Homework #1
Solutions

1. 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)?

   <u>Solution:</u>

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

   (b) Same answer as (a) above.

   (c) Each specific sequence has the probability given above, but in this case there are many possible sequences which satisfy the requirement that we have 4 *A*'s, 2 *T*'s, 2 *G*'s, and 1 *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 over-counted by a factor of 4!, and must divide this out. We can't distinguish the two *T*'s, so we have over-counted by 2!, and must also divide this out. And so on. So the probability of having this composition is

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

2. Suppose you roll a die three times.

   (a) What is the probability of getting a total of two **5**'s 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?

   <u>Solution:</u>

   The probability of getting two fives on three rolls is given by

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

The probability of getting at least two fives is identical to the probability of getting two fives or three fives.

$$p(two\ fives) + p(three\ fives) = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^1 \frac{3!}{2!\,1!} + \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^0 \frac{3!}{3!\,0!} = \frac{15}{216} + \frac{1}{216}$$
$$= \frac{16}{216} = 7.41 \times 10^{-2}$$

3. Consider the probability distribution $p(x) = ax^n$, $0 \le x \le 1$, 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.

Solution:

$$\int_0^1 p(x)dx = 1 \Rightarrow \int_0^1 ax^n\,dx = \frac{ax^{n+1}}{n+1}\Big|_0^1 = \frac{a}{n+1} = 1 \Rightarrow a = n+1$$
$$\langle x \rangle = \int_0^1 xp(x)dx = \int_0^1 (n+1)x^{n+1}dx = \left(\frac{(n+1)x^{n+2}}{n+2}\right)\Big|_0^1 = \frac{n+1}{n+2}$$
$$\langle x^2 \rangle = \int_0^1 x^2 p(x)dx = \int_0^1 (n+1)x^{n+2}dx = (n+1)\left(\frac{x^{n+3}}{n+3}\right)\Big|_0^1 = \frac{n+1}{n+3}$$

So $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$ can be obtained from (b) and (c)

$$\sigma^2 = \frac{n+1}{n+3} - \left(\frac{n+1}{n+2}\right)^2$$

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

Solution:

If you first find the probability, q, that no two students have the same birthday then the probability of tow students having the same birthday is given by

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

The probability that a second student does not have the same birthday as the first is $364/365$. The probability that a third student has a birthday different than the other two is $363/365$, and so on. The probability that no two students out of m students is given by:

$$q = \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\left(\frac{362}{365}\right)\cdots\left(\frac{365-(m-1)}{365}\right) = \frac{N!}{(N-m)!\,N^m}$$

where $N = 365$. Now we apply Stirling's approximation $N! \approx (N/e)^N$

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

Now, we have $m = 25$ students and $N = 365$.

$$q = 0.4163$$

and

$$p = 1 - q = 0.5837$$

Thus, there is a greater than 50% chance that two students out of a class of 25 students have the same birthday!

5. 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?

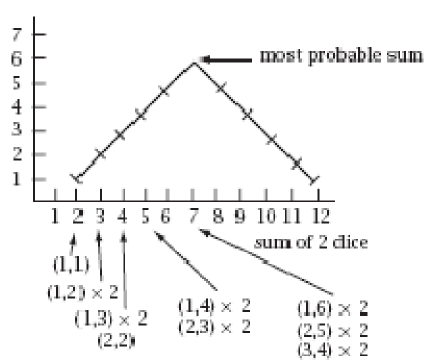
   (b) Given two distinguishable dice, what is the most probable sum of their face values on a given throw of the pair? (That is, which sum between two and twelve has the greatest number of different ways of occurring?)

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

   Solution:

   (a) All rolls are distinguishable so the total possible outcomes is 6^n.

   (b) Number of ways a sum can occur shown graphically



(c) The most probable sum is a 7 and the probability of rolling 7 is given by

$$p(7) = \frac{6}{1+2+3+4+5+6+5+4+3+2+1} = \frac{1}{6} = 0.167$$

6. 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?

Solution:

The probability of four heads in five flips is given by

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

The probability for two green tails is

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

Since the green coin flips and the red coin flips are independent the probability we want is

$$\left(\frac{5}{32}\right)\left(\frac{10}{32}\right) = 4.88 \times 10^{-2}$$

7. 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?

Solution:

$$\langle x \rangle = \int_0^1 x q(x) dx = \int_0^1 c x^2 \, dx = \left[\frac{cx^3}{3}\right]_0^1 = \frac{c}{3}$$

We find c from

$$1 = \int_0^1 q(x) dx = \int_0^1 cx \, dx = \left[\frac{cx^2}{2}\right]_0^1 = \frac{c}{2}$$

so $c = 2$ and

$$\langle x \rangle = \frac{c}{3} = \frac{2}{3}$$

8. 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. 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?

Solution:

Since the fragments are independent

$$p = p(A)p(B)p(C) = (0.01)(0.04)(0.025) = 1 \times 10^{-5}$$

9. Given a flat distribution, from $x = -a$ to $x = a$, with probability distribution $p(x) = 1/(2a)$:
(a) Compute $<x>$.
(b) Compute $<x^2>$.
(c) Compute $<x^3>$.
(d) Compute $<x^4>$.

Solution:

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

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

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

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

10. 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.

Solution:

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

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

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

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

to be at a maximum (for simplicity let $p = ½$). Use Stirling's approximation, $x! \approx \left(\frac{x}{e}\right)^x$. Note that it is easier to find the value of n that maximizes $\ln \mathrm{p}(n)$ than the value that maximizes $\mathrm{p}(n)$. The value of n^* will be the same.

Solution:

For $p = 1/2$ the expression for $p(n)$ becomes

$$\mathrm{p}(n) = \frac{N!}{n!\,(N-n)!}\left(\frac{1}{2}\right)^N$$

and the expression for $\ln p(n)$ is given by

$$\ln p(n) = (N \ln N - N) - (n \ln n - n) - [(N-n)\ln(N-n) - (N-n)] - N \ln 2$$

which simplifies to

$$\ln p(n) = N \ln N - n \ln n - (N-n)\ln(N-n) - N \ln 2$$

Take the derivative of $\ln p(n)$ with respect to n.

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

Optimize the expression for $\ln p(n)$ by setting the derivative equal to zero and solve for n.

$$-\ln n + \ln(N-n) = 0 \;\rightarrow\; \frac{(N-n)}{n} = 1 \;\rightarrow\; n^* = \frac{N}{2}$$

The general result for any given p is $n^* = Np$.

12. 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.

Solution:

$$\frac{p(N)}{p(2N)} = \frac{\left(\frac{(4N)!}{N!\,(3N)!}\right)}{\left(\frac{(4N)!}{(2N)!\,(2N)!}\right)} = \frac{[(2N)!]^2}{N!\,(3N)!}$$

Using Stirling's approximation

$$\frac{p(N)}{p(2N)} \approx \frac{\left[\left(\frac{2N}{e}\right)^{2N}\right]^2}{\left(\frac{N}{e}\right)^N \left(\frac{3N}{e}\right)^{3N}} = \frac{2^{4N} N^{4N}}{3^{3N} N^{4N}} = \left(\frac{2^4}{3^3}\right)^N = \left(\frac{16}{27}\right)^N$$

Note that as $N \to \infty$

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