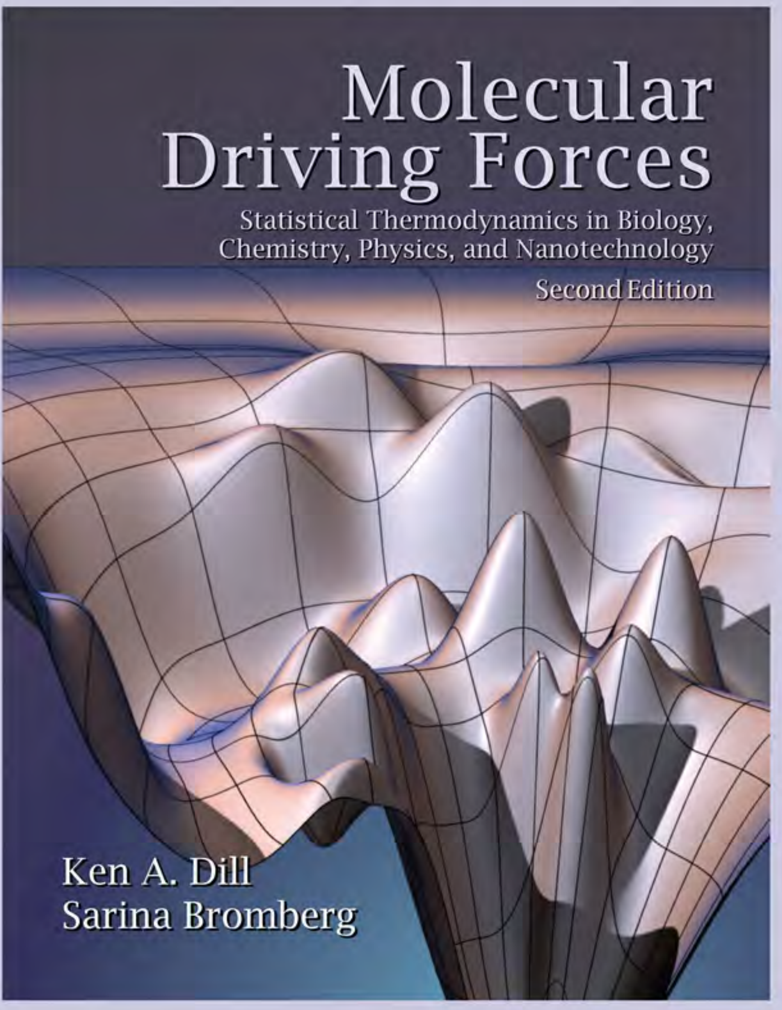
Solutions Manual
Molecular Driving Forces
Second Edition



Ken A. Dill
Sarina Bromberg

Garland Science
Taylor & Francis Group

Chapter 1
Principles of Probability

1. Combining independent probabilities.

You have applied to three medical schools: University of California at San Francisco (UCSF), Duluth School of Mines (DSM), and Harvard (H). You guess that the probabilities you'll be accepted are: $p(\mathrm{UCSF}) = 0.10$, $p(\mathrm{DSM}) = 0.30$, and $p(\mathrm{H}) = 0.50$. Assume that the acceptance events are independent.

(a) What is the probability that you get in somewhere (at least one acceptance)?

(b) What is the probability that you will be accepted by both Harvard and Duluth?

(a) The simplest way to solve this problem is to recall that when probabilities are *independent*, and you want the probability of events A AND B, you can *multiply* them. When events are *mutually exclusive* and you want the probability of events A OR B, you can *add* the probabilities. Therefore we try to structure the problem into an AND and OR problem. We want the probability of getting into H OR DSM or UCSF. But this doesn't help, because these events are not mutually exclusive (mutually exclusive means that if one happens, the other cannot happen). So we try again. The probability of acceptance somewhere, $P(a)$, is $P(a) = 1 - P(r)$, where $P(r)$ is the probability that you're rejected everywhere. (You're either accepted somewhere or you're not.) But this probability *can* be put in the above terms. $P(r) =$ the probability that you're rejected at H AND at DSM AND at UCSF. These events are independent, so we have the answer. The probability of rejection at H is $p(r\mathrm{H}) = 1 - 0.5 = 0.5$. Rejection at DSM is $p(r\mathrm{DSM}) = 1 - 0.3 = 0.7$. Rejection at UCSF is $p(r\mathrm{UCSF}) = 1 - 0.1 = 0.9$. Therefore $P(r) = (0.5)(0.7)(0.9) = 0.315$. Therefore the probability of at least one acceptance $= P(a) = 1 - P(r) = 0.685$.

(b) The simple answer is that this is the intersection of two independent events:

$$\begin{aligned} p(a\mathrm{H})p(a\mathrm{DSM}) &= (0.50)(0.30) \\ &= 0.15. \end{aligned}$$

A more mechanical approach to either part (a) or this part is to write out all the possible circumstances. Rejection and acceptance at H are mutually exclusive. Their probabilities add to one. The same for the other two schools. Therefore all possible circumstances are taken into account by *adding* the mutually exclusive events together, and *multiplying* independent events:

$$[p(a\mathrm{H}) + p(r\mathrm{H})][p(a\mathrm{DSM}) + p(r\mathrm{DSM})][p(a\mathrm{UCSF}) + p(r\mathrm{UCSF})] = 1,$$

or, equivalently,

$$\begin{aligned} = \; & p(a\mathrm{H})p(a\mathrm{DSM})p(a\mathrm{UCSF}) + p(a\mathrm{H})p(a\mathrm{DSM})p(r\mathrm{UCSF}) \\ & + p(a\mathrm{H})p(r\mathrm{DSM})p(a\mathrm{UCSF}) + \cdots, \end{aligned}$$

where the first term is the probability of acceptance at all three, the second term represents acceptance at H and DSM but rejection at UCSF, the third term represents acceptance at H and UCSF but rejection at DSM, etc. Each of these events is mutually exclusive with respect to each other; therefore they are all added. Each individual term represents independent events of, for example, aH *and* aDSM *and* aUCSF. Therefore it is simple to read off the answer in this problem: we want aH *and* aDSM, but notice we don't care about UCSF. This probability is

$$\begin{aligned} p(a\mathrm{H})p(a\mathrm{DSM}) &= p(a\mathrm{H})p(a\mathrm{DSM})[p(a\mathrm{UCSF}) + p(r\mathrm{UCSF})] \\ &= (0.50)(0.30) \\ &= 0.15. \end{aligned}$$

Note that we could have solved part (a) the same way; it would have required adding up all the appropriate possible mutually exclusive events. You can check that it gives the same answer as above (but notice how much more tedious it is).

2. Probabilities of sequences.

Assume that the four bases **A**, **C**, **T**, and **G** occur with equal likelihood in a DNA sequence of nine monomers.

(a) What is the probability of finding the sequence **AAATCGAGT** through random chance?

(b) What is the probability of finding the sequence **AAAAAAAAA** through random chance?

(c) What is the probability of finding any sequence that has four **A**'s, two **T**'s, two **G**'s, and one **C**, such as that in (a)?

(a) Each base occurs with probability 1/4. The probability of an **A** in position 1 is 1/4, that of an **A** in position 2 is 1/4, that of an **A** in position 3 is 1/4, that of a **T** in position 4 is 1/4, and so on. There are nine bases. The probability of this specific sequence is $(1/4)^9 = 3.8 \times 10^{-6}$.

(b) Same answer as (a).

(c) Each specific sequence has the probability given above, but in this case there are many possible sequences that satisfy the requirement that we have four **A**'s, two **T**'s, two **G**'s, and one **C**. How many are there? We start as we have done before, by assuming all nine objects are distinguishable. There are 9! arrangements of nine distinguishable objects in a linear sequence. (The first one can be in any of nine places, the second in any of the remaining eight places, and so on.) But we can't distinguish the four **A**'s, so we have overcounted by a factor of 4!, and must divide this out. We can't distinguish the two **T**'s, so we have overcounted by 2!, and must also divide this out. And so on. So the probability of having this composition is

$$\left[\frac{9!}{4!2!2!1!}\right]\left(\frac{1}{4}\right)^9 = 0.014.$$

3. The probability of a sequence (given a composition).

A scientist has constructed a secret peptide to carry a message. You know only the composition of the peptide, which is six amino acids long. It contains one serine **S**, one threonine **T**, one cysteine **C**, one arginine **R**, and two glutamates **E**. What is the probability that the sequence **SECRET** will occur by chance?

The **S** could be in any one of the six positions with equal likelihood. The probability that it is in position 1 is 1/6. Given that **S** is in the first position, we have two **E**′s, which could occur in any of the remaining five positions. The probability that one of them is in position 2 is 2/5. Given those two letters in position, the probability that the one **C** is in the next of the four remaining positions is 1/4. The probability for the **R** is 1/3. For the remaining **E**, it is 1/2, and for the last **T**, it is 1/1, so the probability is

$$(1/6)(2/5)(1/4)(1/3)(1/2) = 1/360 = \left(\frac{6!}{1!2!1!1!}\right)^{-1}.$$

4. Combining independent probabilities.

You have a fair six-sided die. You want to roll it enough times to ensure that a **2** occurs at least once. What number of rolls k is required to ensure that the probability is at least 2/3 that at least one **2** will appear?

$$\begin{aligned} q &= \frac{5}{6} = \text{probability that a } \mathbf{2} \text{ does not appear on that roll.} \\ q^k &= \text{probability that a } \mathbf{2} \text{ does not appear on } k \text{ INDEPENDENT rolls.} \\ P(k) &= 1 - q^k = \text{probability that at least one } \mathbf{2} \text{ appears on } k \text{ rolls.} \end{aligned}$$

For

$$\begin{aligned} P(k) &\geq \frac{2}{3}, \quad 1 - q^k \geq \frac{2}{3} \implies q^k \leq \frac{1}{3} \implies k \ln q \leq \ln\left(\frac{1}{3}\right) \\ \implies k &\geq \frac{\ln(1/3)}{\ln(5/6)} = 6.03 \end{aligned}$$

Approximately six or more rolls will ensure with probability $P \geq 2/3$ that a **2** will appear.

5. Predicting compositions of independent events.

Suppose you roll a fair six-sided die three times.

(a) What is the probability of getting a **5** twice from all three rolls of the dice?

(b) What is the probability of getting a total of *at least* two **5**'s from all three rolls of the die?

The probability of getting x **5**'s on n rolls of the dice is

$$\left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{n-x} \frac{n!}{x!(n-x)!}.$$

Note that this is a "2-outcome" problem (getting a **5** or not getting a **5**). It is not a "6-outcome" problem.

(a) So the probability of two **5**'s on three dice rolls is

$$\begin{aligned}\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^1 \frac{3!}{2!1!} &= \left(\frac{1}{36}\right)\left(\frac{5}{6}\right)3 \\ &= \frac{15}{216} \\ &= 6.94 \times 10^{-2}.\end{aligned}$$

(b) The probability of getting *at least* two **5**'s is the probability of getting two **5**'s or three **5**'s. Since these two situations are mutually exclusive, we seek

$$\begin{aligned}p(\text{two } \mathbf{5}\text{'s}) + p(\text{three } \mathbf{5}\text{'s}) &= \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right) \frac{3!}{2!1!} + \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^0 \frac{3!}{3!0!} \\ &= \frac{15}{216} + \frac{1}{216} \\ &= \frac{16}{216} \\ &= 7.41 \times 10^{-2}.\end{aligned}$$

6. Computing a mean and variance.

Consider the probability distribution $p(x) = ax^n$, $0 \leq x \leq 1$, for a positive integer n.

(a) Derive an expression for the constant a, to normalize $p(x)$.

(b) Compute the average $\langle x \rangle$ as a function of n.

(c) Compute $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$ as a function of n.

(a)
$$\begin{aligned}\int_0^1 p(x)\,dx &= 1 \Longrightarrow \int_0^1 ax^n\,dx = \left.\frac{ax^{n+1}}{n+1}\right|_0^1 = \frac{a}{n+1} \\ &= 1 \Longrightarrow a = n+1.\end{aligned}$$

(b)
$$\begin{aligned}\langle x \rangle &= \int_0^1 xp(x)\,dx \\ &= \int_0^1 (n+1)x^{n+1}\,dx = \left.\left[\frac{(n+1)x^{n+2}}{n+2}\right]\right|_0^1 = \frac{n+1}{n+2}.\end{aligned}$$

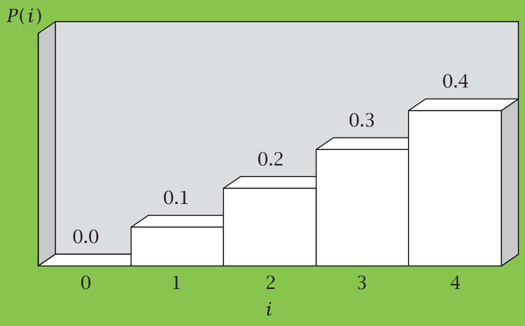
(c)
$$\begin{aligned}\langle x^2 \rangle &= \int_0^1 x^2 p(x)\,dx \\ &= (n+1)\int_0^1 x^{n+2}\,dx \\ &= (n+1)\left.\left(\frac{x^{n+3}}{n+3}\right)\right|_0^1 = \frac{n+1}{n+3}.\end{aligned}$$

So

$$\begin{aligned}\sigma^2 &= \langle x^2 \rangle - \langle x \rangle^2 \\ &= \left(\frac{n+1}{n+3}\right) - \left(\frac{n+1}{n+2}\right)^2.\end{aligned}$$

7. Computing the average of a probability distribution.

Compute the average $\langle i \rangle$ for the probability distribution function shown in the figure below.



A simple probability distribution.

$$
\begin{aligned}
\langle i \rangle &= \sum_{i=0}^{4} ip(i) \\
&= 0(0.0) + 1(0.1) + 2(0.2) + 3(0.3) + 4(0.4) \\
&= 3
\end{aligned}
$$

8. Predicting coincidence.

Your statistical mechanics class has 25 students. What is the probability that at least two classmates have the same birthday?

If you first find the probability q that no two students have the same birthday, then the quantity you want is

$$p(2 \text{ students have same birthday}) = 1 - q$$

The probability that a second student does not have the same birthday as the first is (364/365). The probability that the third student has a birthday different than either of the first two is (363/365), and so on. It is like a sequence problem in which each possible birthday is one card drawn out of a barrel. The probability that no two people have the same birthday, out of m people, is

$$q = \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right)\cdots\left(\frac{365-(m-1)}{365}\right).$$

In factorial notation,

$$q = \frac{N!}{(N-m)!N^m},$$

where $N = 365$. (Incidentally, this expression is identical to the expression for excluded volume in the Flory–Huggins model of polymer solutions (see Chapter 31).) Using Stirling's approximation $x! \approx (x/e)^x$, we get

$$q = \frac{(N/e)^N}{\left(\frac{N-m}{e}\right)^{N-m} N^m}.$$

Collecting together terms in e and dividing the numerator and denominator by N^N gives

$$q = \frac{e^{-m}}{\left(1-\frac{m}{N}\right)^{N-m}}.$$

Substituting $m = 25$ students and $N = 365$ gives

$$q = 0.4163,$$

so

$$\begin{aligned} p &= 1-q \\ &= 0.5837. \end{aligned}$$

There is a better than 50% chance two students will have the same birthday!

9. The distribution of scores on dice.

Suppose that you have n dice, each a different color, all unbiased and six-sided.

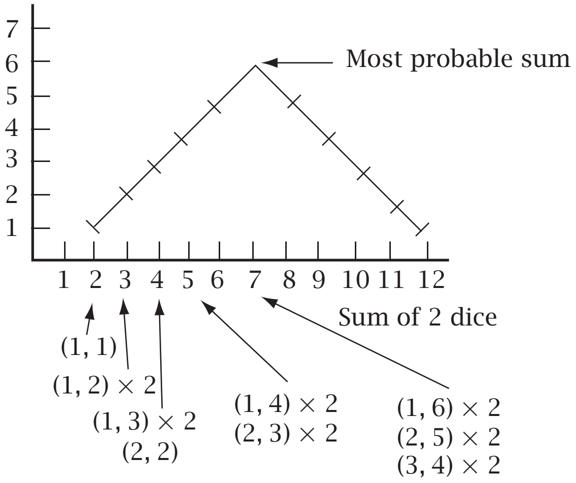
(a) If you roll them all at once, how many distinguishable outcomes are there?

(b) Given two distinguishable dice, what is the most probable sum of their face values on a given throw of the pair? (That is, which sum between 2 and 12 has the greatest number of different ways of occurring?)

(c) What is the probability of the most probable sum?

(a) $$\begin{array}{c} 6 \text{ on one die} \\ 6 \times 6 \text{ on two dice} \\ \vdots \\ 6^n \text{ on } n \text{ dice.} \end{array}$$

(b) Number of ways a sum can occur:



When dice show different numbers, there is a degeneracy of two. When each of the dice has the same number, the degeneracy equals one.

(c) Probability of $7 = p(7) = \dfrac{\text{number of ways of getting 7}}{\text{total number of ways of all outcomes}}$

$$p(7) = \frac{6}{1+2+3+4+5+6+5+4+3+2+1} = \frac{1}{6}.$$

10. The probabilities of identical sequences of amino acids.

You are comparing protein amino acid sequences for homology. You have a 20-letter alphabet (20 different amino acids). Each sequence is a string n letters in length. You have one test sequence and s different data base sequences. You may find any one of the 20 different amino acids at any position in the sequence, independent of what you find at any other position. Let p represent the probability that there will be a 'match' at a given position in the two sequences.

(a) In terms of s, p, and n, how many of the s sequences will be perfect matches (identical residues at every position)?

(b) How many of the s comparisons (of the test sequence against each database sequence) will have exactly one mismatch at any position in the sequences?

(a) For comparing one sequence, each position being assumed independent, the probability of a perfect match of all n residues is
$p^n =$ (number of matched seqs/number of total seqs) $\Longrightarrow$
number of matches in s sequences $= sp^n$.

(b) $n-1$ positions match, so the probability is p^{n-1}; one position doesn't match, which has the probability $(1-p)$; and there are n different positions at which the mismatch could occur; therefore the answer is

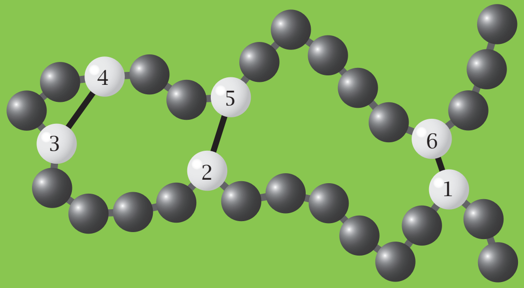
$$sp^{n-1}(1-p)n$$

Note that, in general, for k matches,

$$P(k) = sp^k(1-p)^{n-k}\frac{n!}{k!(n-k)!}. \tag{1}$$

11. The combinatorics of disulfide bond formation.

A protein may contain several cysteines, which may pair together to form disulfide bonds as shown in the figure below. If there is an even number n of cysteines, $n/2$ disulfide bonds can form. How many different disulfide pairing arrangements are possible?



This disulfide bonding configuration with pairs 1–6, 2–5, and 3–4 is one of the many possible pairings. Count all the possible pairing arrangements.

Number the individual sulfhydryl groups along the chain. The first sulfhydryl along the sequence can bond to any of the other $n-1$. This removes two sulfhydryls from consideration. The third sulfhydryl can then bond to any of the remaining $n-3$. Four sulfhydryls are now removed from consideration. The fifth can now bond to any of the remaining $n-5$ sulfhydryls, etc., until all $n/2$ bonds are formed. Thus the total possible number of arrangements of disulfide bonds is a product of $n/2$ terms:

$$D(n) = (n-1)(n-3)(n-5)\cdots 1.$$

Another approach gives an expression that is easier to calculate. Consider placing the sulfhydryls in a sequence. The first place may be occupied by any of n sulfhydryls, the second place by any of $n-1$ sulfhydryls, the third by any of $n-2$ sulfhydryls, etc. Thus, if each sulfhydryl were distinguishable from every other, there would be $n!$ arrangements. However, each sulfhydryl has a mate from which it cannot be distinguished. We must divide by a factor of 2 (per bond) to correct for the indistinguishability of the two ends of each bond. Finally, since we cannot distinguish any of the $n/2$ bonds from any other, we must also divide by $(n/2)!$. Hence the number of arrangements is

$$W(n) = \frac{n!}{2^{n/2}(n/2)!}.$$

Although these two equations were derived in very different ways, they are numerically identical for all n.

12. Predicting combinations of independent events.

If you flip an unbiased green coin and an unbiased red coin five times each, what is the probability of getting four red heads and two green tails?

The probability of four red heads in five coin flips is

$$\left(\frac{1}{2}\right)^5 \left(\frac{5!}{4!1!}\right) = \frac{5}{32}.$$

The probability of two green tails is

$$\left(\frac{1}{2}\right)^5 \frac{5!}{2!3!} = \frac{10}{32}.$$

Since the green coin flips are independent of the red coin flips, the probability we seek is $(5/32)(10/32) = (50/1024) = 4.88 \times 10^{-2}$.

13. A pair of aces.

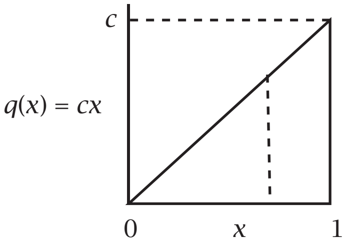
What is the probability of drawing two aces in two random draws without replacement from a full deck of cards?

A deck has 52 cards and four aces. The probability of getting an ace on the first draw is $4/52 = 1/13$. Since you draw without replacement, the probability of getting one of the remaining three aces on the second draw is $3/51$, so the probability of two aces on two draws is

$$\left(\frac{4}{52}\right)\left(\frac{3}{51}\right) = 4.5 \times 10^{-3}.$$

14. Average of a linear function.

What is the average value of x, given a distribution function $q(x) = cx$, where x ranges from zero to one, and $q(x)$ is normalized?



$$\begin{aligned}\langle x\rangle &= \int_0^1 xq(x)\,dx = \int_0^1 cx^2\,dx \\ &= c\left(\frac{x^3}{3}\right)\Big|_0^1 = \frac{c}{3}.\end{aligned}$$

We can also find c:

$$\begin{aligned}1 &= \int_0^1 q(x)\,dx = \int_0^1 cx\,dx \\ &= \left(\frac{cx^2}{2}\right)\Big|_0^1 = \frac{c}{2} = 1.\end{aligned}$$

So,

$$\begin{aligned}c &= 2, \\ \langle x\rangle &= \frac{c}{3} = \frac{2}{3}.\end{aligned}$$

15. The Maxwell–Boltzmann probability distribution function.

According to the kinetic theory of gases, the energies of molecules moving along the x direction are given by $\varepsilon_x = (1/2)mv_x^2$, where m is mass and v_x is the velocity in the x direction. The distribution of particles over velocities is given by the Boltzmann law, $p(v_x) = e^{-mv_x^2/2kT}$. This is the Maxwell–Boltzmann distribution (velocities may range from $-\infty$ to $+\infty$).

(a) Write the probability distribution $p(v_x)$, so that the Maxwell–Boltzmann distribution is correctly normalized.

(b) Compute the average energy $\langle(1/2)mv_x^2\rangle$.

(c) What is the average velocity $\langle v_x\rangle$?

(d) What is the average momentum $\langle mv_x\rangle$?

(a) To write the probability distribution $p(v_x)\,dv_x$ so that the Maxwell–Boltzmann distribution is correctly normalized, we require

$$c\int_{-\infty}^{\infty} e^{-mv_x^2/2kT}\,dv_x = 1$$

From integral tables, we see that

$$I = \int_{-\infty}^{\infty} e^{-ax^2}\,dx = \left(\frac{\pi}{a}\right)^{1/2}.$$

Aside To compute integrals of the form

$$I = \int_{-\infty}^{\infty} e^{-ax^2}\,dx,$$

we use the following trick. It is easy to see that we can write

$$I^2 = \int_{-\infty}^{\infty} e^{-ax^2}\,dx \int_{-\infty}^{\infty} e^{-ay^2}\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-a(x^2+y^2)}\,dx\,dy.$$

This is now an integral over the entire (x, y) plane. Converting to polar coordinates r and θ and recognizing that $r^2 = x^2 + y^2$, the integral becomes

$$I^2 = \int_0^\infty dr\, r \int_0^{2\pi} d\theta\, e^{-ar^2} = \int_0^\infty dr\, re^{-ar^2} \int_0^{2\pi} d\theta = 2\pi \int_0^\infty dr\, re^{-ar^2}$$

Making the substitution $u = -ar^2, du = -2ar\, dr$, we can finish the integral:

$$I^2 = -\frac{\pi}{a} \int_0^{-\infty} du\, e^u = -\frac{\pi}{a}\, e^u \Big|_0^{-\infty} = \frac{\pi}{a}.$$

Hence

$$I = \left(\frac{\pi}{a}\right)^{1/2}.$$

For our integral, $a = m/2kT$.

$$\int_{-\infty}^\infty e^{-mv_x^2/2kT}\, dv_x = \left(\frac{2\pi kT}{m}\right)^{1/2} \implies p(v_x)\, dv_x = \left(\frac{m}{2\pi kT}\right)^{1/2} e^{-mv_x^2/2kT}\, dv_x.$$

(b) To compute the average energy, $\langle (1/2)mv_x^2 \rangle$, we have

$$\left\langle \frac{1}{2}mv_x^2 \right\rangle = \int_{-\infty}^\infty \frac{1}{2}mv_x^2 p(v_x)\, dv_x = \frac{m}{2}\left(\frac{2\pi kT}{m}\right)^{1/2} \int_{-\infty}^\infty v_x^2 e^{-mv_x^2/2kT}\, dv_x$$

Again consulting our table of integrals, we find

$$\int_{-\infty}^\infty x^2 e^{-ax^2}\, dx = \frac{\pi^{1/2}}{2a^{3/2}}$$

Aside: Integrals of the form

$$\int_{-\infty}^\infty x^2 e^{-ax^2}\, dx$$

can be computed by integration by parts. Recall that

$$\int_a^b u\, dv = uv|_a^b - \int_a^b v\, du$$

Choosing the substitutions $u = x$ and $dv = xe^{-ax^2}$, we have $du = dx$ and $v = -\frac{1}{2a}e^{-ax^2}$. Our integral therefore becomes

$$\int_{-\infty}^\infty x^2 e^{-ax^2}\, dx = \left(-\frac{1}{2a}xe^{-ax^2}\right)\Big|_{-\infty}^\infty + \frac{1}{2a}\int_{-\infty}^\infty e^{-ax^2} = \frac{1}{2a}\left(\frac{\pi}{a}\right)^{1/2} = 0 + \frac{\pi^{1/2}}{2a^{3/2}}.$$

Note that we have used the result of the integral from part (a) above.

Therefore

$$\left\langle \frac{1}{2}mv_x^2 \right\rangle = \frac{1}{2}m\left(\frac{\pi^{1/2}}{2a^{3/2}}\right)\left(\frac{m}{2\pi kT}\right)^{1/2} = \frac{1}{4}m\left[\left(\frac{2kT}{m}\right)^{3/2}\right]\left(\frac{m}{\pi^{1/2}kT}\right)^{1/2} = \frac{1}{2}kT.$$

(c) To find the average velocity $\langle v_x \rangle$, we recall that for functions with *odd* symmetry ($f(x) = -f(-x)$), the integral under the curve for negative x cancels with that under the curve for positive x. Using the fact that $p(x) = p(-x)$,

$$\begin{aligned}
\langle v_x \rangle &= \int_{-\infty}^{\infty} v_x p(v_x)\, dv_x \\
&= \int_{-\infty}^{0} v_x p(v_x)\, dv_x + \int_{0}^{\infty} v_x p(v_x)\, dv_x \\
&= \int_{0}^{\infty} (-v_x) p(-v_x)\, dv_x + \int_{0}^{\infty} v_x p(v_x)\, dv_x \\
&= -\int_{0}^{\infty} v_x p(v_x)\, dv_x + \int_{0}^{\infty} v_x p(v_x)\, dv_x. \\
&= 0.
\end{aligned}$$

(d) The average momentum

$$\langle m v_x \rangle = m\langle v_x \rangle = 0,$$

from the result above.

16. Predicting the rate of mutation based on the Poisson probability distribution function.

The evolutionary process of amino acid substitutions in proteins is sometimes described by the Poisson probability distribution function. The probability $p_s(t)$ that exactly s substitutions at a given amino acid position occur over an evolutionary time t is

$$p_s(t) = \frac{e^{-\lambda t}(\lambda t)^s}{s!},$$

where λ is the rate of amino acid substitutions per site per unit time. Fibrinopeptides evolve rapidly: $\lambda_F = 9.0$ substitutions per site per 10^9 years. Lysozyme is intermediate: $\lambda_L \approx 1.0$. Histones evolve slowly: $\lambda_H = 0.010$ substitutions per site per 10^9 years.

(a) What is the probability that a fibrinopeptide has no mutations at a given site in $t = 1$ billion years?

(b) What is the probability that lysozyme has three mutations per site in 100 million years?

(c) We want to determine the expected number of mutations $\langle s \rangle$ that will occur in time t. We will do this in two steps. First, using the fact that probabilities must sum to one, write $\alpha = \sum_{s=0}^{\infty} (\lambda t)^s / s!$ in a simpler form.

(d) Now write an expression for $\langle s \rangle$. Note that

$$\sum_{s=0}^{\infty} \frac{s(\lambda t)^s}{s!} = (\lambda t) \sum_{s=1}^{\infty} \frac{(\lambda t)^{s-1}}{(s-1)!} = \lambda t \alpha.$$

(e) Using your answer to part (d), determine the ratio of the expected number of mutations in a fibrinopeptide to the expected number of mutations in histone protein, $\langle s \rangle_{\text{fib}} / \langle s \rangle_{\text{his}}$.

(a) The probability that a fibrinopeptide has no mutations at a given site in $t = 1$ billion years is

$$\begin{aligned} P_0(t) &= e^{-\lambda_F t} = \exp[-(9.0 \text{ per } 10^9 \text{ years})(10^9 \text{ years})] \\ &= e^{-9} = 1.23 \times 10^{-4}. \end{aligned}$$

(b) For lysozyme,

$$\lambda_L t = (1.0 \text{ per } 10^9 \text{ years})(10^8 \text{ years}) = 0.1.$$

The probability that lysozyme has three mutations per site in 100 million years is then

$$\begin{aligned} P_3(t) &= \frac{e^{-\lambda_L t}(\lambda_L t)^3}{3!} = \frac{(e^{-0.1})(0.1)^3}{6} \\ &= 1.51 \times 10^{-4}. \end{aligned}$$

(c) Since the probabilities sum to 1,

$$\sum_{s=0}^{\infty} P_s(t) = \sum_{s=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^s}{s!} = e^{-\lambda t} \sum_{s=0}^{\infty} \frac{(\lambda t)^s}{s!} = 1.$$

Therefore

$$\alpha = \sum_{s=0}^{\infty} \frac{(\lambda t)^s}{s!} = e^{\lambda t}.$$

(d) $\langle s \rangle = \sum_{s=0}^{\infty} s P_s(t) = \sum_{s=0}^{\infty} \frac{s e^{-\lambda t}(\lambda t)^s}{s!}$

$$= e^{-\lambda t}(\lambda t) \sum_{s=1}^{\infty} \frac{(\lambda t)^{s-1}}{(s-1)!}$$

$$= e^{-\lambda t}(\lambda t) \sum_{h=0}^{\infty} \frac{(\lambda t)^h}{h!} = (\lambda t \alpha) e^{-\lambda t},$$

so

$$\langle s \rangle = \lambda t.$$

(e) $$\frac{\langle s \rangle_{\text{fib}}}{\langle s \rangle_{\text{his}}} = \frac{\lambda_F t}{\lambda_H t} = \frac{\lambda_F}{\lambda_H} = \frac{9}{0.01} = 900.$$

17. Probability in court.

In forensic science, DNA fragments found at the scene of a crime can be compared with DNA fragments from a suspected criminal to determine that the probability that a match occurs by chance. Suppose that DNA fragment A is found in 1% of the population, fragment B is found in 4% of the population, and fragment C is found in 2.5% of the population.

(a) If the three fragments contain independent information, what is the probability that a suspect's DNA will match all three of these fragment characteristics by chance?

(b) Some people believe such a fragment analysis is flawed because different DNA fragments do not represent independent properties. As before, suppose that fragment A occurs in 1% of the population. But now suppose that the conditional probability of B, given that A is $p(B|A) = 0.40$ rather than 0.040, and $p(C|A) = 0.25$ rather than 0.025. There is no additional information about any relationship between B and C. What is the probability of a match now?

(a) Since the fragments are independent,

$$\begin{aligned} p &= p(A)\,p(B)\,p(C) \\ &= (0.01)(0.04)(0.025) = 1 \times 10^{-5}. \end{aligned}$$

(b)
$$\begin{aligned} p &= p(A)\,p(B/A)\,p(C/A) \\ &= (0.01)(0.40)(0.25) = 1 \times 10^{-3}. \end{aligned}$$

18. Flat distribution.

Given a flat distribution, from $x = -a$ to $x = a$, with probability distribution $p(x) = 1/(2a)$:

(a) Compute $\langle x \rangle$.

(b) Compute $\langle x^2 \rangle$.

(c) Compute $\langle x^3 \rangle$.

(d) Compute $\langle x^4 \rangle$.

(a) $$\langle x\rangle = \int_{-a}^{a} xp(x)\,dx = \int_{-a}^{a} \frac{x}{2a}\,dx = \int_{0}^{a} \frac{x}{2a}\,dx - \int_{0}^{a} \frac{x}{2a}\,dx = 0.$$

(b) $$\langle x^2\rangle = \int_{-a}^{a} x^2 p(x)\,dx = \left(\frac{1}{2a}\right)\left(\frac{x^3}{3}\right)\Bigg|_{-a}^{a} = \frac{1}{2a}\left[\frac{a^3}{3} - \left(-\frac{a^3}{3}\right)\right] = \frac{a^2}{3}.$$

(c) By symmetry (as in (a)), $\langle x^3\rangle = 0$. In fact, $\langle x^n\rangle = 0$ for all odd integers n.

(d) $$\langle x^4\rangle = \int_{-a}^{a} x^4 p(x)\,dx = \left(\frac{1}{2a}\right)\frac{x^5}{5}\Bigg|_{-a}^{a} = \frac{a^4}{5}.$$

19. Family probabilities.

Given that there are three children in your family, what is the probability that:

(a) two are boys and one is a girl?

(b) all three are girls?

The probability is about 1/2 for having either a boy or a girl. The binomial distribution shows that the probabilities are:

$$\begin{aligned} \text{3 girls :} \quad & \left(\frac{1}{2}\right)^3 \frac{3!}{3!0!} = \frac{1}{8}, \\ \text{2 girls, 1 boy :} \quad & \left(\frac{1}{2}\right)^3 \frac{3!}{2!1!} = \frac{3}{8}, \\ \text{1 girl, 2 boys :} \quad & \left(\frac{1}{2}\right)^3 \frac{3!}{1!2!} = \frac{3}{8}, \\ \text{3 boys :} \quad & \left(\frac{1}{2}\right)^3 \frac{3!}{0!3!} = \frac{1}{8}. \end{aligned}$$

20. Evolutionary fitness.

Suppose that the probability of having the dominant allele (**D**) in a gene is p and the probability of the recessive allele (**R**) is $q = 1 - p$. You have two alleles, one from each parent.

(a) Write the probabilites of all the possibilities: **DD**, **DR**, and **RR**.

(b) If the fitness of **DD** is f_{DD}, the fitness of **DR** is f_{DR}, and the fitness of **RR** is f_{RR}, write the average fitness in terms of p.

(a)

	D	**R**
D	p^2	pq
R	qp	q^2

DD : p^2
DR : $2pq$
RR : q^2

(b) The average fitness is

$$\begin{aligned} \langle \text{fitness} \rangle &= \sum_i (\text{fitness})_i \text{probability}_i \\ &= f_{\mathrm{DD}} p^2 + 2 f_{\mathrm{DR}} p(1-p) + f_{\mathrm{RR}} (1-p)^2. \end{aligned}$$

21. Ion-channel events.

A biological membrane contains N ion-channel proteins. The fraction of time that any one protein is open to allow ions to flow through is q. Express the probability $P(m, N)$ that m of the channels will be open at any given time.

Channels are either open, with probability q, or closed, with probability $(1 - q)$, so the expression we want is the binomial distribution,

$$P(m, N) = q^m (1-q)^{N-m} \frac{N!}{m!(N-m)!}.$$

22. Joint probabilities: balls in a barrel.

For Example 1.10, two green balls and one red ball drawn from a barrel without replacement:

(a) Compute the probability $p(RG)$ of drawing one red and one green ball in either order.

(b) Compute the probability $p(GG)$ of drawing two green balls.

(a) We have

$$\begin{aligned} p(G_1) &= 2/3, \\ p(R_1) &= 1/3, \\ p(G_2 \mid G_1) &= 1/2, \\ p(G_2 \mid R_1) &= 1, \\ p(R_2 \mid G_1) &= 1/2, \end{aligned}$$

so

$$\begin{aligned} p(RG) &= p(R_1)p(G_2 \mid R_1) + p(G_1)p(R_2 \mid G_1) \\ &= (1/3)(1) + (2/3)(1/2) \\ &= 2/3. \end{aligned}$$

(b) $$\begin{aligned} p(GG) &= p(G_1)p(G_2 \mid G_1) \\ &= (2/3)(1/2) = 1/3. \end{aligned}$$

Note that since $p(RR) = 0$, the quantities in (a) and (b) sum to one.

23. Sports and weather.

The San Francisco football team plays better in fair weather. They have a 70% chance of winning in good weather, but only a 20% chance of winning in bad weather.

(a) If they play in the Super Bowl in Wisconsin and the weatherman predicts a 60% chance of snow that day, what is the probability that San Francisco will win?

(b) Given that San Francisco lost, what is the probability that the weather was bad?

(a) These chances of winning given in the problem are conditional probabilities (i.e., the probability of winning) given that the weather is good or bad. The approach here is to elucidate the four mutually exclusive and collectively exhaustive outcomes, winning and good weather $P(W,G)$, winning and bad weather $P(W,B)$, losing and good weather $P(L,G)$, and losing and bad weather $P(L,B)$. These joint probabilities can be related to the conditional probabilities $P(W|G)$, etc., and the weather probabilities $P(G)$ and $P(B)$ by the following equations:

$$\begin{aligned}
P(W,G) &= P(W|G)P(G) &= (0.7)(0.4) &= 0.28, \\
P(W,B) &= P(W|B)P(B) &= (0.2)(0.6) &= 0.12, \\
P(L,G) &= P(L|G)P(G) &= (0.3)(0.4) &= 0.12, \\
P(L,B) &= P(L|B)P(B) &= (0.8)(0.6) &= 0.48.
\end{aligned}$$

(Note that $P(L|G)$ was computed using the fact that $P(W|G) + P(L|G) = 1$.)

(b)
$$\begin{aligned}
P(B|L) &= \frac{P(L,B)}{P(L)} \\
P(L) &= P(L|G)P(G) + P(L|B)P(B) \\
&= (0.3)(0.4) + (0.8)(0.6) \\
&= 0.6
\end{aligned}$$

Therefore, $P(B|L) = 0.48/0.6 = 0.8$—there is an 80% chance there was bad weather, given that they lost.

24. The Monty Hall Dilemma.

You are a contestant on a game show. There are three closed doors: one hides a car and two hide goats. You point to one door, call it C. The gameshow host, knowing what's behind each door, now opens either door A or B, to show you a goat; say it's door A. To win a car, you now get to make your final choice: should you stick with your original choice C, or should you now switch and choose door B? (*New York Times*, July 21, 1991; *Scientific American*, August 1998.)

A good way to illustrate how people sometimes try to tackle this problem is to consider a similar one: Suppose three cards are lying face down on a table, only one of which is an ace. The first card, A, is turned over, and is *not* an ace, so

$$p(B = \text{ace}) = \frac{\frac{1}{3}}{1-\frac{1}{3}} = \frac{1}{2} \qquad p(C = \text{ace}) = \frac{\frac{1}{3}}{1-\frac{1}{3}} = \frac{1}{2}.$$

So both remaining face-down cards are equally likely to be an ace. Using this type of reasoning, many people will say that switching isn't any more likely to win than staying with the door you initially chose. But these problems are not equivalent – in the Monty Hall case, the host has knowledge of both which door you initially picked and which door contains the car.

A simple way to come to arrive at the correct solution is to break it down into two separate questions:

1. What is the probability of winning if you *don't switch* doors?

If you *don't* switch doors, it means that you make no use of the information given by the host revealing a goat. The only way you can win is if the door you initially chose has the car behind it, and hence the probability of winning is 1/3.

2. What is the probability of winning if you *do switch*?

If you *do* switch doors, the only way you can lose is if the door you initially picked had the car behind it, so you have a 1/3 probability of losing and your probability of winning is therefore 2/3. So you should switch doors.

25. Probabilities of picking cards and rolling dice.

(a) What is the probability of drawing either a queen or a heart in a normal deck of 52 cards?

(b) What is the probability P of getting three 2's and two 4's on 5 independent rolls of a die?

(a) $P(\text{Q of hearts}) = 1/52$, $P(\text{Q not of hearts}) = 3/52$, $P(\text{heart and not a Q}) = 12/52$, so

$$P = (1/52) + (3/52) + (12/52) = 16/52.$$

(b) $P(2) = P(4) = 1/6$, $P(2)^3P(4)^2 = (1/6)^5 = 1/7776 = 1.29 \times 10^{-4}$, so

$$P = \binom{5}{2} 1.29 \times 10^{-4} = 1.29 \times 10^{-3}.$$

26. Probability and translational start codons.

In prokaryotes, translation of mRNA messages into proteins is most often initiated at start codons on the mRNA having the sequence **AUG**. Assume that the mRNA is single-stranded and consists of a sequence of bases, each described by a single letter. The alphabet of letters for mRNA consists of **A**, **C**, **U**, or **G**.

Consider the set of all random pieces of bacterial mRNA of length six bases.

(a) What is the probability of having either no **A**'s or no **U**'s in the mRNA sequence of six base pairs long?

(b) What is the probability of a random piece of mRNA having exactly one **A**, one **U**, and one **G**?

(c) What is the probability of a random piece of mRNA of length six base pairs having an **A** directly followed by a **U** directly followed by a **G**; in other words, having an **AUG** in the sequence?

(d) What is the total number of random pieces of mRNA of length six base pairs that have exactly one **A**, exactly one **U**, and exactly one **G**, with **A** appearing first, then the **U**, then the **G**? (e.g., **AXXUXG**).

(a) $P(\text{no } \mathbf{A}\text{'s oR no } \mathbf{U}\text{'s}) = P(\text{no } \mathbf{A}) + P(\text{no } \mathbf{U}) - P(\text{no } \mathbf{A} \text{ aND no } \mathbf{U})$

$$= \left(\frac{3}{4}\right)^6 + \left(\frac{3}{4}\right)^6 - \left(\frac{1}{2}\right)^6$$
$$= 0.34$$

(b) Use the multinomial probability distribution (1.33):

$$W = \frac{6!}{1!1!1!3!}, \ P = \left(\frac{1}{4}\right)^6 \frac{6!}{1!1!1!3!} = 0.0293.$$

(c) Let us find $W(\mathbf{AUG})$:
6 positions total:
3 positions are fixed: $W_f = 1^3$,
3 positions are variable: $W_v = 4^3$,
4 positions **AUG** can take in the sequence: $W_{\text{pos}} = 3$

$$\underline{\mathbf{A}}\ \underline{\mathbf{U}}\ \underline{\mathbf{G}}\ \_\ \_\ \_$$
$$\_\ \underline{\mathbf{A}}\ \underline{\mathbf{U}}\ \underline{\mathbf{G}}\ \_\ \_$$
$$\_\ \_\ \underline{\mathbf{A}}\ \underline{\mathbf{U}}\ \underline{\mathbf{G}}\ \_$$

Note: We do not count this next **AUG** position because it has already been included in the first $\underline{\mathbf{A}}\ \underline{\mathbf{U}}\ \underline{\mathbf{G}}\ \_\ \_\ \_$ when we count the multiplicity of the variable positions:

$$\_\ \_\ \_\ \underline{\mathbf{A}}\ \underline{\mathbf{U}}\ \underline{\mathbf{G}}$$

Because W_{total} is the product of independent sources of multiplicity,

$$W(\mathbf{AUG}) = W_f W_v W_{\text{pos}} = 1^3 \cdot 4^3 \cdot 3,$$

$$P(\mathbf{AUG}) = \frac{W_f}{W_{\text{total}}} = \frac{3 \cdot 4^3}{4^6} = \frac{3}{4^3} = \frac{3}{64} \approx 0.0469.$$

(d) We will calculate W for one **A**, one **U**, and one **G** for any ordering of the **A**, **U**, **G** with respect to each other. Then we will divide this result by the number of ways of ordering **A**, **U**, **G** with respect to each other, because we wish only to count the cases where **A** appears before **U** appears before **G**. We have

$$W(1\mathbf{A}, 1\mathbf{U}, 1\mathbf{G}) = \frac{6!}{1!1!1!3!}.$$

The number of ways of ordering **A**, **U**, **G** with respect to each other is 3!, and we have

$$W(1\mathbf{A}, 1\mathbf{U}, 1\mathbf{G}) = \frac{6!/(1!1!1!3!)}{3!} = \frac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1} = 20.$$

27. DNA synthesis.

Suppose that upon synthesizing a molecule of DNA, you introduce a wrong base pair, on average, every 1000 base pairs. Suppose you synthesize a strand of DNA that is 1000 bases long.

(a) Calculate and draw a bar graph indicating the yield (probability) of each product DNA, containing 0, 1, 2, and 3 mutations (wrong base pairs).

(b) Calculate how many combinations of DNA sequences of 1000 bases contain exactly 2 mutant bases.

(c) What is the probability of having specifically the 500th base pair and the 888th base pair mutated in the pool of DNA that has only two mutations?

(d) What is the probability of having two mutations side-by-side in the pool of DNA that has only two mutations?

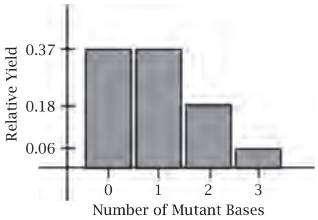
(a) Let letter R stand for a right base and M stand for a mutant one.

$$p(1000R, 1000) = (0.999)^{1000}(0.001)^{0}\frac{1000!}{1000!} \approx 0.368,$$

$$p(999R, 1M, 1000) = (0.999)^{999}(0.001)^{1}\frac{1000!}{999!1!} \approx 0.368,$$

$$p(998R, 2M, 1000) = (0.999)^{998}(0.001)^{2}\frac{1000!}{998!2!} \approx 0.18,$$

$$p(997R, 2M, 1000) = (0.999)^{997}(0.001)^{3}\frac{1000!}{997!3!} \approx 0.06.$$



(b) $$W = \frac{1000!}{998!2!}$$

(c) The probability of having a DNA strand with only 2 mutated bases is 0.18. The probability of having mutations at places 500 and 888 of the strand is

$$p(...M(500)......M(888)...) = 0.18 \times \frac{998!2!}{1000!} = 0.18 \times 2 \times 10^{-6} = 3.6 \times 10^{-7}.$$

(d) $$p(...MM...) = 0.18 \times \frac{998!2!}{1000!} \times 999 = 0.18 \times \frac{1}{500} = 3.6 \times 10^{-4}$$

28. Presidential election.

Two candidates are running for president. Candidate A has already received 80 electoral votes and only needs 35 more to win. Candidate B already has 50 votes, and needs 65 more to win.

Five states remain to be counted. Winning a state gives a candidate 20 votes; losing gives the candidate zero votes. Assume both candidates otherwise have equal chances to win in those 5 states:

(a) Write an expression for $W_{A,\,\text{total}}$, the number of ways A can succeed at winning 40 more electoral votes.

(b) Write the corresponding expression for $W_{B,\,\text{total}}$.

(c) What is the probability candidate A beats candidate B?

(a) For candidate A, this is like flipping a coin 5 times. Each head is like winning 20 points; each tail is zero. Candidate A can win by getting 2 more more states ('heads'), so

$$W_{A,\text{total}} = \frac{5!}{2!3!} + \frac{5!}{3!2!} + \frac{5!}{4!1!} + \frac{5!}{5!0!} = 10 + 10 + 5 + 1 = 26.$$

(b) Candidate B can win by getting 3 or more states, so

$$W_{B,\,\text{total}} = \frac{5!}{4!1!} + \frac{5!}{5!0!} = 5 + 1 = 6.$$

(c) $$P(A\,\text{wins}) = \frac{W_{A,\text{total}}}{W_{A,\text{total}} + W_{B,\text{total}}} = \frac{26}{26+6} = 0.81.$$

Chapter 2
Extremum Principles Predict Equilibria

1. A lattice gas.

How many arrangements are there of 15 indistinguishable lattice gas particles distributed on:

(a) $V = 20$ sites?

(b) $V = 16$ sites?

(c) $V = 15$ sites?

$$
\begin{aligned}
&\text{(a)} \qquad W(N=15, V=20) &&= \frac{20!}{15!5!} = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}{5 \cdot 4 \cdot 3 \cdot 2} = 15,504. \\
&\text{(b)} \qquad W(N=15, V=16) &&= \frac{16!}{15!1!} = 16. \\
&\text{(c)} \qquad W(N=15, V=15) &&= \frac{15!}{15!0!} = 1.
\end{aligned}
$$

2. Maximum of binomial distribution.

Find the value $n = n^*$ that causes the function

$$W = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$

to be at a maximum, for constants p and N. Use Stirling's approximation, $x! \simeq (x/e)^x$. Note that it is easier to find the value of n that maximizes $\ln W$ than the value that maximizes W. The value of n^* will be the same.

W is maximal where $\ln W$ is maximal, and

$$\ln W = n \ln p + (N-n)\ln(1-p) + \ln N! - \ln n! - \ln(N-n)!.$$

Now using Stirling's approximation, $\ln N! \approx N \ln N - N$, we obtain

$$\begin{aligned}
\ln W &\approx n \ln p + (N-n)\ln(1-p) + (N\ln N - N) - (n \ln n - n) \\
&\quad -[(N-n)\ln(N-n) - (N-n)] \\
&= n \ln p + (N-n)\ln(1-p) + N \ln N - n \ln n - (N-n)\ln(N-n)
\end{aligned}$$

This function is maximal where

$$\frac{d \ln W}{dn} = 0.$$

We have

$$\begin{aligned}
\frac{d \ln W}{dn} &= \ln p - \ln(1-p) - \left(n \cdot \frac{1}{n} + \ln n\right) \\
&\quad - \left[(N-n)\cdot\frac{1}{(N-n)}(-1) + \ln(N-n)\cdot(-1)\right] \\
&= \ln\left(\frac{p}{1-p}\right) - \ln n + \ln(N-n) - 1 + 1
\end{aligned}$$

We add $\ln N - \ln N$ to the right-hand side and rearrange terms to allow us to write $\ln(n/N)$ and $\ln[(N-n)/N]$:

$$\begin{aligned}
\frac{d\ln W}{dn} &= \ln\left(\frac{p}{1-p}\right) - \ln\left(\frac{n}{N}\right) + \ln\left(\frac{N-n}{N}\right) = 0, \\
\ln\left(\frac{n^*}{N-n^*}\right) &= \ln\left(\frac{p}{1-p}\right), \\
\frac{n^*}{N-n^*} &= \frac{p}{1-p}, \\
n^*\left(1+\frac{p}{1-p}\right) &= \frac{Np}{1-p}, \\
\frac{n^*}{N} &= \frac{p/(1-p)}{1/1-p} = p.
\end{aligned}$$

3. Finding extrema.

$$V(x) = \frac{1}{3}x^3 + \frac{5}{2}x^2 - 24x$$

(a) Where is the maximum?

(b) Where is the minimum?

To find the extrema, determine the values $x = x^*$ that cause the derivative to equal zero:

$$\begin{aligned}
\left.\frac{dV}{dx}\right|_{x^*} &= \left.(x^2+5x-24)\right|_{x^*} = (x^*-3)(x^*+8) = 0 \\
\implies & \quad x^* = 3, -8.
\end{aligned}$$

To determine whether the extrema are maxima or minima, evaluate the second derivative at the x^* points:

$$\frac{d^2V}{dx^2} = 2x + 5,$$

$$\left.\frac{d^2V}{dx^2}\right|_{x^*=3} = 2(3) + 5 = 11.$$

Since this value is positive, $x^* = 3$ is a minimum.

$$\left.\frac{d^2V}{dx^2}\right|_{x^*=-8} = 2(-8) + 5 = -9.$$

This value is negative, so $x^* = -8$ is a maximum.

4. The binomial distribution narrows as N increases.

Flip a coin $4N$ times. The most probable number of heads is $2N$, and its probability is $p(2N)$. If the probability of observing N heads is $p(N)$, show that the ratio $p(N)/p(2N)$ diminishes as N increases.

$$\begin{aligned}
\frac{p(N)}{p(2N)} &= \frac{\left(\dfrac{(4N)!}{(N!)(3N)!}\right)}{\left(\dfrac{(4N)!}{(2N)!(2N)!}\right)} \\
&= \frac{[(2N)!]^2}{N!(3N)!} \approx \frac{\left[(2N/e)^{2N}\right]^2}{(N/e)^N\,(3N/e)^{3N}} \\
&= \frac{2^{4N}N^{4N}}{3^{3N}N^{4N}} \\
&= \left(\frac{2^4}{3^3}\right)^N = \left(\frac{16}{27}\right)^N .
\end{aligned}$$

Note that as $N \to \infty$,

$$\lim_{N\to\infty}\frac{p(N)}{p(2N)} = \lim_{N\to\infty}\left(\frac{16}{27}\right)^N = 0.$$

5. De-mixing is improbable.

Using the diffusion model of Example 2.3, with $2V$ lattice sites on each side of a permeable wall and a total of $2V$ white particles and $2V$ black particles, show that perfect de-mixing (all white on one side, all black on the other) becomes increasingly improbable as V increases.

The ratio r of the perfectly demixed to perfectly mixed configurations is

$$r = \frac{1}{\left(\frac{2V!}{V!V!}\right)\left(\frac{2V!}{V!V!}\right)} = \left(\frac{V!V!}{(2V)!}\right)^2 \approx \left(\frac{(V/e)^{2V}}{(2V/e)^{2V}}\right)^2 = \left(\frac{1}{2^4}\right)^V = \left(\frac{1}{16}\right)^V.$$

Note that as $V \to \infty$,

$$\lim_{r\to\infty} r = \lim_{N\to\infty} \left(\frac{1}{16}\right)^N = 0.$$

6. Stable states.

For the energy function $V(\theta) = \cos\theta$ for $0 \leq \theta \leq 2\pi$, find the values $\theta = \theta_s$ that identify stable equilibria, and the values $\theta = \theta_u$ that identify unstable equilibria.

$$\frac{dV(\theta)}{d\theta} = -\sin\theta.$$

This derivative is zero when $\sin\theta$ is zero. This occurs at $\theta = n\pi$, where n is an integer. To determine which of these points are maxima and which are minima, we need to compute the second derivative:

$$\frac{d^2V}{d\theta^2} = -\cos\theta.$$

This is negative for all even-numbered multiples of π and positive for all odd-numbered multiples. Therefore, the unstable equilibria are given by $\theta_u = 2k\pi$ and the stable equilibria by $\theta_s = \pi + 2k\pi$, where k is an integer.

7. One-dimensional lattice.

You have a one-dimensional lattice that contains N_A particles of type A and N_B particles of type B. They completely fill the lattice, so the number of sites is $N_A + N_B$. Write an expression for the multiplicity $W(N_A, N_B)$, the number of distinguishable arrangements of the particles on the lattice.

$$W(N_A, N_B) = \binom{N_A + N_B}{N_A} = \binom{N_A + N_B}{N_B} = \frac{(N_A + N_B)!}{N_A!N_B!}.$$