A Brief Review of Probability Theory

E. J. Maginn, J. K. Shah

Department of Chemical and Biomolecular Engineering
University of Notre Dame
Notre Dame, IN 46556 USA

Monte Carlo Workshop
Universidade Federal do Rio de Janeiro
Rio de Janeiro Brazil

Simple Probabilities

- Classical probability rule: count all the points in a given sample space and assign them equal probabilities
- Given W points in the sample space, the probability of each point is $1/W$
- $p_i = \frac{1}{W}$
- Example: Two outcomes of a coin toss (H or T)
  $p_H = 1/2$
  $P_T = 1/2$

Simple Probabilities

- Classical probability rule: count all the points in a given sample space and assign them equal probabilities
- Given W points in the sample space, the probability of each point is $1/W$
- $p_i = \frac{1}{W}$
- Example: Two outcomes of a coin toss (H or T)
  $p_H = 1/2$
  $P_T = 1/2$

Example: coin toss

- If we toss a coin N times, how many total outcomes are there?

Example

- If we toss a coin N times, how many total outcomes are there?
- If $N = 1$, 2 outcomes
- $N = 2$, 4 outcomes (HH, TT, HT, TH)
- $N = 3$, 8 outcomes (HHH, HTT, $\cdots$ TTT)
- In general, 2^N outcomes
- Key: each outcome counted separately even though we may not be able to distinguish one "head" from another
- Probability of any *one event*

$$p_i = \frac{1}{2^N} \tag{1}$$

Example

- If we toss a coin N times, how many total outcomes are there?
- If $N = 1$, 2 outcomes
- $N = 2$, 4 outcomes (HH, TT, HT, TH)
- $N = 3$, 8 outcomes (HHH, HTT, $\cdots$ TTT)
- In general, 2^N outcomes
- Key: each outcome counted separately even though we may not be able to distinguish one “head” from another
- Probability of any *one event*

$$p_i = \frac{1}{2^N} \tag{1}$$

Example

- If we toss a coin N times, how many total outcomes are there?
- If $N = 1$, 2 outcomes
- $N = 2$, 4 outcomes (HH, TT, HT, TH)
- $N = 3$, 8 outcomes (HHH, HTT, $\cdots$ TTT)
- In general, 2^N outcomes
- Key: each outcome counted separately even though we may not be able to distinguish one "head" from another
- Probability of any *one event*

$$p_i = \frac{1}{2^N} \tag{1}$$

Example

- If we toss a coin N times, how many total outcomes are there?
- If $N = 1$, 2 outcomes
- $N = 2$, 4 outcomes (HH, TT, HT, TH)
- $N = 3$, 8 outcomes (HHH, HTT, $\cdots$ TTT)
- In general, 2^N outcomes
- Key: each outcome counted separately even though we may not be able to distinguish one "head" from another
- Probability of any *one event*

$$p_i = \frac{1}{2^N} \tag{1}$$

Enumeration of Simple Events

1. We have enumerated all possible *simple events* and assign equal probabilities to each
2. *Simple event*: **the outcome of a trial that doesn't depend on any other event**
3. Coin flipping works very well. Other cases?
4. Predicting sex of a baby
   - How many outcomes?
   - 2
   - What is the probability any given birth will be a boy?
   - Should be $p_{boys} = 0.5$ BUT $p_{boys} = 0.51$ for many countries.

What is wrong with our analysis?

Enumeration of Simple Events

1. We have enumerated all possible *simple events* and assign equal probabilities to each
2. *Simple event*: **the outcome of a trial that doesn't depend on any other event**
3. Coin flipping works very well. Other cases?
4. Predicting sex of a baby
   - How many outcomes?
   - 2
   - What is the probability any given birth will be a boy?
   - Should be $p_{boys} = 0.5$ BUT $p_{boys} = 0.51$ for many countries.

What is wrong with our analysis?

Enumeration of Simple Events

1. We have enumerated all possible *simple events* and assign equal probabilities to each
2. *Simple event*: **the outcome of a trial that doesn't depend on any other event**
3. Coin flipping works very well. Other cases?
4. Predicting sex of a baby
   - How many outcomes?
   - 2
   - What is the probability any given birth will be a boy?
   - Should be $p_{boys} = 0.5$ BUT $p_{boys} = 0.51$ for many countries.

What is wrong with our analysis?

Enumeration of Simple Events

1. We have enumerated all possible *simple events* and assign equal probabilities to each
2. *Simple event*: **the outcome of a trial that doesn't depend on any other event**
3. Coin flipping works very well. Other cases?
4. Predicting sex of a baby
   - How many outcomes?
   - 2
   - What is the probability any given birth will be a boy?
   - Should be $p_{boys} = 0.5$ BUT $p_{boys} = 0.51$ for many countries.

What is wrong with our analysis?

Enumeration of Simple Events

1. We have enumerated all possible *simple events* and assign equal probabilities to each
2. *Simple event*: **the outcome of a trial that doesn't depend on any other event**
3. Coin flipping works very well. Other cases?
4. Predicting sex of a baby
   - How many outcomes?
   - 2
   - What is the probability any given birth will be a boy?
   - Should be $p_{boys} = 0.5$ BUT $p_{boys} = 0.51$ for many countries.

What is wrong with our analysis?

Another Example

- Will it rain today?
- Two possible outcomes (in Rio): rain (R) or shine (S)
- In South Bend, we have rain, shine, snow, and sleet
- Is $p_{rain} = 0.5$? Why not?

Simple vs Compound Events

- The problem: births, weather, and other such things are *not simple events*
- They are *compound events*
- Compound event: A collection of simple events
- Outcome depends on product of many simple event probabilities

Must be careful and test any *a priori* assignment of probabilities we make.

Statistical Probability

- *Statistical probability* assignment of probabilities to events by measuring the relative frequency of occurrence
- Example: Statistical probability of a person's eye color
- How to determine? Go make the measurements!
- Graduate students examine 1000 people, report 601 brown, 251 blue, and 148 green

$$p_{brown} = 0.601,\ p_{blue} = 0.251, and\ p_{green} = 0.148$$

Is this "correct"?

Statistical Probability

- *Statistical probability* assignment of probabilities to events by measuring the relative frequency of occurrence
- Example: Statistical probability of a person's eye color
- How to determine? Go make the measurements!
- Graduate students examine 1000 people, report 601 brown, 251 blue, and 148 green

$$p_{brown} = 0.601, \ p_{blue} = 0.251, and \ p_{green} = 0.148$$

Is this "correct"?

Statistical Probability Example

As usual, the advisor of these students isn't satisfied. He tells them to go out and make more measurements

- New sample size is 10,000
- Results

$$n_{brown} = 6,205, \; n_{blue} = 2,688, \; n_{green} = 1,107$$

- Revised probabilities

$$p_{brown} = 0.6205, \; p_{blue} = 0.2688, \textit{and } p_{green} = 0.1107$$

Is this "correct"?

Statistical Probability Example

As usual, the advisor of these students isn't satisfied. He tells them to go out and make more measurements

- New sample size is 10,000
- Results

$$n_{brown} = 6,205, \; n_{blue} = 2,688, \; n_{green} = 1,107$$

- Revised probabilities

$$p_{brown} = 0.6205, \; p_{blue} = 0.2688, \textit{and } p_{green} = 0.1107$$

Is this "correct"?

Statistical Probability Example

As usual, the advisor of these students isn't satisfied. He tells them to go out and make more measurements

- New sample size is 10,000
- Results

$$n_{brown} = 6,205, \; n_{blue} = 2,688, \; n_{green} = 1,107$$

- Revised probabilities

$$p_{brown} = 0.6205, \; p_{blue} = 0.2688, and \; p_{green} = 0.1107$$

Is this "correct"?

Statistical Probability

- Relative frequencies are (n_i/N), where n_i is the number of occurrences and N are the total number of samples
- If relative frequencies tend toward a constant as N goes to infinity, then this limit is defined as the *statistical probability*, p_i

$$p_i = \lim_{N\to\infty} \left(\frac{n_i}{N}\right) \tag{2}$$

- In practice, (n_i/N) will fluctuate, but should converge
- For random events, fluctuations in statistical probability diminish as $N^{-1/2}$
- Larger N, the more accurate is p_i.

Probability

- Key point 1: statistical probabilities are only accurate when the same event occurs many times, so that the ratio (n_i/N) can converge
- It is incorrect to speak of the "statistical probability" of a team winning a football match, since the contest will only occur once
- Each game is played under different conditions
- Key point 2: Statistical probabilities only have meaning when (n_i/N) tends to a limit
- What is the statistical probability the Dow Jones Industrial Average will close at 10,000 tonight?
- The market is not converging, so there is no "statistical probability" associated with it

Probability

- Key point 1: statistical probabilities are only accurate when the same event occurs many times, so that the ratio (n_i/N) can converge
- It is incorrect to speak of the “statistical probability” of a team winning a football match, since the contest will only occur once
- Each game is played under different conditions
- Key point 2: Statistical probabilities only have meaning when (n_i/N) tends to a limit
- What is the statistical probability the Dow Jones Industrial Average will close at 10,000 tonight?
- The market is not converging, so there is no “statistical probability” associated with it

Axioms of Probability Theory

1. All probabilities are either *zero or positive numbers*
   Easy to see from definition: $p_i = n_i/N$
2. All probabilities are *less than or equal to one*
   Also easy to see from the definition
3. For *mutually exclusive* events, the probability of either event i occurring or j occurring is

$$p_{i+j} = p_i + p_j \tag{3}$$

   where the notation p_{i+j} refers to the probability that either event i *or* j occur.

Axioms of Probability Theory

1. All probabilities are either *zero or positive numbers*
   Easy to see from definition: $p_i = n_i/N$
2. All probabilities are *less than or equal to one*
   Also easy to see from the definition
3. For *mutually exclusive* events, the probability of either event i occurring or j occurring is

$$p_{i+j} = p_i + p_j \tag{3}$$

   where the notation p_{i+j} refers to the probability that either event i *or* j occur.

Axioms of Probability Theory

1. All probabilities are either *zero or positive numbers*
   Easy to see from definition: $p_i = n_i/N$
2. All probabilities are *less than or equal to one*
   Also easy to see from the definition
3. For *mutually exclusive* events, the probability of either event i occurring or j occurring is

$$p_{i+j} = p_i + p_j \tag{3}$$

   where the notation p_{i+j} refers to the probability that either event i *or* j occur.

Axioms of Probability Theory

1. All probabilities are either *zero or positive numbers*
   Easy to see from definition: $p_i = n_i/N$
2. All probabilities are *less than or equal to one*
   Also easy to see from the definition
3. For *mutually exclusive* events, the probability of either event i occurring or j occurring is

$$p_{i+j} = p_i + p_j \tag{3}$$

   where the notation p_{i+j} refers to the probability that either event i *or* j occur.

Axioms of Probability Theory

1. All probabilities are either *zero or positive numbers*
   Easy to see from definition: $p_i = n_i/N$
2. All probabilities are *less than or equal to one*
   Also easy to see from the definition
3. For *mutually exclusive* events, the probability of either event i occurring or j occurring is

$$p_{i+j} = p_i + p_j \tag{3}$$

   where the notation p_{i+j} refers to the probability that either event i *or* j occur.

Distributions

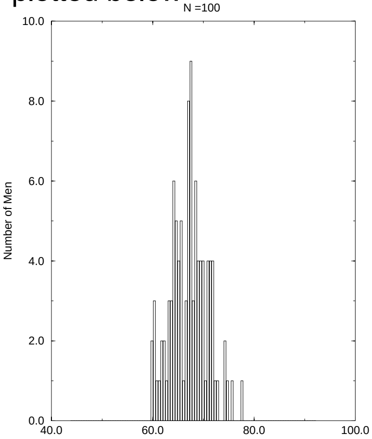
- Let's go back to our team of graduate students
- Now the advisor wants to know the height of the average American adult male
- The students measure heights (in inches, this is the US)
- After measuring 100 men, they compute the average height using

$$\langle h \rangle = \frac{1}{N} \sum_{i=1}^{N} h_i \tag{4}$$

  where $\langle h \rangle$ is the average height, N is the total number of men (samples) measured, and h_i is the result of measurement i.
- They obtain a value of $\langle h \rangle = 67.4$ inches

Height Example

The raw data are plotted below



Height Example, cont.

- The ruler has a resolution of 1/2 in, so heights were *binned*
- Thus, the *actual* formula used for the mean was

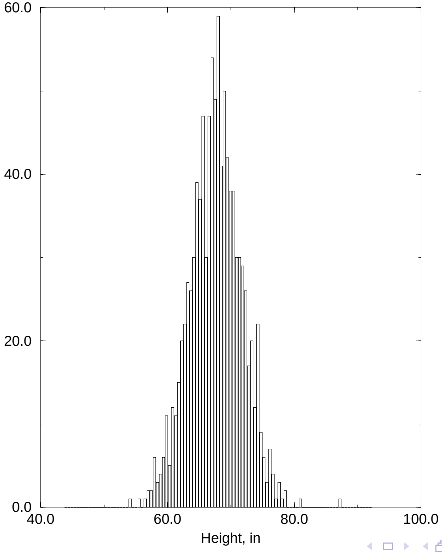
$$\langle h \rangle = \frac{\sum_{i=1}^{nbins} n_i h_i}{N} \tag{5}$$

- $N = \sum_{i=1}^{nbins} n_i$ is the sum over all the bins of the number of men having a height within some discrete bin width
- In the limit of $N \to \infty$, eqn 5 goes to

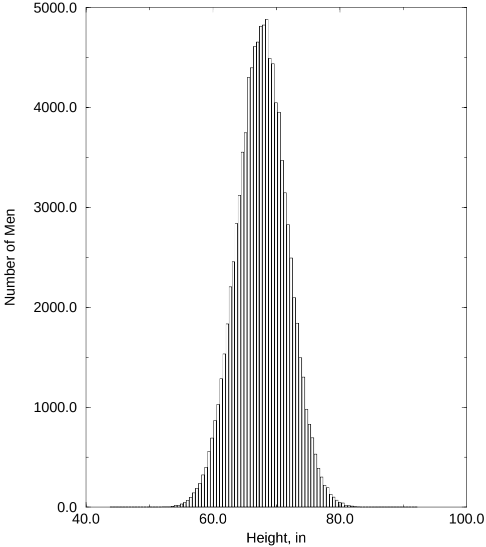
$$\langle h \rangle = \sum_{i=1}^{nbins} p_i h_i \tag{6}$$

- p_i is the *statistical probability* that a man is of height h_i

The graduate students are so excited they take more samples with $N = 1000$



...and $N = 100,000$


5000.0
4000.0
3000.0
2000.0
1000.0
0.0
Number of Men
40.0
60.0
80.0
100.0
Height, in

Distributions

- As the number of samples increases, the distribution becomes smoother and smoother
- Given a ruler with fine resolution dh, $p_i \rightarrow p(h)dh$
- dh is the differential “bin width”
- $p(h)$ is a smooth function of h
- Continuous curve $p(h)$ is the *probability density distribution*
- Sum over bins becomes an integral

$$\sum_{i=1}^{nbins} p_i = 1 \tag{7}$$

becomes

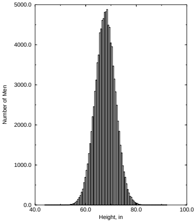
$$\int_{-\infty}^{+\infty} p(h)dh = 1 \tag{8}$$

Continuous Distributions

$$\int_{-\infty}^{+\infty} p(h)dh = 1 \tag{9}$$

We write the above equation in the general case; for the example we were talking about, the lower limit would obviously be zero. Note that $p(h)$ must be a probability *density* distribution with units of $(h)^{-1}$.

Gaussian Distribution



- The distribution which fits the data above is the most important probability distribution in statistical mechanics
- The *Gaussian* or *normal* distribution

$$p(h) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2}\left(\frac{h-<h>}{\sigma}\right)^2\right] \qquad (10)$$

- Two parameters: the mean value ($<h>$) and standard deviation (σ)

Discrete Probability Distributions

- Let $F(x)$ be the value of a discrete function at x
- If there are M possible values of F $(F(x_1), F(x_2), \ldots F(x_M))$ with probabilities P $(P(x_1), P(x_2), \ldots P(x_M))$ then

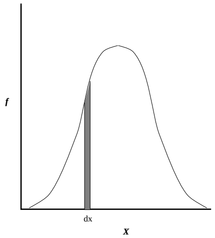
$$\langle F(x) \rangle = \frac{\sum_{j=1}^{M} P(x_j)F(x_j)}{\sum_{j=1}^{M} P(x_j)} \tag{11}$$

- $P(x)$: *discrete distribution*; $F(x)$: *discrete random variable*.
- Since P is a probability, we know it is normalized.

$$\sum_{j=1}^{M} P(x_j) = 1 \tag{12}$$

$$\langle F(x) \rangle = \sum_{j=1}^{M} P(x_j)F(x_j) \tag{13}$$

Continuous Distributions

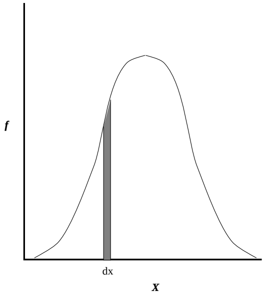


- Let f be a *continuous distribution function* of events depending on x
- Let $\rho\, dx$ be probability of event occurring in region dx

$$\rho\, dx = \frac{f\, dx}{\int_{-\infty}^{+\infty} f\, dx} \tag{14}$$

- ρ is the *probability density*

Continuous Distributions



- Normalized *probabilities* $\int_{-\infty}^{+\infty} \rho \, dx = 1$
- Averages are calculated as follows

$$\langle F \rangle = \frac{\int F \, f(x) \, dx}{\int f(x) \, dx} \tag{15}$$

Gaussian Distribution

We have already encountered a Gaussian distribution. Using the symbols for this section, the Gaussian distribution has the form

$$P(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[\frac{-(x - \langle x \rangle)^2}{2\sigma^2}\right] \tag{16}$$

Summary

- Simple probabilities
- Statistical probability derived from measuring frequency of occurence of eents
- Relative frequencies of events must converge for large number of samples to have a statistical probability
- Discrete and continuous probability distributions
- Weight individual occurences by distribution to get average
- Most important is the Gaussian distribution

$$P(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[\frac{-(x - \langle x \rangle)^2}{2\sigma^2}\right] \tag{17}$$