Exponential, Gamma, and Power Law Distributions in Information Flow on a Construction Site

Walid F. Nasrallah, M.ASCE[1]; and Rana Bou-Matar[2]

Abstract: This paper examines the statistical distributions of interarrival and response times for construction-site correspondence. Data from a number of construction projects are analyzed and hypotheses are proposed and tested to link the probability distribution to the type of correspondence. Although the commonly assumed exponential distribution of interarrival times is found to be accurate for all types of correspondence, the response times for different types of correspondence follow different distributions depending on the type of correspondence. In particular, a power law relationship is observed between incidence frequency and response time for requests for information. Knowing the statistical distributions for a class of events helps managers forecast future work and manage risk. Simulation models used by practitioners and researchers for various project management goals are also improved by incorporating appropriate statistical distributions for generating various events.

DOI: 10.1061/(ASCE)0733-9364(2008)134:6(442)

CE Database subject headings: Construction management; Information management; Information systems; Statistics; Simulation models.

Introduction

Consider a modern building construction site; say the new headquarters of a high-tech corporation. Dozens of contractors maintain trailers on the site, and others have only occasional presence. Some have contracts with the owner of the building, others have contracts with each other, and some may provide one-time services or products at negotiated or spot rates. All parties must constantly exchange information to avoid getting in each other's way, literally or figuratively. This information exchange takes on many different forms.

Although the scope and duration of the work is clearly delineated, with allowance for changes along the way, delays will often occur. Some of these delays will be inevitable, such as those caused by weather or by external interruption of needed supplies or capital. Other delays indicate some problem in flow of information. For example, since the nature of construction does not lend itself to minute advance planning of every detail, many decisions that have great influence on the course of the work will be made during the course of the project. The party that is authorized to make a decision is usually different from the party that will do the work, so construction professionals often need to let their resources lie idle while waiting for a vital piece of information that needs to travel through some organizational process. Information-shortage delays are particularly insidious and hard to prevent with the usual tools for dealing with physical constraints. It is often straightforward to add more workers or more equipment or more access space to immediately increase productivity in a well-predictable manner when faced with physical constraints. By contrast, the processes followed in the coordination of information exchange on a construction site are harder to analyze (Chan and Kumaraswamy 1998; Odeh and Battanieh 2002).

Two common ways of anticipating and preparing for events that are hard to analyze are statistical forecasting and Monte Carlo simulation. In the former, certain statistical metrics are computed for a series of past events, and the metrics are used to anticipate the equivalent metric for future events in the same series. The most familiar metrics are mean and standard deviation. Most usage of these two metrics involves an implicit assumption that the values measured follow a Gaussian (normal) probability density function. A probability density function is needed to answer questions such as "What proportion of the future events will have a value greater than some threshold?" For events that follow a normal distribution, the answer can be found on most scientific calculators by expressing the threshold as the number of standard deviations above (or below) the mean. For example, if the mean length of a safety-related work stoppage is 4 h, and the standard deviation is 1 h, then the chances of a work stoppage exceeding 8 h is calculated by expressing the 8-h threshold of interest as 4 SDs ($4 \times 1 = 4$) above the mean (4). Since the normal distribution occurs so widely in nature, it is usually safe to assume that an event whose magnitude is over 4 SDs above the mean will occur no more than three times in 100,000 as the Gaussian distribution function predicts. But, it is often wise to check that the assumption of a normal distribution is borne out by the facts. For example, it is widely acknowledged that the bankruptcy rate of financial firms during the market crashes of the 1980s and 1990s was exacerbated by an assumption that price movements follow a normal distribution. Examining price movements of different

[1]Assistant Professor, Engineering Management Program, American Univ. of Beirut, Beirut 1107-2020, Lebanon (corresponding author). E-mail: walid.nasrallah@aub.edu.lb

[2]Project Management Consultant, Davis Langdon & Seah International, Company: Davis Langdon, Box 13-5422, Beirut 1102-2040, Lebanon. E-mail: rana.boumatar@davislangdonlb.com

Note. Discussion open until November 1, 2008. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on January 23, 2006; approved on April 4, 2007. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 134, No. 6, June 1, 2008. ©ASCE, ISSN 0733-9364/2008/6-442–450/$25.00.

Table 1. Numbers of Data Point in Each Set Analyzed

Project number	RFIs	Shop drawings	Submittals	Multiple trades	Location
1	248	91	91	No	Kuwait
2	392	0	0	11 trades, 2 sets	Kuwait
3	58	0	0	No	California
4	230	0	0	No	California
5	54	0	0	No	California
6	291	0	0	No	Lebanon
7	115	0	0	No	Lebanon
8	332	0	0	4 trades (125, 83, 62, 62)	Saudi Arabia
9	0	62	0	No	Syria

markets historically reveals that the movements are not actually distributed normally, but are more accurately described by a power law (Mandelbrot and Hudson 2004). For example, the probability of a price movement n times the size of the mean would be n raised to some constant negative power, such as -1.5.

The goal of this research is to help statistical forecasting of information-related construction delays by looking for the best underlying statistical distribution information-exchange events in historical data. Knowing these statistical distributions will also aid in the development of more accurate Monte Carlo simulations, by generating the random events in the simulation from the appropriate underlying distribution.

Background

Simulation in construction has mostly been utilized for operations-level planning, but there are a few places where simulation of a whole project is required. Simulation of a complete project based on randomized durations of activities in a critical-path schedule is recognized as the most reliable way to estimate total project duration, but this method is not widely used because it is more complicated than the project evaluation and review technique (PERT) method that it has replaced. Isidore and Back (2002) proposed that the utility of simulation goes beyond simply predicting a range of project completion times, since the same simulations can be fed probability distributions of both cost and time. This multisimulation analysis technique (MSAT), coupled with cheaper computational power, can lead to more widespread use of simulation in construction planning, provided that the hidden assumption of a normal/Gaussian or a beta distribution is addressed.

Assigning probability distributions to durations of construction events is of value both as a stand-alone predictive tool and as an input into whole-project simulations. It has been convincingly demonstrated that the beta distribution is best suited for fitting construction durations (AbouRizk and Halpin 1992; AbouRizk et al. 1994; Schexnayder et al. 2005). There is also an ongoing debate about how well different goodness-of-fit tests predict the fit between construction duration data (Maio et al. 2000).

More recently, some simulations have included both a schedule of hard activities and organizational tasks that involve information exchange (Jin et al. 1995; Christensen et al. 1997; Salazar-Kish 2001). The original duration of an activity is given by the schedule, but may be increased by such organizational processes such as overwork, communication breakdown, and need for upper-level decisions in midactivity. This method, called the virtual design team (VDT), was originally intended for studying design work, but has been adapted for high-level construction modeling by Salazar-Kish (2001), and is most accurately described by Levitt et al. (1994) and Jin and Levitt (1996). What is lacking so far is an empirically validated set of probability distributions to be used as input into such simulations. Such distributions, while valuable as a stand-alone management tool, may also be of use in running different models such as neural network enhanced PERT (Lu 2002).

Research Question

To measure how often a failure or delay occurs in the dissemination of information on construction projects, raw data were obtained from logs of requests for information (RFIs), submittals, and shop drawing approvals from several construction projects [see Bou-Matar (2005) for more detailed data and analyses]. The project managers who made these logs available for the purposes of this investigation wanted to ensure anonymity of the project and firms, but Table 1 gives some information about the types of projects. The projects came from five different countries [united states (California), Saudi Arabia, Kuwait, Syria, and Lebanon], and included a mix of pubic sector and private sector. The country of origin was significant in determining the number of working days between events, since there are 3 different weekend-day regimes among these four countries. The weekend is observed on:

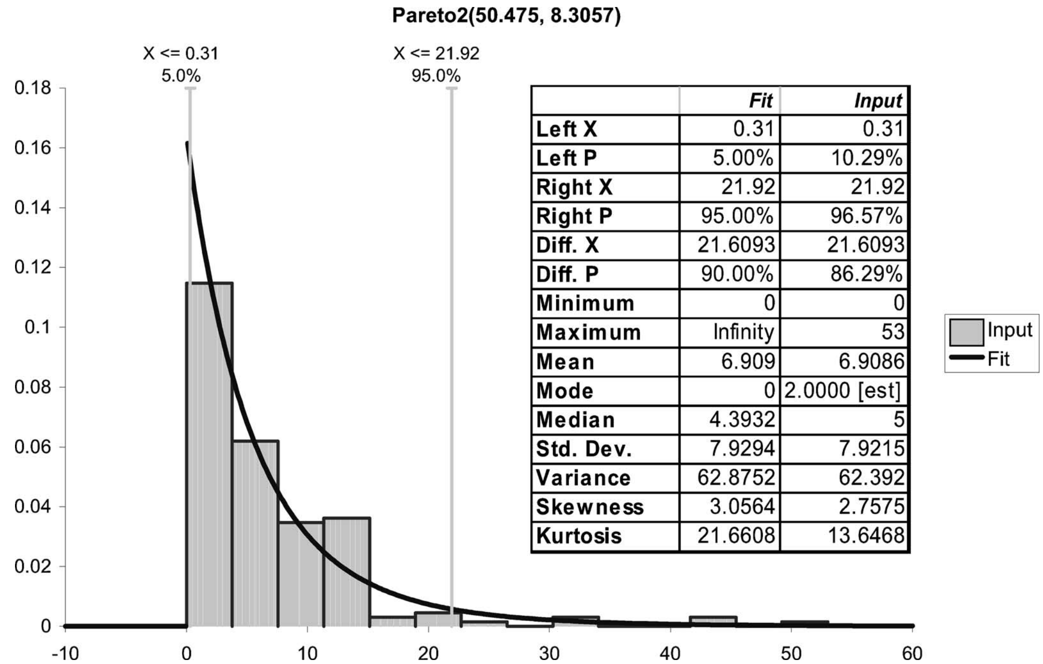
1. Saturday and Sunday in the United States;
2. Friday only in Syria, Kuwait, and Saudi Arabia; and
3. Sunday only in Lebanon.

Methodology

The logs were rendered in the form of "sent date" and "response date" for each communication, omitting the subject of the communication to preserve anonymity. On some of the larger projects, the trade to which the communication was pertinent was recorded to check for statistical relevance.

The raw data were transformed to an "interarrival" series and a "response time" series using spreadsheet formulas. Custom equations were developed to exclude the weekend days (as determined by the location of the project) in order to transform the count of calendar days to a count of working days. This transformation was imperfect because holidays and other closure dates were not considered due to lack of data.

The transformed data were then copied into the "Best Fit"



	Fit	Input
Left X	0.31	0.31
Left P	5.00%	10.29%
Right X	21.92	21.92
Right P	95.00%	96.57%
Diff. X	21.6093	21.6093
Diff. P	90.00%	86.29%
Minimum	0	0
Maximum	Infinity	53
Mean	6.909	6.9086
Mode	0	2.0000 [est]
Median	4.3932	5
Std. Dev.	7.9294	7.9215
Variance	62.8752	62.392
Skewness	3.0564	2.7575
Kurtosis	21.6608	13.6468

Fig. 1. Sample BestFit output (Project 4 RFI response times)

application by Palisade Corporation (Version 4.5.2). The software plotted the values in each series (e.g., 3 days between arrivals) against the frequency (e.g., seven submittals arrive 3 days after the previous submittal). The software then computed parameters for a number of possible probability distributions so as to give the closest fit between the actual data series and the curve generated from the parameter and distribution type. Several types of data distributions were tried, excluding only those distributions that allowed negative values. The type of distribution that, given its best possible parameter value, was able to track the original data better than the others was recorded for each data series. When the distribution type was the same for several series of the same origin, such as "interarrival for time for mechanical submittals" then the distribution type was deemed to be suitable for modeling future data series of the same provenance. In other words, the process that underlies the generation of these data points was sufficiently similar across projects to leave the same statistical signature. Of course the value of the parameter, which determines the exact shape of the curve, will vary from project to project. For practical purposes, estimating the value of the parameter would be part of the forecasting process for each project.

For some data series, there was no single such distribution across projects, which meant that no support was found for any hypothesis about the type of process that gave rise to the data. Fig. 1 shows sample of the output of best fit; analogous output for the other projects can be found in Bou-Matar (2005).

Goodness-of-Fit Tests

There are several tests for the closeness of fit between a data distribution curve and a series of data points (Maio et al. 2000). For this paper, we used the Kolmogorov–Smirnof (K-S) statistic, which measures the maximal vertical distance between the cumulative distribution of the actual data and the cumulative distribution of the theoretical curve being fit. This was deemed a more apt test of fit than the Anderson–Darling (A-D) test, which is based on the area between the two cumulative distribution curves. The frequency distribution of the A-D test is more sensitive to differences that are large in magnitude but few in number, which is a disadvantage in predicting response times to construction correspondence. If one letter was delayed for 3 months while 300 were responded to in 2 days or less, then we would want to give more weight to the 300 than to the one when predicting how likely the next letter is to be answered in 1 day.

Tests that use the chi distribution and the student-T distribution, which are more familiar to casual users of statistics, rely on a comparison with a sample randomly drawn from a population that has a normal distribution. Since the data were clearly too skewed to warrant an assumption of normal distribution, in addition to being often being bimodal, these tests would not have been valid for the purposes of this study.

Results and Discussion

Since the presentation of the results partially depends on the conclusions reached from previous results, we present the hypotheses, results, and discussion in one section, divided by type of observation.

Interarrival Times

The first hypothesis we tested was about the time between arrivals of any communication. In a random (Monte Carlo) simulation, it is easiest to assume that events such as the arrival of a RFI or a submittal is as likely to happen at any moment as it is at any other. The mathematical logic behind thinking that this approximation might be accurate stems from the derivation of the exponential distribution from the limiting case when many different and independent factors give rise to a large number of events. As the number of different independent sources of events, in this case the sending of a RFI or submittal, gets very large, the distribution of interarrival times approaches the exponential.

J. Constr. Eng. Manage., 2008, 134(6): 442-450

In contrast, the reasons for sending a RFI or submittal on a certain day might be few, leading to a more specific distribution of interarrival times, with perhaps bursts of requests around a contractual deadline followed by long periods of quiet. In summary we have:

Hypothesis H1$_0$. The frequency distribution of interarrival times for all construction communications follows an exponential curve better than it follows alternative distributions.

The converse hypothesis can be stated as follows:

Hypothesis H1$_1$. Different types of communication follow different interarrival frequency distributions, depending on the type of communication.

The data in Table 2 clearly bears out H1$_0$. (The lowest values of the KS statistic indicate best fit). Observations about the "bursty" nature of internet traffic (e.g., Huberman and Lukose 1997; Harris et al. 2000) are not duplicated in the construction data set. This gives a strong indication that such patterns do not aid in the prediction or simulation of construction information flow. The naïve assumption of random arrivals is as accurate as any.

Response Times

In a simulation such as that of Jin and Levitt (1996), the responses to a request for communication are delayed by a queuing process. When there are more requests pending, the parties responsible for responding have to finish processing the earlier (or more urgent) requests before getting to the rest. More generally, events that require an information-bearing response are highly skewed towards the left hand side of the distribution, i.e., their kurtosis and skew put them in the "exponential zone" as described by AbouRizk and Halpin (1992). They do not ideally take more time to process than the journey time of the message. In the projects studied, these messages were hand carried, giving a 1-business-day round trip time.

This leads to our first hypothesis:

Hypothesis H2$_1$. Response time is correlated to number of pending requests of the same type. The null hypothesis H2$_0$ would be that there is no such correlation. As can be seen from Table 3, no significant pattern of correlation was found. The correlation coefficients between back-log and calendar-day delay ranged from 0.49 and –0.43. This can be a reflection of one of two conditions, both of which the writers know from experience to be true of a construction site. First of all, one professional often fulfils several duties, so the back-log of one type of communication is not an accurate indication of how busy that person is. Second, the time that it takes to answer a RFI or to approve a submittal is usually very short compared to the time that it takes to route the request or submittal to the person who has the answer or the authority to approve it. The result is that the expected wait time for a response cannot be predicted based on how many similar items are outstanding. This lack of correlation may be old news to experienced practitioners, but should be taken into account by those management educators and by researchers who design simulation models to study and improve construction processes.

Given that many different factors play into the response time of a communication, it may stand to reason that a single exponential service time can also offer the best fit for the data on response times.

Hypothesis H3$_0$. The frequency distribution of response times for all construction communications follows an exponential curve better than it follows alternative distributions. According to queuing theory, observing such a distribution would indicate low load on the server, or in the extreme to what is known in operations research as the "infinite server" or "machine repair" problem. The converse hypothesis is:

Hypothesis H3$_1$ The frequency distribution of response times for different types of construction communications follow different functions depending on the type of communication. The data give a mixed picture as follows: The results for submittals and shop drawings are mixed, but seem to point towards a gamma distribution as the most suitable way to model the process (Table 4). The gamma distribution is derived from a basic model of a series of randomly timed events that must occur in sequence.

The results for RFIs tell a different story. The distribution that best fits the response times of most RFIs is the Pareto 2 (Table 5). This distribution is a typical fat tail or power law distribution with two parameters: b and q. The probability that the response time is greater than some value t is given by $1-F(t)=(1+t/b)^{-q}$. The Pareto 2 was proposed by Italian economist Vilfredo Pareto in 1896 to fit data on the distribution of income (Kleiber and Kotz 2003), and is also known as the Lomax 1. It differs from the Pareto 1 distribution, for which $1-F(t)=(t/b)^{-q}$, and which Pareto fit to wealth distribution with $q=1.5$ and b represents minimum wealth.

Working and Calendar Days

The exclusion of nonworking days from the calculation of interarrival and response times seems to be a reasonable way to increase the accuracy of predictions. Work that takes 3 days should logically happen on, say, a Monday, Tuesday, and Wednesday, but if started on Friday before a long weekend, it should proceed on Friday, Tuesday, and Wednesday, leading to an apparent lag of 5 days instead of 3. One would therefore want to check:

Hypothesis H4$_1$. The results of fitting a distribution curve to data that are calibrated in working days should show more consistency between projects than results that use calendar days.

In fact, our results showed no such difference. The analyses using working days were virtually indistinguishable from those using calendar days in all cases, as can be seen from Table 6, which is almost identical to it's analog, Table 2. This may be due to the fact that we only excluded weekend days, and not holidays and other work stoppages. As with the correlation of response time with back-log, the fact that the work in question is done by people with different duties may also account for the lack of any distinction between working days and calendar days. For example, work that takes 3 h may wait for several days while other tasks are performed, some due to different projects. Incidence of a weekend in the middle of this waiting period would have a much smaller effect on the wait time than other factors not considered in this study.

Statistical Check

Since the distributions tested had parameters generated from the data itself, standardized hypothesis tests do not apply. In other words, since the hypothesis is of the form "these data are generated from a process that follows an exponential distribution" and not of the form "these data are generated from a process that follows an exponential distribution with parameter lambda," there is no sense in asking the probability that a process with a randomly distributed output of a different mean could have produced the same data. Instead, we can check how well the data that seem

J. Constr. Eng. Manage., 2008, 134(6): 442-450

Table 2. K-S Fitness Score of Different Distributions of Interarrival Times

	Data set																		
	Project number 1			Project number 1											Project number 8				
Distrib.	RFI	Shop drawings	Submittals	RFI (noncivil)	RFI (civil)	Submittals civil 1	Submittals civil 2	Submittals non civil 1	Submittals non civil 2	RFI proj. number 3	RFI proj. number 4	RFI proj. number 5	RFI proj. number 6	RFI proj. number 7	RFI (civil)	RFI (electrical)	RFI (installation)	RFI (mechanical)	Shop drawings proj. number 9
BetaGeneral	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
ChiSq	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Erlang	**0.33**	—	—	**0.54**	**0.32**	**0.45**	—	**0.52**	**0.70**	**0.52**	**0.39**	—	—	**0.44**	—	**0.21**	**0.20**	**0.10**	—
Expon	**0.33**	**0.69**	**0.69**	**0.54**	**0.32**	**0.45**	**0.59**	**0.52**	**0.70**	**0.52**	**0.39**	**0.33**	**0.38**	**0.44**	**0.30**	**0.21**	**0.20**	**0.10**	**0.41**
Gamma	0.41	**0.69**	**0.69**	**0.54**	0.39	0.47	—	0.54	**0.70**	0.55	0.43	—	—	0.49	—	0.25	0.26	0.17	—
InvGauss	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
LogLogistic	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Lognorm	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Lognorm2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Pareto	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Pareto2	0.42	**0.69**	—	—	—	—	—	—	—	**0.52**	**0.39**	—	—	—	**0.30**	**0.21**	**0.20**	0.11	—
Pearson5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Pearson6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Rayleigh	0.50	**0.69**	**0.69**	0.73	0.48	0.66	0.96	0.62	0.73	0.62	0.51	0.65	0.63	0.53	0.58	0.48	0.43	0.37	0.61
Triang	0.54	**0.69**	**0.69**	0.82	0.54	0.70	0.95	0.64	0.76	0.59	0.54	0.68	0.78	0.47	0.59	0.52	0.34	0.42	0.55
Uniform	0.69	**0.69**	**0.69**	0.88	0.69	0.80	0.96	0.73	0.82	0.66	0.71	0.79	0.84	0.53	0.72	0.65	0.52	0.59	0.62
Weibull	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Note: Calendar days, best fit per column is boldface.

Table 3. Correlation Between Response Time and Back-Log at Arrival Time

Distrib.	Project number 1			Project number 2							RFI proj. number 4	RFI proj. number 5	RFI proj. number 6	RFI proj. number 7	Project number 8				Shop drawings proj. number 9
	RFI	Shop drawings	Submittals	RFI (noncivil)	RFI (civil)	Submittals civil 1	Submittals civil 2	Submittals non civil 1	Submittals non civil 1	RFI proj. number 3 / Submittals non civil 2					RFI (civil)	RFI (electrical)	RFI (installation)	RFI (mechanical)	
Correlation coefficient	−0.14	−0.38	−0.43	−0.19	0.082	0.012	0.004	0.24	0.49	−0.14	0.11	0.019	−0.09	0.02	−0.03	−0.01	0.06	0.275	−0.05

to follow a power law distribution fit a straight line on a log-log scale. Fig. 2 shows one such plot for Project 1, and it is seen that the straight line approximation is suitable, with a correlation coefficient $R^2=0.71$, for the whole data set and $R^2=0.85$ if low-frequency items are grouped and outliers removed.

Practical Application

The knowledge that a Pareto 2 distribution accurately describes the expected wait time for a RFI can immediately be put to use. Lacking advance knowledge of how many steps a particular RFI must pass through before it is answered, a contractor can still make a more accurate prediction of response times by using the Pareto 2 distribution with the parameters derived from past data from the same project. For example, if $b=2$ and $q=2.4$ from past observations, then the probability that a response to a future RFI will take more than 10 days is given by $(2/(2+10))^{2.4}=0.014$. An exponential distribution for the same data would yield a mean of $l=b/(q-1)=2/1.4=1.43$, for which the probability of taking more than 10 days is given by $\exp(-10/1.43)=0.0009$.

If, instead, of matching the mean, we match the probability of getting a response in 1 day or less, the figures are even less accurate. $(2/(2+1))^{2.4}=0.3779=$probability of getting a response in more than 1 day, and the parameter that gives the same figure with the exponential distribution is $l=-1/\ln(0.3779)=1.028$. With that parameter, the probability predicted by the exponential distribution for taking over 10 days would be $\exp(-10/1.028)=0.00006$.

These results can be used immediately by practitioners for everyday planning, as well as by researchers who want to build better models of why construction takes as long as it does.

Conclusions

1. The interarrival times of both RFIs and submittals can be very well fitted with an exponential distribution. The exponential distribution measures the time to the first occurrence of an equally likely random event. From a practical standpoint, this implies that a letter is as likely to arrive on any given day as it is on any other day;
2. For the most part, the time to process a communication is independent of the total number outstanding. This suggests that even for those communications processed on-site, people budget their time and support staff to accommodate demand, and delays are for the most part not due to lack of processing capacity. In other words, a pure information-processing view does not tell the whole story. Queuing back-logs do still occur, but the data suggest that we know very little about when or why. The nonroutine nature of construction work suggests that information flows on a construction site are more likely to be held up by factors that cannot be mitigated by reorganization;
3. The response times for submittals and shop drawings mostly follow a gamma distribution. This distribution mathematically derives from assuming a fixed number of steps that need to be followed, with an exponentially distributed duration for each step. Compared to what we know about how submittals and shop drawings are approved, this seems to accurately describe the process; and
4. RFI response times follow a power law. One explanation for

J. Constr. Eng. Manage., 2008, 134(6): 442-450

Table 4. K-S Fitness of Response Time Distributions for Submittals and Shop Drawings

Distrib.	Data set: Project 1 shop dwgs	Project 1 submittals	Proj. 2 civil 1 submittals	Proj. 2 civil 2 submittals	Proj. 2 noncivil 1 submittals	Proj. 2 noncivil 2 submittals	Project 9 shop dwgs
BetaGeneral	—	—	—	—	—	—	—
ChiSq	—	0.363	—	0.116	0.199	—	0.370
Erlang	0.202	0.207	0.132	0.092	0.131	—	—
Expon	0.202	0.494	0.132	0.321	0.140	0.280	0.202
Gamma	**0.129**	0.207	**0.121**	**0.089**	0.087	**0.181**	—
InvGauss	—	0.194	—	0.119	**0.061**	—	0.134
LogLogistic	—	0.205	—	0.121	0.074	—	0.119
Lognorm	—	—	—	0.121	0.066	—	**0.115**
Lognorm2	—	**0.193**	—	—	—	—	—
Pareto	—	—	—	—	—	—	—
Pareto2	—	—	—	—	—	—	—
Pearson5	—	0.206	—	0.143	0.093	—	—
Pearson6	—	—	—	0.121	0.072	—	0.121
Rayleigh	0.293	0.407	0.322	0.123	0.317	0.185	0.505
Triang	0.356	0.337	0.189	0.501	0.420	0.277	0.428
Uniform	0.551	0.402	0.349	0.662	0.583	0.231	0.532
Weibull	—	0.227	—	0.124	0.088	—	0.144

Note: Best fit per column is boldface.

Table 5. K-S Fitness of Response Time Distributions for RFIs

Distrib.	Data set: Project 1	Project 2 noncivil	Project 2 civil	Project 3	Project 4	Project 5	Project 6	Project 7	Project 8 civil	Project 8 electrical	Project 8 installation	Project 8 mechanical
BetaGeneral	—	—	—	—	—	—	—	—	—	—	—	—
ChiSq	—	—	—	—	—	—	—	—	—	—	0.245	—
Erlang	—	**0.108**	0.087	—	0.109	0.181	0.104	—	**0.134**	0.155	0.128	—
Expon	0.225	**0.108**	0.087	0.171	0.109	0.181	0.104	0.382	**0.134**	0.155	0.128	0.232
Gamma	—	0.131	0.104	—	0.184	0.221	**0.103**	—	0.159	0.183	0.132	—
InvGauss	—	—	—	—	—	—	—	—	—	—	0.092	—
LogLogistic	—	—	—	—	—	—	—	—	—	—	0.103	—
Lognorm	—	—	—	—	—	—	—	—	—	—	0.107	—
Lognorm2	—	—	—	—	—	—	—	—	—	—	—	—
Pareto	—	—	—	—	—	—	—	—	—	—	—	—
Pareto2	**0.185**	0.113	**0.056**	**0.097**	**0.103**	**0.151**	0.119	**0.148**	—	**0.154**	—	**0.133**
Pearson5	—	—	—	—	—	—	—	—	—	—	**0.088**	—
Pearson6	—	—	—	—	—	—	—	—	—	—	0.098	—
Rayleigh	0.654	0.377	0.388	0.512	0.356	0.486	0.388	0.721	0.422	0.460	0.381	0.598
Triang	0.749	0.428	0.444	0.533	0.459	0.569	0.606	0.669	0.368	0.448	0.332	0.569
Uniform	0.816	0.587	0.595	0.660	0.656	0.704	0.741	0.749	0.487	0.602	0.381	0.68
Weibull	—	—	—	—	—	—	—	—	—	—	0.116	—

Note: Best fit per column is boldface.

J. Constr. Eng. Manage., 2008, 134(6): 442-450

Table 6. K-S for Fitness Scores of Different Distributions of Working Days between Arrivals

.	Data set																		
	Project number 1			Project number 2											Project number 8				
Distrib.	RFI	Shop drawings	Submittals	RFI (noncivil)	RFI (civil)	Submittals civil 1	Submittals civil 2	Submittals non civil 1	Submittals non civil 2	RFI proj. number 3	RFI proj. number 4	RFI proj. number 5	RFI proj. number 6	RFI proj. number 7	RFI (civil)	RFI (electrical)	RFI (installation)	RFI (mechanical)	Shop drawings proj. number 9
BetaGeneral	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
ChiSq	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Erlang	**0.33**	—	—	**0.55**	**0.32**	**0.45**	—	**0.53**	**0.70**	**0.45**	**0.36**	—	**0.38**	**0.44**	—	**0.21**	**0.20**	**0.11**	—
Expon	**0.33**	**0.69**	**0.69**	**0.55**	**0.32**	**0.45**	**0.57**	**0.53**	**0.70**	**0.45**	**0.36**	**0.33**	**0.38**	**0.44**	**0.30**	**0.21**	**0.20**	**0.11**	**0.41**
Gamma	0.44	**0.69**	**0.69**	**0.55**	0.41	0.49	—	0.56	**0.70**	0.52	0.37	—	0.39	0.48	—	0.27	0.28	0.17	—
InvGauss	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
LogLogistic	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Lognorm	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Lognorm2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Pareto	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Pareto2	0.43	**0.69**	**0.69**	**0.55**	**0.32**	—	—	—	—	—	–	—	—	—	—	**0.21**	—	0.12	—
Pearson5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Pearson6	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Rayleigh	0.52	**0.69**	**0.69**	0.73	0.51	0.67	0.96	0.62	0.73	0.59	0.61	0.65	0.65	0.52	0.57	0.50	0.43	0.37	0.59
Triang	0.55	**0.69**	**0.69**	0.82	0.56	0.71	0.95	0.66	0.78	0.52	0.73	0.68	0.78	0.45	0.60	0.52	0.33	0.42	0.54
Uniform	0.69	**0.69**	**0.69**	0.89	0.70	0.81	0.96	0.73	0.82	0.61	0.82	0.78	0.85	0.55	0.72	0.65	0.50	0.58	0.63
Weibull	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—

Note: Calendar days, best fit per column is boldface.

J. Constr. Eng. Manage., 2008, 134(6): 442-450

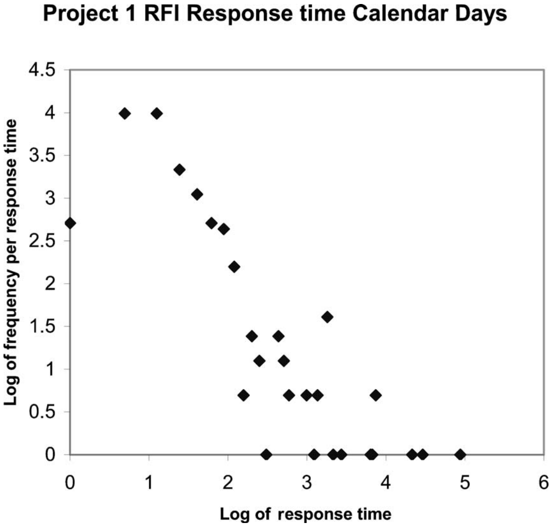


Fig. 2. Log-Log plot of RFI response time frequency (Project 1)

this is the unpredictability of the number of steps that different types of RFIs have to go through, but many more possible explanations for the phenomenon of power laws in can be found (e.g., Mandelbrot and Hudson 2004).

References

AbouRizk, S. M., and Halpin, D. W. (1992). "Statistical properties of construction duration data." *J. Constr. Eng. Manage.*, 118(3), 525–544.

AbouRizk, S. M., Halpin, D. W., and Wilson, J. R. (1994). "Fitting beta distributions based on sample data." *J. Constr. Eng. Manage.*, 120(2), 288–305.

Bou-Matar, R. S. (2005). "Statistical analysis of information processing tasks in a construction project." Master's thesis, American University of Beirut, Beirut, Lebanon.

Christensen, L., Christiansen, T. R., Jin, Y., Kunz, J., and Levitt, R. E. (1997). "Object-oriented enterprise modeling and simulation of aec projects." *Microcomput. Civ. Eng.*, 12, 157–170.

Harris, C. M., Brill, P. H., and Fischer, M. J. (2000). "Internet-type queues with power-tailed interarrival times and computational methods for their analysis." *INFORMS J. Comput.*, 12(4), 261–271.

Huberman, B. A., and Lukose, R. M. (1997). "Social dilemmas and internet congestion." *Science*, 277, 535–537.

Isidore, L., and Back, W. E. (2002). "Multiple simulation analysis for probabilistic cost and schedule intergration." *J. Constr. Eng. Manage.*, 128(3), 211–219.

Jin, Y., and Levitt, R. E. (1996). "The virtual design team: A computational model of project organizations." *Comput. Math. Org. Theory*, 2(2), 171–196.

Jin, Y., Levitt, R. E., Christiansen, T. R., and Kunz, J. C. (1995). "The virtual design team: A computer simulation framework for studying organizational aspects of concurrent design." *Simulation*, 64(3), 160–174.

Kleiber, C., and Kotz, S. (2003). *Statistical size distributions in economics and actuarial sciences*, Wiley, New York.

Kumaraswamy, M. M., and Chan, D. W. M. (1998). "Contributors to construction delays." *Constr. Manage. Econom.*, 16(1), 17–29.

Levitt, R. E., Cohen, G. P., Kunz, J. C., Nass, C., Christiansen, T. R., and Jin, Y. (1994). "The virtual design team: Simulating how organization structure and information processing tools affect team performance." *Computational organization theory*, K. M. Carley and M. J. Prietula, eds., Lawrence Erlbaum Associates, Hillsdale, N.J., 1–18.

Lu, M. (2002). "Enhancing project evaluation and review technique simulation through artificial neural network-based input modeling." *J. Constr. Eng. Manage.*, 128(5), 438–445.

Maio, C., Schexnayder, C., Knutson, K., and Weber, S. (2000). "Probability distribution functions for construction simulation." *J. Constr. Eng. Manage.*, 126(4), 285–292.

Mandelbrot, B. B., and Hudson, R. L. (2004). *The (mis)behaviour of markets: A fractal view of risk, ruin and reward*, Profile Books, London.

Odeh, A., and Battanieh, H. (2002). "Causes of construction delay: Traditional contracts." *Int. J. Proj. Manage.*, 20, 67–73.

Salazar-Kish, J. M. (2001). "Modeling concurrency trade-offs and their effects on project duration and rework." Ph.D. thesis, Stanford Univ., Stanford, Calif.

Schexnayder, C., Knutson, K., and Fente, J. (2005). "Describing a beta probability distribution function for construction simulation." *J. Constr. Eng. Manage.*, 131(2), 221–229.

J. Constr. Eng. Manage., 2008, 134(6): 442-450