Phys 366 Mathematical Methods of Physics Fall 2016

Homework Assignment #10 Due Tuesday, November 29
(30 points) (at lecture)

10.3 (10 points) Challenge problem. **Random walks and their generalizations.** Suppose you are walking on a one-dimensional line. At each of a set of equally spaced times, you have a probability p to take a step to the right and a probability $q = 1 - p$ to take a step to the left. We are interested in the probability $P_N(n)$ that after N total steps, you have taken n steps to the right. This formulation is called a *random walk*.

The random-walk scenario is equivalent to polling N people regarding their opinion on a binary (yes-no) question, where the probability of answering yes is p and the probability of answering no is $q = 1 - p$. In this polling situation, we are interesting in the same probability, $P_N(n)$, which is the probability that n out of N people answer yes.

(a) *Show* that

$$P_N(n) = \binom{N}{n} p^n q^{N-n} \ .$$

For obvious reasons, this probability distribution is called the binomial distribution. It is normalized so that the probability of everything is one, i.e., $1 = \sum_{n=0}^{N} P_N(n)$, by the binomial theorem.

The mean (or average) number of steps to the right after N total steps is

$$\bar{n} = \sum_{n=0}^{N} n P_N(n) \ .$$

More generally, one can define the kth moment of the distribution by

$$\overline{n^k} = \sum_{n=0}^{N} n^k P_N(n) \ .$$

One is often interested in the *variance* of a probability distribution,

$$(\Delta n)^2 = \overline{(n - \bar{n})^2} = \overline{n^2} - \bar{n}^2 \ .$$

You should think about why that second equality is true before going on. The square root of the variance, Δn, called the *standard deviation* or *uncertainty*, is a measure of the spread of the distribution away from the mean value.

(b) *Show* that

$$\overline{n^k} = \left(p \frac{\partial}{\partial p} \right)^k (p+q)^N \bigg|_{p+q=1} \ ,$$

and use this to *show* that $\bar{n} = Np$ and $(\Delta n)^2 = Npq = \bar{n}(1-p)$. (Hint: This is easy, but you will need to allow p and q to be independent variables while doing the derivatives, before setting $p + q = 1$ after doing the derivatives.)

The key thing about this result is the following. Suppose you were trying to estimate p from the number of steps to right. You would estimate p as n/N, since the average of this quantity is p. The uncertainty in your estimate would be $\Delta n/N = \sqrt{p(1-p)}/\sqrt{N}$. This $1/\sqrt{N}$ improvement in your estimate is the most important result in statistics and is the foundation for all of polling.

Suppose now that you are out in a rainstorm. You are interested in the number of rain drops that hit your umbrella in a time T and how that number fluctuates. Assume that the average rate at which rain drops hit the umbrella is R. Divide the interval T up into N very short intervals $dt = T/N$; we can make dt short enough (ultimately we take the limit in which dt is infinitesimal) that the only possibilities during each interval are a single drop, which occurs with the small probability $dp = Rdt$, or no drop, which occurs with near-certain probability $1 - dp = 1 - Rdt$. The tiny intervals are like the steps in a random walk, so the probability of n drops in the time T is

$$P_T(n) = \frac{N!}{n!(N-n)!}(Rdt)^n(1-Rdt)^{N-n} \ .$$

Notice that since the probability for a drop in a given interval is very small, this distribution is essentially zero unless $n \ll N = T/dt$.

(c) By judicious use of approximations, *show* that in the limit $dt \to 0$, i.e., $N = T/dt \to \infty$ (we are taking the limit $dt \to 0$ and $N \to \infty$, with $T = Ndt$ held constant), this distribution becomes

$$P_T(n) = e^{-RT}\frac{(RT)^n}{n!} \ .$$

This distribution is called the *Poisson distribution*. *Show* that it is normalized and that its mean and variance are given by $\bar{n} = RT$ and $(\Delta n)^2 = \bar{n}$.

There is really one further thing that we should do, and that is to show that the binomial distribution becomes a normalized Gaussian distribution when N is large. What one shows is that near the peak of the binomial distribution at $n = Np$, the logarithm of the binomial distribution is well approximated by a quadratic function of $n - Np$; exponentiating this quadratic expression then gives a Gaussian probability distribution. But I'm sure you think you've done enough work already—I do, too—and if you're interested, you can see the derivation of the Gaussian as part (d) in the solution.