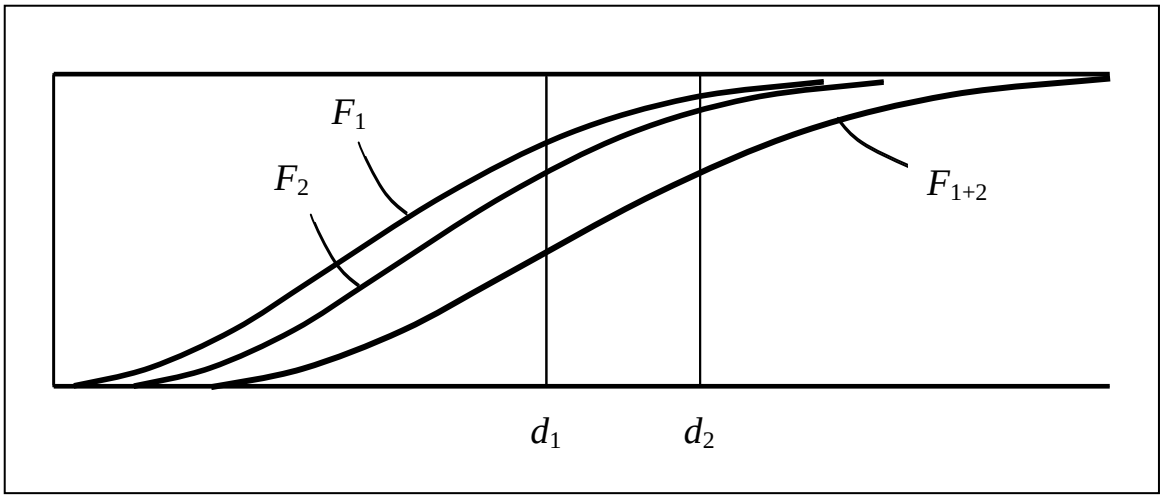

Figure RN6.1

Parenthetically, in Chapter 7, we introduce *predictive Gantt charts*. A predictive Gantt chart provides distributions for starting times and completion times and shows due date performance graphically. Essentially, Figures 6.2 and RN6.1 incorporate all the ingredients of a predictive Gantt chart. Our proof highlights the usefulness of predictive Gantt charts for stochastic analysis.

To continue, calculating expected tardiness by tail areas can sometimes be performed without simulation. For instance, when processing times are normal it is easy to obtain distributions such as F_{1+2} by convolution. In such cases, we can perform API tests by comparing the expected weighted cost of the two possible sequences. If job 1 comes first, the total weighted cost is given by: $w_1T_{1,1} + w_2T_{1+2,2}$ whereas if job 2 comes first, the total weighted cost is given by: $w_1T_{1+2,1} + w_2T_{2,2}$. Therefore, job 1 can come first if:

$$w_1T_{1,1} + w_2T_{1+2,2} \leq w_1T_{1+2,1} + w_2T_{2,2}$$

or

$$w_1(T_{1+2,1} - T_{1,1}) \geq w_2(T_{1+2,2} - T_{2,2})$$

This condition can take the place of the WMDD dispatching heuristic that we introduced in Chapter 4. Job 1 comes first if it satisfies the condition for any selection of another job as job 2. Stochastic dominance is not required. It can also be used within branches in a B&B application.

We invoked the normal distribution because it is easy to calculate convolutions for it and to calculate tail areas thereafter: by (B.14), $E(T_j) = \sigma[\varphi(w) - w\Phi(-w)]$, where φ is the standard normal density function, $\Phi(w)$ is the standard normal cdf, $w = (d - \mu)/\sigma$, μ is the mean, and σ is the standard deviation of the processing time. The service level in this case is given by $\Phi(w)$. Although $\Phi(w)$ is elliptic, it is as good as analytic in the practical sense (because very precise calculations are available by appropriate series); e.g., the Excel function NORMSDIST(w) can be used in calculations. Therefore, it is

possible to use the normal tail result for the purpose of solving tardiness problems by B&B or for DP.

Furthermore, it is equally easy to calculate tail areas for the lognormal distribution. Let μ and s be the mean of the lognormal distribution and the standard deviation of its core normal (see Appendix A for the relationship between these parameters). Define $z = \ln(d/\mu)/s + s/2$, which implies a service level of $\Phi(z)$. Then it can be shown that the expected tardiness is $\mu\Phi(s - z) - d\Phi(-z)$. The only problem with the lognormal, however, is that we don't have convenient convolutions for it. Nonetheless, there is one important special case for which we can use the lognormal distribution without resorting to approximations, and that is when processing times are linearly associated but with stochastic variation restricted to the common element, Q. Another important case that can be approximated very well is when each element is distributed lognormal with the same s, and Q is also lognormal. In that case, we can invoke the lognormal central limit theorem (see Appendix A) to obtain approximate convolutions. When all processing times are lognormal with the same s, then they are stochastically ordered.[†] In such case, we can also apply Theorem RN6.1. In Trietsch et al. (2010) we report that linearly-associated lognormal distributions provided a good fit for processing times in two project environments in Armenia. We believe that it should be useful in a much wider context as well. Thus, solving for this particular distribution has a validated practical application.

Returning to our tail-based analysis, one might think that such observations, at least with respect to the better-known normal model, would have led to the application of B&B or DP to the stochastic T-problem and the stochastic T_w-problem. Nonetheless, that does not seem to be the case (except by the sample-based approach that we cited above). Thus, we believe that the application of these tools to stochastic scheduling is a ripe area for research. Similarly, these stochastic problems can be addressed by various search heuristics with little adaptation. For example, the current best search technique for the T_w-problem seems to be *dynasearch* (Congram et al., 2002; Grosso et al., 2004). As discussed in our Research Notes for Chapter 4, Dynasearch is a neighborhood search approach that is based on searching various combinations of pair interchanges and has been shown to be much more effective than regular pair-interchange search heuristics that consume the same search time. On the one hand, in the deterministic T_w-problem case, the fastest application of dynasearch utilizes shortcuts that go beyond just using the analogue of Theorem RN6.1. On the other hand, it is conceptually easy to adapt *basic* dynasearch—without shortcuts—to the stochastic version. Specifically, the effects of independent PIs are additive, which, as we discussed in RN4, is the main requirement for dynasearch to be potentially effective. (By this criterion, dynasearch could also be applied to a stored sample.) We may also be able to identify useful shortcuts that apply in the stochastic case (including Theorem RN6.1). Thus, the application of dynasearch and other search techniques to our problem is highly likely to bear fruit. Nonetheless, it is also important to establish the size limit of problems that can be solved to optimality (e.g., by B&B or DP), so we should not limit our attention to search heuristics.

[†] In our Research Notes for Appendix A we show that lognormals with the same s are stochastically ordered in the likelihood ratio sense, which is a slightly stronger result.

References

Box, G.E.P. and N.R. Draper (1987) *Empirical Model-Building and Response Surfaces*, Wiley.

Banerjee, B.P. (1965) "Single Facility Sequencing with Random Execution Times," *Operations Research* 13, 358-364.

Congram, R.K., C.N. Potts and S.L. van de Velde (2002) "An Iterated Dynasearch Algorithm for the Single-Machine Total Weighted Tardiness Scheduling Problem," *INFORMS Journal on Computing* 14, 52-57.

Crabill, T.B. and W.L. Maxwell (1969) "Single Machine Sequencing with Random Processing Times and Random Due-Dates," *Naval Research Logistics Quarterly* 16, 549-555.

Dodin, B. (1996) "Determining the Optimal Sequences and the Distributional Properties of their Completion Times in Stochastic Flow Shops," *Computers and Operations Research* 23, 829-843.

Grosso, A., F. Della Croce, R. Tadei (2004) "An enhanced dynasearch neighborhood for the single-machine total weighted tardiness scheduling problem," *Operations Research Letters* 32, 68-72.

Gurkan, G., A.Y. Ozge and S. Robinson (1994), "Sample-path Optimization in Simulation," *Proceedings of the 1994 Winter Simulation Conference*, 247-254.

Gutjahr, W.J, A. Hellmayr and G.Ch. Pflug (1999) "Optimal Stochastic Single-Machine-Tardiness Scheduling by Stochastic Branch-and-Bound," *European Journal of Operational Research* 117, 396-413.

Healy, K. and L.W. Schruben (1991) "Retrospective Simulation Response Optimization," *Proceedings of the 1991 Winter Simulation Conference* 901-906.

Hodgson, T.J. (1977) "A Note on Single Machine Sequencing with Random Processing Times," *Management Science* 23, 1144-1146.

Kleywegt, A.J., A. Shapiro and T. Homem-De-Mello (2001) "The Sample Average Approximation Method for Stochastic Discrete Optimization," *SIAM Journal of Optimization* 12, 479-502.

Ross, S.M. (1996) *Stochastic Processes*, 2nd Ed., Wiley.

Rothkopf, M.H. (1966) "Scheduling with Random Service Times," *Management Science* 12, 707-713.

Portougal, V. and Trietsch, D. (1998) "Makespan-Related Criteria for Comparing Schedules in Stochastic Environments," *Journal of the Operational Research Society* 49, 1188−95.

Shapiro, A. and T. Homem-de-Mello (1998) "A Simulation-Based Approach to Two-Stage Stochastic Programming with Recourse," *Mathematical Programming* 81, 301-325.

Trietsch, D. and F. Quiroga (2009) "Balancing Stocahstic Resource Criticalities Hierarchically for Optimal Economic Performance and Growth," *Quality Technology and Quantitative Management* 6. URL: <http://web2.cc.nctu.edu.tw/~qtqm/qtqmpapers/2009V6N2/2009V6N2_F2.pdf>.

Trietsch, D., L. Mazmanyan, L. Gevorgyan and K.R. Baker (2010), A New Stochastic Engine for PERT (working paper).

Verweij, B., S. Ahmed, A.J. Kleywegt, G. Nemhauser and A. Shapiro (2003) "The Sample Average Approximation Method Applied to Stochastic Routing Problems: A Computational Study," *Computational Optimization and Applications* 24, 289-333.