33-341 — Thermal Physics I

Department of Physics, Carnegie Mellon University, Fall Term 2018, Deserno

Problem sheet #7

20. The Central limit theorem *(5 points, due on Monday)*

Let $X_1, \ldots, X_n$ be n *independent* random variables with *identical distribution* $p_X(x)$, which has mean μ and variance $\sigma^2 < \infty$.

1. Consider the centered and normalized random variables $Y_i = (X_i - \mu)/\sigma$. What are their mean and variance? Let $\tilde{p}_Y(k)$ be the characteristic function of the p-density $p_Y(x)$. What does its Taylor expansion up to order $\mathcal{O}(k^3)$ look like?

2. Now define the new (and curiously normalized) random variable $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$, with p-density $p_{Z_n}(z)$ and characteristic function $\tilde{p}_{Z_n}(k)$. Show that in the limit of large n you get

$$\lim_{n\to\infty} \tilde{p}_{Z_n}(k) \;=\; \mathrm{e}^{-\frac{1}{2}k^2} \,. \tag{1}$$

   This implies $p_{Z_n}(z) \longrightarrow \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2}$, *i. e.*, a Gaussian with zero mean and unit variance (which you do not need to show).

   Hint: you will need a beautiful representation for the exponential function: $\lim_{n\to\infty}[1 + z/n + o(z/n)]^n = \mathrm{e}^z$, *where* $o(z)$ *is any term that satisfies* $\lim_{z\to 0} o(z)/z = 0$. *Also: recall some of the tricks we have learned about characteristic functions in problems 17 and 19!*

   Note: This is (a version of) the amazing **Central Limit Theorem**: The distribution of the $\sqrt{n}$-normalized sum of the centered X_i becomes a Gaussian with zero mean and unit variance, *independent of the actual distribution of the* X_i (as long as their variance is finite). It also implies that for increasing n the p-density of the arithmetic mean, $\overline{X_i} = \frac{1}{n}(X_1 + \cdots + X_n) = \mu + \frac{\sigma}{\sqrt{n}} Z_n$ converges against a Gaussian centered around μ with variance $\sigma/\sqrt{n}$. Hence, the *error of the mean* also becomes Gaussian and decreases like $1/\sqrt{n}$. The CLT explains why the Gaussian distribution is "normal".

3. Finally, an illustration. Let X be a random number chosen betwen $-a$ and $+a$ with even probability. What is its mean and variance? Calculate its characteristic function $\tilde{p}_X(k)$. What now is the characteristic function belonging to the associated Z_n defined above? For the special case where $\sigma_X^2 = 1$, plot $\tilde{p}_{Z_n}(k)$ for $n \in \{1, 2, 3, 4, 5, 10, 20\}$ and compare it to the characteristic function of a Gaussian with zero mean and unit variance!

21. Other strange averages *(5 points, due on Tuesday)*

Consider a set of *positive* numbers $\{X_1, X_2, \ldots, X_n\}$. There are multiple ways to define an *average* from this set, the most obvious one being the arithmetic mean $\overline{X}_n^{(\mathrm{a})} := (X_1 + X_2 + \cdots + X_n)/n$. But there are others. For instance, the *harmonic* mean $\overline{X}_n^{(\mathrm{h})}$ results from averaging the inverses, and then taking the inverse of that:

$$\overline{X}_n^{(\mathrm{h})} := \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_n}} \,. \tag{2}$$

Use Jensen's inequality to show that $\overline{X}_n^{(\mathrm{h})} \le \overline{X}_n^{(\mathrm{a})}$.
Hint: Jensen's inequality relies on having (a) a probability distribution (discrete or continuous—it doesn't matter) and (b) some function that is known to be convex (or concave, which flips the sign of the inequality). Hence: try to find what that distribution is, and then try to find which function to use. In other words, you need to rework the problem in the form "function of average versus average of function", with a suitably defined average and a suitably defined function.

22. Pumping gas *(5 points, due on Wednesday)*

When Alice needs to go to the gas station, she always purchases gasoline for a fixed amount of money. When Bob needs to get gas, he always fills up the whole tank. Considering that gas prices fluctuate, show that these two strategies differ economically! Try to estimate how much better the cheaper strategy is.
Hint: assume that whenever Alice or Bob go to the gas station, the "price per mile", p_i, *is a random variable with some unknown distribution. Calculate the total fuel cost of Bob after* N *visits to the gas station, and the total number of miles Alice reaches after* N *visits. Then calculate the effective average price per mile after* N *visits for Alice and Bob. Now remember Jensen.*

23. Inexact differentials and integrating factors *(5 points, due on Friday)*

Show that the differential $đf = y\,\mathrm{d}x - x\,\mathrm{d}y$ is not exact. Find a differential equation for an integrating factor $r(x,y)$ that would result in an exact differential $\mathrm{d}F = r(x,y)\,đf$. Find a solution by using the ansatz $r(x,y) = X(x)$!