ATTACHMENT C

NCHRP Project 22-12(3)

Recommended Guidelines for the Selection of Test Levels 2 Through 5 Bridge Railings

Predicting Life-Crash Costs using RSAP

Introduction

RSAP predicts the expected crash cost on each segment for each alternative over the design life of the project. The prediction can usually be interpreted as representing the expected cost at the mid-life conditions (i.e., the ADT and traffic mix at the middle of the design life). The following sections present an example based on concrete New Jersey median barriers which assess how accurately RSAP predicts the expected crash costs in comparison to several real-world cases.

As will be shown shortly and as illustrated in Figure 1, the crash cost observed in any particular year is a random variable. Figure 1 shows crash data collected on a portion of the New Jersey Turnpike where continuous TL5 concrete median barriers are used as will be discussed later. At this point, however, it is useful to notice several things about the variation of the observed crash data represented by the dots in Figure 1. Notice that the horizontal axis scale is logarithmic and that the general shape of the curve indicated by the dots is a familiar "S" shaped curve which hints that some type of exponential form will be required. A distribution for crash costs would be expected to have the following characteristics:

- Crash costs may never be negative,
- The most frequently observed segment crash cost (i.e., the mode) will usually be zero or very small,
- The data will be dominated by lower and moderate costs with the very occasional high crash-costs on a particular segment in a particular year and
- Unlike normal distributions, the percentile of the mean of a distribution in an exponential form is not known *a priori* but is a function of the distribution parameters.

These characteristics suggest that three good choices for modeling crash costs are the lognormal distribution, the Weibull distribution or the Gamma family of distributions.[Ang07] Each of these distributions share the characteristics listed above and they are also commonly used in many other areas of risk and failure analysis.

First, each of the candidate distributions will be discussed in terms of their utility for modeling crash costs. Next, three sets of real-world crash data will be examined to determine which of the distributions best matches the crash data. Once the distributions are known, the parameters can be estimated from the observed crash data and the 95 percent confidence intervals for the mean values can be calculated and compared to the RSAP predicted value. If the RSAP prediction lies within the 95th percent confidence interval of the mean for the observed data and at least one of the theoretical distributions, RSAP can be considered validated for that case since the RSAP prediction cannot be statistically rejected as coming from the same distribution.

Lognormal Distribution

The cumulative density function, F(x); mean E(x); median, Md(x); and mode Mo(x) of the lognormal distribution are given by:

$$F(x) = \frac{1}{2} + \frac{1}{2} erf\left[\frac{\ln x - \mu}{\sqrt{2\sigma^2}}\right]$$
$$E(x) = e^{\mu + \frac{\sigma^2}{2}}$$
$$Md(x) = e^{\mu}$$
$$Mo(x) = e^{\mu - \sigma^2}$$

Where:

x	=	The segment crash cost in a particular year,
μ	=	The mean of the log of the observed crash costs,
σ	=	The standard deviation of the log of the observed crash costs and
erf	=	Complimentary error function.

The crash cost predicted by RSAP corresponds to the expected value of the crash cost over the life of the project. As shown above, the expected value is a function of the mean and standard deviation of the distribution of the logarithms of the crash costs on each segment in each year. In a lognormal distribution the logarithms of the crash cost are presumed to be normally distributed. μ and σ are easily calculated from the observed data and can then be compared to the RSAP prediction. The mean value of an observed sample is presumed to be a random normal variable and, therefore, the 95^{th} percent confidence intervals can be calculated. For the lognormal distribution a naïve approach to calculating the 95^{th} percentiles will not always “cover” the mean so it is better to use an alternate method. Herein, a modified Cox method as follows: [cox]

$$E(X)_{95} = e^{\left\{\mu + \frac{\sigma^2}{2} \pm z_{95,n}\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^4}{2(n-1)}}\right\}}$$

where the terms are as defined above and n is the number of segment-year observations and z is the standard z score for n samples and 95 percent confidence. If the RSAP predicted expected value is between the values given by the modified Cox method for the lognormal distribution, the RSAP model is valid.

Weibull Distribution

The Weibull distribution is a frequently used probability distribution used in failure and risk analyses. [Ang07] Important characteristics of this distribution are given by the following equations:

$$F(x, A, B) = 1 - e^{-\left[\frac{x}{B}\right]^A}$$

$$E(x) = B\Gamma\left[1 + \frac{1}{A}\right]$$
$$Md(x) = B(ln2)^{1/A}$$
$$E(x)_{95} = E(x) \pm t_{n,95}\frac{\sigma}{\sqrt{n}}$$

Where the parameters are the same as defined above and Γ is the gamma function, $t_{n,95}$ is the t statistics for the 95th percentile with n segment-years of data and A and B are parameters of the distribution that need to be fitted to the observed data. There is no closed-form expression for the mode for values of A less than 1 and, as will be shown in the example, A for crash costs is generally less than 1. If the RSAP predicted expected value falls within the 95th percentile confidence range of the mean of the Weibull distribution, the RSAP model is valid.

Gamma Distribution

The Gamma distribution is actually a broad family of distributions that are widely used in failure and risk analyses. [Ang07] Like the Weibull distribution, the model is very flexible and can be used to develop detailed fits of observed data. Important characteristics of this distribution are given by the following equations:

$$F(x, A, B) = \frac{1}{\Gamma(A)}\gamma(A, \frac{x}{B})$$
$$E(x) = B\mathrm{A}$$
$$E(x)_{95} = E(x) \pm t_{n,95}\frac{\sigma}{\sqrt{n}}$$

where the parameters are the same as defined above. There is no general closed-form expression for either the median or the mode for the gamma distribution.

Concrete Median Barrier Example

A TL5 New Jersey shape concrete median barrier in the center of a 27-ft wide median on a divided highway is examined to illustrate the distribution of segment-life crash costs. The section examined is a 69-mi long portion of the New Jersey Turnpike which uses TL5 concrete median barriers or bridge rails continuously for the entire distance in the study section. In examining the crash data, the cases were carefully reviewed to include only cases that involved:

- A collision with the concrete median barrier,
- A non-barrier related terrain rollover or
- A median cross-over event.

Since these are the only events modeled in RSAP for this case, it is important to exclude any extraneous cases that involve collisions with items not explicitly modeled in RSAP. For example, if there were a collision with a bridge pier in the observed data this would be excluded since bridge piers were not included in the RSAP model of this highway. This case is a particularly useful validation case since the only possible crashes for leftward encroachments are striking the median barrier so there are no confounding crashes.

Each section of roadway was divided into one-mile segments and the New Jersey crash data for 2005 through 2007 were used to determine the severity and number of crashes on each segment in each year. The segment crash cost was then calculated using the 2009 FHWA crash cost recommendations (i.e., a fatal crash is estimated to have a comprehensive cost of $6,000,000).[Miller] This resulted in 207 segment-years of data which are shown as dots in Figure 1.

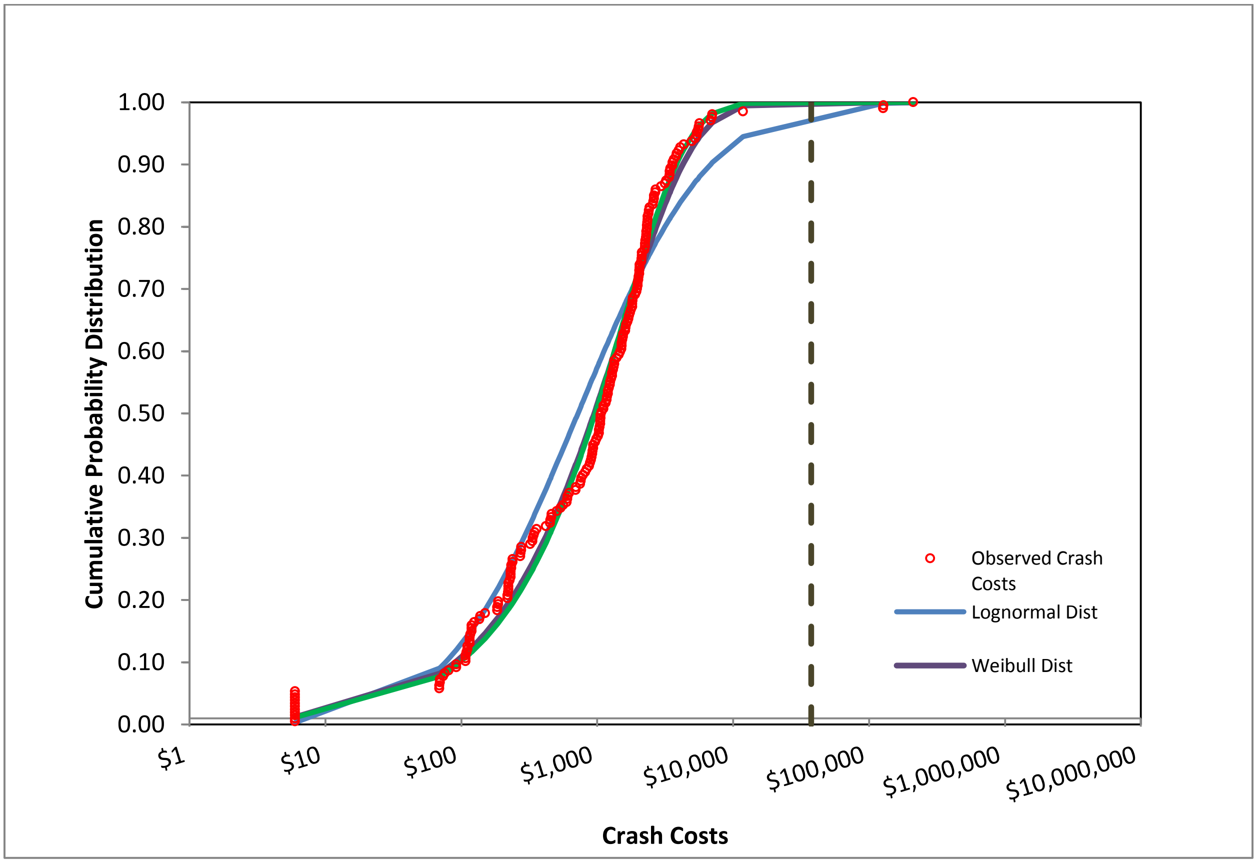

Figure 1. Observed and RSAP predicted crash costs for a 27-ft wide median with TL5 concrete median barriers.

RSAP was used to estimate the expected value of the crash costs for both the before (i.e., no median barrier) and after (i.e., median barrier installed) cases. A median barrier has been present on this section of the New Jersey Turnpike for many decades so "before" data is not readily available but the2005-2007 crash data can be used to assess the accuracy of the RSAP prediction of the expected crash cost in the "after" condition. Similarly, 34-miles of the highway use a four-lane cross-section and the remaining portion uses a six-lane cross-section. In order to group all the segments together the results were examined on a million-vehicle-miles-traveled (MVMT) basis rather than a crash cost/mi/yr basis.

In fact, Figure 1 shows that there is a cluster of $1 crash costs at the far left of the observed data. These data actually represent segments with zero crash costs but a value of $1 was used since the ln(0) is undefined. In fact, the zero values observed should more properly be considered "unobservable" since crashes may have occurred but were of such low severity that they were not reported to the police. The estimated comprehensive cost for property damage only crash is $2000 so any crash resulting is less than $2000 of costs would likely not be counted. A zero crash cost on a particular segment in a particular year could, therefore, have a cost anywhere between zero and $2000 but it would not be recorded by the police so the segment crash cost is not observable. In fact, there could be several crashes on the segment with crash costs less than $2000 so unless at least one crash has a crash cost greater than $2000 the observed segment crash cost for that year will be zero. On the other hand, a fatal crash has a comprehensive cost of $6,000,000. These are very rare but when they do occur they have a dramatic effect on the observed mean crash cost. As shown in Figure 1, there are a few segment-years where there were severe crashes that result in data points at the far right of the distribution.

The RSAP analysis resulted in an expected crash cost of $37,462 for a mid-life ADT of 58,887 vehicles/day and 21.5 MVMT. In order to include as many segments from the actual crash data as possible it is convenient to represent this in terms of vehicle-miles-travelled so the RSAP results correspond to an expected crash cost of 1,743 $/MVMT.

Table 1. Statistical properties of the concrete median barrier validation example.

Distribution	Segment-Years	Distribution Parameters		Lower 95th Percentile Crash Cost	Expected Crash Cost	RSAP Predicted Crash Cost	Upper 95th Percentile Crash Cost	Median	Mode	Coefficient of Determination	Critical K-S Statistic = 0.0945
Lognormal	207	μ=6.58	σ=1.75	$1,061	$1,730	$1,743	$2,269	$721	$125	0.95	0.13
Weibull	207	A=0.80	B=1501.00	$1,389	$1,701		$2,012	$949	NA	0.99	0.06
Gamma	207	A=0.78	B=2000.00	$1,319	$1,560		$1,801	NA	NA	0.99	0.07
Data	207	NA	NA	$1,108	$3,716		$6,324	NA	NA	NA	NA

Table 1 shows the statistical results for the concrete median barrier example. First, parameters for the three distributions (i.e., lognormal, Weibull and gamma) were estimated using the maximum likelihood method and the observed data. Goodness of fit was evaluated using the coefficient of determination (i.e., R^2) and the Kolmogorov-Smirnov (i.e., K-S) test. As shown in Table 1, all three distributions result in very high R^2 values but the lognormal distribution did not pass the K-S test at the 95 percent confidence level. This indicates that the Weibul and Gamma distributions are better fits for the observed data than the lognormal distribution and this is also confirmed visually by examining Figure 1. The Weibull and Gamma curves plot essentially on top of each other and fit the data quite well. This is not surprising since the Weibull and Gamma distribution have very similar forms.

Next, the RSAP prediction of $1,743/MVMT is compared to the observed data and the three distributions. A 95th percentile confidence range on the mean was calculated for the observed data and the three distributions as shown in Table 1. The narrowest range results from the Gamma distribution and is only $481 wide. The 95 percent confidence range for the observed data is quite large, about $6300. This shows that finding the best fitting distribution essentially adds information about the shape of the curve making the range narrower. The RSAP prediction falls within the 95th percentile confidence interval for the mean for all three distributions and for the observed data so the RSAP prediction is validated for this example. In statistical terms, there is no basis to reject the hypothesis that the RSAP predictions does not come from the same distribution of means as the observed data and the theoretical distributions.

The number of events predicted by RSAP can also be compared to the number of events observed on the test section. As shown in Table 2, RSAP predicts a total of 0.1953 (i.e., 0.1843+0.0102+0.0008) median barrier collisions/MVMT whereas 0.1555 median barrier collisions/MVMT were actually observed. RSAP should over predict the number of collisions because the observed data does not include unreported and under-reported collisions whereas RSAP does. There were 0.0074 median cross-overs/MVMT predicted and none were observed and while 0.0047 “terrain” rollovers were predicted none were actually observed. On balance the RSAP estimate of the number and types of events were quite similar to what was actually observed on the 69-mile test section of the New Jersey Turnpike.

Table 2. Observed and predicted frequency of events.

Collision Type	**RSAP**	**Observed**
Median Barrier Collisions		
No rollover	0.1843	0.1510
Penetrated/rolled or vaulted	0.0008	0.0045
Rollover on traffic side	0.0102	0.0000
Median Crossovers	0.0074	0.0000
Terrain Rollovers	0.0047	0.0000

A comparison of the observed crash costs and events for the 69-mile section of the New Jersey Turnpike indicate that the RSAP model is a valid model of the performance of the median barrier. The observed events in Table 2 also show that this TL-5 concrete barrier is a very safe median barrier. Further, this example indicates that crash costs on this portion of the New Jersey Turnpike conform to a Weibull distribution and a Gamma distribution with parameters estimated from the data.

Estimating Extreme Values with RSAP

The foregoing example illustrates that crash costs are distributed as a Weibull or Gamma distribution and that both of these distributions work equally well. Traditionally RSAP only reports the mean crash cost but in developing the mean, tens of thousands of crashes are simulated. Each individual crash could occur in any year and in combination with any other modeled crash.

The research team developed two add-on subroutines to the standard RSAP. RSAP saves the crash severity of every modeled crash and then writes them to a list. Using the mean number of crashes per year and the list of crash costs a Monte Carlo simulation was performed to create a sample of the segment-life crash costs over the design life of the project. Once the crash costs had been estimated in each year using a Monte Carlo simulation the parameter distributions (i.e., Weibull and Gamma) were estimated using the maximum likelihood method. The result is an estimate of the distribution of crash costs over the life of the project which allows the analyst to determine precisely the probability of experiencing a cost greater than any arbitrary value. Since penetrations of a bridge rail are a rare (i.e., generally around one or two percent of reported crashes) but costly events, the probability of experiencing a very large crash cost can be determined for use in the later analysis.

References

Ang A. H-S. Ang and W. H. Tang, Probability Concepts in Engineering, 2nd Edition, John Wiley and Sons, Inc., New York, 2007.

Cox U. Olsson, "Confidence intervals for the mean of a log-normal distribution," Journal of Statistics Education, Vol. 13, No. 1, 2005.