# Assignment 2: Probability and Counting

Due Tuesday July 3 at 11:59PM under Fernando Rendon's door

**Preface:** The basic methods of probability and counting (i.e., combinatorics) will prove to be crucial in our subsequent development of statistical physics. This problem set is meant to provide practice in these methods.

1. **Big $\mathcal{O}$ notation and Taylor series**
   When expanding functions as Taylor series to a certain order, we use "Big O" notation to denote higher order terms that are not explicitly written down. For example, the Taylor series of the exponential function can be written as

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots, \tag{1}$$

   where "$+\cdots$" denotes all the terms besides the first four non-zero terms in the Taylor series, but we could also write Eq.(1), using "Big O" notation as

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \mathcal{O}(x^4) \tag{2}$$

   where the expression "$\mathcal{O}(x^4)$" represents all terms raised to the power of four or higher in the expansion.

   Using Taylor series, expand the following functions up to the first three non-zero terms and use Big O notation to represent the remaining terms.

   (a) $\sin(x)\cos(x)$
   (b) $\tan^2(x)$
   (c) $\ln(1 + x^2)$

   *You can look up and use the Taylor series for the three basic trigonometric functions and the logarithmic function.*

2. **Number of Hands of Poker**
   For a deck with 52 cards the values 2, 3, 4, 5, 6, 7, 8, 9, 10, $\mathbb{J}$, $\mathbb{Q}$, $\mathbb{K}$, and $\mathbb{A}$ constitute the *rank* of a card. Whether the card is club (♣), diamonds (♦), hearts (♥), or spades (♠) constitutes the card's *suit*.

   Say we draw 5 cards from this 52 card deck. There are 52 ways we can choose the first card, 51 ways to choose the second card, 50 ways for the third, 49 ways for the fourth and 48 ways for the fifth. There are 5! ways to reorder the chosen cards such that we obtain the same hand. Therefore there are

$$\frac{52 \times 51 \times 50 \times 49 \times 48}{5!} = 2,598,960 \tag{3}$$

   possible hands.

   Using similar reasoning we can count the possible hands in a game of poker.

   - **Straight Flush:** A poker hand containing cards of sequential rank and all of the same suit. The rank $\mathbb{A}$ can act as both the low rank and the high rank so that both $\mathbb{A}$♥ 2♥ 3♥ 4♥ 5♥ and 10♦ $\mathbb{J}$♦ $\mathbb{Q}$♦ $\mathbb{K}$♦ $\mathbb{A}$♦ are straight flushes.
     There are

$$40 \times 1 \times 1 \times 1 \times 1 = 40 \tag{4}$$

     possible straight flushes.

- **Full House:** A poker hand containing three cards of one rank and two cards of another rank. *Example:* $\mathbb{A}$♥ $\mathbb{A}$♠ $\mathbb{A}$♦ 8♥ 8♣.
  There are
  $$\frac{52 \times 3 \times 2 \times 48 \times 3}{3! \times 2!} = 3,744 \tag{5}$$
  possible full-house hands.
- **Flush:** A poker hand containing five cards all of the same suit but *not* of sequential rank (i.e., not a straight flush). *Example:* 4♠ 6♠ $\mathbb{K}$♠ 9♠ 2♠.
  There are
  $$\frac{52 \times 12 \times 11 \times 10 \times 9}{5!} - 40 = 5,108 \tag{6}$$
  possible flush hands
- **Straight:** A poker hand containing five cards of sequential rank but *not* all of the same suit (i.e., not a straight flush). *Example:* 2♥ 3♠ 4♦ 5♥ 6♣.
  There are
  $$40 \times 4 \times 4 \times 4 \times 4 - 40 = 10,200 \tag{7}$$
  possible straight hands

For each of the hands listed above, explain how the computed value correctly counts the number of possible hands.

3. **Mean and variance for various distributions**
   For a random variable $n$, which takes on a particular integer value $n = n'$ with probability $p_{n'}$, we can compute the mean and variance of the random variable with the formulas
   $$\langle n \rangle = \sum_n n \,\text{Prob}(n)\,, \qquad \sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2, \tag{8}$$
   where the summation runs over all possible values of $n$ for the random variable.
   Compute the mean and variance of $n$ for the following probability distribution
   (a) **Binomial distribution:** An unfair coin with a probability $p$ of getting heads and a probability $1-p$ of getting tails (in one flip) is flipped $N$ times. The probability of getting $n \le N$ heads is
   $$\text{Prob}(n) = \binom{N}{n} p^n (1-p)^{N-n}. \tag{9}$$
   *The calculation of $\langle n \rangle$ should not have to take more than five lines. After you compute $\langle n \rangle$, to compute $\langle n^2 \rangle$, it will be easier to first compute $\langle n(n-1) \rangle$.*

4. **Gaussian integral**
   We will evaluate the integral
   $$\int_{-\infty}^{\infty} dx \, \exp\left(-\alpha x^2 + \beta x + \gamma\right) \tag{10}$$
   in two steps. *Note: In what follows we use the notation where the exponential function $e^x$ is written as $e^x = \exp(x)$.*
   (a) Let us say that integrating $e^{-x^2}$ from $-\infty$ to $+\infty$ has the value
   $$\int_{-\infty}^{\infty} dx \, \exp\left(-x^2\right) = c_0, \tag{11}$$

for some real number $c_0$[1].Using $u$-substitution compute, in terms of $c_0$, the integral

$$\int_{-\infty}^{\infty} dx \, \exp\left(-\alpha x^2 + \beta x + \gamma\right). \tag{12}$$

*Hint: You should complete the square in the argument of the exponential and then choose your "u" to be* $x - \beta/2\alpha$

(b) In this problem we will compute the integral of $e^{-x^2}$ from $x = -\infty$ to $x = \infty$. We have the integral

$$I = \int_{-\infty}^{\infty} dx \, \exp\left(-x^2\right). \tag{13}$$

Squaring this integral we have

$$I^2 = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \, \exp\left(-x^2 - y^2\right) = \int_0^{2\pi} d\phi \int_0^{\infty} dr \, r \, \exp\left(-r^2\right), \tag{14}$$

where in the final equality we replaced the differential *cartesian* area element $dx\,dy$ with the differential *polar* area element $d\theta\,dr\,r$ where $r = \sqrt{x^2 + y^2}$ and $\tan\theta = y/x$. Compute the double integral in Eq.(14) in polar coordinates, and use the result to determine the value of Eq.(13).

(c) What is the value of Eq.(10)?

5. **The 68, 95, 99.7 Rule**
To obtain a simple quantitative interpretation of the standard deviation for the Gaussian distribution, we can numerically integrate the distribution between various multiples of standard deviations.

Say we have have a continuous random variable $x$ whose probability density function is a Gaussian distribution. Most generally, the probability density function has the form

$$p(x) = \frac{e^{-(x-x_0)^2/2\sigma_0^2}}{\sqrt{2\pi\sigma_0^2}} \tag{15}$$

where $x_0$ is the mean of the distribution and $\sigma_0$ is the standard deviation.

(a) For a positive integer $k$, write the probability of finding $x$ between $x_0 - k\sigma_0$ and $x_0 + k\sigma_0$ as a definite integral. We will denote this probability as

$$\text{Prob}(x_0 - k\sigma_0 \leq x \leq x_0 + k\sigma_0). \tag{16}$$

(b) Using $u$-substitution write the integral in (a) without the parameters $x_0$ and $\sigma_0$. You should end up with a definite integral which is only a function of $k$ (*There will still be some factor which is a function of* $\pi$.).

(c) Using the result in (b) and *WolframAlpha* (or any other numerical integration software), numerically compute

i. $\text{Prob}(x_0 - \sigma_0 \leq x \leq x_0 + \sigma_0)$
ii. $\text{Prob}(x_0 - 2\sigma_0 \leq x \leq x_0 + 2\sigma_0)$
iii. $\text{Prob}(x_0 - 3\sigma_0 \leq x \leq x_0 + 3\sigma_0)$ (Include a print out of your code or numerical integration with the problem set)

(d) Qualitatively, what do the results in (c) tell us?

[1]We are using the notation where the exponential function $e^x$ is written as $\exp(x)$.