# Complex Systems 899: Homework 6

1. **Generating functions:** Consider a probability distribution $P(k)$ over the nonnegative integers $k$. We have seen that the generating function $g(z) = \sum_{k=0}^{\infty} P(k)z^k$ plays a role somewhat similar to the Fourier transform for continuous distributions.

   (a) Show that $g(1) = 1$ and that $g(z)$ is a non-decreasing function of $z$ for $z \geq 0$.

   (b) Write down or derive an expression for the distribution of a quantity $m = k_1 + k_2$ where $k_1, k_2$ are drawn from the same distribution $P(k)$.

   (c) Thus show that the generating function $g^{(2)}(z)$ for the distribution of $m$ is given by $g^{(2)}(z) = [g(z)]^2$, where $g(z)$ is the generating function for the distribution of $k_1, k_2$.

   (d) Generalize this result to show that the distribution of the sum of $n$ identically distributed integers has a generating function $g^{(n)}(z) = [g(z)]^n$.

   (e) Hence show that the equivalent of the central limit theorem for nonnegative integers is that the distribution of the sum of $n$ identically distributed integers tends to the binomial distribution as $n \to \infty$.

2. **Maximum likelihood:** Suppose we have $n$ real numbers $\{x_1 \ldots x_n\}$ that represent measurements of some quantity. We believe (rightly or wrongly) that they follow a Gaussian distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

   Assuming a uniform prior probability distribution on the mean and variance, show that the maximum likelihood fit of this Gaussian model to the observed data has a mean and variance equal to the naive estimates:

$$\mu = \frac{1}{n}\sum_i x_i, \qquad \sigma^2 = \frac{1}{n}\sum_i (x_i - \mu)^2.$$

   This is a special property of the Gaussian distribution. It doesn't work for most distributions, meaning the maximum likelihood value of the mean is usually not what you think it is.

3. **Power-law distribution:** Assume the sales numbers of books follow a perfect power-law distribution with exponent $\alpha$ over their entire range.

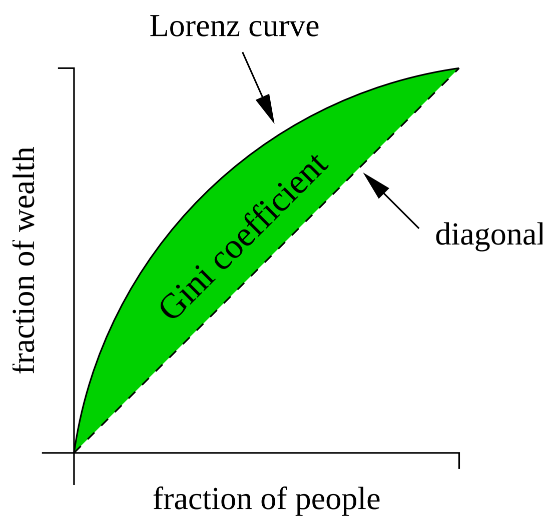
   (a) By approximating sales numbers as a continuous quantity rather than an integer, give a rough expression for the expected sales of the best selling book in a year as a function of the total number of books sold in the same year.

(b) The total number of books sold in the US during 2005 is reported by Nielsen Bookscan to be 709.8 million. Assuming the value $\alpha = 3.5$ given in the handout for the exponent of the book sales distribution, estimate the sales of the number one bestseller for last year. The actual number one bestseller, *Harry Potter and the Half-Blood Prince*, sold 7.02 million copies. What conclusion(s) do you draw? (Hint: There are many reasonable answers to this question, any of which I'm happy to entertain. "This theory is garbage," however, is not one of them.)

4. **The 80/20 rule and the Gini coefficient:**

   (a) If personal wealth is exactly power-law distributed, what value of the exponent would we need for the "80/20" rule to be obeyed exactly?

   (b) The Gini coefficient $g$ is a measure of wealth inequality. It is defined as the area between the Lorenz curve and the diagonal on the Lorenz plot:




   Calculate an expression for the Gini coefficient for power-law distributed wealth with exponent $\alpha$.

   (c) Find, by any convenient means, a reasonably current value for the Gini coefficient for wealth in the United States. (Please state where your value comes from.) Hence make an estimate of the exponent of the power law for wealth in America.