10.34 Numerical Methods Applied to Chemical Eng.

Lectures 28 & 30: Probability Theory

1 Experiments, Events, and Probabilities

1.1 Experiments

The notion of probability is defined with respect to an underlying *experiment*. An experiment can be, for example, flipping a coin or measuring the temperature in a chemical reactor. An experiment is defined by the *set of outcomes*, E. If the experiment is flipping a coin, then $E = \{\text{heads}, \text{tails}\}$. If the experiment is measuring the temperature in a reactor T in degrees K, then E is the set of all nonnegative real numbers, $E = \mathbb{R}_+$.

A single performance of an experiment is called a *trial*. In each trial, a single outcome $\xi \in E$ is observed.

1.2 Events

A subset of $A \subset E$ is called an *event*. For example, in the coin-flipping experiment, {heads} is an event. In the temperature measurement experiment, $\{100\}$ is an event interpreted as '*the measured temperature is 100 K*'. The interval $[270, 290]$ is an event interpreted as '*the measured temperature is in the interval* $[270, 290]$ K'.

Consider a single trial in which the outcome $\xi \in E$ is observed. An event A is said to have occurred during this trial if ξ is an element of A, $\xi \in A$. For example, if the measurement of the temperature of the reactor was $\xi = 285$, then the event $A = [270, 290]$ = '*the measured temperature is in the interval* $[270, 290]$ K' occurred, but the event $B = \{100\}$ = '*the measured temperature is 100 K*' did not occur.

The probabilities of events are of interest (see Section 1.4). For example, $\Pr([270, 290])$ is the probability that, in a single trial, the value of the temperature is between 270 and 290.

1.3 Joint Events

Often of interest are events that are combinations of other events. For example, an engineer may want to know the probability that the temperature of a polymerization reactor is within its normal operating range, such as the probability that the reactor temperature is greater than 300 K *and* less than 400 K. In this case, there are two events, $A = (300, +\infty)$ and $B = [0, 400)$, and of interest is the probability that *both* events occur in single trial. This probability is denoted by $\Pr(A \text{ and } B)$, where $(A \text{ and } B)$ is the event $(300, +\infty) \cap [0, 400) = (300, 400)$. This operation is also called the *intersection* of A and B.

Though not the case in this example, in general it is possible for the event $(A \text{ and } B)$ to be the empty set. This happens if there are no outcomes that are in both events A and B, in which case A and B are said to be *mutually exclusive* events.

Another type of joint event is (A or B), which occurs if A occurs, B occurs, or both occur. In the example above, (A or B) $= (300, +\infty) \cup [0, 400) = [0, +\infty) = E$. Of course, it is not always the case that (A or B) $= E$.

The operation (A or B) is also called the *union* of A and B.

1.4 Probabilities

The standard physical interpretation of the probability of an event is: if the reactor temperature is measured N times and the number of times the temperature is observed to fall between 270 and 290 is $N_{[270,290]}$, then

$$\Pr([270, 290]) = \lim_{N \to \infty} \frac{N_{[270,290]}}{N}.$$

The *probability* obeys the three basic rules:

1. For any event A, the probability of event A satisfies $\Pr(A) \geq 0$.
2. The probability of the event consisting of all possible outcomes is one, $\Pr(E) = 1$.
3. If the events A and B are mutually exclusive, then $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)$.

Every other result in probability theory, intuitive or otherwise, can be derived from these three rules without any further assumptions. One fundamental result that follows from these properties is: For any events A and B,

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B).$$

1.5 Conditional Probabilities

For two events A and B, the *conditional probability of A given B* is the probability that the event A will occur, given that event B is known to occur or to have occurred. This conditional probability is related to the intersection of the two sets by

$$\Pr(A|B) \equiv \frac{\Pr(B \text{ and } A)}{\Pr(B)}.$$

As an example, suppose that event A is '*the measured temperature is 300*' and event B is '*the measured temperature is less than 400.*' Then $\Pr(A|B)$ is the probability that the temperature is 300, given that the temperature is less than 400, which could be computed from the probability of the intersection of A and B and the probability of B (if known). Sometimes the conditional probability can be computed without knowing the exact value of the probabilities of either event A or B. For example, $\Pr(B|A)$ is the probability that the temperature is less than 400, given that the temperature is 300, which is equal to 1 as seen by applying the above formula:

$$\Pr(A \text{ and } B) = \Pr(\{300\} \cap [0, 400]) = \Pr(\{300\}) = \Pr(A).$$

Then

$$\Pr(B|A) = \frac{\Pr(A \text{ and } B)}{\Pr(A)} = \frac{\Pr(A)}{\Pr(A)} = 1.$$

This conditional probability was computable without knowing the value of the probability of event A.

Rearranging the expression for the conditional probability gives that

$$\Pr(B \text{ and } A) = \Pr(B)\Pr(A|B).$$

The events A and B are said to be *independent* if

$$\Pr(B \text{ and } A) = \Pr(B)\Pr(A),$$

which implies that the events A and B are independent if and only if $\Pr(A|B) = \Pr(A)$. That is, the probability of the occurrence of event A is not affected by whether the event B has occurred. This result is consistent with our intuition as to the meaning of two events being independent.

1.6 The Total Probability Theorem

Suppose that $A_1, \ldots, A_n$ are mutually exclusive events. n events are *mutually exclusive* if at most one of them can occur in a single trial. Suppose further that these events are *exhaustive*, which means that together they account for every possible outcome. In other words, for any trial $\xi \in E$, $\xi \in A_k$ for at least one k because the A_i are exhaustive, and $\xi \notin A_j$ for all $j \neq k$ because the A_i are mutually exclusive.

For such a set of events, the *Total Probability Theorem* states that: For any event B,

$$\Pr(B) = \Pr(B|A_1)\Pr(A_1) + \Pr(B|A_2)\Pr(A_2) + \cdots + \Pr(B|A_n)\Pr(A_n).$$

1.7 Bayes' Theorem

For two events A_i and B, taking the intersection of two sets is independent of the order of the two sets,

$$\Pr(A_i \text{ and } B) = \Pr(B \text{ and } A_i).$$

Applying the expression for conditional probability to both sides of this equation results in

$$\Pr(A_i|B) = \Pr(B|A_i)\frac{\Pr(A_i)}{\Pr(B)}. \tag{1}$$

Now suppose that event A_i is a member of a collection of events, $A_1, \ldots, A_n$, which is mutually exclusive and exhaustive. The Total Probability Theorem indicates that

$$\Pr(B) = \Pr(B|A_1)\Pr(A_1) + \Pr(B|A_2)\Pr(A_2) + \cdots + \Pr(B|A_n)\Pr(A_n).$$

Substituting this equation into the denominator of (1) gives *Bayes' Theorem*:

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\Pr(B|A_1)\Pr(A_1) + \Pr(B|A_2)\Pr(A_2) + \cdots + \Pr(B|A_n)\Pr(A_n)}.$$

Bayes' theorem provides a very useful framework for using experimental measurements to update knowledge about a physical problem, which is often applied in the design of prognostic/diagnostic systems and in the estimation of states or model parameters from experimental data.

2 Random Variables

In the example of measuring the temperature of a reactor T, the measured temperature is an example of a *random variable*. Formally, a random variable is any function of the outcome of an experiment, $g : E \to \mathbb{R}$. Since the measured temperature T was the outcome of the experiment, it is the simplest example of a random variable; $T(\xi) = \xi$. As another example, the value of a rate constant calculated from the temperature measurement,

$$k(\xi) = A \exp\left(-\frac{E}{R\xi}\right),$$

is also a random variable.

In general, let x denote a random variable on a set of outcomes E. Performing a single trial produces an outcome $\xi \in E$ that is a *realization* or *sample* of the random variable x, $x(\xi) = \hat{x} \in \mathbb{R}$. Since the value of a random variable depends on the outcome of a trial, the events can be defined in terms of random variables. For example, the subset of E defined by $\{\xi : x(\xi) \geq 5\}$ is an event, which is interpreted as '*the outcome of the trial ξ satisfies* $x(\xi) \geq 5$'. For the above example of the rate constant, this event would be interpreted as '*the measured temperature predicts a rate constant greater than 5*'. These events also have probabilities, which are denoted by

$$\Pr(x \geq 5) = \Pr(\{\xi : x(\xi) \geq 5\}).$$

2.1 Cumulative Probability and Probability Density

The *cumulative probability function*[1] F_x of a random variable is defined as

$$F_x(\hat{x}) = \Pr(-\infty \leq x \leq \hat{x}).$$

Given some $\hat{x}$, this function returns the probability that the random variable x will take a value less than $\hat{x}$ in a single trial. F_x is a nondecreasing function with $F_x(-\infty) = 0$ and $F_x(+\infty) = 1$.

The *probability density function (PDF)* of the random variable x is a function p_x defined by: For every $\hat{x} \in \mathbb{R}$, the probability that, in a given trial, $x(\xi)$ is in the infinitesimal interval $[\hat{x}, \hat{x} + dx]$ is

$$\Pr([\hat{x}, \hat{x} + dx]) = p_x(\hat{x})dx.$$

Figure 1 shows a probability density function for x.

If x can only take discrete values, such as integer values, then for any such value $\hat{x}$, $p_x(\hat{x})$ is simply the probability that the result of a single experiment will be $\hat{x}$. In general a random variable can take on a continuum of values, such as the measurement of temperature T in the above example. In such cases, it is incorrect to call $p_x(\hat{x})$ the *probability* of observing the value $\hat{x}$. Instead, $p_x(\hat{x})$ is called the *probability density* of $\hat{x}$.

The PDF p_x satisfies

$$\int_{-\infty}^{+\infty} p_x(x')dx' = 1,$$

and is related to the cumulative probability function by

$$F_x(\hat{x}) = \int_{-\infty}^{\hat{x}} p_x(x')dx'.$$

[1] Also known as the *cumulative distribution function* or *cumulative distribution function*.

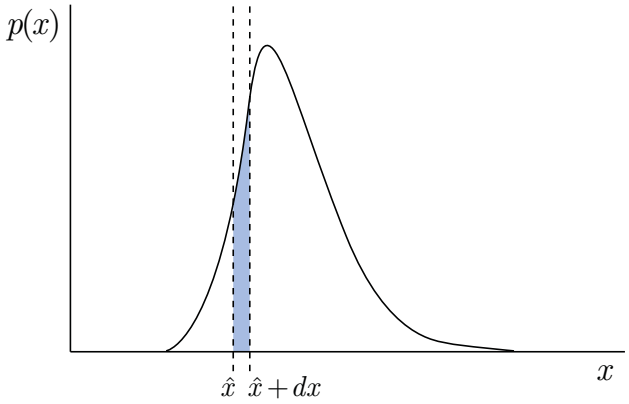

Figure 1: Probability density function for x.

Furthermore,

$$\Pr(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} p_x(x')dx'.$$

2.2 Working with Random Variables

Strictly speaking, the PDF p_x of a random variable x is defined in terms of the probabilities of events in the underlying set of outcomes E. Because the PDF specifies the behavior of x completely, it is common in applications to refer to a random variable with a specified PDF, without defining or otherwise referring to the underlying set of outcomes E.

In this context, it is common to say that to *sample* the PDF p_x is to get a value $\hat{x} \in \mathbb{R}$, which is shorthand for the more formal description: a trial is performed to get an outcome $\xi \in E$ and define $\hat{x} = x(\xi)$. Sampling a PDF p_x means to generate in some way numbers $\hat{x}$ preferentially in regions where p_x is large. That is, for a large number N of samples, the number of samples lying in a given interval $[x_1, x_2]$, $N_{[x_1,x_2]}$, should obey

$$\frac{N_{[x_1,x_2]}}{N} \approx \Pr(x_1 \leq \hat{x} \leq x_2) = \int_{x_1}^{x_2} p_x(x')dx'.$$

2.3 Mean and Variance

Let f be a function of the random variable x. The *expected value* of f is defined as

$$\langle f \rangle = \int_{-\infty}^{+\infty} f(x')p_x(x')dx'.$$

The expected value of x, which is known as the *mean* of x, is

$$\langle x \rangle = \int_{-\infty}^{+\infty} x'p_x(x')dx'.$$

The *variance* of x, σ_x^2, is defined as the expected value of $(x - \langle x \rangle)^2$,

$$\begin{aligned}
\sigma_x^2 &= \langle (x - \langle x \rangle)^2 \rangle, \\
\sigma_x^2 &= \int_{-\infty}^{+\infty} (x' - \langle x \rangle)^2 p_x(x') dx', \\
\sigma_x^2 &= \int_{-\infty}^{+\infty} (x')^2 p_x(x') dx' - 2\langle x \rangle \int_{-\infty}^{+\infty} x' p_x(x') dx' + \langle x \rangle^2 \int_{-\infty}^{+\infty} p_x(x') dx', \\
\sigma_x^2 &= \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle x \rangle^2, \\
\sigma_x^2 &= \langle x^2 \rangle - \langle x \rangle^2.
\end{aligned}$$

The variance is a measure of the deviation of samples of the PDF p_x from the mean $\langle x \rangle$. The variance of f is

$$\begin{aligned}
\sigma_f^2 &= \langle (f(x) - \langle f \rangle)^2 \rangle, \\
&= \int_{-\infty}^{+\infty} (f(x') - \langle f \rangle)^2 p_x(x') dx', \\
&= \langle f^2 \rangle - \langle f \rangle^2.
\end{aligned}$$

The *standard deviations* of x and f are $\sigma_x = \sqrt{\sigma_x^2}$ and $\sigma_f = \sqrt{\sigma_f^2}$.

2.4 Notable PDFs

The *normal* or *Gaussian* PDF is

$$p_x(\hat{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\hat{x}-\eta)^2/2\sigma^2}.$$

The mean and standard deviation of a random variable x with this PDF are $\langle x \rangle = \eta$ and $\sigma_x = \sigma$.

The *uniform* PDF on an interval $[a, b] \subset \mathbb{R}$ is

$$p_x(\hat{x}) = \begin{cases} \dfrac{1}{b-a} & \text{if } \hat{x} \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

A random variable x with this PDF has zero probability of being observed outside of the interval $[a, b]$, and the probability of being observed in any infinitesimal interval $[\hat{x}, \hat{x} + dx]$ is equal for all $\hat{x} \in [a, b)$. The mean and standard deviation of the uniform distribution are $\frac{b+a}{2}$ and $\frac{1}{3}(\frac{b-a}{2})^2$.

3 Random Vectors

A *random vector* is a vector of random variables

$$\mathbf{x} = (x_1, \ldots, x_n),$$

which some references write as a row and some as a column. In the most natural case, random vectors arise from experiments where the outcomes are vector-valued. For example, the experiment may be to measure the velocity of a particle, which has three components $\mathbf{v} = (v_x, v_y, v_z)$.

3.1 Joint Cumulative Probability and Probability Density

The *joint cumulative probability function* of a random vector $\mathbf{x}$ is defined as

$$F_{\mathbf{x}}(\hat{\mathbf{x}}) = \Pr(x_1 \leq \hat{x}_1 \text{ and } x_2 \leq \hat{x}_2 \text{ and } \cdots \text{ and } x_n \leq \hat{x}_n),$$

where $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n)$.

Similarly, the *joint PDF* of the random vector $\mathbf{x}$ is a function $p_{\mathbf{x}}$ defined by: The probability that, in a given trial, $\mathbf{x}(\xi)$ is in the infinitesimal interval

$$[\hat{\mathbf{x}}, \hat{\mathbf{x}} + d\mathbf{x}] \equiv [\hat{x}_1, \hat{x}_1 + dx_1] \times \cdots \times [\hat{x}_n, \hat{x}_n + dx_n]$$

is

$$\Pr([\hat{\mathbf{x}}, \hat{\mathbf{x}} + d\mathbf{x}]) = p_{\mathbf{x}}(\hat{\mathbf{x}}) d\mathbf{x}.$$

The PDF $p_{\mathbf{x}}(\mathbf{x})$ satisfies

$$\int_{\mathbb{R}^n} p_{\mathbf{x}}(\mathbf{x}') d\mathbf{x}' = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_{\mathbf{x}}(x_1', \cdots, x_n') dx_1' \cdots dx_n' = 1,$$

and is related to the joint cumulative probability function by

$$F_{\mathbf{x}}(\hat{\mathbf{x}}) = \int_{-\infty}^{\hat{x}_1} \cdots \int_{-\infty}^{\hat{x}_n} p_{\mathbf{x}}(x_1', \cdots, x_n') dx_1' \cdots dx_n'.$$

Furthermore, for any region $\Omega \subset \mathbb{R}^n$,

$$\Pr(\mathbf{x} \in \Omega) = \int_{\Omega} p_{\mathbf{x}}(\mathbf{x}') d\mathbf{x}'.$$

The (scalar) random variables $x_1, \ldots, x_n$ are said to be *independent* if

$$p_{\mathbf{x}}(\hat{\mathbf{x}}) = p_{x_1}(\hat{x}_1) p_{x_2}(\hat{x}_2) \cdots p_{x_n}(\hat{x}_n).$$

3.2 Mean, Correlation and Covariance

The mean of a random vector $\mathbf{x}$ is the vector with i^{th} component

$$\langle x_i \rangle = \int_{\mathbb{R}^n} x_i' p_{\mathbf{x}}(\mathbf{x}') d\mathbf{x}'.$$

This vector is denoted compactly as

$$\langle \mathbf{x} \rangle = \int_{\mathbb{R}^n} \mathbf{x}' p_{\mathbf{x}}(\mathbf{x}') d\mathbf{x}'.$$

The *covariance* of the two (scalar) random variables x_i and x_j is

$$\begin{aligned} C_{ij} &= \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle, \\ &= \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle. \end{aligned}$$

The variance of x_i is $\sigma_{x_i}^2 = C_{ii}$. The *covariance* matrix of the random vector $\mathbf{x}$ is

$$\mathbf{C} = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}.$$

The random vector $\mathbf{x}$ is *uncorrelated* if $\mathbf{C}$ is diagonal. If the elements of $\mathbf{x}$ are independent, then $\mathbf{x}$ is uncorrelated.

3.3 Sample Averages and The Central Limit Theorem

Suppose that x is a scalar random variable with PDF p_x and f is a function of x. The *sample average* of f with N samples is

$$S_f^N = \frac{1}{N}\sum_{i=1}^{N} f(x_i),$$

where $x_1, \ldots, x_N$ are N samples from p_x (equivalently, values of the random variable x in N trials, $x_i = x(\xi_i)$).

The sample average is a function of N samples from p_x. An equivalent interpretation is that the sample average is a function of N independent random variables, all of which have the same PDF p_x. That is, S_f^N is a function of a random vector $\mathbf{x} = (x_1, \ldots, x_n)$ with

$$\begin{aligned} p_{\mathbf{x}}(\hat{\mathbf{x}}) &= p_{x_1}(\hat{x}_1)p_{x_2}(\hat{x}_2)\cdots p_{x_N}(\hat{x}_N), \\ p_{\mathbf{x}}(\hat{\mathbf{x}}) &= p_x(\hat{x}_1)p_x(\hat{x}_2)\cdots p_x(\hat{x}_N). \end{aligned}$$

From this interpretation of S_f^N, its mean and variance are defined as

$$\begin{aligned} \langle S_f^N \rangle &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} S_f^N(x_1, \cdots, x_N) p_x(x_1) \cdots p_x(x_N) dx_1 \cdots dx_N, \\ \sigma^2_{S_f^N} &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left(S_f^N(x_1, \cdots, x_N) - \langle S_f^N \rangle\right)^2 p_x(x_1) \cdots p_x(x_N) dx_1 \cdots dx_N, \\ \sigma^2_{S_f^N} &= \langle (S_f^N)^2 \rangle - \langle S_f^N \rangle^2. \end{aligned}$$

It can be shown that

$$\begin{aligned} \langle S_f^N \rangle &= \langle f \rangle, \\ \sigma^2_{S_f^N} &= \frac{1}{N}\sigma_f^2. \end{aligned}$$

In words, the sample average of f based on N samples is itself a random variable, with mean being the same as the mean of f, and variance being $1/N$ the variance of f. The *Central Limit Theorem* states that the distribution of S_f^N approaches the normal distribution, regardless of the PDF of x, as N approaches infinity.

4 Random Number Generators

Almost all programming environments, including MATLAB, have built-in random number generators for computations with random variables. These subroutines generate samples of a specified PDF. MATLAB has a random number generator `rand` that samples the uniform PDF on the open interval $(0, 1)$ and a random number generator `randn` that samples the normal PDF with mean 0 and standard deviation 1. It is possible to modify the output of these subroutines to produce samples of a random variable with any desired PDF p_x through a variety of methods.

Algorithms for generating random numbers are not studied in this course, but we do make two notes:

1. The generated numbers are not actually random, but are designed to appear random in certain respects, which make them useful for certain applications. If simulations involving random numbers become important to your thesis/job, you should study the available methods and make sure that you are using a random number generator that is appropriate for your application.

2. Generating good random numbers is computationally expensive. When writing computer programs whose main cost involves calls to a random number generator, which occurs in Monte Carlo methods described later in this class, one goal is to design the algorithm to avoid performing any unnecessary calls to the random number generator.

10.34 Numerical Methods Applied to Chemical Engineering
Fall 2015