INTRODUCTION TO PROBABILISTIC SEISMIC HAZARD ANALYSIS

JACK W. BAKER

Preferred citation for this document:

Baker, Jack W. (2013) Probabilistic Seismic Hazard Analysis. *White Paper Version 2.0.1*, 79 pp.

TUFTE-LATEX.GOOGLECODE.COM

Contents

1 Introduction **13**

2 An overview of PSHA **15**
2.1 Deterministic versus probabilistic approaches 15
2.1.1 Variability in the design event 15
2.1.2 Variability of ground motion intensity 16
2.1.3 Can we use a deterministic approach, given these uncertainties? 18
2.2 Probabilistic seismic hazard analysis calculations 18
2.2.1 Identify earthquake sources 19
2.2.2 Identify earthquake magnitudes 19
2.2.3 Identify earthquake distances 24
2.2.4 Ground motion intensity 27
2.2.5 Combine all information 32
2.3 Example PSHA calculations 33

3 Extensions of PSHA **43**
3.1 Deaggregation 43
3.2 Bounds on considered magnitudes and distances 48
3.3 Rates, probabilities and return periods 49
3.4 The uniform hazard spectrum 52
3.5 Joint distributions of two intensity measures 52

4 Conclusions **59**

A Review of probability **61**
A.1 Random events 61
A.2 Conditional probabilities 62
A.3 Random variables 66
A.4 Expectations and moments 73

B Further study **75**

Bibliography **77**

List of Figures

1.1 Quantification of the possibility of intense ground shaking at a site. 14

2.1 (a) Map view of an illustrative site, with two nearby sources capable of producing earthquakes. (b) Predicted median response spectra from the two earthquake events, illustrating that the event producing the maximum response spectra may vary depending upon the period of interest (prediction obtained from the model of Campbell & Bozorgnia (2008). 16

2.2 Example site at the center of an area source, with potential earthquakes at zero distance from the site. 16

2.3 Observed spectral acceleration values from the 1999 Chi-Chi, Taiwan earthquake, illustrating variability in ground motion intensity. The predicted distribution comes from the model of Campbell & Bozorgnia (2008) 17

2.4 Schematic illustration of the basic five steps in probabilistic seismic hazard analysis. (a) Identify earthquake sources. (b) Characterize the distribution of earthquake magnitudes from each source. (c) Characterize the distribution of source-to-site distances from each source. (d) Predict the resulting distribution of ground motion intensity. (e) Combine information from parts a-d to compute the annual rate of exceeding a given ground motion intensity. 20

2.5 Typical distribution of observed earthquake magnitudes, along with Gutenberg-Richter and bounded Gutenberg-Richter recurrence laws fit to the observations. 22

2.6 Illustration of discretization of a continuous magnitude distribution for a source with a truncated Gutenberg-Richter distribution, a minimum considered magnitude of 5, a maximum magnitude of 8, and a b parameter of 1. (a) Continuous probability density function from equation 2.5. (b) Discrete probabilities from equation 2.6. 23

2.7 Illustration of an example area source. 25
2.8 PDF and CDF of the source-to-site distance for future earthquakes from the example area source. 26
2.9 Illustration of an example line source. 26
2.10 PDF and CDF of the source-to-site distance for future earthquakes from the example line source. 27
2.11 Graphical depiction of the example ground motion prediction model for a magnitude 6.5 earthquake, and the probability of *PGA* >1g at several source-to-site distances. . 30
2.12 Map view of the example site, with one earthquake source. 33
2.13 *PGA* hazard curve for the single-source site. 35
2.14 Map view of an example site with two earthquake sources. 35
2.15 *PGA* hazard curve for the example two-source site. . . . 37
2.16 Map view of an example site with one source producing earthquakes having a variety of magnitudes. 38
2.17 *PGA* hazard curve for the example site with one source and a Gutenberg-Richter magnitude distribution. 40

3.1 Deaggregation results associated with the example calculation of Section 2.3. (a) Magnitude distribution, given *PGA* >0.2g. (b) Magnitude distribution, given *PGA* >1g. . . . 46
3.2 Deaggregation results associated with the example calculation of Section 2.3, with magnitudes discretized into 0.5-unit intervals. (a) Magnitude distribution, given *PGA* >0.2g. (b) Magnitude distribution, given *PGA* >1g. 47
3.3 Example disaggregation for *SA*(1.0s) at a site in Palo Alto, California (USGS, 2008) 48
3.4 Hazard curves computed for the example site from Section 2.3, using several choices for the minimum considered magnitude. 50
3.5 Probability of occurrence of an event in time t, as a function of the expected number of occurrences, λt. 51
3.6 Combining hazard curves from individual periods to generate a uniform hazard spectrum with a $4^{*}10^{-4}$ rate of exceedance for a site in Los Angeles. (a) Hazard curve for *SA*(0.3s), with UHS point identified. (b) Hazard curve for *SA*(1s), with UHS point identified. (c) Uniform hazard spectrum, based on a series of calculations like those in (a) and (b). 53
3.7 Schematic illustration of the joint distribution of *PGA* and *SA*. 57

A.1 Venn diagram illustrating a sample space and events. . . 61

A.2 Schematic illustration of the events E_1 and E_2. The shaded region depicts the area corresponding to event E_1E_2. . . 62
A.3 Schematic illustration of the total probability theorem. . 65
A.4 Example descriptions of a discrete random variable. (a) Probability mass function. (b) Cumulative distribution function. 67
A.5 Plot of a continuous probability density function. The area of the shaded rectangle, $f_X(x)\,dx$, represents the probability of the random variable X taking values between x and $x + dx$. 68
A.6 Joint normal probability density function. 72

List of Tables

2.1 Magnitude probabilities for a source with a truncated Gutenberg-Richter distribution, a minimum considered magnitude of 5, a maximum magnitude of 8, and a b parameter of 1. The numbers in this table were computed using equations 2.4 and 2.6. 23

2.2 *PGA* probabilities associated with a magnitude 5 earthquake at 10 km. 31

2.3 Probabilities used to compute $\lambda(PGA{>}0.2g)$. 39

2.4 Probabilities used to compute $\lambda(PGA > 1g)$. 39

3.1 Deaggregation calculations for the Section 2.3 example. 45

3.2 Deaggregation calculations for the Section 2.3 example, with magnitudes discretized into 0.5-unit intervals. 47

3.3 *PGA* probabilities associated with a magnitude 5 earthquake at 10 km, and an $SA(0.5s)$ value of 0.2g. 56

A.1 Standard normal cumulative distribution function. . . . 71

Acknowledgements

Thanks to Marcello Bianchini, Tom Hanks, Ting Ling, Nirmal Jayaram and Seok Goo Song for pointing out errors and making suggestions to improve this document.

This document is based in part on a report written for the US Nuclear Regulatory Commission. Financial support for the writing of that material is greatly appreciated.

1
Introduction

The goal of many earthquake engineering analyses is to ensure that a structure can withstand a given level of ground shaking while maintaining a desired level of performance. But what level of ground shaking should be used to perform this analysis? There is a great deal of uncertainty about the location, size, and resulting shaking intensity of future earthquakes. Probabilistic Seismic Hazard Analysis (PSHA) aims to quantify these uncertainties, and combine them to produce an explicit description of the distribution of future shaking that may occur at a site.

In order to assess risk to a structure from earthquake shaking, we must first determine the annual probability (or rate) of exceeding some level of earthquake ground shaking at a site, for a range of intensity levels. Information of this type could be summarized as shown in Figure 1.1, which shows that low levels of intensity are exceeded relatively often, while high intensities are rare. If one was willing to observe earthquake shaking at a site for thousands of years, it would be possible to obtain this entire curve experimentally. That is the approach often used for assessing flood risk, but for seismic risk this is not possible because we do not have enough observations to extrapolate to the low rates of interest. In addition, we have to consider uncertainties in the size, location, and resulting shaking intensity caused by an earthquake, unlike the case of floods where we typically only worry about the size of the flood event. Because of these challenges, our seismic hazard data must be obtained by mathematically combining models for the location and size of potential future earthquakes with predictions of the potential shaking intensity caused by these future earthquakes. The mathematical approach for performing this calculation is known as Probabilistic Seismic Hazard Analysis, or PSHA.

The purpose of this document is to discuss the calculations involved in PSHA, and the motivation for using this approach. Because many models and data sources are combined to create results like

those shown in Figure 1.1, the PSHA approach can seem opaque. But when examined more carefully, the approach is actually rather intuitive. Once understood and properly implemented, PSHA is flexible enough to accommodate a variety of users' needs, and quantitative so that it can incorporate all knowledge about seismic activity and resulting ground shaking at a site.

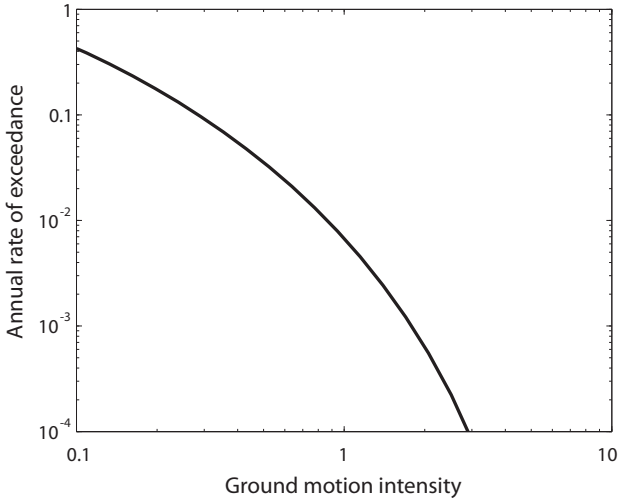


Figure 1.1: Quantification of the possibility of intense ground shaking at a site.

Probability calculations are a critical part of the procedures described here, so a basic knowledge of probability and its associated notation is required to study this topic. A review of the concepts and notation used in this document is provided for interested readers in Appendix A.

2
An overview of PSHA

2.1 Deterministic versus probabilistic approaches

The somewhat complicated probabilistic evaluation could be avoided if it was possible to identify a "worst-case" ground motion and evaluate the facility of interest under that ground motion. This line of thinking motivates an approach known as deterministic hazard analysis, but we will see that conceptual problems arise quickly and are difficult to overcome.

2.1.1 Variability in the design event

A designer looking to choose a worst-case ground motion would first want to look for the maximum magnitude event that could occur on the closest possible fault. This is simple to state in theory, but several difficulties arise in practice. Consider first the hypothetical site shown in Figure 2.1a, which is located 10 km from a fault capable of producing an earthquake with a maximum magnitude of 6.5. It is also located 30 km from a fault capable of producing a magnitude 7.5 earthquake. The median predicted response spectra from those two events are shown in Figure 2.1b. As seen in that figure, the small-magnitude nearby event produces larger spectral acceleration amplitudes at short periods, but the larger-magnitude event produces larger amplitudes at long periods. So, while one could take the envelope of the two spectra, there is not a single "worst-case" event that produces the maximum spectral acceleration amplitudes at all periods.

While the site shown in Figure 2.1a produces some challenges in terms of identifying a worst-case event, an even greater challenges arise when faults near a site are not obvious and so the seismic source is quantified as an areal source capable of producing earthquakes at any location, as shown in Figure 2.2. In this case, the worst-case event has to be the one with the maximum conceivable

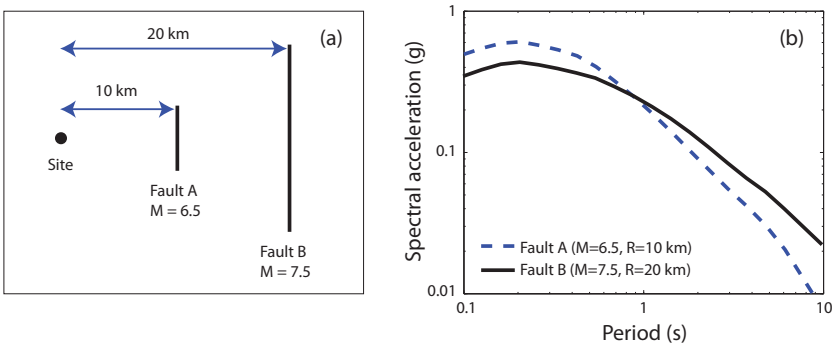


Figure 2.1: (a) Map view of an illustrative site, with two nearby sources capable of producing earthquakes. (b) Predicted median response spectra from the two earthquake events, illustrating that the event producing the maximum response spectra may vary depending upon the period of interest (prediction obtained from the model of Campbell & Bozorgnia (2008).

magnitude, at a location directly below the site of interest (i.e., with a distance of 0 km). This is clearly the maximum event, no matter how unlikely its occurrence may be. For example, in parts of the Eastern United States, especially near the past Charleston or New Madrid earthquakes, one can quite feasibly hypothesize the occurrence of magnitude 7.5 or 8 earthquakes immediately below a site, although that event may occur very rarely.

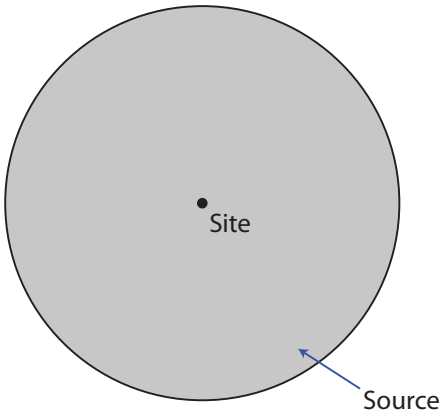


Figure 2.2: Example site at the center of an area source, with potential earthquakes at zero distance from the site.

2.1.2 *Variability of ground motion intensity*

While the choice of a "worst-case" earthquake can be difficult and subjective, as discussed in the previous section, an even greater problem with deterministic hazard analysis is the choice of worst-case ground motion intensity associated with that earthquake. The response spectra plotted in Figure 2.1 are the median [1] spectra predicted by empirical models calibrated to recorded ground motions. But recorded ground motions show a very large amount of scatter around those median predictions. By definition, the median predictions shown in Figure 2.1b are exceeded in 50% of observed ground

[1] There is considerable opportunity for confusion when referring to means and medians of predicted ground motion intensity. Ground motion predictions models, such as the one used to make Figure 2.3, provide the mean and standard deviation of the *natural logarithm* of spectral acceleration (lnSA) or peak ground acceleration (lnPGA). These lnSA values are normally distributed, which means that the non-logarithmic SA values are lognormally distributed. The exponential of the *mean* lnSa value can be shown to equal the median SA value. It is easiest to work with lnSA values in the calculations that follow, so this text will often refer to mean lnSA values rather than median SA values. Plots such as Figure 2.3 will show non-logarithmic SA, because the units are more intuitive, but the axis will always be in logarithmic scale so that the visual effect is identical to if one was viewing a plot of lnSA.

motions having the given magnitudes and distances.

An example of the large scatter around those ground motion prediction models is seen in Figure 2.3, which shows spectral acceleration values at 1 second that were observed in a past earthquake (1999 Chi-Chi, Taiwan), plotted versus the closest distance from the earthquake rupture to the recording site. Note that observations at distances between 1 and 3 km vary between 0.15g and more than 1g—nearly an order of magnitude. Also plotted are the mean predicted ln*SA* values, along with bounds illustrating one standard deviation above and below that mean. The scatter of the log of spectral accelerations around the mean prediction is well-represented by a normal distribution (leading to symmetric scatter in Figure 2.3, which is plotted in logarithmic scale).

The one-standard-deviation bounds should enclose about 2/3 of the observed values if the variations are normally distributed, and that is the case here. To account for this scatter, deterministic hazard analyses sometimes specify a "mean plus one standard deviation" response spectra, but even that will be exceeded 16% of the time[2]. The normal distribution for scatter implies that there is no theoretical upper bound on the amplitude of ground motion that might be produced at a given magnitude and distance[3].

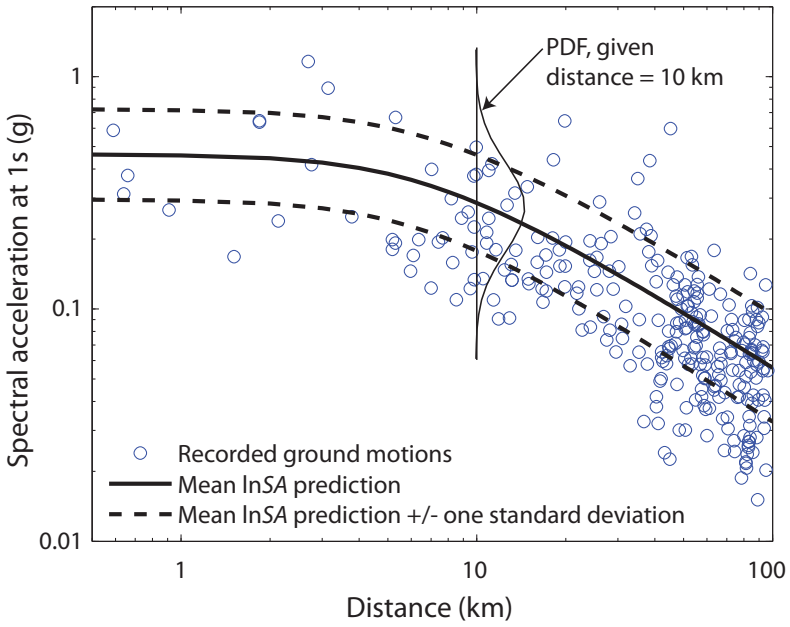


Figure 2.3: Observed spectral acceleration values from the 1999 Chi-Chi, Taiwan earthquake, illustrating variability in ground motion intensity. The predicted distribution comes from the model of Campbell & Bozorgnia (2008)

[2] This number comes from considering normally-distributed residuals. The probability of a normal random variable being more than one standard deviation greater than its mean (i.e., 1 - $\Phi(2.1)$) is 0.16.

[3] There is some true physical upper bound on ground motion intensity caused by an inability of the earth to carry more intense seismic waves without shattering or otherwise failing. Current research suggests that this limit may be important to structures designed for extremely intense ground motions, such as nuclear waste repositories, but has no practical impact on more common structures such as buildings or bridges, which are analyzed for ground motion intensities that are exceeded once every few thousand years. Thus, the assumption of no theoretical upper bound is reasonable and appropriate in most cases.

2.1.3 *Can we use a deterministic approach, given these uncertainties?*

Given these challenges, it is clear that whatever deterministic design earthquake and ground motion intensity is eventually selected, it is not a true "worst-case" event, as a larger earthquake or ground motion can always plausibly be proposed. Without a true worst-case event to consider, we are left to identify a "reasonably large" event. That is often done by choosing a nearby large-magnitude event, and then identifying some level of reasonable intensity associated with this event. While it is possible to proceed using this type of approach, two issues should be made clear. 1) The resulting ground motion is not a "worst-case" ground motion. 2) The result may be very sensitive to decisions made with regard to the chosen scenario magnitude and ground motion intensity. An event chosen in this manner was historically described as a "Maximum Credible Earthquake," or MCE. More recently, however, the acronym has been retained but taken to mean "Maximum Considered Earthquake," in recognition of the fact that larger earthquakes (and larger ground motion intensities) are likely to be credible as well. This "worst-case" thinking will be abandoned for the remainder of the document, although the problems identified here will serve as a useful motivation for thinking about probability-based alternatives.

2.2 *Probabilistic seismic hazard analysis calculations*

In this section, we will describe a probability-based framework capable of addressing the concerns identified above. Rather than ignoring the uncertainties present in the problem, this approach incorporates them into calculations of potential ground motion intensity. While incorporation of uncertainties adds some complexity to the procedure, the resulting calculations are much more defensible for use in engineering decision-making for reducing risks.

With PSHA, we are no longer searching for an elusive worst-case ground motion intensity. Rather, we will consider all possible earthquake events and resulting ground motions, along with their associated probabilities of occurrence, in order to find the level of ground motion intensity exceeded with some tolerably low rate. At its most basic level, PSHA is composed of five steps.

1. Identify all earthquake sources capable of producing damaging ground motions.

2. Characterize the distribution of earthquake magnitudes (the rates at which earthquakes of various magnitudes are expected to occur).

3. Characterize the distribution of source-to-site distances associated with potential earthquakes.
4. Predict the resulting distribution of ground motion intensity as a function of earthquake magnitude, distance, etc.
5. Combine uncertainties in earthquake size, location and ground motion intensity, using a calculation known as the total probability theorem.

These steps will be explained in more detail below.

The end result of these calculations will be a full distribution of levels of ground shaking intensity, and their associated rates of exceedance. The illusion of a worst-case ground motion will be removed, and replaced by identification of occurrence frequencies for the full range of ground motion intensities of potential interest. These results can then be used to identify a ground motion intensity having an acceptably small probability of being exceeded.

2.2.1 *Identify earthquake sources*

In contrast to the deterministic thinking above, which focused only on the largest possible earthquake event, here we are interested in all earthquake sources capable of producing damaging ground motions at a site. These sources could be faults, which are typically planar surfaces identified through various means such as observations of past earthquake locations and geological evidence. If individual faults are not identifiable (as in the less seismically active regions of the eastern United States), then earthquake sources may be described by areal regions in which earthquakes may occur anywhere. Once all possible sources are identified, we can identify the distribution of magnitudes and source-to-site distances associated with earthquakes from each source.

2.2.2 *Identify earthquake magnitudes*

Tectonic faults are capable of producing earthquakes of various sizes (i.e., magnitudes). Gutenberg & Richter (1944) first studied observations of earthquake magnitudes, and noted that the number of earthquakes in a region greater than a given size generally follows a particular distribution

$$\log_{10} \lambda_m = a - bm \tag{2.1}$$

where λ_m is the rate of earthquakes with magnitudes greater than m, and a and b are constants. This equation is called the *Gutenberg-Richter recurrence law*. Figure 2.5 illustrates typical observations from

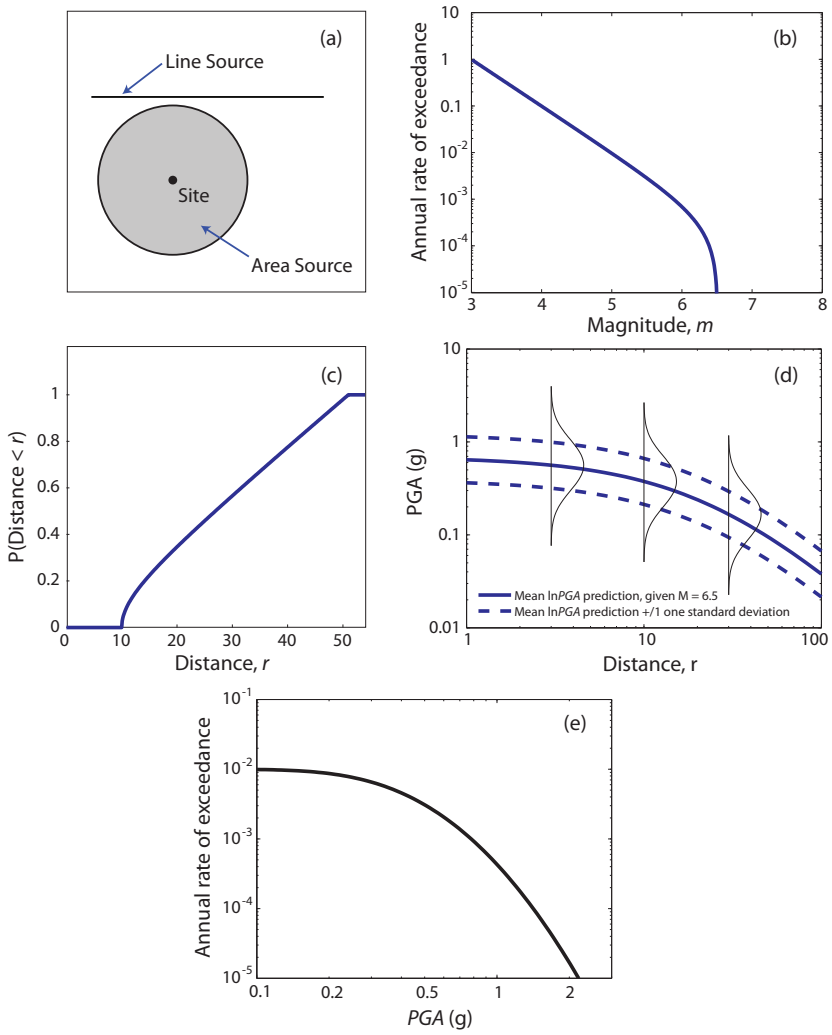


Figure 2.4: Schematic illustration of the basic five steps in probabilistic seismic hazard analysis. (a) Identify earthquake sources. (b) Characterize the distribution of earthquake magnitudes from each source. (c) Characterize the distribution of source-to-site distances from each source. (d) Predict the resulting distribution of ground motion intensity. (e) Combine information from parts a-d to compute the annual rate of exceeding a given ground motion intensity.

a fault or region, along with the Gutenberg-Richter recurrence law given by equation 2.1.

The a and b constants from equation 2.1 are estimated using statistical analysis of historical observations, with additional constraining data provided by other types of geological evidence[4]. The a value indicates the overall rate of earthquakes in a region, and the b value indicates the relative ratio of small and large magnitudes (typical b values are approximately equal to 1).

Equation 2.1 can also be used to compute a cumulative distribution function[5] (CDF) for the magnitudes of earthquakes that are larger than some minimum magnitude m_{min} (this conditioning is used because earthquakes smaller than m_{min} will be ignored in later calculations due to their lack of engineering importance).

$$\begin{aligned} F_M(m) &= P(M \leq m | M > m_{\min}) \\ &= \frac{\text{Rate of earthquakes with } m_{\min} < M \leq m}{\text{Rate of earthquakes with } m_{\min} < M} \\ &= \frac{\lambda_{m_{\min}} - \lambda_m}{\lambda_{m_{\min}}} \\ &= \frac{10^{a-bm_{\min}} - 10^{a-bm}}{10^{a-bm_{\min}}} \\ &= 1 - 10^{-b(m-m_{\min})}, \qquad m > m_{\min} \end{aligned} \tag{2.2}$$

where $F_M(m)$ denotes the cumulative distribution function for M. One can compute the probability density function (PDF) for M by taking the derivative of the CDF

$$\begin{aligned} f_M(m) &= \frac{d}{dm} F_M(m) \\ &= \frac{d}{dm}\left[1 - 10^{-b(m-m_{\min})}\right] \\ &= b\ln(10) 10^{-b(m-m_{\min})}, \qquad m > m_{\min} \end{aligned} \tag{2.3}$$

where $f_M(m)$ denotes the probability density function for M.

Note that the PDF given in equation 2.3 relies on the Gutenberg-Richter law of equation 2.1, which theoretically predicts magnitudes with no upper limit, although physical constraints make this unrealistic. There is generally some limit on the upper bound of earthquake magnitudes in a region, due to the finite size of the source faults (earthquake magnitude is related to the area of the seismic rupture). If a maximum magnitude can be determined, then equation 2.2 becomes

$$F_M(m) = \frac{1 - 10^{-b(m-m_{\min})}}{1 - 10^{-b(m_{\max}-m_{\min})}}, \qquad m_{\min} < m < m_{\max} \tag{2.4}$$

[4] Note that some care is needed during this process to ensure that no problems are caused by using historical data that underestimates the rate of small earthquakes due to the use of less sensitive instruments in the past. Methods have been developed to address this issue (e.g. Weichert, 1980), but are not considered further in this document.

[5] Probability tools such as cumulative distribution functions and probability density functions are necessary for much of the analysis that follows. See Appendix A for a review of this material.

and equation 2.3 becomes

$$f_M(m) = \frac{b \ln(2.10) 10^{-b(m-m_{\min})}}{1 - 10^{-b(m_{\max}-m_{\min})}}, \qquad m_{\min} < m < m_{\max} \tag{2.5}$$

where m_{max} is the maximum earthquake that a given source can produce. This limited magnitude distribution is termed a *bounded Gutenberg-Richter recurrence law*. Example observations of earthquake magnitudes are shown in Figure 2.5, along with Gutenberg-Richter and bounded Gutenberg-Richter recurrence laws fit to the data.

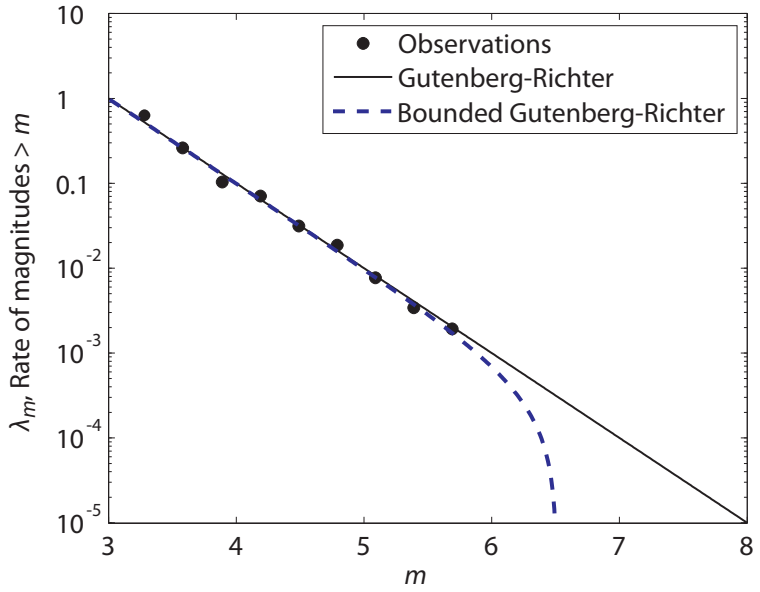


Figure 2.5: Typical distribution of observed earthquake magnitudes, along with Gutenberg-Richter and bounded Gutenberg-Richter recurrence laws fit to the observations.

For our later PSHA equations, we will convert the continuous distribution of magnitudes into a discrete set of magnitudes. For example, consider a source with a minimum considered magnitude of 5, a maximum magnitude of 8, and a b parameter equal to 1. Table 2.1 lists probabilities of interest for this example source. The first column lists 13 magnitude values between 5 and 8. The second column lists the cumulative distribution function, as computed using equation 2.4. The third column lists probabilities of occurrence of these discrete set of magnitudes, assuming that they are the only possible magnitudes; they are computed as follows

$$P(M = m_j) = F_M(m_{j+1}) - F_M(m_j) \tag{2.6}$$

where m_j are the discrete set of magnitudes, ordered so that $m_j < m_{j+1}$. This calculation assigns the probabilities associated with all magnitudes between m_j and m_{j+1} to the discrete value m_j. As long as the discrete magnitudes are closely spaced, the approximation will not affect numerical results. Magnitudes are spaced at intervals of

0.25 for illustration in Table 2.1 so that the table is not too lengthy, but a practical PSHA analysis might use a magnitude spacing of 0.1 or less.

m_j	$F_M(m_j)$	$P(M = m_j)$
5.00	0.0000	0.4381
5.25	0.4381	0.2464
5.50	0.6845	0.1385
5.75	0.8230	0.0779
6.00	0.9009	0.0438
6.25	0.9447	0.0246
6.50	0.9693	0.0139
6.75	0.9832	0.0078
7.00	0.9910	0.0044
7.25	0.9954	0.0024
7.50	0.9978	0.0014
7.75	0.9992	0.0008
8.00	1.0000	0.0000

Table 2.1: Magnitude probabilities for a source with a truncated Gutenberg-Richter distribution, a minimum considered magnitude of 5, a maximum magnitude of 8, and a b parameter of 1. The numbers in this table were computed using equations 2.4 and 2.6.

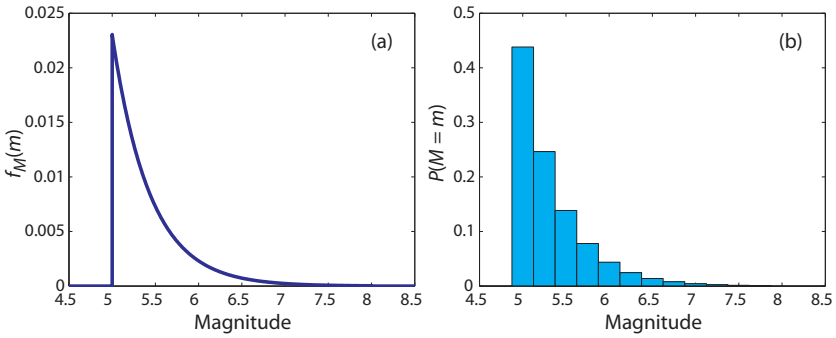


Figure 2.6: Illustration of discretization of a continuous magnitude distribution for a source with a truncated Gutenberg-Richter distribution, a minimum considered magnitude of 5, a maximum magnitude of 8, and a b parameter of 1. (a) Continuous probability density function from equation 2.5. (b) Discrete probabilities from equation 2.6.

An aside: The Gutenberg-Richter models above are not the only models proposed for describing the distribution of earthquake magnitudes. One common alternative is the Characteristic Earthquake model, which proposes that some faults have repeated occurrences of a characteristic earthquake with a reasonably consistent magnitude (Schwartz & Coppersmith, 1984). This characteristic magnitude occurs more often than predicted by the Gutenberg-Richter models proposed above. All that is required to adopt an alternative recurrence model is to replace equation 2.5 with a suitably modified probability density function (e.g. Lomnitz-Adler & Lomnitz, 1979; Youngs & Coppersmith, 1985). All other PSHA equations remain identical.

2.2.3 *Identify earthquake distances*

To predict ground shaking at a site, it is also necessary to model the distribution of distances from earthquakes to the site of interest. For a given earthquake source, it is generally assumed that earthquakes will occur with equal probability at any location on the fault[6]. Given that locations are uniformly distributed, it is generally simple to identify the distribution of source-to-site distances using only the geometry of the source. Example calculations are shown in this section for an area source and a line source.

[6] In a few special cases, analysts use non-uniform distributions for future earthquake locations based on models for stress transfer and time-dependent earthquake occurence. Those situations will not be considered here.

An aside: While "distance" sounds like a well-defined term, there are in fact several definitions commonly used in PSHA. One can use distance to the epicenter or hypocenter, distance to the closest point on the rupture surface, or distance to the closest point on the surface projection of the rupture. Note that some distance definitions account for the depth of the rupture, while others consider only distance from the surface projection of the rupture. Note also that epicenter- and hypocenter-based definitions need only consider the location of rupture initiation; some other definitions need to explicitly account for the fact that ruptures occur over a plane rather than at a single point in space. The analyst's choice of distance definition will depend upon the required input to the ground motion prediction model. Here we will consider only distance to the epicenter, for simplicity.

Example: Area source

Consider a site located in an area source. The source produces earthquakes randomly and with equal likelihood anywhere within 100 km of the site. (In actuality, the source may be larger, but is typically truncated at some distance beyond which earthquakes are not expected to cause damage at the site.) Area sources are often used in practice to account for "background" seismicity, or for earthquakes that are not associated with any specific fault. The example source is illustrated in Figure 2.7.

We can easily compute a probabilistic description for the distances to earthquakes in this case by noting that if the earthquakes are equally likely to occur anywhere, then the probability of an epicenter being located at a distance of less than r is equal to the area of

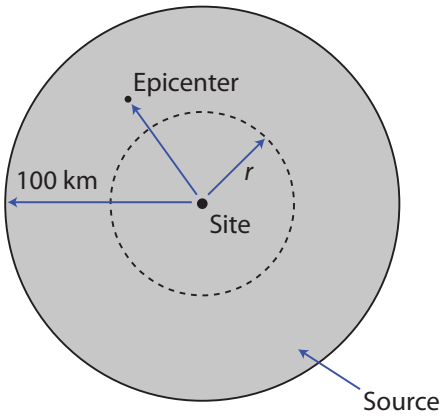


Figure 2.7: Illustration of an example area source.

a circle of radius r, divided by the area of the circle of radius 100

$$\begin{aligned} F_R(r) &= P(R \leq r) \\ &= \frac{\text{area of circle with radius } r}{\text{area of circle with radius } 100} \\ &= \frac{\pi r^2}{\pi(100)^2} \\ &= \frac{r^2}{10,000} \end{aligned} \tag{2.7}$$

Equation 2.7 is only valid, however, for distance (r) values between 0 and 100 km. Accounting for other ranges gives the more complete description

$$F_R(r) = \begin{cases} 0 & \text{if } r < 0 \\ \frac{r^2}{10,000} & \text{if } 0 \leq r < 100 \\ 1 & \text{if } r \geq 100 \end{cases} \tag{2.8}$$

We can also find the PDF for the distance by taking a derivative of the CDF

$$f_R(r) = \frac{d}{dr} F_R(r) = \begin{cases} \frac{r}{5000} & \text{if } 0 \leq r < 100 \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

Plots of this PDF and CDF are shown in Figure 2.8. We see that distances close to 0 km are possible but unlikely, because there are few locations in the source that are associated with such small distances. Unlike the deterministic "worst-case" distance of 0 km, the PSHA calculations below will use these distributions to account for the differing probabilities of observing earthquakes at various distances.

Example: Line source

Earthquake sources are also sometimes quantified as line sources[7].

[7] It is also common to treat the earth's structure in three dimensions, meaning that faults will be represented as planes rather than lines. The examples in this document will all be two-dimensional for simplicity, but the extension to three dimensions is straightforward.

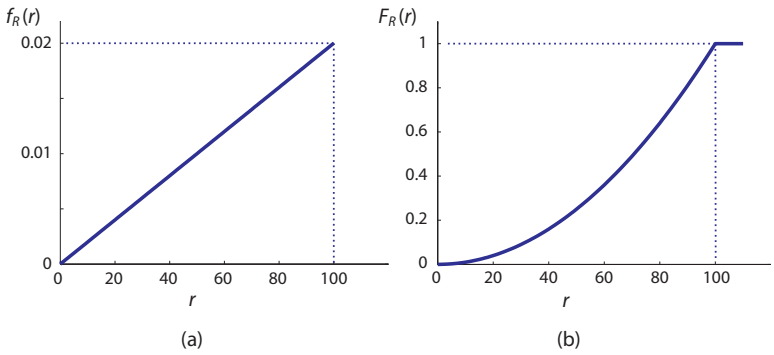


Figure 2.8: PDF and CDF of the source-to-site distance for future earthquakes from the example area source.

This is particularly appropriate for modeling identified faults that exist on the boundary of two tectonic plates (as is the case in much of coastal California).

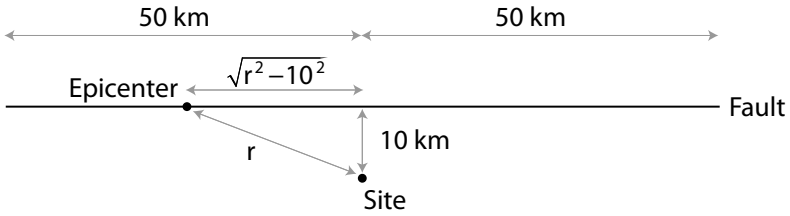


Figure 2.9: Illustration of an example line source.

Consider a 100 km fault, modeled as a line source in Figure 2.9, with a site located 10 km from the center of the fault. We again assume that earthquake epicenters are equally likely at all locations. In this case, the probability of observing a distance of less than r is equal to the fraction of the fault located within a radius of r. Using the Pythagorean theorem, we can see that the distance from the center of the fault to a point a distance r from the site is $\sqrt{r^2 - 10^2}$.

Using this information, we can then compute the CDF of R

$$\begin{aligned} F_R(r) &= P(R \leq r) \\ &= \frac{\text{length of fault within distance } r}{\text{total length of fault}} \\ &= \frac{2\sqrt{r^2 - 10^2}}{100} \end{aligned} \tag{2.10}$$

but that equation is only true for distances of less than 10 km or greater than 51 km. Distances outside that range are not possible on

this fault, so the complete CDF is

$$F_R(r) = \begin{cases} 0 & \text{if } r < 10 \\ \frac{2\sqrt{r^2-10^2}}{100} & \text{if } 10 \le r < 51 \\ 1 & \text{if } r \ge 51 \end{cases} \tag{2.11}$$

The PDF can be obtained from the derivative of the CDF

$$f_R(r) = \frac{d}{dr}F_R(r) = \begin{cases} \frac{r}{50\sqrt{r^2-100}} & \text{if } 10 \le r < 51 \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

The PDF and CDF are plotted in Figure 2.10.

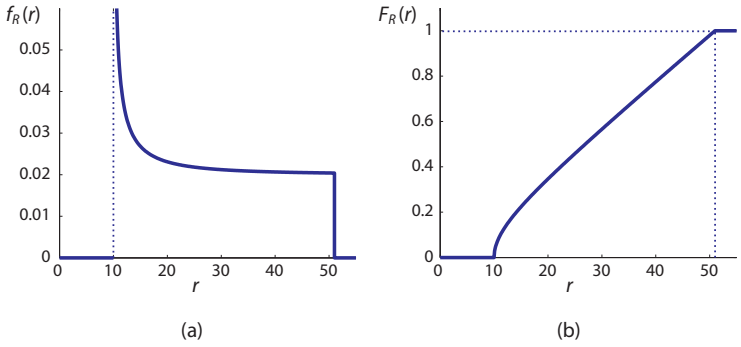


Figure 2.10: PDF and CDF of the source-to-site distance for future earthquakes from the example line source.

These two example sources are shown to provide simple examples of distance distributions. Distributions for more irregular sources can also be computed in the same manner. These distributions are important inputs for the calculations that follow.

2.2.4 *Ground motion intensity*

We have now quantified the distribution of potential earthquake magnitudes and locations, but we are interested in analyzing ground motions—not earthquakes. The next step is therefore a ground motion prediction model[8]. These models predict the probability distribution of ground motion intensity, as a function of many predictor variables such as the earthquake's magnitude, distance, faulting mechanism, the near-surface site conditions, the potential presence of directivity effects, etc. Because the number of predictor variables is large, we often write that the model predicts ground motion intensity given "magnitude, distance, etc."

Ground motion prediction models are generally developed using statistical regression on observations from large libraries of observed ground motion intensities. For example, spectral acceleration (SA) values at 1 second observed in the 1999 Chi-Chi, Taiwan, earthquake were shown previously in Figure 2.3, along with lines showing the

[8] These models were called "attenuation models" or "attenuation relations" until recently. Those names have fallen out of favor, however, because the prediction model accounts for a great number of effects, of which attenuation is only one.

predicted mean (and mean +/- one standard deviation) of the lnSA values from an example ground motion prediction model (Campbell & Bozorgnia, 2008). That prediction model, like other modern models, was fit to thousands of observed intensities from dozens of past earthquakes.

It is apparent from Figure 2.3 that there is significant scatter in observed ground motion intensities, even after accounting for the effect of magnitude, distance, etc. Thus, these predictive models must provide a probability distribution for intensities, rather than just a single intensity. This is important, because our later PSHA calculations need to account for the possibility of unlikely outcomes such as extreme intensities much larger than the predicted mean (Bommer & Abrahamson, 2006).

To describe this probability distribution, prediction models take the following general form:

$$\ln IM = \overline{\ln IM}(M, R, \theta) + \sigma(M, R, \theta) \cdot \varepsilon \tag{2.13}$$

where lnIM is the natural log of the ground motion *intensity measure* of interest (such as spectral acceleration at a given period); this lnIM is modeled as a random variable, and has been seen to be well-represented by a normal distribution. The terms $\overline{\ln IM}(M, R, \theta)$ and $\sigma(M, R, \theta)$ are the outputs of the ground motion prediction model; they are the predicted mean and standard deviation, respectively, of lnIM. These terms are both functions of the earthquake's magnitude (M), distance (R) and other parameters (generically referred to as θ). Finally, ε is a standard normal random variable that represents the observed variability in lnIM. Positive values of ε produce larger-than-average values of lnIM, while negative values of ε produce smaller-than-average values of lnIM.

Over decades of development and refinement, the prediction models for $\overline{\ln IM}(M, R, \theta)$ and $\sigma(M, R, \theta)$ have become complex, consisting of many terms and tables containing dozens of coefficients. These modern models are no longer simple to calculate using pencil and paper, so here we will use an older and much simpler (but obsolete) model to illustrate the example calculations. The approach is identical when using modern prediction models, but this simple model keeps us from being distracted by tedious arithmetic.

Cornell et al. (1979) proposed the following predictive model for the mean of log peak ground acceleration (in units of g)

$$\overline{\ln PGA} = -0.152 + 0.859M - 1.803 \ln(R + 25) \tag{2.14}$$

The standard deviation of lnPGA was 0.57 in this model, and was constant for all magnitudes and distances. The natural logarithm of PGA was seen to be normally distributed, so we can compute the

probability of exceeding any *PGA* level using knowledge of this mean and standard deviation

$$P(PGA > x|m,r) = 1 - \Phi\left(\frac{\ln x - \overline{\ln PGA}}{\sigma_{\ln PGA}}\right) \tag{2.15}$$

where $\Phi()$ is the standard normal cumulative distribution function, as shown in Table A.1 on page 71. Modern prediction models also provide a mean and standard deviation to be used in equation 2.15, so the general procedure is identical when using newer models; the equations for predicting the mean and standard deviation are just more complicated.

Equation 2.15 used the cumulative distribution function to compute $P(PGA > x|m,r)$, but sometimes it may be useful to use an alternate formulation incorporating the probability density function for *PGA*. Noting that the cumulative distribution function is equivalent to an integral of the probability density function (equation A.27), we can also write

$$P(PGA > x|m,r) = \int_x^\infty f_{PGA}(u)\,du \tag{2.16}$$

where $f_{PGA}(u)$ is the probability density function of *PGA*, given m and r. Unlike the cumulative distribution function $\Phi()$, $f_{PGA}(u)$ can actually be written out analytically. Substituting in this PDF gives

$$P(PGA > x|m,r) = \int_x^\infty \frac{1}{\sigma_{\ln PGA}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln u - \overline{\ln PGA}}{\sigma_{\ln PGA}}\right)^2\right) du \tag{2.17}$$

This integral can then be evaluated numerically within the PSHA software.

To connect these equations to a visual display of ground motion predictions, consider Figure 2.11, which shows *PGA* predictions for a magnitude 6.5 earthquake, as a function of distance. The mean and the mean +/- one standard deviation of the Cornell et al. prediction is plotted for distances between 1 and 100 km. At distances of 3, 10 and 30 km, the entire PDF of the predicted normal distribution is also superimposed. Suppose we were interested in the probability of *PGA* >1g. At those three distances, equation 2.14 gives predicted means of -0.5765, -0.9788 and -1.7937, respectively[9]. At all three distances, the standard deviation of ln*PGA* is 0.57. So we can use equation 2.15 to

[9] All of the example calculations will provide answers with more significant figures than should reasonably be used or reported. This is done to allow readers to reproduce the example calculations exactly, and because many answers are intermediate results for later calculations.

compute the probabilities of exceedance as

$$P(PGA > 1|6.5,3) = 1 - \Phi\left(\frac{\ln 1 - (-0.5765)}{0.57}\right) = 0.16 \qquad (2.18)$$

$$P(PGA > 1|6.5,10) = 1 - \Phi\left(\frac{\ln 1 - (-0.9788)}{0.57}\right) = 0.043 \qquad (2.19)$$

$$P(PGA > 1|6.5,30) = 1 - \Phi\left(\frac{\ln 1 - (-1.7937)}{0.57}\right) = 0.0008 \qquad (2.20)$$

These probabilities correspond to the fraction of the corresponding PDFs in Figure 2.11 that are shaded. This visual interpretation may provide some intuition regarding the previous equations, which may unfamiliar to readers who do not regularly perform probability calculations.

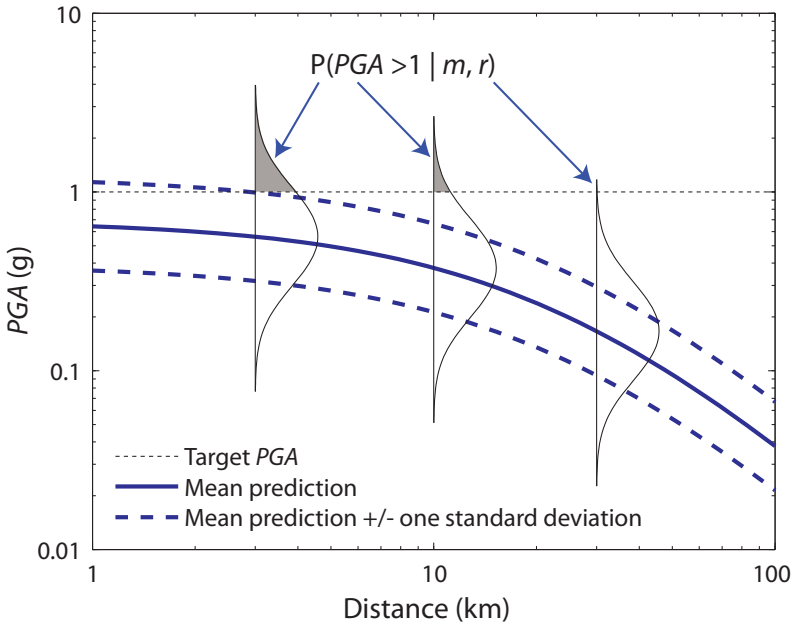


Figure 2.11: Graphical depiction of the example ground motion prediction model for a magnitude 6.5 earthquake, and the probability of *PGA* >1g at several source-to-site distances.

Let us consider a second example, which will provide intermediate results for some calculations that follow. Assume magnitude 5 earthquake has occurred at a distance of 10 km. The Cornell at al. ground motion prediction model provides a mean ln*PGA* of -2.2673, and a standard deviation of 0.57. The first column of Table 2.2 lists a series of *PGA* values of potential interest. The second column lists the probabilities of exceeding those *PGA* values, using equation 2.15. The third column lists the probability of equaling those PGA values, using the same discretization approach we used previously for the continuous magnitude distribution

$$P(PGA = x_j) = P(PGA > x_j) - P(PGA > x_{j+1}) \qquad (2.21)$$

x_j	$P(PGA > x_j)$	$P(PGA = x_j)$
0.20	0.12418	0.06312
0.25	0.06106	0.03004
0.30	0.03102	0.01471
0.35	0.01631	0.00745
0.40	0.00886	0.00390
0.45	0.00496	0.00211
0.50	0.00285	0.00117
0.55	0.00168	0.00067
0.60	0.00101	0.00039
0.65	0.00062	0.00024
0.70	0.00038	0.00014
0.75	0.00024	0.00009
0.80	0.00015	0.00006
0.85	0.00009	0.00004
0.90	0.00005	0.00002
0.95	0.00003	0.00002
1.00	0.00001	0.00001

Table 2.2: *PGA* probabilities associated with a magnitude 5 earthquake at 10 km.

An aside: At first glance, one might wonder whether the large variability must necessarily be a part of prediction models, or whether it is a fundamental error caused by over-simplifications or inappropriate mixing of observational data. The uncertainty arises because we are trying to predict a highly complex phenomenon (ground shaking intensity at a site) using very simplified predictive parameters (magnitude, distance, and perhaps a few other parameters). Earthquake rupture is a complex spatial-temporal process, and here we represent it simply by "magnitude," which measures the total seismic energy released in the complex rupture. Similarly, nonlinear wave scattering and propagation through a complex structure such as the earth is represented by simply the distance between the source and the site. It is true that if we had more detailed knowledge of the rupture and propagation process, we might predict ground shaking intensity with less uncertainty. But we don't always have detailed models for rupture and wave propagation from past earthquakes to use in calibrating predictive models, and even if we were able to develop complex predictive models, then our predictions of future earthquake events would have to be much more detailed than simply predicting their distributions of magnitudes and distances. Ground motion prediction equations of the type used to produce Figure 2.3 have evolved over a period of 40 years, and are now developed using thousands of observed ground motions and are constrained using many theo-

retical and physical insights. This level of refinement suggests that there is little hope of this scatter being reduced significantly in the near future. The scatter is best thought of as an inherent variability in the earthquake hazard environment that must be accounted for when identifying a design ground motion intensity.

2.2.5 *Combine all information*

With the above information in place, we can now combine it using the PSHA equations. We will first consider two intermediate calculations as we build towards a PSHA equation that considers multiple sources.

First, let us compute the probability of exceeding an *IM* intensity level x, *given* occurrence of a future earthquake from a single source. The ground motion prediction model of Section 2.2.4 allows us to compute the probability of exceeding that *IM* level for a given magnitude and distance. The magnitude and distance of the future earthquake are not yet known, but we can find their probability distributions using Sections 2.2.2 and 2.2.3. We then combine this information using the total probability theorem

$$P(IM > x) = \int_{m_{\min}}^{m_{\max}} \int_{0}^{r_{\max}} P(IM > x|m,r) f_M(m) f_R(r)\, dr\, dm \qquad (2.22)$$

where $P(IM > x|m,r)$ comes from the ground motion model, $f_M(m)$ and $f_R(r)$ are our PDFs for magnitude and distance, and we integrate over all considered magnitudes and distances[10]. The integration operation adds up the conditional probabilities of exceedance associated with all possible magnitudes and distances (with the PDFs weighting each conditional exceedance probability by the probability of occurrence of the associated magnitude and distance).

[10] More generally, we should use a joint distribution for magnitude and distance, $f_{M,R}(m,r)$, rather than the product of their marginal distances $f_M(m)f_R(r)$. The above formulation is correct only if the magnitudes and distances of events are independent. The above formulation is helpful, however, for easily incorporating the magnitude and distance distributions computed earlier.

Equation 2.22 is a probability of exceedance given and earthquake, and does not include any information about how often earthquakes occur on the source of interest. We can make a simple modification to that equation, to compute the rate of *IM* >x, rather than the probability of *IM* >x given occurrence of an earthquake.

$$\lambda(IM > x) = \lambda(M > m_{\min}) \int_{m_{\min}}^{m_{\max}} \int_{0}^{r_{\max}} P(IM > x|m,r) f_M(m) f_R(r)\, dr\, dm \qquad (2.23)$$

where $\lambda(M > m_{\min})$ is the rate of occurrence of earthquakes greater than m_{min} from the source, and $\lambda(IM > x)$ is the rate of *IM* >x.

To generalize the analysis further, we would like to consider cases with more than one source. Recognizing that the rate of *IM* >x when

considering all sources is simply the sum of the rates of $IM > x$ from each individual source, we can write

$$\lambda(IM > x) = \sum_{i=1}^{n_{sources}} \lambda(M_i > m_{\min}) \int_{m_{\min}}^{m_{\max}} \int_{0}^{r_{\max}} P(IM > x|m,r) f_{M_i}(m) f_{R_i}(r)\, dr\, dm \tag{2.24}$$

where $n_{sources}$ is the number of sources considered, and M_i & R_i denote the magnitude & distance distributions for source i. Since we will nearly always be performing this calculation on a computer, it is practical to discretize our continuous distributions for M and R, and convert the integrals into discrete summations, as follows

$$\lambda(IM > x) = \sum_{i=1}^{n_{sources}} \lambda(M_i > m_{\min}) \sum_{j=1}^{n_M} \sum_{k=1}^{n_R} P(IM > x|m_j, r_k) P(M_i = m_j) P(R_i = r_k) \tag{2.25}$$

where the range of possible M_i and R_i have been discretized into n_M and n_R intervals, respectively, using the discretization technique discussed earlier.

Equation 2.24 (or, equivalently, 2.25) is the equation most commonly associated with PSHA. It has integrated our knowledge about rates of occurrence of earthquakes, the possible magnitudes and distances of those earthquakes, and the distribution of ground shaking intensity due to those earthquakes. Each of those inputs can be determined through scientific studies of past earthquakes and processing of observed data. The end result—the rate of exceeding *IM* levels of varying intensity—is very useful for engineering decision making, and can be determined even for rare (low exceedance-rate) intensities that are not possible to determine through direct observation. In the next section, we will perform some example calculations to demonstrate how this equation is used in practice.

2.3 *Example PSHA calculations*

To illustrate the procedure described in the previous section, several numerical examples will be performed below, starting from basic calculations and building to more complex cases. These examples will compute rates of exceeding varying levels of Peak Ground Acceleration, using the procedures described above.

Example: a source with a single magnitude and distance

We first start with a simple site shown in Figure 2.3. There is a single fault (Fault A) that produces only magnitude 6.5 earthquakes. We assume that this earthquake will rupture the entire fault, so that the source-to-site distance is exactly 10 km in every case (that is,

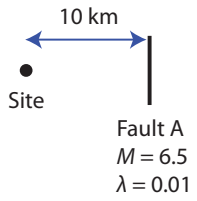


Figure 2.12: Map view of the example site, with one earthquake source.

we will not consider random locations). Assume also that earthquakes with this magnitude and distance occur at a rate of $\lambda = 0.01$ times per year. Although the magnitudes and distances of all future earthquakes are fixed at 6.5 and 10 km, respectively, we still expect variations in observed peak ground accelerations at the site, due to differences from event to event that are not captured by our simple measures of magnitude and distance.

Using the Cornell et al. model presented in equation 2.14, we predict a median PGA of 0.3758g (i.e., a mean $lnPGA$ of -0.979), and log standard deviation of 0.57. We can easily compute that the probability of exceeding 0.3758g given an earthquake is 0.5, since this is the median predicted value. The annual rate of exceeding 0.3758g is thus $0.5 * 0.01 = 0.005$ per year. This quick calculation is done to develop intuition regarding the calculations, but we can also use the more formal equations presented above. When considering equation 2.25, we see that $P(M = 6.5) = 1$, $P(R = 10) = 1$ and $\lambda(M_i > m_{\min}) = 0.01$. There is only one magnitude, distance and source to consider, so all of the summations are replaced by a single term. We thus have

$$\begin{aligned}\lambda(PGA > x) &= \lambda(M > m_{\min})P(PGA > x|6.5, 10)P(M = 6.5)P(R = 10) \\ &= 0.01P(PGA > x|6.5, 10)\end{aligned} \tag{2.26}$$

Next, since we know the mean and standard deviation of lnPGA, we can compute the probability of exceeding a given PGA value using equation 2.15

$$\begin{aligned}P(PGA > x|6.5, 10) &= 1 - \Phi\left(\frac{\ln x - \overline{\ln PGA}}{\sigma_{\ln PGA}}\right) \\ &= 1 - \Phi\left(\frac{\ln x - \ln(0.3758)}{0.57}\right)\end{aligned} \tag{2.27}$$

We can use Table A.1 to evaluate the standard normal cumulative distribution function $\Phi()$. For example, the probability of PGA >1g is

$$P(PGA > 1g|6.5, 10) = 1 - \Phi(1.72) = 0.044 \tag{2.28}$$

Substituting this into equation 2.26, we can find the annual rate of exceeding 1g

$$\lambda(PGA > 1g) = 0.01\, P(PGA > 1g|6.5, 10) = 0.00044 \tag{2.29}$$

By repeating these calculations for many PGA levels, one can construct the curve shown in Figure 2.13. This "ground motion hazard curve" for PGA summarizes the rates of exceeding a variety of PGA levels. The two calculations performed explicitly above (for PGA >0.3758g and PGA >1g) are labeled on this figure as well. Note that

because both axes often cover several orders of magnitude, and the y axis contains very small values, one or both axes of ground motion hazard curves are often plotted in log scale.

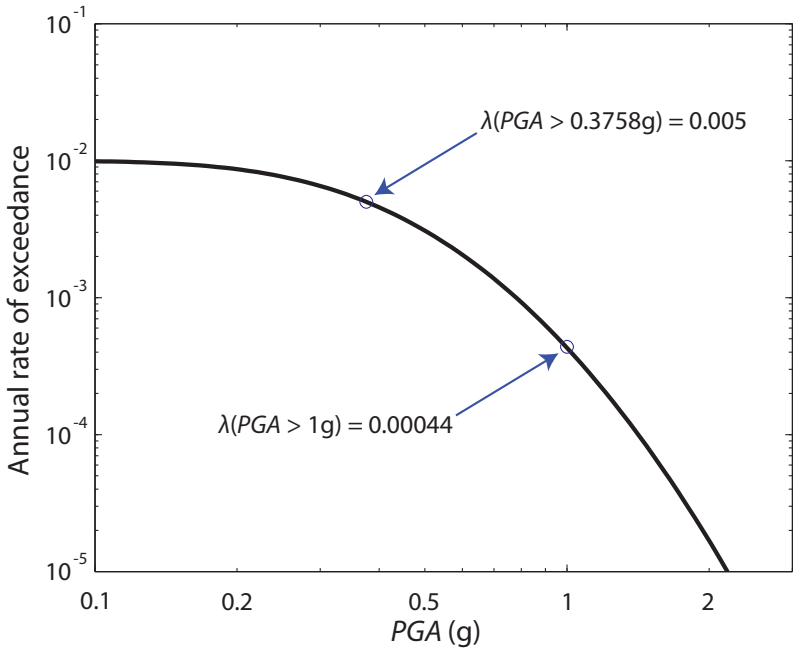


Figure 2.13: *PGA* hazard curve for the single-source site.

This example demonstrates the essence of a PSHA calculation. All that remains to perform a more realistic calculation is to consider a range of feasible magnitudes and distances, rather than just the single event in this hypothetical example.

Example: two magnitudes and distances

Before moving to an example with a continuous range of magnitudes, let us first try another hypothetical example with only two possible magnitude/distance combinations. The first source, "Fault A," is identical to the source in the immediately preceding example. The second source, "Fault B," produces only magnitude 7.5 earthquakes at a distance of 20 km. The earthquake on Fault B occurs at a rate of $\lambda = 0.002$ times per year. The map of this example site is shown in Figure 2.3. We will continue using the Cornell et al. model presented in equation 2.14 to predict *PGA* at the site.

The previous example to quantified the hazard from Fault A, so let us focus on calculating the hazard from Fault B. Using the Cornell et al. ground motion model, we predict a median *PGA* of 0.5639g if the earthquake on Fault B occurs, and a log standard deviation of 0.57. Now consider the two *PGA* values considered in the previous example. The probability of *PGA* >0.3758g, given an earthquake on

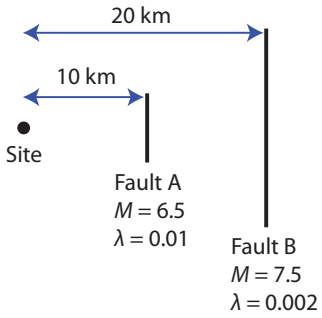


Figure 2.14: Map view of an example site with two earthquake sources.

Fault B, is

$$\begin{aligned} P(PGA > 0.3758g|7.5,20) &= 1 - \Phi\left(\frac{\ln(0.3758) - \ln(0.5639)}{0.57}\right) \\ &= 1 - \Phi(-0.71) \\ &= 0.761 \end{aligned} \tag{2.30}$$

We can then multiply this probability by the rate of occurrence of earthquakes on Fault B (0.02), to get the rate of *PGA* >0.3758g due to earthquakes on Fault B. But the PSHA formula of equation 2.25 includes a summation over all sources, so we add these probabilities to the corresponding probabilities for Fault A to find the overall rate of *PGA* >0.3758g

$$\begin{aligned} \lambda(PGA > 0.3758g) &= \overbrace{0.01\,P(PGA > 0.3758|6.5,10)}^{\text{(Fault A)}} &&+ \overbrace{0.002\,P(PGA > 0.3758|7.5,20)}^{\text{(Fault B)}} \\ &= 0.01(0.5) &&+ 0.002(0.761) \\ &= 0.00500 &&+ 0.00152 \\ &= 0.00652 \end{aligned} \tag{2.31}$$

Similarly, we can compute the probability of *PGA* >1g, given an earthquake on Fault B

$$\begin{aligned} P(PGA > 1g|7.5,20) &= 1 - \Phi\left(\frac{\ln(1) - \ln(0.5639)}{0.57}\right) \\ &= 1 - \Phi(1.01) \\ &= 0.158 \end{aligned} \tag{2.32}$$

and then combine this with our other known information to compute the overall rate of *PGA* >1g

$$\begin{aligned} \lambda(PGA > 1g) &= \overbrace{0.01\,P(PGA > 1|6.5,10)}^{\text{(Fault A)}} &&+ \overbrace{0.002\,P(PGA > 1|7.5,20)}^{\text{(Fault B)}} \\ &= 0.01(0.043) &&+ 0.002(0.158) \\ &= 0.000430 &&+ 0.000316 \\ &= 0.000746 \end{aligned} \tag{2.33}$$

The two rates computed above are plotted in Figure 2.15, along with rates for all other *PGA* levels. Also shown in Figure 2.15 are curves showing the rates of exceedance for the two individual faults. The intermediate rate calculations shown above (0.005 and 0.00152 for *PGA*>0.3758g, and 0.0043 and 0.00316 for *PGA*>1g) are also noted with circles on Figure 2.15 for reference. A few observations can be made regarding this figure and its associated calculations. First, note that the hazard curve for Fault A in the figure is identical to

the curve in Figure 2.13. We have simply added the additional rates of exceedance due to Fault B in order to get the total hazard curve shown in Figure 2.15. Second, we can note that the relative contributions of the two faults to the ground motion hazard vary depending upon the *PGA* level of interest. At relatively lower *PGA* values such as in the calculation of equation 2.31, Fault A contributes much more to the overall rate of exceedance. This is because it has a higher rate of earthquakes. At larger *PGA* levels such as in the calculation of equation 2.33, the relative contributions of the two faults are close to equal; this is because the smaller-magnitude earthquakes from Fault A have a low probability of causing high *PGA* values, even thought they are more frequent than the larger-magnitude earthquakes from Fault B. We see in Figure 2.15 that for *PGA* values greater than about 1.5g, Fault B actually makes a greater contribution to the hazard than Fault A, even though its rate of producing earthquakes is only 1/5 of Fault A's. This is a typical situation for real-life PSHA calculations as well: low-intensity shaking is generally dominated by frequent small earthquakes, while high-intensity shaking is caused primarily by large-magnitude rare earthquakes.

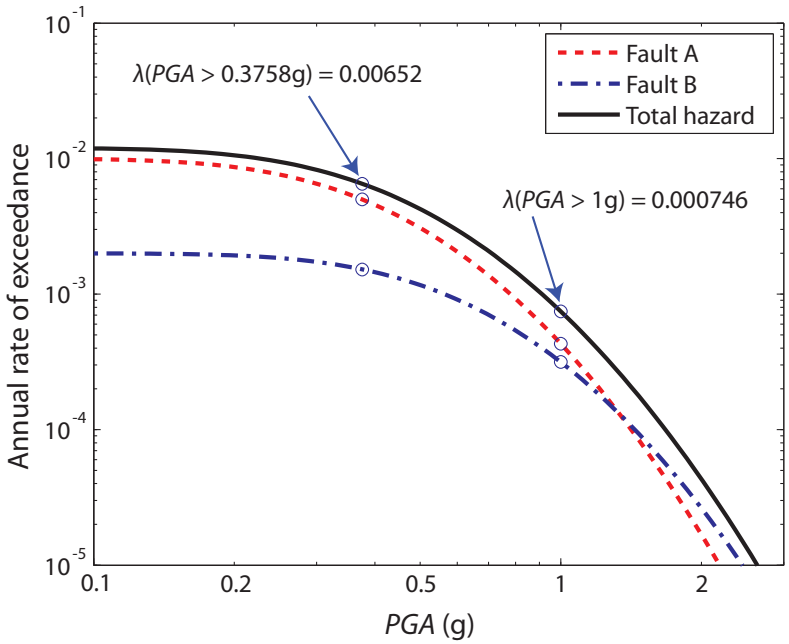


Figure 2.15: *PGA* hazard curve for the example two-source site.

Example: point source with Gutenberg-Richter magnitudes

In this example we will now consider a source that is capable of producing earthquakes with a variety of magnitudes. The source produces events with $M \geq 5$ at a rate of 0.02 events per year. The

distribution of those earthquakes follows the bounded Gutenberg-Richter model, with a maximum magnitude of 8 and "b" value of 1. We will thus use equation 2.5 to describe the PDF of earthquake magnitudes, with $m_{min} = 5$ and $m_{max} = 8$. We again assume that all earthquakes occur at the same distance of 10 km, so that we can simplify the PSHA summations.

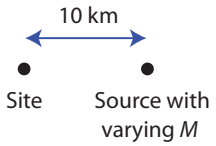


Figure 2.16: Map view of an example site with one source producing earthquakes having a variety of magnitudes.

We will use equation 2.25 to perform the PSHA calculation for peak ground acceleration, using the Cornell et al. ground motion model from the previous examples. Noting that there is only one source and one distance, and substituting $P(R = 10) = 1$ and $\lambda(M_i > m_{\min}) = 0.02$, we get

$$\lambda(PGA > x) = 0.02 \sum_{j=1}^{n_M} P(PGA > x | m_j, 10) P(M = m_j) \qquad (2.34)$$

To compute this rate of exceeding some *PGA* level, we simply need to compute the probabilities of observing various earthquake magnitudes, compute the probabilities of exceeding our *PGA* level given those magnitudes, and then sum the product of those two terms evaluated for the range of feasible magnitudes. Table 2.3 shows those probabilities for calculations of $\lambda(PGA > 0.2g)$. The first column lists the discrete set of magnitudes considered (a magnitude increment of 0.25 has been used). The second column lists the probabilities of observing those magnitudes, as computed using equation 2.6 (note these probabilities are identical to those in Table 2.1, because the assumed magnitude distribution is the same). The third column lists the probability of *PGA* >0.2g, given occurrence of an earthquake having the specified magnitude. This is computed, as was done in the earlier examples, by evaluating the Cornell et al. ground motion model using each magnitude value. The fourth column lists the products of the second and third columns. We see that equation 2.34 is simply a summation of the terms in the fourth column, multiplied by the rate of occurrence of earthquakes. The sum of the fourth column is 0.269, so the rate of *PGA*>0.2g at the site of interest is $\lambda(PGA > 0.2) = 0.02(0.269) = 0.0054$.

We continue the hazard analysis by repeating the calculations of Table 2.3 for more *PGA* values. In Table 2.4, the same calculation is repeated for *PGA* >1g. Here we see that the summation of the right-hand column is 0.0048, so the rate of *PGA* >1g at the site of interest is $\lambda(PGA > 1) = 0.02(0.0048) = 9.6 \cdot 10^{-5}$.

By repeating this calculation for many more *PGA* values, we can create the *ground motion hazard curve* shown in Figure 2.17. The two individual rates of exceedance calculated above are labeled on this curve.

Comparing Table 2.3 to Table 2.4, we can make several observa-

m_j	$P(M = m_j)$	$P(PGA > 0.2 \mid m_j, 10)$	$P(PGA > 0.2 \mid m_j, 10) \cdot P(M = m_j)$
5.00	0.4381	0.1242	0.0544
5.25	0.2464	0.2185	0.0538
5.50	0.1385	0.3443	0.0477
5.75	0.0779	0.4905	0.0382
6.00	0.0438	0.6379	0.0279
6.25	0.0246	0.7672	0.0189
6.50	0.0139	0.8657	0.0120
6.75	0.0078	0.9310	0.0073
7.00	0.0044	0.9686	0.0042
7.25	0.0025	0.9873	0.0024
7.50	0.0014	0.9955	0.0014
7.75	0.0008	0.9986	0.0008
			Sum = 0.269

Table 2.3: Probabilities used to compute $\lambda(PGA{>}0.2g)$.

m_j	$P(M = m_j)$	$P(PGA > 1 \mid m_j, 10)$	$P(PGA > 1 \mid m_j, 10) \cdot P(M = m_j)$
5.00	0.4381	0.0000	0.0000
5.25	0.2464	0.0002	0.0000
5.50	0.1385	0.0006	0.0001
5.75	0.0779	0.0022	0.0002
6.00	0.0438	0.0067	0.0003
6.25	0.0246	0.0181	0.0004
6.50	0.0139	0.0430	0.0006
6.75	0.0078	0.0901	0.0007
7.00	0.0044	0.1676	0.0007
7.25	0.0025	0.2786	0.0007
7.50	0.0014	0.4168	0.0006
7.75	0.0008	0.5662	0.0004
			Sum = 0.0048

Table 2.4: Probabilities used to compute $\lambda(PGA > 1g)$.

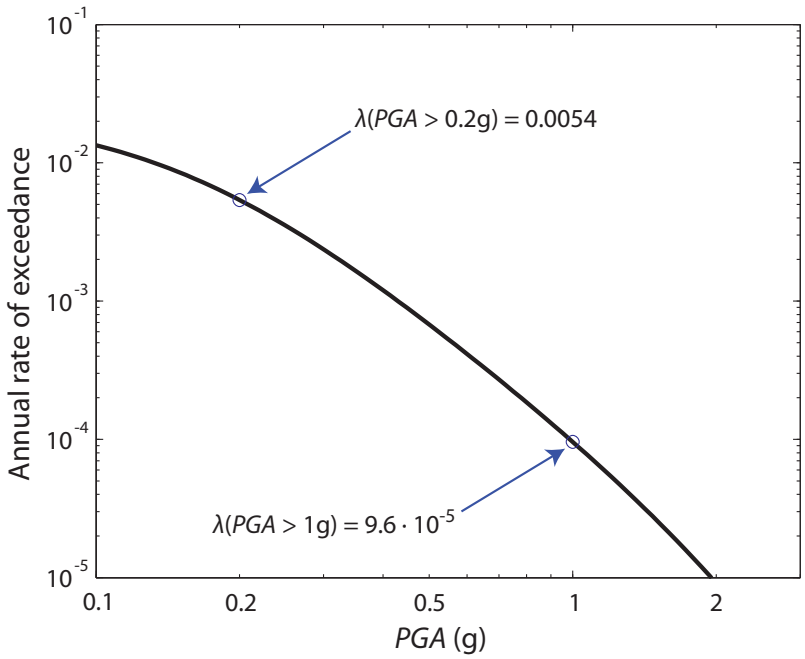


Figure 2.17: *PGA* hazard curve for the example site with one source and a Gutenberg-Richter magnitude distribution.

tions. The first two columns of both tables are identical, as they are only describing the earthquake magnitudes and so are not affected by changes in the PGA level of interest. We see that all probabilities in the third column are much larger in Table 2.3 than in Table 2.4: the PGA threshold was lower in Table 2.3, so the probability of exceeding the threshold is therefore higher.

In Table 2.4, the probability of $PGA > 1g$ is equal to zero for the smallest magnitude considered. This means that considering even smaller magnitudes would have no impact on our final answer, because smaller magnitude earthquakes have effectively zero probability of causing a PGA greater than $1g$. In Table 2.3, however, we see that even magnitude 5 earthquakes have a non-zero probability of causing $PGA > 0.2g$; this is somewhat worrisome because lower magnitude earthquakes could also cause $PGA > 0.2g$, so including them would have changed our answer. This suggests that the choice of the minimum considered earthquake can be important in some cases. We will return to this issue in Section 3.2.

By looking at the right-hand column of these tables, we can also identify the magnitudes that make the greatest contribution to the probability of exceeding the PGA level of interest. Each number in this column is a product of the probability of occurrence of some magnitude and the probability of exceedance of the PGA given that magnitude, which is equal to the probability of both events occurring (given that an earthquake has occurred). In Table 2.3, we see that

the probabilities are largest for the small magnitudes; this is because small-magnitude earthquakes are much more likely to occur than large-magnitude earthquakes (as seen in column 2), *and* because these small-magnitude earthquakes have a reasonable probability of causing $PGA > 0.2g$. In Table 2.4, on the other hand, these small magnitude earthquakes have very small probabilities in the fourth column because they have a very small probability of causing $PGA > 1g$. In Table 2.4, the moderate- to large-magnitude earthquakes have the highest probabilities in the fourth column, because even though they are relatively rare, they are the only ones with significant probabilities of causing $PGA > 1g$. Intuitively, we can imagine that this information would be useful for identifying the earthquake scenarios most likely to damage a structure at the site of interest. We will revisit this information in a more quantitative manner in Section 3.1 below.

For more complex sites than the simple cases shown in the above examples, the PSHA summations can quickly get lengthy. For this reason, PSHA is performed using computer software in all practical analysis cases. The software's purpose is to perform the calculations shown here, but for more complicated cases involving many earthquake sources, while also using modern ground motion prediction models that are much more complex than the one used here. Note that in the example above we used a relatively wide magnitude spacing of 0.25 units in our discretization in order to keep the length of Table 2.3 reasonably short. When performing these calculations in a computer program, it is also easy to use a fine discretization of the magnitudes and distances of interest.

3
Extensions of PSHA

The PSHA inputs and primary PSHA calculations were described in the previous section. To fully utilize the information provided by those calculations, several extensions are often used. These extensions are described in the following sections.

3.1 Deaggregation

One of the primary advantages of PSHA—that it accounts for all possible earthquake sources in an area when computing seismic hazard—is also a disadvantage. Once the PSHA computations are complete, a natural question to ask is "which earthquake scenario is most likely to cause *PGA*>x?" Because we have aggregated all scenarios together in the PSHA calculations, the answer is not immediately obvious. We saw in the example calculations above, however, that some of the intermediate calculation results indicated the relative contribution of different earthquake sources and magnitudes to the rate of exceedance of a given ground motion intensity. Here we will formalize those calculations, through a process known as deaggregation[1] (Bazzurro & Cornell, 1999; McGuire, 1995).

[1] The calculations shown in this section are known as both "deaggregation" and "disaggregation." No universal terminology has yet been adopted. Disaggregation is the only one of the two words that is found in a dictionary, but deaggregation is currently used more often.

Let us start with magnitude deaggregation. In this case, we are interested in the probability that an earthquake's magnitude is equal to m, given that a ground motion of $IM > x$ has occurred. Intuitively, this is equal to the rate of earthquakes with $IM > x$ and $M = m$, divided by the rate of all earthquakes with $IM > x$

$$P(M = m | IM > x) = \frac{\lambda(IM > x, M = m)}{\lambda(IM > x)} \tag{3.1}$$

This relationship can also be derived more rigorously, as an application of Bayes' rule. The denominator of this equation is exactly what we have computed previously in equation 2.25 (i.e., our primary PSHA equation). The numerator is very similar, except that it specifies occurrence of a given causal magnitude, rather than integrating

over all magnitudes as we did in the previous PSHA equations. To compute this numerator, we simply omit the summation over M from equation 2.25

$$\lambda(IM > x, M = m) = \sum_{i=1}^{n_{sources}} \lambda(M_i > m_{\min}) \sum_{k=1}^{n_{R_i}} P(IM > x|m, r_k) P(M_i = m) P(R_i = r_k) \tag{3.2}$$

Example 1

To illustrate, let us consider again the example of Section 2.3. This site had two sources, and we might be interested in the relative contributions of the two sources to exceedance of a given intensity level. Consider first the case PGA >0.3758g, since we have previously computed some needed probabilities for that PGA value. Looking at equation 2.31 from that example, we can see that

$$\begin{aligned} \lambda(PGA > 0.3758g, M = 6.5) &= 0.01\, P(PGA > 0.3758|6.5, 10) \\ &= 0.005 \end{aligned} \tag{3.3}$$

$$\begin{aligned} \lambda(PGA > 0.3758g, M = 7.5) &= 0.002\, P(PGA > 0.3758|7.5, 20) \\ &= 0.00152 \end{aligned} \tag{3.4}$$

$$\lambda(PGA > 0.3758g) = 0.00652 \tag{3.5}$$

Plugging those three numbers into equation 3.1 gives $P(M = 6.5|PGA > 0.3758g) = 0.77$ and $P(M = 7.5|PGA > 0.3758g) = 0.23$. Repeating the same calculations using the results of equation 2.33 gives $P(M = 6.5|PGA > 1g) = 0.58$ and $P(M = 7.5|PGA > 1g) = 0.42$. So we see, for the relatively lower PGA value of 0.3758g, that the smaller and more active fault has a high probability of being the causal fault. At the larger PGA intensity of 1g, the less active but larger fault makes a greater contribution to exceedance of the PGA. This is consistent with the qualitative observations made at the end of the original example calculation. And note that these quantitative calculations are actually very simple. The probabilities we have computed are exactly proportional to the rates we previously computed for $\lambda(IM > x, M = m)$; all we have done here is normalize those rates to get probabilities that sum to one.

Example 2

Let us consider the example of Section 2.3 as a slightly more realistic application of deaggregation. This site had a single source producing earthquake magnitudes with a Gutenberg-Richter distribution. We

can use deaggregation to find the probability that some *PGA* level was exceeded by an earthquake with a given magnitude. Let us first consider *PGA* values greater than 0.2g. Referring to equation 3.1, we see that we have previously computed $\lambda(PGA > 0.2) = 0.0054$. Further, since this site has only a single source and a single distance, equation 3.2 simplifies to

$$\lambda(IM > x, M = m) = \lambda(M > m_{\min})P(IM > x|m, 10)P(M = m) \quad (3.6)$$

For this example, $\lambda(M > m_{\min}) = 0.02$, and the remaining two terms are provided in Table 1.3 (in fact, their product is given in the far right column). So we see that all of the needed inputs for a deaggregation calculation are already computed as part of the basic PSHA calculation. To demonstrate this calculation, let us find the probability of a *PGA* >0.2g ground motion being caused by a magnitude 5 earthquake.

$$P(M = 5|PGA > 0.2) = \frac{\lambda(PGA > 0.2, M = 5)}{\lambda(PGA > 0.2)} = \frac{0.02(0.4381)(0.1242)}{0.0054} = 0.20 \quad (3.7)$$

This deaggregation data is computed for other magnitudes and summarized in Table 2.1. The first two columns list magnitudes of interest and their probabilities of occurrence in this example (and are identical to the probabilities computed earlier in Table 1.1). The third column shows the probability of $PGA > 0.2$ associated with each of the magnitudes. The fourth column shows $\lambda(PGA > 0.2, M = m_j)$, computed using equation 3.2 and noting that $\lambda(M_i > m_{\min}) = 0.02$ for this problem. The fifth column shows the deaggregation probability, computed using equation 3.1, and noting that $\lambda(PGA > 0.2) = 0.0054$.

Table 3.1: Deaggregation calculations for the Section 2.3 example.

m_j	$P(M = m_j)$	$P(PGA > 0.2\|m, 10)$	$\lambda(PGA > 0.2, M = m_j)$	$P(M = m_j\|PGA > 0.2)$
5.00	0.4381	0.1242	0.0011	0.2022
5.25	0.2464	0.2185	0.0011	0.2000
5.50	0.1385	0.3443	0.0010	0.1773
5.75	0.0779	0.4905	0.0008	0.1420
6.00	0.0438	0.6379	0.0006	0.1039
6.25	0.0246	0.7672	0.0004	0.0702
6.50	0.0139	0.8657	0.0002	0.0446
6.75	0.0078	0.9310	0.0001	0.0270
7.00	0.0044	0.9686	0.0001	0.0158
7.25	0.0025	0.9873	0.0000	0.0090
7.50	0.0014	0.9955	0.0000	0.0051
7.75	0.0008	0.9986	0.0000	0.0029
8.00	0.0000	0.9996	0.0000	0.0000

The deaggregation values shown in the fifth column of Table 2.1 are plotted in Figure 3.1a. We easily see in this figure that small-

magnitude events have the highest probability of causing *PGA* >0.2g ground motions, matching the intuitive observations that were made during that example calculation. Figure3.1b shows the same deaggregation plot, but conditioned on *PGA* >1g. We see that large-magnitude events have the highest probability of causing exceedance of this large amplitude, because the frequent small events have a very low probability of causing such a large amplitude.

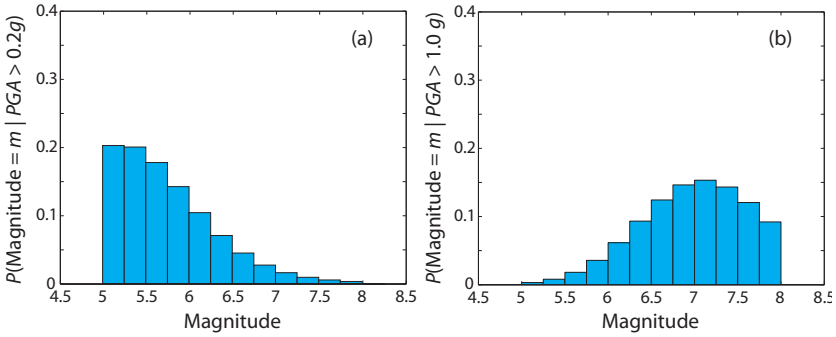


Figure 3.1: Deaggregation results associated with the example calculation of Section 2.3. (a) Magnitude distribution, given *PGA* >0.2g. (b) Magnitude distribution, given *PGA* >1g.

We see in Figure 3.1 that the deaggregation results vary as the intensity level of interest changes. The deaggregation results will also vary if one studies a different measure of ground motion intensity. That is, the events that cause extreme *PGA* levels will differ from the events that cause extreme spectral acceleration levels at long periods. Our intuition in this respect may benefit from reviewing Figure 1.1b, which shows the median response spectra from two different earthquake events. In that figure, the large-magnitude event produces larger spectral acceleration values at long periods, while the small-magnitude nearby event produces larger spectral acceleration values at short periods. This will intuitively lead to different events dominating the deaggregation results at those associated periods, and that is what we see in practice with real sites as well. Thus, deaggregation helps us see that there is no single earthquake event that is the design earthquake for every situation at a given site. The earthquake of interest will depend upon the ground motion intensity measure of interest, as well as the intensity level of interest.

The results shown in Figure 3.1 were computed using the same magnitude bins as were used for the original PSHA calculations, so that we did not have to re-compute as many probabilities. But this is not representative of common practice in real calculations. A typical PSHA calculation will use a finer magnitude discretization for the basic PSHA summation (perhaps a 0.1 magnitude interval), and a much coarser discretization for the deaggregation computations (perhaps a 0.5 magnitude interval). This is because the original summation is never output, and so can be finely discretized to maximize

accuracy. The deaggregation output, however, is often presented in tabular form, and so is more coarsely discretized to minimize the length of output tables. Keep in mind that here we are deaggregating on only magnitude, but when deaggregating to find the probabilities of combinations of various magnitudes, distances, etc., the output can quickly become much lengthier. To illustrate the use of a more coarse discretization, and to provide data that will be used later, Table 2.2 and Figure 3.2 show results identical to Table 2.1 and Figure 3.1, but with a coarser discretization of magnitudes, into intervals of width 0.5. The only change needed to produce these calculations is to modify the $P(M = m_j)$ calculation to account for the fact that a larger interval of magnitudes is being assigned to each discrete magnitude. By comparing Figure 3.1 and Figure 3.2, we see that the locations and shapes of these magnitude distributions are still clear using the coarser discretization. For qualitative evaluations of causal earthquakes, coarse discretizations are thus useful. For quantitative calculations that use these results, it is important to consider potential errors caused by coarsely discretizing the deaggregation results.

m_j	$P(M = m_j)$	$P(PGA > 0.2 \mid m, 10)$	$\lambda(PGA > 0.2, M = m_j)$	$P(M = m_j \mid PGA > 0.2)$
5.0	0.6845	0.1242	0.0017	0.3685
5.5	0.2164	0.3443	0.0015	0.3230
6.0	0.0684	0.6379	0.0009	0.1892
6.5	0.0216	0.8657	0.0004	0.0812
7.0	0.0068	0.9686	0.0001	0.0287
7.5	0.0022	0.9955	0.0000	0.0093
8.0	0.0000	0.9996	0.0000	0.0000

Table 3.2: Deaggregation calculations for the Section 2.3 example, with magnitudes discretized into 0.5-unit intervals.

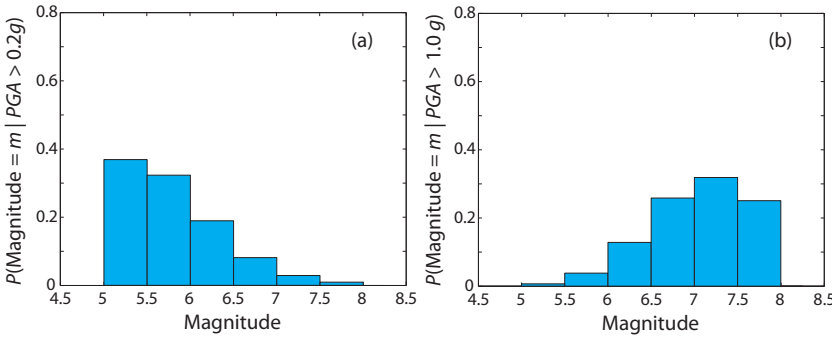


Figure 3.2: Deaggregation results associated with the example calculation of Section 2.3, with magnitudes discretized into 0.5-unit intervals. (a) Magnitude distribution, given *PGA* >0.2g. (b) Magnitude distribution, given *PGA* >1g.

The above deaggregation results focus solely on the conditional distribution of magnitude. The same calculation can be done to find the conditional distribution of distance, by simply modifying equation 3.2 to have a summation over magnitudes but not over distances. One can also find the conditional *joint* distribution of magnitudes and

distances, using the following equation

$$P(M = m, R = r|IM > x) = \frac{\lambda(IM > x, M = m, R = r)}{\lambda(IM > x)} \tag{3.8}$$

where the numerator of equation 3.8 is computed using the basic PSHA equation but not summing over either M or R

$$\lambda(IM > x, M = m, R = r) = \sum_{i=1}^{n_{sources}} \lambda(M_i > m_{\min}) P(IM > x|m_j, r_k) P(M_i = m) P(R_i = r) \tag{3.9}$$

An example of this conditional distribution of M and R given IM >x is shown in Figure 3.3.

These deaggregation calculations are a critical part of many PSHA analyses, and deaggregation results should be provided as part of the output from any PSHA calculation. The U.S. Geological Survey, which performs PSHA for the United States that is incorporated into building codes, provides deaggregation results alongside the basic PSHA output (*https://geohazards.usgs.gov/deaggint/2008/*). An example is shown on Figure 3.3 for a site in Palo Alto, California.

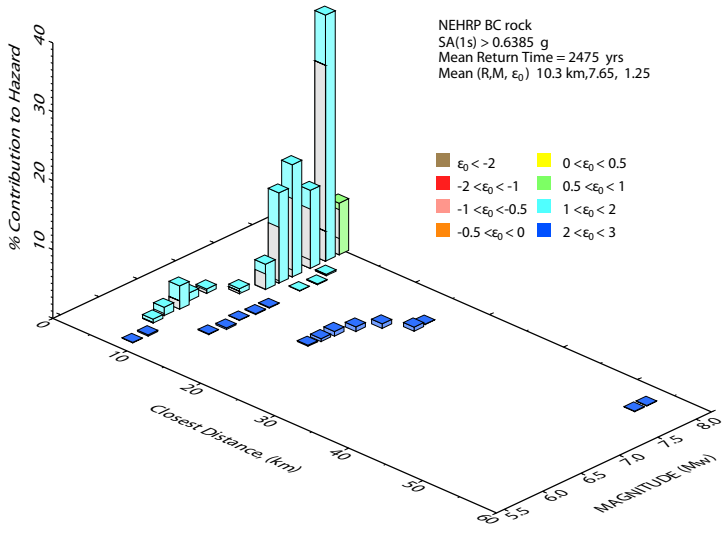


Figure 3.3: Example disaggregation for $SA(1.0s)$ at a site in Palo Alto, California (USGS, 2008) .

3.2 *Bounds on considered magnitudes and distances*

For practical reasons, not all earthquake magnitudes are considered in PSHA calculations. Typically, only earthquakes with magnitudes greater than approximately 4.5 or 5 are considered. This is chosen as a conservative value, for which the omitted small-magnitude earthquakes are not believed to be capable of damaging structures, and

thus not relevant for seismic risk calculations. This also reduces the size of the calculations. The exact magnitude at which an earthquake is no longer damaging is not obvious, however, and unfortunately the choice of cutoff magnitude can significantly affect some PSHA results.

To illustrate, let us consider again the example calculation from Section 2.3. In this example, there was a point source that produced magnitudes having a truncated Gutenberg-Richter distribution. In the original calculation, we assumed a minimum magnitude of 5.0. But what would happen to our results if we chose a different minimum magnitude? Figure 3.4 shows the hazard curves for that site using three choices of minimum magnitude[2]. The case with $m_{min} = 5.0$ is identical to the result shown in Figure2.17. We see that the exceedance rates for small *PGA* values vary dramatically, but the rate exceedance rates for large *PGA* values are nearly identical. This is not surprising, given the deaggregation results of Figure 3.1. In Figure 3.1b, we see that small earthquakes make almost no contribution to exceedances of 1g, so the minimum considered magnitude will not impact that calculation. In Figure 3.1a, on the other hand, we see that small magnitude earthquakes make a significant contribution to exceedances of 0.2g. Thus, while it might be conceptually reasonable to omit non-damaging small-magnitude earthquakes from the PSHA calculation, it is also clear that the results may be sensitive to the actual choice of minimum magnitude.

An aside: There is typically a restriction placed on the maximum considered distance in equation 2.24 as well. That choice typically has no significant impact on the PSHA results, however, as long as the maximum distance is several hundred kilometers. The choice of maximum distance will not be considered further here.

[2] The calculations to produce the three ground motion hazard curves are identical to those used in section 2.3. The only changes needed are to adjust m_{min}, and to adjust the corresponding rate of occurrence of $M>m_{min}$. The appropriate rate of occurrence was found by using equation 2.1 to find the a value that gives the specified rate of $M>5$, and then using that a value to solve for the rate of $M>m_{min}$ using the alternatem_{min} choices.

3.3 *Rates, probabilities and return periods*

The text above has focused only on rates of exceeding a given ground motion intensity. Sometimes, PSHA results are also formulated in terms of probabilities or return periods of exceedance. The return period is defined as the reciprocal of the rate of occurrence. For example, if a given ground motion intensity has a 0.01 annual rate of occurrence, then the return period is equal to 1/0.01=100 years. This does not imply that the ground motion will be exceeded exactly once every 100 years, but rather that the average (or mean) time between exceedances is 100 years. For this reason, the reciprocal of the exceedance rate is more precisely termed the *mean* return period. While "mean return period" or simply "return period" are commonly used names used to refer to the reciprocal of the rate of occurrence, one

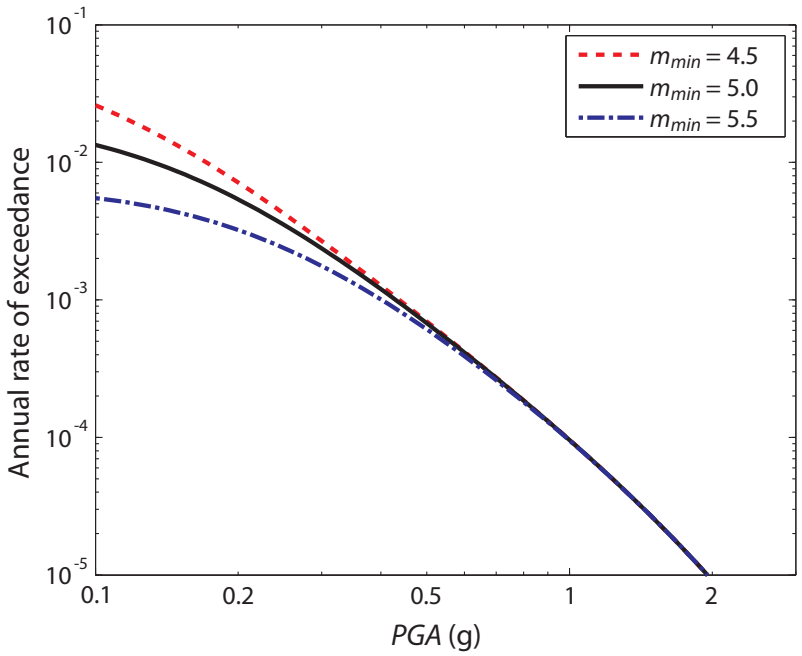


Figure 3.4: Hazard curves computed for the example site from Section 2.3, using several choices for the minimum considered magnitude.

may avoid some confusion regarding the implied time between exceedances by simply reporting rates rather than return periods.

For a given rate of exceedance, one can also compute a probability of exceeding a given ground motion intensity within a given window of time. This calculation requires further information regarding the probability distribution of time between occurrences of earthquakes. This distribution is nearly always assumed to be "Poissonian," for three reasons: it results in simple mathematical equations, it appears to match observations in most cases, and more complicated models typically do not impact the final results significantly. The Poisson model assumes that occurrences of earthquakes are independent in time (that is, the probability of an earthquake in a window of time is related only to the size of the window, and is independent of anything such as the time since the most recent occurrence), and that the probability of more than one occurrence in a very short interval is negligible. *Under the assumption of Poissonian occurrences*, the probability of observing at least one event in a period of time t is equal to

$$P(\text{at least one event in time } t) = 1 - e^{-\lambda t} \tag{3.10}$$

where λ is the rate of occurrence of events. A plot of this relationship is shown in Figure 3.5.

If λt is small (less than approximately 0.1), then the probability can also be approximated by

$$P(\text{at least one event in time } t) = 1 - e^{-\lambda t} \cong \lambda t \tag{3.11}$$

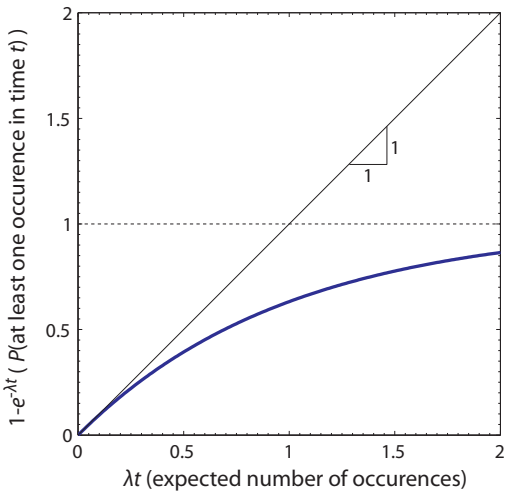


Figure 3.5: Probability of occurrence of an event in time t, as a function of the expected number of occurrences, λt.

This approximation comes from taking the first term of a Taylor series expansion of $1\text{-}e^{-\lambda t}$. The accuracy of the approximation can be seen in Figure 3.5, where the plot follows a straight line with a slope of 1 for λt values less than 0.1.

Using the above calculations, PSHA results are converted between rates of exceedance, probabilities of exceedance, and return periods. There are two important caveats to these conversions that should be kept in mind:

1. The conversion between rates of exceedance and probabilities of exceedance is almost always made by assuming a Poissonian occurrence of earthquakes (whether or not this has been stated explicitly by the analyst).

2. Probabilities of exceedance and rates of exceedance are only equivalent if the probability level of interest is small (i.e., less than 0.1).

An Aside: Many proposals have been made to predict the recurrence of earthquakes using models other than the Poisson model (e.g., Anagnos & Kiremidjian, 1984). One summary of the sensitivity of computed probabilities to the choice of earthquake recurrence model is given by Cornell & Winterstein (1988), who also describe some alternative recurrence models. Cornell and Winterstein found that the Poisson model is accurate for PSHA unless the seismic hazard is dominated by a single source, the time since that source's last event is greater than the average time between events, and the source has strong "characteristic time" behavior.

3.4 *The uniform hazard spectrum*

A common goal of probabilistic seismic hazard analysis is to identify a design response spectrum to use for structural or geotechnical analysis. One approach for developing a spectrum is to compute a uniform hazard spectrum (UHS). This spectrum is developed by first performing the above PSHA calculations for spectral accelerations at a range of periods. Then, a target rate of exceedance is chosen, and for each period the spectral acceleration amplitude corresponding to that rate is identified. Those spectral acceleration amplitudes are then plotted versus their periods, as illustrated in Figure 3.6.

This spectrum is called a uniform hazard spectrum because every ordinate has an equal rate of being exceeded. But it should be clear that this spectrum is an envelope of separate spectral acceleration values at different periods, each of which may have come from a different earthquake event[3]. This mixing of events to create a spectrum has sometimes been used to criticize the entire PSHA procedure. But it is important to recognize that a UHS is merely one way to use the output of PSHA. None of the calculations prior to this section required the use of a UHS, and it is quite possible to productively use PSHA results without ever computing a UHS. In fact, the Commission (1997) relies on the deaggregation results discussed earlier to identify scenario magnitudes and distances and then computes a design spectral shape based on those magnitudes and distances. In that procedure, no uniform hazard spectrum is needed.

[3] For example, consider Figure 2.1b as an illustration of a case where high-frequency ground motions may be caused by one source, and low frequency ground motions by another. Even in the case where there is only a single source, it is not necessary for extreme (i.e., larger-than-median) ground motion amplitudes to occur simultaneously in the same ground motion.

Some other design procedures (e.g., for design of buildings) use the UHS, but in those cases one can simply remember the manner in which a UHS is computed in order to avoid misinterpreting the results as the spectrum from some single ground motion. Although the uniform hazard spectrum is not required for PSHA calculations, it is a commonly-used procedure. It is also a good example of the various ways in which PSHA calculations can be adopted for various uses. For these reasons, it warrants at least a brief mention in any summary of PSHA.

3.5 *Joint distributions of two intensity measures*

While the uniform hazard spectrum calculation of the previous section provides one way of combining multiple ground motion intensity measures, it is not probabilistically rigorous (i.e., it does not compute the probability of simultaneous occurrence these parameters). But it is possible to account for joint predictions, and some of the necessary tools are discussed briefly here.

Logarithms of pairs of spectral acceleration values (and presum-

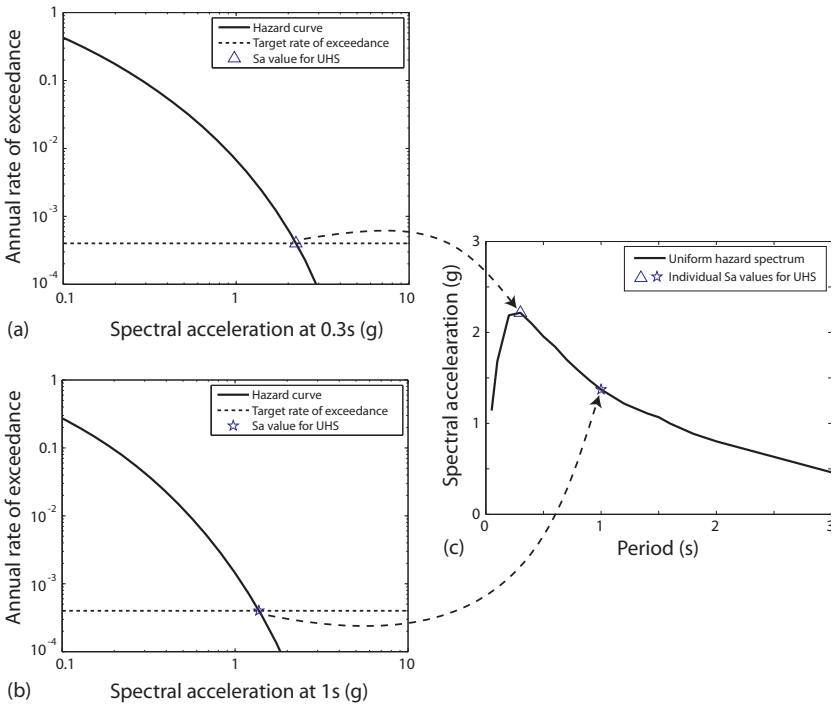


Figure 3.6: Combining hazard curves from individual periods to generate a uniform hazard spectrum with a $4^{*}10^{-4}$ rate of exceedance for a site in Los Angeles. (a) Hazard curve for $SA(0.3s)$, with UHS point identified. (b) Hazard curve for $SA(1s)$, with UHS point identified. (c) Uniform hazard spectrum, based on a series of calculations like those in (a) and (b).

ably also PGA values) have been shown to have a joint normal distribution (Jayaram & Baker, 2008), so calculations of joint distributions of two intensity measures becomes reasonably simple in this special case. In the case of a joint normal distribution, conditional distributions of one intensity measure parameter, given the other, are also normally distributed, and can be computed using only a linear correlation coefficient between the two parameters (see Section A.3 for a few further details, and Benjamin and Cornell, 1970, for a more complete discussion).

Let us consider joint predictions of PGA and spectral acceleration at a period of 0.5 seconds ($SA(0.5s)$), given a magnitude 5 earthquake at a distance of 10 km. Abrahamson & Silva (1997) provide the following predictions for the mean and standard deviation of $\ln SA$

$$\overline{\ln SA} = -2.7207 \tag{3.12}$$

$$\sigma_{\ln SA} = 0.80 \tag{3.13}$$

Note that a few more parameters than just magnitude and distance are needed to obtain this $\ln SA$ prediction; here we have also assumed a rock site and a strike-slip mechanism for the earthquake. The median of (non-log) SA is simply the exponential of this number, which in this case is 0.065 g.

Looking back to Section 2.2.4, we recall that the mean and standard deviation of lnPGA for this event was

$$\overline{\ln PGA} = -2.2673 \tag{3.14}$$

$$\sigma_{\ln PGA} = 0.57 \tag{3.15}$$

This mean lnPGA translates to a median PGA of 0.104g.

The only thing needed further to compute the joint distribution of PGA and SA is the correlation coefficient between the two (typically referred to using the Greek letter ρ). These correlation coefficients have been determined in a manner similar to the way that ground motion prediction models are calibrated; several documents provide these coefficients (e.g., Baker & Cornell, 2006; Baker & Jayaram, 2008), and estimate a ρ of approximately 0.7 for this case.

Now let us consider a prediction of the distribution of PGA, given knowledge of the SA(0.5s) value for a ground motion coming from the specified earthquake event. Because of the joint normality of lnPGA and lnSA, we can write the conditional mean and standard deviation of lnPGA as simply

$$\overline{\ln(PGA|SA)} = \overline{\ln(PGA)} + \rho\varepsilon_{SA}\sigma_{\ln PGA} \tag{3.16}$$

$$\sigma_{\ln PGA|SA} = \sqrt{1-\rho^2}\sigma_{\ln PGA} \tag{3.17}$$

where all parameters have been defined above except ε_{SA}. That parameter is the number of standard deviations by which a given lnSA value differs from its mean predicted value. Mathematically, it can be written

$$\varepsilon_{SA} = \frac{\ln x - \overline{\ln(SA)}}{\sigma_{\ln SA}} \tag{3.18}$$

where x is the observed SA value, and the other terms are the mean and standard deviation from the original ground motion prediction model.

Now imagine that we have observed an SA value of 0.2g from the magnitude 5 earthquake at a distance of 10 km. Using equation 3.18, we find

$$\begin{aligned}\varepsilon_{SA} &= \frac{\ln x - \overline{\ln(SA)}}{\sigma_{\ln SA}} \\ &= \frac{\ln 0.2 - (-2.7207)}{0.8} \\ &= 1.4\end{aligned} \tag{3.19}$$

That is, the observed spectral acceleration is 1.4 standard deviations larger than the mean predicted value associated with this earthquake.

If SA was larger than its mean, and SA and PGA are correlated, then knowledge of this large SA value should increase our predictions of PGA for the given ground motion. We make this increased prediction using equation 3.16

$$\begin{aligned}\overline{\ln(PGA|SA)} &= \overline{\ln(PGA)} + \rho\varepsilon_{SA}\sigma_{\ln PGA} \\ &= -2.2673 + 0.7(1.4)(0.57) \\ &= -1.7124 \end{aligned} \tag{3.20}$$

Taking an exponential of this number tells us that the median conditional PGA is 0.18g (a significant increase from the median prediction of 0.104g we made before we had observed SA=0.2g).

Knowledge of SA should also decrease our uncertainty in PGA, and this is reflected in equation 3.17

$$\begin{aligned}\sigma_{\ln PGA|SA} &= \sqrt{1-\rho^2}\sigma_{\ln PGA} \\ &= \sqrt{1-0.7^2}(0.57) \\ &= 0.41 \end{aligned} \tag{3.21}$$

Using the updated conditional mean and standard deviation of PGA, we can now predict the probability of exceeding different PGA values conditional upon $SA(0.5s) = 0.2g$, by using equation 2.15 with our updated conditional mean and standard deviation. Some sample results are summarized in Table 2.3. The first column lists a series of PGA values of potential interest. The second column lists the probability of exceeding those PGA values, given a magnitude 5 earthquake at a distance of 10 km, but *not yet conditioned on any observed SA value*. That is, the second column was computed using the original mean and standard deviation from equations 3.14 and 3.15. Note that this calculation is identical to the one from Table 1.2. In the third column of Table 2.3, we compute probabilities of exceeding the same PGA values, but this time conditioned upon knowledge that $SA(0.5s) = 0.2g$. That is, we evaluate equation 2.15 with our new conditional mean and standard deviation. Examining the second and third columns of this table, two interesting features are apparent. First, the probability of exceeding low PGA values has increased significantly, because we now know that the correlated parameter $SA(0.5s)$ is larger than usual for this event. Second, we see that the probability of exceeding very large PGA values has actually decreased. The decrease is because knowledge of SA has reduced our uncertainty in PGA. Although we know that SA is larger than its mean prediction, we have also eliminated the possibility that SA is even more extreme than the observed value, so the most extreme PGA values actually become less likely. Finally, in the fourth column

of the table, we compute the conditional probability of PGA equaling the various values of interest, using equation 2.21

x_j	$P(PGA > x_j)$	$P(PGA > x_j \mid SA = 0.2)$	$P(PGA = x_j \mid SA = 0.2)$
0.05	0.8994	0.9992	0.0727
0.1	0.5247	0.9265	0.5263
0.2	0.1242	0.4002	0.2943
0.3	0.0311	0.1058	0.0806
0.4	0.0089	0.0253	0.0191
0.5	0.0029	0.0061	0.0046
0.6	0.0010	0.0016	0.0011
0.7	0.0004	0.0004	0.0003
0.8	0.0002	0.0001	0.0001
0.9	0.0001	0.0000	0.0000

Table 3.3: PGA probabilities associated with a magnitude 5 earthquake at 10 km, and an SA(0.5s) value of 0.2g.

To aid in intuitive understanding of these calculations, Figure 3.7 shows a schematic illustration of the joint distribution referred to above. The horizontal axes represent the range of (log) PGA and SA values that might result from earthquakes with a given magnitude and distance. The contour lines illustrate the contours of the joint distribution of PGA and SA. The centroid and spread of these contours with respect to each horizontal axis are specified by the mean and standard deviation from ground motion prediction models. The correlation between the two parameters is reflected by the elliptical shape of the contours, which means that large values of lnSA are likely to be associated with large values of lnPGA. What we are interested in here is the distribution of PGA, given that we have observed some SA value. This is represented by cuts through the joint distribution. The conditional distributions at two cuts (i.e., two potential SA values) are shown on the vertical axis of the figure. The probability of exceeding some PGA value x_1 is represented by the area of the conditional distribution to the right of x_1. We see from the two cuts that as the observed lnSA value gets larger, the probability of exceeding x_1 also gets larger. That is the effect we also saw in the third column of Table 3.3.

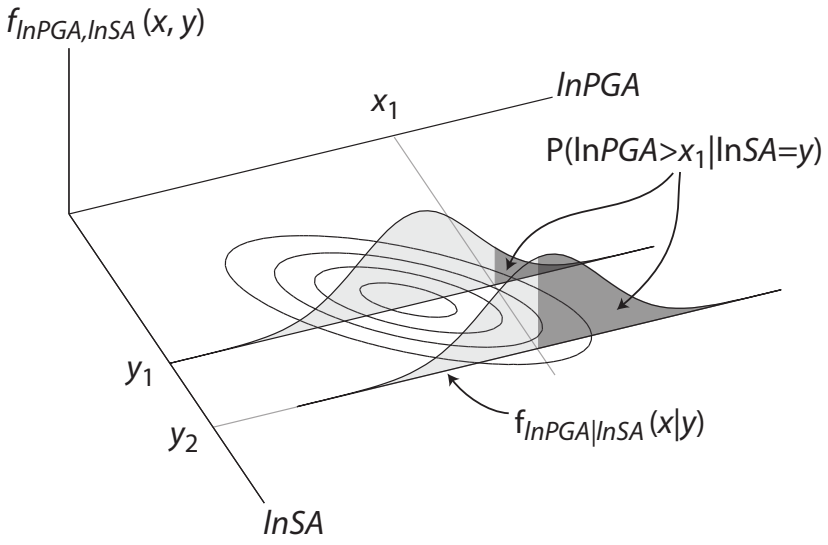


Figure 3.7: Schematic illustration of the joint distribution of *PGA* and *SA*.

4
Conclusions

We have now completed an overview of probabilistic seismic hazard analysis (PSHA) and several extensions of the basic methodology. Example calculations have been presented to illustrate how the computations are performed in practice. With these tools, one can quantify the risk of ground motion shaking at a site, given knowledge about seismic sources surrounding the site.

Having now considered the many sources of uncertainty present when predicting future shaking at a site, it is hopefully clear to the reader why deterministic approaches to seismic hazard analysis can be unsatisfying. It should be clear that there is no such thing as a deterministic "worst-case" ground motion, and that attempts to identify an alternate deterministic ground motion necessitate making decisions that may be arbitrary and hard to justify.

PSHA is fundamentally an accounting method that lets one combine diverse sources of data regarding occurrence rates of earthquakes, the size of future earthquakes, and propagation of seismic shaking through the earth. It would be impossible to model the distribution of future earthquake shaking at a site through direct observation, because one would have to wait thousands or millions of years to collect enough observations to make a reasonable inference regarding rare ground motions. But, by incorporating many sources of data into the calculations, it becomes possible to project out to these low probabilities with scientifically-defensible and reproducible models.

The basic PSHA calculation, and its required inputs, was discussed in Section 2. In Section 3, several extensions were presented, such as deaggregation and uniform hazard spectra. There is also a vast literature regarding the accurate estimation of the many inputs, such as occurrence rates of earthquakes and their magnitude distributions, which was not discussed here. References for further study on these topics are provided in Appendix B for the interested reader.

A
Review of probability

Probability is so fundamental to probabilistic seismic hazard analysis that the word appears in its title. The calculations in this document thus rely heavily on the use of probability concepts and notation. These probabilistic tools allow us to move through calculations without having to stop and derive intermediate steps. The notational conventions allow us to easily describe the behavior of uncertain quantities. It is recognized that these concepts and notations are not familiar to all readers, however, so this section is intended to provide a brief overview of the needed material. Readers desiring more details may benefit from reviewing a textbook dedicated specifically to practical applications of probability concepts (e.g. Ang & Tang, 2007; Benjamin & Cornell, 1970; Ross, 2004).

A.1 Random events

The most basic building block of probability calculations is the *random event*: an event having more than one possible outcome. The *sample space* (denoted S) is the collection of all possible outcomes of a random event. Any subset of the sample space is called an *event*, and denoted E. Sample spaces and events are often illustrated graphically using Venn diagrams, as illustrated in Figure A.1.

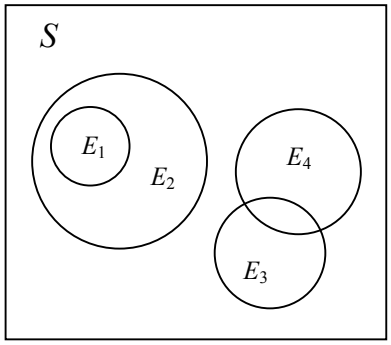


Figure A.1: Venn diagram illustrating a sample space and events.

For example, the number obtained from rolling of a die is a random event. The sample space for this example is $S = \{1, 2, 3, 4, 5, 6\}$. The outcomes in the event that the number is odd are $E_1 = \{1, 3, 5\}$. The outcomes in the event that the number is greater than three are $E_2 = \{4, 5, 6\}$

We are commonly interested in two operations on events. The first is the *union* of E_1 and E_2, denoted by the symbol $\cup$. $E_1 \cup E_2$ is the event that contains all outcomes in either E_1 or E_2. The second is the intersection. $E_1 E_2$ (also denoted $E_1 \cap E_2$) is the event that contains all outcomes in both E_1 and E_2. For example, continuing the die illustration from above, $E_1 \cup E_2 = \{1, 3, 4, 5, 6\}$ and $E_1 \cap E_2 = \{5\}$.

Special events

There are a few special terms and special events that are often useful for probability calculations:

The *certain event* is an event that contains all possible outcomes in the sample space. The sample space S is the certain event.

The *null event*, denoted φ, is an event that contains no outcomes.

Events E_1 and E_2 are *mutually exclusive* when they have no common outcomes. $E_1 E_2 \quad = \quad \varphi$if E_1 and E_2 are mutually exclusive.

Events E_1, $E_2 \ldots E_n$ are *collectively exhaustive* when their union contains every possible outcome of the random event (i.e., $E_1 \cup E_2 \cup \cdots \cup E_n = S$).

The *complementary event*, $\overline{E_1}$, of an event E_1, contains all outcomes in the sample space that are not in event E_1. It should be clear that, by this definition, $\overline{E_1} \cup E_1 = S$ and $\overline{E_1} E_1 = \phi$. That is, $\overline{E_1}$ and E_1 are mutually exclusive and collectively exhaustive.

Axioms of probability

We will be interested in the probabilities of occurrence of various events. These probabilities must follow three axioms of probability:

$$0 \leq P(E) \leq 1 \tag{A.1}$$

$$P(S) = 1 \tag{A.2}$$

$P(E_1 \cup E_2) = P(E_1) + P(E_2)$, for mutually exclusive events E_1 and E_2(56)

These axioms form the building blocks of all other probability calculations. It is very easy to derive additional laws using these axioms, and the previously-defined events. For example,

$$P(\bar{E}) = 1 - P(E) \tag{A.3}$$

$$P(\phi) = 0 \tag{A.4}$$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 E_2) \tag{A.5}$$

A.2 Conditional probabilities

The probability of the event E_1 may depend upon the occurrence of another event E_2. The conditional probability $P(E_1 \mid E_2)$ is defined as the probability that event E_1 has occurred, given that event E_2 has occurred. That is, we are computing the probability of E_1, if we restrict our sample space to only those outcomes in event E_2. Figure A.2 may be useful as the reader thinks about this concept.

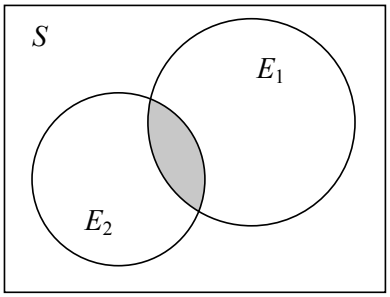


Figure A.2: Schematic illustration of the events E_1 and E_2. The shaded region depicts the area corresponding to event $E_1 E_2$.

We can deduce the following from Figure A.2

$$P(E_1|E_2) = \begin{cases} \frac{P(E_1E_2)}{P(E_2)} & \text{if } P(E_2) > 0 \\ 0 & \text{if } P(E_2) = 0 \end{cases} \tag{A.6}$$

Rearranging this equation gives

$$P(E_1E_2) = P(E_1|E_2)P(E_2) \tag{A.7}$$

Independence

Conditional probabilities give rise to the concept of independence. We say that two events are *independent* if they are not related probabilistically in any way. More precisely, we say that events E_1 and E_2 are independent if

$$P(E_1|E_2) = P(E_1) \tag{A.8}$$

That is, the probability of E_1 is not in any way affected by knowledge of the occurrence of E_2. Substituting equation 62 into equation 61 gives

$$P(E_1E_2) = P(E_1)P(E_2) \tag{A.9}$$

which is an equivalent way of stating that E_1 and E_2 are independent. Note that equations 62 and 63 are true if and only if E_1 and E_2 are independent.

Total Probability Theorem

Consider an event A and a set of mutually exclusive, collectively exhaustive events E_1, $E_2 \ldots E_n$. The total probability theorem states that

$$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i) \tag{A.10}$$

In words, this tells us that we can compute the probability of Aif we know the probability of the E_i's, and know the probability of A, given each of these E_i's. The schematic illustration in Figure A.3 may help to understand what is being computed. At first glance, the utility of this calculation may not be obvious, but it is critical to many engineering calculations where the probability of A is difficult to determine directly, but where the problem can be broken down into several pieces whose probabilities can be computed.

Consider the following example, to illustrate the value of this calculation. You have been asked to compute the probability that Building X collapses during the next earthquake in the region. You do not know with certainty if the next earthquake will be strong, medium or weak, but seismologists have estimated the following probabilities:

$$\begin{aligned} P(strong) &= 0.01 \\ P(medium) &= 0.1 \\ P(weak) &= 0.89 \end{aligned} \tag{A.11}$$

Additionally, structural engineers have performed analyses and estimated the following:

$$\begin{aligned} P(collapse|strong) &= 0.9 \\ P(collapse|medium) &= 0.2 \\ P(collapse|weak) &= 0.01 \end{aligned} \tag{A.12}$$

Referring to equation 64, the "A" in that equation is the event that the building collapses, and the E_i's are the events that the earthquake is strong, medium or weak. We can therefore compute the probability of collapse as

$$\begin{aligned} P(collapse) &= P(collapse|strong)P(strong) \\ &\quad + P(collapse|medium)P(medium) \\ &\quad + P(collapse|weak)P(weak) \\ &= 0.9(0.01) + 0.2(0.1) + 0.01(0.89) \\ &= 0.0379 \end{aligned} \tag{A.13}$$

The total probability theorem allows one to break the problem into two parts (size of the earthquake and capacity of the building), compute probabilities for those parts, and then re-combine them to answer the original question. This not only facilitates solution of the problem in pieces, but it allows different specialists (e.g., seismologists and engineers) to work on different aspects of the problem.

Probabilistic seismic hazard analysis is a direct application of the total probability theorem (except that it uses random variables, discussed below, instead of random events). The distribution of earthquake magnitudes and distances are studied independently of the conditional distribution of ground motion intensity, and this probabilistic framework allows us to re-combine the various sources of information in a rigorous manner.

Bayes' Rule

Consider an event A and a set of mutually exclusive, collectively exhaustive events E_1, $E_2 \ldots E_n$. From our conditional probability equations above, we can write

$$P(AE_j) = P(E_j|A)P(A) = P(A|E_j)P(E_j) \tag{A.14}$$

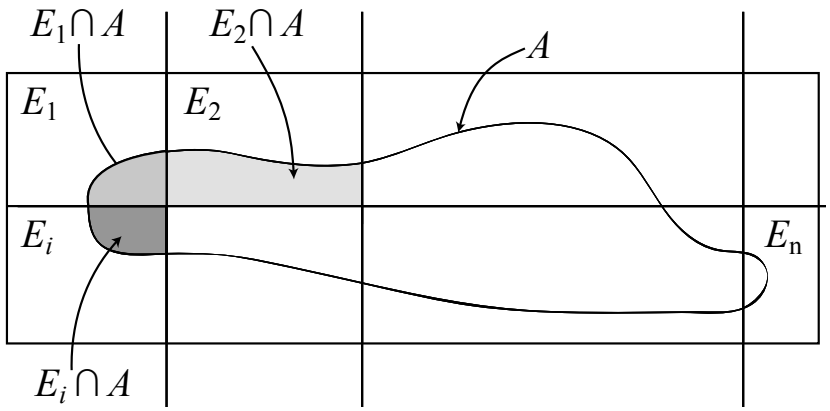


Figure A.3: Schematic illustration of the total probability theorem.

Rearranging the last two terms gives

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{P(A)} \tag{A.15}$$

This equation is known as Bayes' Rule. An alternate form is based on substituting equation 64 for the total probability theorem in place of $P(A)$ in the denominator of equation 67.

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{\sum_{i=1}^{n} P(A|E_i)P(E_i)} \tag{A.16}$$

The utility of these equations lies in their ability to compute conditional probabilities, when you only know probabilities related to conditioning in the reverse order of what is desired. That is, you would like to compute $P(A \mid B)$ but only know $P(B \mid A)$. This is exactly the type of calculation used in the deaggregation calculations of Section 3.1.

To provide a simple illustration of how this equation works, consider again the example used to illustrate the total probability theorem. Suppose you have just learned that an earthquake occurred and building X collapsed. You don't yet know the size of the earthquake, and would like to compute the probability that it was a strong earthquake. Using equation 67, you can write

$$P(strong|collapse) = \frac{P(collapse|strong)P(strong)}{P(collapse)} \tag{A.17}$$

Substituting the numbers from above, you then find that

$$P(strong|collapse) = \frac{0.9(0.01)}{0.0379} = 0.24 \tag{A.18}$$

It is not obvious intuitively how large that probability would be, because strong earthquakes have a high probability of causing collapse,

but they are also extremely unlikely to occur. Like the Total Probability Theorem, Bayes' Rule provides a valuable calculation approach for combining pieces of information to compute a probability that may be difficult to determine directly.

A.3 Random variables

Here we will introduce an important concept and an associated important notation. A random variable is a numerical variable whose specific value cannot be predicted with certainty before the occurrence of an "event" (in our context, this might be the magnitude of a future earthquake). Examples of random variables relevant to PSHA are the time to the next earthquake in a region, the magnitude of a future earthquake, the distance form a future earthquake to a site, ground shaking intensity at a site, etc.

We need a notation to refer to both the random variable itself, and to refer to specific numerical values which that random variable might take. Standard convention is to denote a random variable with an uppercase letter, and denote the values it can take on by the same letter in lowercase. That is, x_1, x_2, x_3, ... denote possible numerical outcomes of the random variable X. We can then talk about probabilities of the random variable taking those outcomes (i.e., $P(X = x_1)$).

Discrete and continuous random variables

We can in general treat all random variables using the same tools, with the exception of distinguishing between discrete and continuous random variables. If the number of values a random variable can take are countable, the random variable is called *discrete*. An example of a discrete random variable is the number of earthquakes occurring in a region in some specified period of time. The probability distribution for a discrete random variable can be quantified by a probability mass function (PMF), defined as

$$p_X(x) = P(X = x) \tag{A.19}$$

The cumulative distribution function (CDF) is defined as the probability of the event that the random variable takes a value less than or equal to the value of the argument

$$F_X(x) = P(X \leq x) \tag{A.20}$$

The probability mass function and cumulative distribution function have a one-to-one relationship

$$F_X(a) = \sum_{\text{all } x_i \leq a} p_X(x_i) \tag{A.21}$$

Examples of the PMF and CDF of a discrete random variable are shown in Figure A.4.

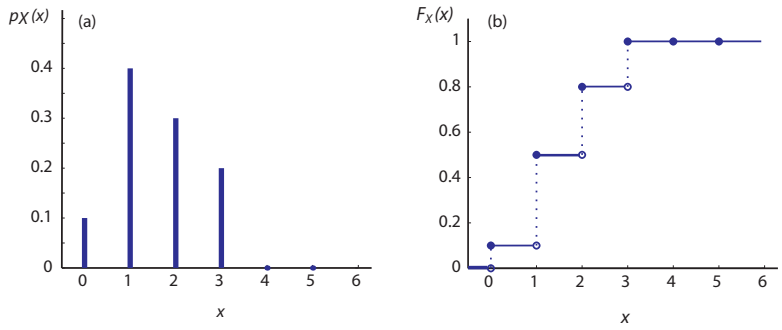


Figure A.4: Example descriptions of a discrete random variable. (a) Probability mass function. (b) Cumulative distribution function.

In many cases we are interested in the probability of $X > x$, rather than the $X \leq x$ addressed by the CDF. But noting that those two outcomes are mutually exclusive and collectively exhaustive events, we can use the previous axioms of probability to see that $P(X > x) = 1 - P(X \leq x)$.

In contrast to discrete random variables, continuous random variables can take any value on the real axis (although they don't have to). Because there are an infinite number of possible realizations, the probability that a continuous random variable X will take on any single value x is zero. This forces us to use an alternate approach for calculating probabilities. We define the *probability density function* (PDF) using the following

$$f_X(x)\,dx = P(x < X \leq x + dx) \tag{A.22}$$

where dx is a differential element of infinitesimal length. An illustration of the PDF and related probability calculation is given in Figure A.5. We can compute the probability that the outcome of X is between a and b by "summing" (integrating) the probability density function over the interval of interest

$$P(a < X \leq b) = \int_a^b f_X(x)\,dx \tag{A.23}$$

Note that in many of the PSHA equations above, we have approximated continuous random variables by discrete random variables, for ease of numerical integration. In those cases, we have replaced the infinitesimal dx by a finite Δx, so that equation A.23 becomes:

$$p_{\widetilde{X}}(x) = f_X(x)\,\Delta x = P(x < X \leq x + \Delta x) \tag{A.24}$$

where $p_{\widetilde{X}}(x)$ is the probability mass function for $\widetilde{X}$, the discretized version of the continuous random variable X. Reference to Figure A.5

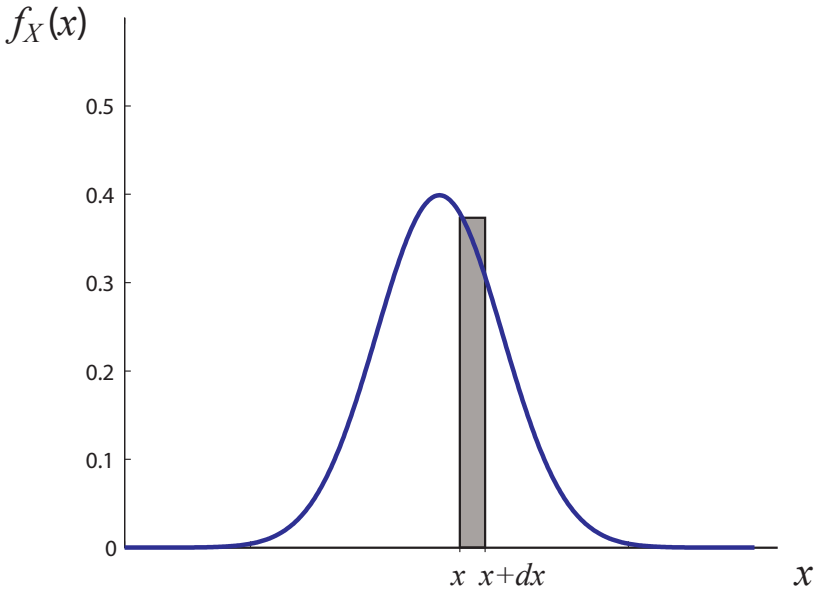


Figure A.5: Plot of a continuous probability density function. The area of the shaded rectangle, $f_X(x)\,dx$, represents the probability of the random variable X taking values between x and $x + dx$.

should help the reader understand that the probabilities of any outcome between x and $x + \Delta x$ have been assigned to the discrete value x.

Another way to describe a continuous random variable is with a *cumulative distribution function* (CDF)

$$F_X(x) = P(X \leq x) \tag{A.25}$$

The PDF and CDF are related by the following

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u)\,du \tag{A.26}$$

$$f_X(x) = \frac{d}{dx} F_X(x) \tag{A.27}$$

Note that the CDF of continuous and discrete random variables has the same definition. This is because probabilities of outcomes within an interval are identically defined for discrete and continuous outcomes. Plots of continuous cumulative distribution functions are seen in the body of the document (e.g., Figure 2.8b and Figure 2.10b).

Comments on notion

This PMF/PDF/CDF notation allows us to compactly and precisely describe probabilities of outcomes of random variables. Note that the following conventions have been used:

1. The initial letter indicates the type of probability being described (i.e., "p" for PMFs, "f" for PDFs, and "F" for CDFs).

2. The subscript denotes the random variable (e.g., "X"), and thus is always a capital letter.

3. The argument in parentheses indicates the numerical value being considered (e.g., "x"), and is thus either a lower-case letter or a numeric value (e.g., $F_X(2.2) = P(X \leq 2)$).

It is worth noting that these conventions are not chosen arbitrarily here or unique to PSHA. They are used almost universally in all probability papers and books, regardless of the field of application.

Conditional distributions

We are often interested in conditional probability distributions of random variables. We can adopt all of the results from Section A.2 if we recognize that the random variable X exceeding some value x is an event. So we can adapt equation A.7, for example, to write

$$\begin{aligned} f_{X|Y}(x|y)dx &= P(x < X \leq x + dx | y < Y \leq y + dy) \\ &= \frac{P(x < X \leq x + dx \cap y < Y \leq y + dy)}{P(y < Y \leq y + dy)} \end{aligned} \tag{A.28}$$

where the notation $f_{X|Y}(x|y)$ is introduced to denote the conditional probability density function of X, given that the random variable Y has taken value y. If we further introduce the following notation for the joint probability density function of X and Y

$$f_{X,Y}(x,y)\,dx\,dy = P(x < X \leq x + dx \cap y < Y \leq y + dy) \tag{A.29}$$

then equationA.29 becomes

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \tag{A.30}$$

Similarly, equation A.10 can be used to show that random variables X and Y are said to be independent if and only if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \tag{A.31}$$

Another example is the PSHA equations above that use integrals over the random variables for magnitude and distance; these are the random-variable analog of the total probability theorem introduced earlier for events.

These types of manipulations, which are only briefly introduced here, are very useful for computing probabilities of outcomes of random variables, conditional upon knowledge of other probabilistically-dependent random variables.

The normal distribution

One particular type of random variable plays an important role in the calculations above, so we will treat it carefully here. A random variable is said to have a "normal" (or "Gaussian") distribution if it has the following PDF

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \tag{A.32}$$

where μ_x and σ_x denote the mean value and standard deviation, respectively, of X. This PDF forms the familiar "bell curve" seen above in Figure A.5. This is one of the most common distributions in engineering, and has been found to describe very accurately the distribution of the logarithm of ground motion intensity associated with a given earthquake magnitude and distance. Because of that, we often want to find the probability that a normally-distributed random variable X takes a value less than x. From above, we know that we can find this probability by integrating the PDF over the region of interest

$$\begin{aligned} P(X \leq x) &= \int_{-\infty}^{x} f_X(u)\, du \\ &= \int_{-\infty}^{x} \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{u-\mu_X}{\sigma_X}\right)^2\right) du \end{aligned} \tag{A.33}$$

Unfortunately, there is no analytic solution to this integral. But because it is so commonly used, we tabulate its values, as shown in Table A.1. To keep this table reasonably small in size, we use two tricks. First, we summarize values for the "standard" normal distribution, where standard denotes that the random variable has a mean value (μ_x) of 0 and a standard deviation (σ_x) of 1. So the CDF becomes

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du \tag{A.34}$$

Because the CDF of the standard normal random variable is so common, we give it the special notation $P(X \leq x) = \Phi(x)$.

If the random variable of interest, X, is normal but not standard normal, then we can transform it into a standard normal random variable as follows

$$U = \frac{X-\mu_X}{\sigma_X} \tag{A.35}$$

where U is a standard normal random variable. This means that we can use the right-hand side of equation 87 as an argument for the

standard normal CDF table. That is

$$P(X \leq x) = \Phi\left(\frac{X - \mu_X}{\sigma_X}\right) \tag{A.36}$$

where Table A.1 provides values of $\Phi(\ )$.

A second trick used to manage the size of the standard normal CDF table is to note that the normal PDF is symmetric about zero. This means that

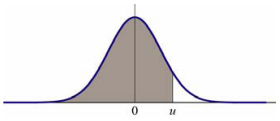
$$\Phi(-u) = 1 - \Phi(u) \tag{A.37}$$

so the CDF value a negative number can be determined from the CDF value for the corresponding positive number. Thus, the table is tabulated for only positive values of u. The identity of equation 89 might be intuitively clear if one views the standard normal PDF illustrated at the top of Table A.1.

Table A.1: Standard normal cumulative distribution function.

$$\Phi(u) = F_U(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-1/2 x^2}\, dx$$

$$\Phi(-u) = 1 - \Phi(u)$$



u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3	0.998650									
3.5	0.999767									
4	0.999968									
4.5	0.9999966									
5	0.9999997									

The normal distribution can be generalized to the case of more

than one random variable. Two random variables are said to have a joint normal distribution if they have the following joint PDF

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{z}{2(1-\rho^2)}\right\} \tag{A.38}$$

where ρ is the correlation coefficient between X and Y, and

$$z = \frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \tag{A.39}$$

A plot of this joint PDF is shown in Figure A.6.

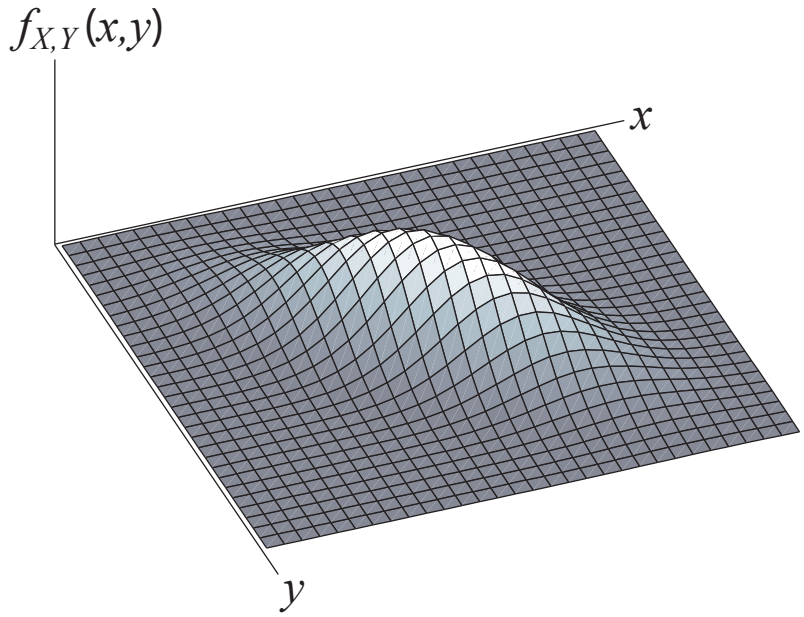


Figure A.6: Joint normal probability density function.

An important property of random variables having this distribution is that if X and Y are joint normal, then their marginal distributions ($f_X(x)$ and $f_Y(y)$) are normal, and their conditional distributions are also normal. Specifically, the distribution of X given $Y = y$ is normal, with conditional mean

$$\mu_{X|Y=y} = \mu_X + \rho \cdot \sigma_X \left(\frac{y-\mu_Y}{\sigma_Y}\right) \tag{A.40}$$

and conditional standard deviation

$$\sigma_{X|Y=y} = \sigma_X\sqrt{1-\rho^2} \tag{A.41}$$

These properties are convenient when computing joint distributions of ground motion parameters.

A.4 Expectations and moments

A random variable is completely defined by its PMF or PDF (for discrete and continuous random variables, respectively). Sometimes, however, it is convenient to use measures that describe general features of the distribution, such as its "average" value, breadth of feasible values, and whether the distribution has a heavier tail to the left or right. We can measure these properties using moments of a random variable, and they are often more convenient to work with for engineering applications.

The *mean* is the most common moment, and is used to describe the central location of a random variable. The mean of X is denoted μ_X or $E[X]$. It can be calculated for a discrete random variable as

$$\mu_X = \sum_{\text{all } i} x_i \, p_X(x_i) \tag{A.42}$$

and for a continuous random variable as

$$\mu_X = \int_{\text{all } x} x \, f_X(x) \, dx \tag{A.43}$$

Note that this is equal to the center of gravity of the PMF or PDF. The equations may be recognizable to some readers as being very similar to centroid calculations.

The variation of values to be expected from a random variable can be measured using the *variance*, denoted σ_X^2 or $Var[X]$. It is calculated for a discrete random variable as

$$\sigma_X^2 = \sum_{\text{all } i} (x_i - \mu_x)^2 \, p_X(x_i) \tag{A.44}$$

and for a continuous random variable as

$$\sigma_X^2 = \int_{\text{all } x} (x - \mu_x)^2 \, f_X(x) \, dx \tag{A.45}$$

This the moment of inertia of the PDF (or PMF) about the mean.

The square root of the variance is known as the *standard deviation*, and is denoted σ_X. It is often preferred to the variance when reporting a description of a random variable, because it has the same units as the random variable itself (unlike the variance, whose units are the original units squared).

Means and variances are special cases of the expectation operation. The expectation of $g(X)$ is defined for discrete random variables as

$$E[g(X)] = \sum_{\text{all } i} g(x_i) \, p_X(x_i) \tag{A.46}$$

and for continuous random variables as

$$E[g(X)] = \int_{\text{all } x} g(x)\, f_X(x)\, dx \tag{A.47}$$

The mean value is the special case of expectation where $g(X) = X$, and the variance is the special case where $g(X) = (X - \mu_X)^2$. These are special cases of what are called *moments* of random variables, but we will restrict the discussion here to those two special cases.

Finally, note that the normal random variable described above uses the mean and standard deviation explicitly as parameters in its PDF. So given knowledge of the mean and standard deviation of a normal random variable, one knows its complete distribution. This is not the case for random variables in general, but it is one of the reasons why the normal random variable is convenient to work with.

B
Further study

Below is a list of important papers and summary books that would be valuable for those interested in further study. References are grouped by type, with a short description of their scope.

Origins and development of PSHA

- Cornell (1968) wrote the seminal document describing the concept of PSHA.
- McGuire (2007) wrote history and summary of the early development of PSHA.

Books and summary papers

- Kramer (1996) is a geotechnical-engineering focused book, with chapter 4 devoted exclusively to PSHA. This chapter also contains many references regarding the development of PSHA, and estimation of the parameters needed as inputs to the calculations.
- McGuire (2004) is a monograph focused on probabilistic estimation of losses from earthquakes, with a significant portion devoted to PSHA. The monograph provides more information on practical estimation of the needed input parameters, and discusses several advanced topics omitted from this report.
- Reiter (1990) is another book describing PSHA. The authors insights are particularly focused on nuclear applications, though many of the concepts are more general. This book also makes a significant effort to compare deterministic and probabilistic seismic hazard analysis methods.
- Abrahamson (2006) is a fairly recent summary of current challenges and opportunities in PSHA.

Bibliography

Abrahamson, N A. 2006. Seismic hazard assessment: problems with current practice and future developments. *Page 17 of: First European Conference on Earthquake Engineering and Seismology*.

Abrahamson, N. A., & Silva, W.J. 1997. Empirical response spectral attenuation relations for shallow crustal earthquakes. *Seismological Research Letters*, **68**(1), 94–127.

Anagnos, T., & Kiremidjian, A. S. 1984. Stochastic time-predictable model for earthquake occurrences. *Bulletin of the Seismological Society of America*, **74**(6), 2593–2611.

Ang, Alfredo Hua-Sing, & Tang, Wilson H. 2007. *Probability concepts in engineering emphasis on applications in civil & environmental engineering*. New York: Wiley.

Baker, Jack W, & Cornell, C Allin. 2006. Correlation of response spectral values for multi-component ground motions. *Bulletin of the Seismological Society of America*, **96**(1), 215–227.

Baker, Jack W, & Jayaram, Nirmal. 2008. Correlation of spectral acceleration values from NGA ground motion models. *Earthquake Spectra*, **24**(1), 299–317.

Bazzurro, Paolo, & Cornell, C. Allin. 1999. Disaggregation of Seismic Hazard. *Bulletin of the Seismological Society of America*, **89**(2), 501–520.

Benjamin, Jack R., & Cornell, Carl Allin. 1970. *Probability, Statistics, and Decision for Civil Engineers*. New York: McGraw-Hill.

Bommer, Julian J., & Abrahamson, Norman A. 2006. Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates? *Bulletin of the Seismological Society of America*, **96**(6), 1967–1977.

Campbell, Kenneth W., & Bozorgnia, Yousef. 2008. NGA Ground Motion Model for the Geometric Mean Horizontal Component of

PGA, PGV, PGD and 5% Damped Linear Elastic Response Spectra for Periods Ranging from 0.01 to 10 s. *Earthquake Spectra*, **24**(1), 139–171.

Commission, Nuclear Regulatory. 1997. *Identification and Characterization of Seismic Sources and Determination of Safe Shutdown Earthquake Ground Motion*. Regulatory Guide 1.165. Nuclear Regulatory Commission.

Cornell, C. A., & Winterstein, Steven R. 1988. Temporal and magnitude dependence in earthquake recurrence models. *Bulletin of the Seismological Society of America*, **78**(4), 1522–1537.

Cornell, C Allin. 1968. Engineering Seismic Risk Analysis. *Bulletin of the Seismological Society of America*, **58**(5), 1583–1606.

Cornell, C. Allin, Banon, Hooshang, & Shakal, Anthony F. 1979. Seismic motion and response prediction alternatives. *Earthquake Engineering & Structural Dynamics*, **7**(4), 295–315.

Gutenberg, B., & Richter, C. F. 1944. Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, **34**(4), 185–188.

Jayaram, Nirmal, & Baker, Jack W. 2008. Statistical Tests of the Joint Distribution of Spectral Acceleration Values. *Bulletin of the Seismological Society of America*, **98**(5), 2231–2243.

Kramer, Steven Lawrence. 1996. *Geotechnical earthquake engineering*. Upper Saddle River, N.J.: Prentice Hall.

Lomnitz-Adler, Jorge, & Lomnitz, Cinna. 1979. A modified form of the Gutenberg-Richter magnitude-frequency relation. *Bulletin of the Seismological Society of America*, **69**(4), 1209–1214.

McGuire, Robin K. 1995. Probabilistic Seismic Hazard Analysis and Design Earthquakes: Closing the Loop. *Bulletin of the Seismological Society of America*, **85**(5), 1275–1284.

McGuire, Robin K. 2004. *Seismic Hazard and Risk Analysis*. Berkeley: Earthquake Engineering Research Institute.

McGuire, Robin K. 2007. Probabilistic seismic hazard analysis: Early history. *Earthquake Engineering & Structural Dynamics*, **37**(3), 329 – 338.

Reiter, L. 1990. *Earthquake hazard analysis: Issues and insights*. New York: Columbia University Press.

Ross, Sheldon M. 2004. *Introduction to probability and statistics for engineers and scientists*. Vol. 3rd. Amsterdam ; Boston: Elsevier Academic Press.

Schwartz, D. P., & Coppersmith, Kevin J. 1984. Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas fault zones. *Journal of geophysical research*, **89**(B7), 5681–5698.

USGS. 2008. *Interactive Deaggregation Tools*.

Weichert, Dieter H. 1980. Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin of the Seismological Society of America*, **70**(4), 1337–1346.

Youngs, R. R., & Coppersmith, K.J. 1985. Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard analysis. *Bulletin of the Seismological Society of America*, **75**(4), 939–964.