Problem Set 1 - Probability and Statistics Review

(Dated: PHYS403, Spring 2024)

I. Gaussian integral identities [10pts]

1. **Basic (Zero mean, unit variance) Gaussian integral:** This section will show a neat trick to compute Gaussian integrals. Follow the steps described to calculate: $I_0 = \int_{-\infty}^{\infty} dx e^{-x^2/2}$. First compute the square of I_0, by writing it as a double integral $I_0^2 = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy e^{-(x^2+y^2)/2}$. Evaluate this integral by converting to cylindrical polar coordinates: $r = \sqrt{x^2+y^2}$, $\phi = \tan^{-1}(y/x)$, and changing variables for the radial integration to $u = r^2/2$. Then take the square root to find I_0.

2. **Gaussian with mean and variance:** Next, compute $I(\mu, \sigma) = \int_{-\infty}^{\infty} dx e^{-(x-\mu)^2/2\sigma^2}$, by shifting and rescaling the integration variable to turn the integral into the form of I_0 above.

3. **Fourier transform of a Gaussian:** Compute the moment generating function for the Gaussian: $Z(\kappa) = \int dx e^{\kappa x} \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$, by completing the square inside the exponent, and using your formula from the previous part to evaluate the integral. [1]

II. Generating functions [10pts]

For the exponential distribution: $p(x) = \frac{1}{\mathcal{N}} \begin{cases} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$. (where $\lambda > 0$ is a positive constant):

1. Find the constant, $\mathcal{N}$, that properly normalizes the distribution

2. Compute the moment generating function, and, from this, find a closed-form expression for the moments, $\langle x^j \rangle$ for any integer j. In addition to the expression for general j, write out the first four moments ($j = 1, 2, 3, 4$) explicitly.

3. Compute the cumulant generating function, and, from this, find a closed-form expression for the cumulants, $\langle x^j \rangle_c$ for any integer j. In addition to the expression for general j, write out the first four cumulants ($j = 1, 2, 3, 4$) explicitly. [2]

---

[1] Note that I chose to look at the MGF with a real-valued argument, κ rather than the complex argument, $-ik$, that we considered in class. Since the resulting Z is analytic, the results for general complex argument can be obtained by just substituting $\kappa \to a + ib$ for any real a, b.

[2] Hint: use the Taylor expansion formula for $\log(1-x) = -\sum_{j=1}^{\infty} \frac{x^j}{j}$.

III. Entropy [5pts]

Find the entropy, $h[p]$, of the Gaussian distribution:

$$p(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$$

as a function of σ and μ.

Suppose you draw a (very) large number of samples from this distribution with $\sigma = 1$, and that, when optimally encoded, this data can be stored in a 1 gigabyte (GB) file. Then, suppose you take the same number of samples, but from a different gaussian distribution with the same mean, but larger standard-deviation: $\sigma = 4$. How much hard-drive space do you expect to need to store the $\sigma = 4$ data set?

IV. Estimation [5pts]

Suppose you make a plot, with statistical error bars reflecting the standard error, from some data you took during an experiment where each point on the plot represents an average over 100 repetitions or "shots" of the experiment. Assuming each shot is statistically independent, approximately how many shots would be required to reduce the error bars by a factor of $\frac{1}{10}\times$ to obtain a less-noisy plot? Be sure to (briefly) explain your reasoning.