

4 Probability for Seismic Hazard Analyses

This chapter gives a brief review of the basic concepts and models in probability theory that are commonly used in seismic hazard analysis. For more in depth discussions of probability theory, refer to an introductory probability text book.

This chapter also includes a discussion of aleatory variability and epistemic uncertainty which are commonly used terms in seismic hazard analyses. These concepts are not included in introductory probability textbooks, so a more extensive discussion is given of these concepts.

4.1 Discrete vs Continuous Variables

There are two types of random variables used in seismic hazard analyses: discrete and continuous. Discrete variables have a finite number of possible values, whereas, continuous variables have an infinite number of possible values. For example, a variable that can be any integer value from 1 to 10 is a discrete variable. There are 10 possible values that the variable could have. In contrast, an example of a continuous variable is a number between 0 and 10. There are an infinite number of possible values.

If we take a random sample from a distribution of discrete variables, then the probability of selecting a particular value x_i is given by fraction of the total population that has a value of x_i . For example, if the integer numbers from 1 to 10 are equally likely, then the probability of selecting a 5 is 1/10. The notation $P(x_i)$ will be used for the probability of observing the value x_i .

A property for discrete variables is that the sum of the probabilities of all possible values is unity. If there are N possible values of x_i , then

$$\sum_{i=1}^N P(x_i) = 1 \tag{4.1}$$

If we take a random sample from a distribution of continuous variables, there are an infinite number of possible values. The probability of selecting a particular value, x , is zero. To handle continuous variables, probability density functions are used instead of probabilities. The probability density function (pdf) for the random variable x will be denoted $f(x)$. The pdf gives the relative likelihood of different values of a continuous variable. For small dx , the probability density function is related to probability by

$$P(x - \frac{dx}{2} < x < x + \frac{dx}{2}) = f(x) dx \quad (4.2)$$

where $f(x)$ has the units of $1/x$.

For discrete variables, the probabilities sum to unity (eq. 4.1). For continuous variables, the corresponding constraint is that the pdf integrates to unity:

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (4.3)$$

The probability of observing a value of x greater than some value z is given by

$$P(x > z) = \int_z^{\infty} f(x) dx \quad (4.4)$$

And the probability of observing a value of x between two values, z_1 and z_2 is given by

$$P(z_1 \leq x \leq z_2) = \int_{z_1}^{z_2} f(x) dx \quad (4.5)$$

For continuous variables, we can compute probabilities only when defining a range of values (e.g. $z_1 < x < z_2$) or for exceeding (or not exceeding) a value (e.g. $x > z$).

4.2 Conditional Probability

When working with more than one random variable at a time, we need to consider conditional probabilities. First, for discrete variables, the probability of x_i occurring

given that y_j has occurred is called the conditional probability and is denoted $P(x_i|y_j)$. The vertical bar denotes “given”. The joint probability that both x_i and y_j occur is given by

$$P(x_i, y_j) = P(y_j) P(x_i|y_j) \quad (4.6)$$

If the variables x and y are independent, then $P(x_i|y_j) = P(x_i)$. For continuous variables, the joint probability density function is given by

$$f(x, y) = f(y) f(x|y) \quad (4.7)$$

and if the x and y are independent, then $f(x, y) = f(x)$.

4.3 Mean, Median, and Mode

There are three different measures of an “average” that are commonly used in seismic hazard analyses: mean, median, and mode. In general, these three different measures of “average” do not lead to the same value. The normal distribution is a special case for which the mean, median, and mode are all equal.

4.3.1 Mean

For discrete variables, the mean is the weighted average of the x_i values where the weights are the probabilities of the x_i values.

$$Mean[x] = \sum_{i=1}^N x_i P(x_i) \quad (4.8)$$

For continuous variables, we replace the probability with the density function, $f(x)$, times the step size, dx .

$$Mean[x] = \int_{-\infty}^{\infty} x f(x) dx \quad (4.9)$$

In some cases, a different parameter, g , will be a function of the random variable, x . In that case, then the mean of $g(x)$ is given by

$$Mean[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (4.10)$$

4.3.2 Median

The median is the value for which there is a 50% chance of a larger (or smaller) value. the probability of exceeding the value is 0.5 (e.g. there is a 50% chance of observing a value larger than the median).

For discrete variables, the median is determined by summing the $P(x_i)$ until $P(x_i) \geq 0.5$. The smallest x_i for which the sum of the probabilities is greater than or equal to 0.5 is the median.

For continuous variables, the median is the z value such that

$$\int_z^{\infty} f_x(x) dx = 0.5 \quad (4.11)$$

4.3.3 Mode

The mode is the value that is most likely to occur. For discrete variables, it is the value with the largest probability. For continuous variables, it is the value which gives the maximum of the pdf.

4.3.4 Example

An example of the these different averages is given below using discrete variables. Table 4-1 gives the probabilities from a discretized lognormal distribution (see 4.5.4) for peak acceleration. As shown in Figure 4-1, the lognormal distribution is skewed to the right so the mean is greater than then median. For this distribution, the mode is at 0.20g, the median is 0.25g, and the mean is 0.30g.

Table 4-1. Example of a discrete probability distribution for peak acceleration (see Figure 4-1).

PGA (g)	Probability	Cumulative Probability
0.05	0.02233	0.022
0.10	0.10160	0.124
0.15	0.15211	0.276
0.20	0.15420	0.430
0.25	0.13280	0.563
0.30	0.10594	0.669
0.35	0.08136	0.750
0.40	0.06135	0.812
0.45	0.04588	0.858
0.50	0.03424	0.892
0.55	0.02558	0.917
0.60	0.01918	0.937
0.65	0.01444	0.951
0.70	0.01093	0.962
0.75	0.00832	0.970
0.80	0.00637	0.977
0.85	0.00490	0.982
0.90	0.00379	0.985
0.95	0.00295	0.988
1.00	0.00231	0.991
1.05	0.00182	0.992
1.10	0.00144	0.994
1.15	0.00114	0.995
1.20	0.00091	0.996
1.25	0.00073	0.997
1.30	0.00059	0.997
1.35	0.00048	0.998
1.40	0.00039	0.998
1.45	0.00031	0.998
1.50	0.00026	0.999

4.4 Standard Deviation and C.V.

The standard deviation is a measure of the range of random variables. It is usually denoted σ . Formally, the standard deviation is the square root of the mean of the differences between the values and the mean of the values. That is

$$\sigma_x = \sqrt{\text{Mean}[(x - \bar{x})^2]} \quad (4.12)$$

where \bar{x} is the mean of x . The variance is the square of the standard deviation.

In some cases, the term “standard error” is used rather than standard deviation. A population has a standard deviation. The standard error is an estimate of the standard deviation of the population computed from a sample from the population. It is common to see these terms used interchangeably.

The coefficient of variation (C.V.) is the standard deviation divided by the mean

$$C.V._x = \frac{\sigma_x}{\text{Mean}[x]} \quad (4.13)$$

4.5 Probability Distributions Commonly Used in Seismic Hazard Analyses

There are several forms of probability density functions for continuous variables that are commonly used in seismic hazard analyses. These are presented below. The equations for the probability density functions are summarized in Table 4-3 at the end of this section.

4.5.1 Uniform Distribution

The uniform distribution is used to describe random variables for which all values are equally likely (over a finite range of values). An example of a random variable that is often modelled by a uniform distribution is the location of an earthquake along a fault. If the random variable is restricted to the range 0 to L (Figure 4-2a), then the pdf for the uniform distribution is given by

$$f^U(x) = \begin{cases} 1/L & \text{for } 0 \leq x \leq L \\ 0 & \text{otherwise} \end{cases} \quad (4.14)$$

4.5.2 Normal Distribution

The normal distribution is a commonly used distribution in modelling natural processes. In seismic hazard analysis, the normal distribution is commonly used for the distribution of magnitudes of an earthquake given a rupture dimension (e.g. the conditional distribution of M given rupture area). For a normal distribution, the pdf is given by

$$f^N(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(\frac{-(x - \bar{x})^2}{2\sigma_x^2}\right) \quad (4.15)$$

where σ_x is the standard deviation and \bar{x} is the mean.

In some cases, the normal distribution will be truncated at a maximum number of standard deviations (Figure 4-2b). If the distribution is truncated, then the pdf needs to be renormalized such that it integrates to unity. If the distribution is truncated at both the high and low ends, then the truncated normal distribution is given by

$$f^{TN}(x) = \begin{cases} \frac{1}{(2\Phi(nsig_{\max}) - 1)} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(\frac{-(x - \bar{x})^2}{2\sigma_x^2}\right) & \text{for } \left| \frac{x - \bar{x}}{\sigma_x} \right| \leq nsig_{\max} \\ 0 & \text{for } \left| \frac{x - \bar{x}}{\sigma_x} \right| > nsig_{\max} \end{cases} \quad (4.16)$$

where $nsig_{\max}$ is the maximum number of standard deviations above or below the mean and $\Phi(x)$ is the standard cumulative normal distribution. Tables of $\Phi(x)$ are given in most probability textbooks. Some values of $\Phi(x)$ are listed in Table 4-2 for typical values of $nsig_{\max}$ used in seismic hazard analysis.

Table 4-2. Tabulated values of the cumulative normal probability distribution

$nsig_{max}$	$\Phi(nsig_{max})$
2.0	0.977
2.5	0.994
3.0	0.999

4.5.3 Exponential Distribution

For an exponential distribution, the pdf is given by

$$f^E(x) = \alpha \exp(-\alpha x) \quad (4.17)$$

for $x > 0$. The exponential distribution has only one parameter, α , which has units of $1/x$.

In seismic hazard analyses, the most common use of the exponential distribution is for the distribution of earthquake magnitudes on a source. Since there is a maximum magnitude for the source and also a minimum magnitude of engineering interest, the exponential distribution is typically truncated at both the high and low ends (Figure 4-2c). If the exponential distribution is truncated at the x_{min} and x_{max} , then the renormalized pdf is given by

$$f^{TE}(x) = \begin{cases} 0 & \text{for } x < x_{min} \\ \frac{\alpha \exp(-\alpha(x - x_{min}))}{1 - \exp(-\alpha(x_{max} - x_{min}))} & \text{for } x_{min} \leq x \leq x_{max} \\ 0 & \text{for } x > x_{max} \end{cases} \quad (4.18)$$

4.5.4 Lognormal Distribution

Another commonly used distribution in seismic hazard analysis is the lognormal distribution. A lognormal distribution indicates that the logarithm of the random variable, x , is normally distributed. The lognormal distribution is typically used for ground motion parameter values (e.g. spectral acceleration) and for source-scaling relations (e.g. rupture area as a function of the earthquake magnitude). It is also used to model the recurrence intervals of large earthquakes on faults. The lognormal pdf is given by

$$f^{LN}(x) = \frac{1}{\sqrt{2\pi}\sigma_{\ln x} x} \exp\left[-\frac{(\ln(x) - \ln(\mu))^2}{2\sigma_{\ln x}^2}\right] \quad (4.19)$$

where μ is the median value (in units of x) and $\sigma_{\ln x}$ is the standard deviation in natural log units. The lognormal distribution is skewed to high values so that the mean value is greater than the median (Figure 4-2d). For a lognormal distribution, the relation between the mean and the median is

$$Mean[x] = \exp(\sigma_{\ln x}^2 / 2) \mu \quad (4.20)$$

The relation between the C.V. and $\sigma_{\ln x}$ can be given by

$$(4.21)$$

When used to model the distribution of ground motions, the log-normal distribution will often be truncated in terms of the maximum number of standard deviations ($nsig_{\max}$) or in terms of the maximum value of the ground motion (x_{\max}) as shown in Figure 4-3. A truncation based on x_{\max} reflects the physical limitations on the ground motions; whereas, a truncation based on $nsig_{\max}$ reflects limitations of the statistical model.

If the distribution is truncated, then the lognormal distribution needs to be renormalized. For a given median, μ , and standard deviation, $\sigma_{\ln x}$, the renormalization will depend on which truncation controls (e.g. is it controlled by the maximum number of standard

deviations or the maximum absolute level of the ground motion). The truncated lognormal pdf is given by

$$f^{TLN}(x) = \begin{cases} \frac{1}{\Phi(c)} \frac{1}{\sqrt{2\pi}\sigma_{\ln x} x} \exp\left(-\frac{(\ln(x) - \ln(\mu))^2}{2\sigma_{\ln x}^2}\right) & \text{for } x \leq \mu \exp(\sigma_{\ln x} c) \\ 0 & \text{otherwise} \end{cases} \quad (4.22)$$

where

$$c = \min\left(\frac{\ln(x_{\max}) - \ln(\mu)}{\sigma_{\ln x}}, \text{nsig}_{\max}\right) \quad (4.23)$$

4.5.5 Delta Function Distribution

The simplest distribution that will be considered is the delta function distribution. The delta function is a mathematical tool that causes a continuous variable to have only one possible value (Figure 4-2e). A delta function is given by

$$\delta(x) dx = \begin{cases} 1 & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.24)$$

A property of a delta function is

$$\int_{-\infty}^{\infty} \delta(x - c) g(x) dx = g(c) \quad (4.25)$$

That is, for a delta function, there is no variability of the parameter x. As shown by eq. (4.25), delta functions reduce integrals to single values.

4.5.6 Composite Distributions

Composite distributions can also be used which are combinations of different distributions. The requirement is that the composite distribution is renormalized such that it integrates to unity.

An example of a composite distribution commonly used in seismic hazard analysis is the Youngs and Coppersmith (1985) characteristic earthquake model for the distribution of earthquake magnitudes on a fault (see Chapter 6). This distribution is a combination of an exponential distribution for small to moderate magnitude earthquakes and a uniform distribution for large magnitude earthquakes (Figure 4-4). This composite distribution can be written as

$$f(x) = c_1 f^{TE}(x) + c_2 f^U(x) \quad (4.26)$$

where $c_1 + c_2 = 1$.

Table 4-2. Summary of probability density functions commonly used in PSHA.

Distribution	Probability Density Function
Uniform	$f^U(x) = \begin{cases} 1/L & \text{for } 0 \leq x \leq L \\ 0 & \text{otherwise} \end{cases}$
Truncated Normal	$f^{TN}(x) = \begin{cases} \frac{1}{2\Phi(nsig_{\max})} \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(\frac{-(x-\bar{x})^2}{2\sigma_x^2}\right) & \text{for } \frac{x-\bar{x}}{\sigma_x} \leq nsig_{\max} \\ 0 & \text{for } \frac{x-\bar{x}}{\sigma_x} > nsig_{\max} \end{cases}$
Truncated Exponential	$f^{TE}(x) = \begin{cases} 0 & \text{for } x < x_{\min} \\ \frac{\alpha \exp(-\alpha(x-x_{\min}))}{1 - \exp(-\alpha(x_{\max}-x_{\min}))} & \text{for } x_{\min} \leq x \leq x_{\max} \\ 0 & \text{for } x > x_{\max} \end{cases}$
Truncated Lognormal	$f^{TLN}(x) = \begin{cases} \frac{1}{\Phi(c)} \frac{1}{\sqrt{2\pi}\sigma_{\ln x} x} \exp\left(\frac{-(\ln(x) - \ln(\mu))^2}{2\sigma_{\ln x}^2}\right) & \text{for } x \leq \mu \exp(\sigma_{\ln x} c) \\ 0 & \text{otherwise} \end{cases}$ <p>Where</p> $c = \min\left(\frac{\ln(x_{\max}) - \ln(\mu)}{\sigma_{\ln x}}, nsig_{\max}\right)$
Delta Function	$f^\delta(x) = \delta(x - c)$

4.6 Poisson Process

If a random process generates events at some average rate, ν , and the occurrence of an event does not depend on the time since the last event, then it is a Poisson process. For a Poisson process with an average rate of events ν , the probability of observing n events in time interval T is given by

$$P(n | T) = \frac{\exp(-\nu T)(\nu T)^n}{n!} \quad (4.27)$$

where n is an integer. In seismic hazard analysis, we will typically be concerned with the probability of one or more events in a given time interval (e.g. $n \geq 1$). By definition, the sum of the probabilities of all values of n must be unity (eq. 4.1). Therefore, the probability of one or more events is just one minus the probability of $n=0$ and is given by

$$P(n \geq 1 | T) = 1 - P(n = 0 | T) = 1 - \exp(-\nu T) \quad (4.28)$$

A common assumption used in PSHA is that the occurrence of earthquakes is a Poisson process. That is, there is no memory of past earthquakes, so the chance of an earthquake occurring in a given year does not depend on how long it has been since the last earthquake. If the occurrence of earthquakes follows a Poisson process then the occurrence of peak ground motions also follows a Poisson process. The Poisson process model provides one method for converting rates of ground motions to probabilities over a specified time interval. Alternative earthquake probability models will be discussed in Chapter 11.

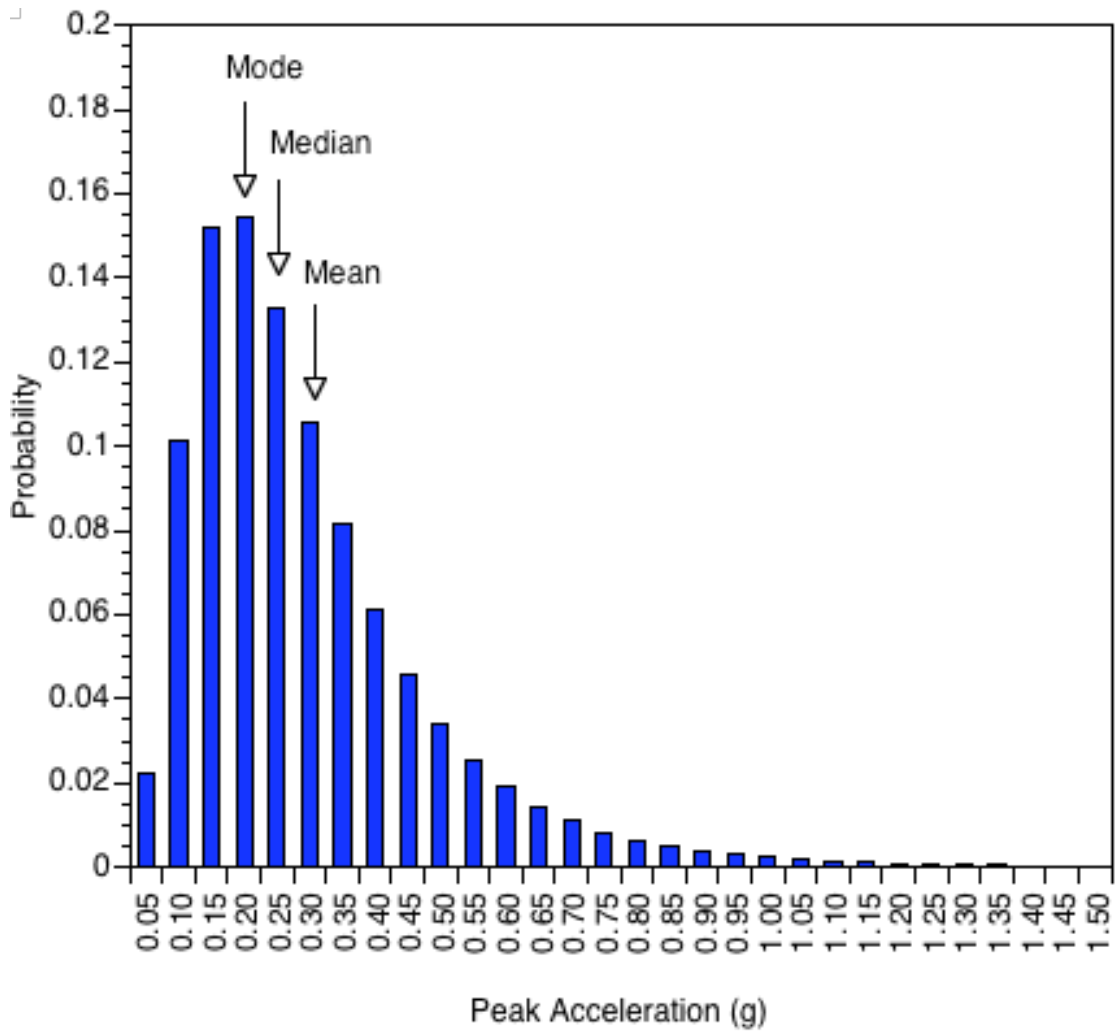


Figure 4-1. Example of the different measures of average: mode, median, and mean.

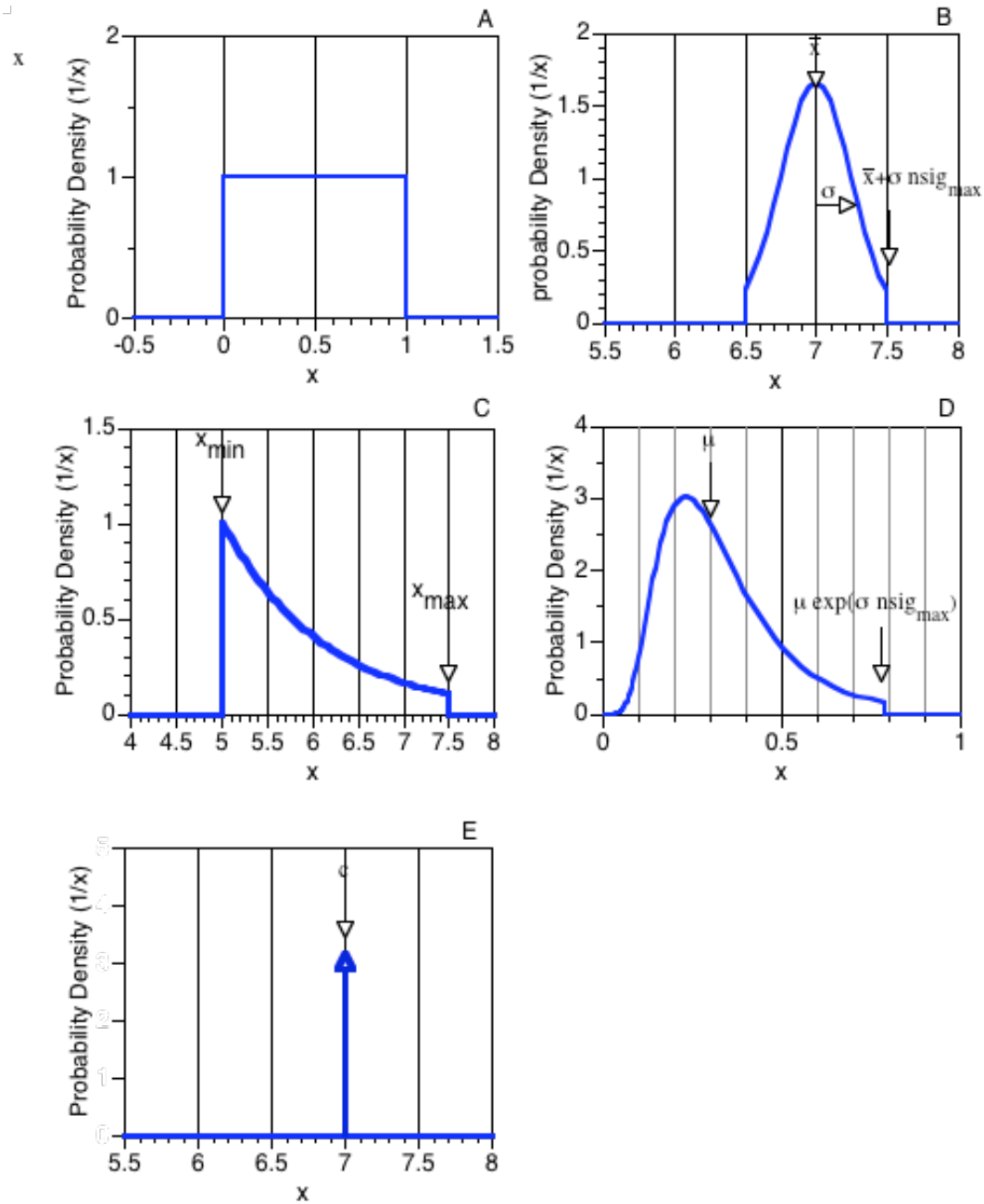


Figure 4-2. Probability density functions commonly used in seismic hazard analyses. (A) Uniform. (B) Truncated normal. (C) Truncated exponential. (D) Truncated log-normal. (E) Delta function.

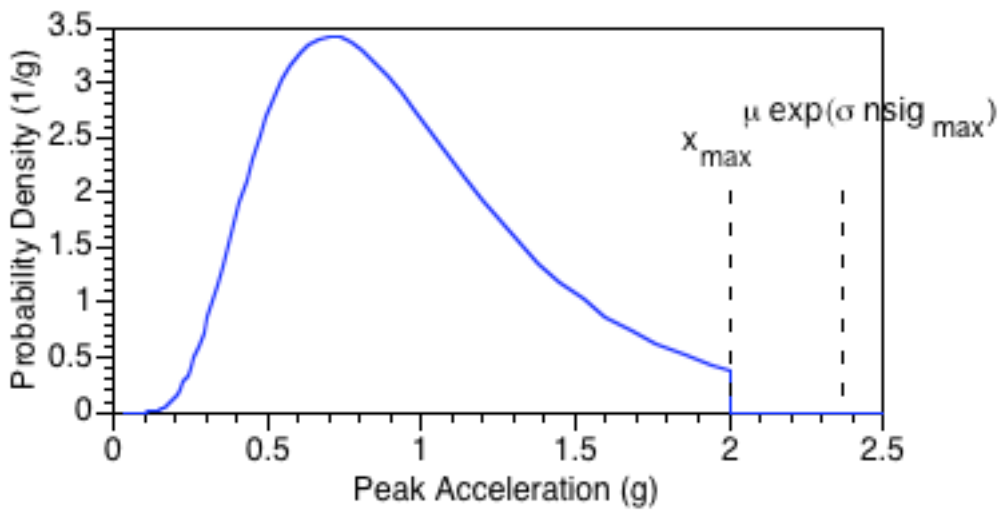
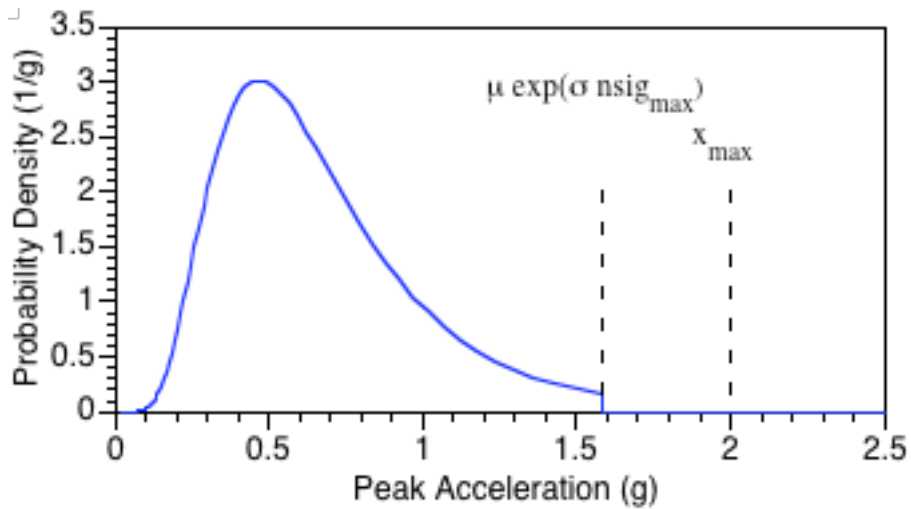


Figure 4-3. The truncation may be controlled by the maximum number of standard deviations (upper frame) or the maximum value of the parameter (lower frame).

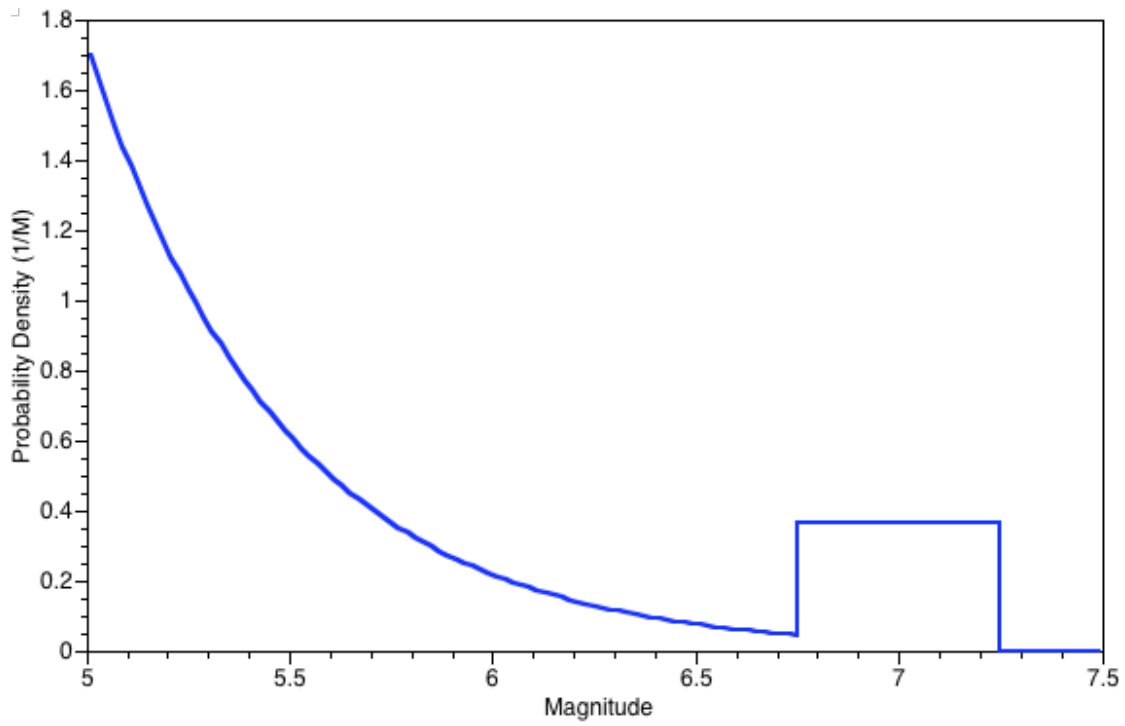


Figure 4-4. Example of a composite distribution: Youngs and Coppersmith (1984) characteristic earthquake model.