Probability problem set
(4 Problems, 60 possible points)

Due on: Feb. 6, 5:00 pm

(Return to my mailbox in N212, hand to me personally, or email me TeX'd solutions)

> Note: An important part of science is communicating your understanding to other people. That is to say, a solution that may be technically correct but which I (the grader) cannot understand is not much better than an incorrect solution. So, please answer the following questions neatly, clearly, and logically – Thanks!

Problem 1 (10 points): Simple mutual information calculation

Suppose you have random variables X_1 and X_2, with a joint PDF of

$$p(x_1, x_2) \propto \exp\left(-\frac{a_{11}x_1^2}{2} - \frac{a_{22}x_2^2}{2} - \frac{a_{12}x_1x_2}{2}\right).$$

Calculate the mutual information, $I(X, Y)$.

Problem 2 (15 points): Unbiased estimation

In this section we'll be deriving probability distributions associated with velocities (say, of a gas particle) under various circumstances, by generalizing the discrete "biased estimation of probabilities section of the notes. In both cases, let's think of a random variable v that can take any value $-\infty \leq v \leq \infty$.

Constrained speed:

Find the unbiased estimate of the probability distribution, $p_1(v)$, subject to the constraint that you know the speed: $\langle |v| \rangle = c$

Constrained kinetic energy:

Find the unbiased estimate of the probability distribution, $p_2(v)$, subject to the constraint that you know the average kinetic energy: $\langle mv^2/2 \rangle = mc^2/2$.

Generalization [worth zero points! Just think about it]:

It sure seems like, from the above two examples, that when the first n moments of a distribution are specified, the unbiased estimate is an exponential of an nth-order polynomial. Is this the case?

Problem 3 (25 points): Manipulating random variables

I've been working on a problem where I want to figure out the elastic properties (bulk modulus, shear modulus, etc.) of a disordered colloidal solid just by looking through a microscope and watching the colloids fluctuate in space. After working hard I come up with an argument that if I measure a particular funny looking quantity, Λ_{xy} (defined below) over windows of size $L = 2R$, then the variance in that variable will be proportional to the inverse shear modulus:

$$\text{var}\left[\Lambda_{xy}\right] = \frac{kT}{4L^2}(G(L))^{-1},$$

and if I let L get bigger and bigger $G(L)$ will approach the shear modulus of the system. (You can tell there was hard work involved by the factor of 4: somebody did more than dimensional analysis for this problem!).

You have no problem with the derivation to get the above expression, but you think in my actual experiment I'm measuring pure noise. You set out to prove it:

Needed definitions

The definition of the quantity of interest is:

$$\Lambda_{xy} = \sum_i^n \Delta_j \left(Ay_i + Bx_i\right),$$

where $\{x_i, y_i\}$ is the spatial position of particle i, there are n particles in the observation window, Δ_j is the x-component of the displacement of particle i from one time point to the next, and A and B are structural quantities related to the relative arrangements of particles in the observation window:

$$A = \frac{a}{ac - b^2} \qquad ; \qquad B = \frac{-b}{ac - b^2}$$

$$a = \sum_i^n (x_i)^2 \quad b = \textstyle\sum_i^n (x_i y_i) \quad c = \displaystyle\sum_i^n (y_i)^2$$

Part A: Warm-up

Let X be a random variable with a Gaussian probability density function with mean $\lambda = 0$ and variance $\sigma^2 = 1$. Consider the random variable $Y = X^2$. Using the change-of-variables formula, write down the probability density function, $p_Y(y)$. Plot it on a log-linear scale.

Part B: Structural quantities

You first want to think about A and B as random variables, and decide to slowly build up to what their distributions are:

Step 1: Assume that each of the particle positions are (1) uncorrelated with each other in the observation window and (2) uniformly distributed in the range $[-R, R]$, i.e.

$$p_{x_i}(x) = \begin{cases} \frac{1}{2R} & |x| \leq R \\ 0 & |x| > R \end{cases} .$$

What is the probability distribution of the quantities x^2 and y^2? What about the quantity xy? Using the central limit theorem, write down an estimate of the probability density function for a, b, and c. Write your answers in terms of Gaussians with n, R, and all the right numerical factors.

Step 2: You next tackle the denominators, $ac - b^2$. Since this appears in the denominator, you decide to make some simplifying assumptions (can you think of any justifications?) and approximate

$$ac - b^2 = \frac{(a+c)^2}{4} - \frac{(a-c)^2}{4} - b^2 \approx \frac{(a+c)^2}{4}.$$

If the mean of $a + c$ had been zero I would have asked you to do some extra derivations, but it's not. Instead, I'll tell you that $(a + c)^2/4$ is a "non-central χ^2 variable with number of summed parameters $k = 1$" (wiki link). Given that information, what is your approximation for the following:

$$\begin{aligned} \left\langle ac - b^2 \right\rangle &= ? \\ \mathrm{var}\left[ac - b^2\right] &= ? \end{aligned}$$

Step 3: You're finally ready to estimate properties of A and B, which involve ratios of random variables. Assuming the numerator and denominator have vanishing covariance, you approximate:

$$\begin{aligned} \left\langle \frac{X}{Y} \right\rangle &= \frac{\langle X \rangle}{\langle Y \rangle} + \frac{\langle X \rangle \,\mathrm{var}\left[Y\right]}{\langle Y \rangle^3} + \cdots \\ \mathrm{var}\left[\frac{X}{Y}\right] &= \frac{\mathrm{var}\left[X\right]}{\langle Y \rangle^2} + \frac{\langle X \rangle^2 \,\mathrm{var}\left[Y\right]}{\langle Y \rangle^4} + \cdots \end{aligned}$$

From all of that, what are your estimates for:

$$\begin{aligned} \langle A \rangle = ? \qquad & \langle B \rangle = ? \\ \mathrm{var}\left[A\right] = ? \qquad & \mathrm{var}\left[B\right] = ? \end{aligned}$$

Part C: Combining everything!

You are firmly convinced that the Δ_j are just Gaussian distributed with zero mean and variance (set by the temperature, basically) σ_Δ^2, uncorrelated with the structural random variables. Using the facts for a collection of random variables X_i the variance behaves like

$$\begin{aligned} \mathrm{var}\left[\sum_i X_i\right] &= \sum_i \mathrm{var}\left[X_i\right] \\ \mathrm{var}\left[\prod_i X_i\right] &= \prod_i \left(\mathrm{var}\left[X_i\right] + \langle X_i\rangle^2\right) - \prod_i \langle X_i\rangle^2 , \end{aligned}$$

and assuming that the contribution from each of the j particles contributes identically and independently, give an expression for

$$\mathrm{var}\left[\Lambda_{xy}\right] = \mathrm{var}\left[\sum_i^n \Delta_j \left(Ay_i + Bx_i\right)\right]$$

Problem 4 (10 points): Information, compression, and entropy

At the end of class on Tuesday, Jan 28 I showed the result of a numerical experiment where I generated a bunch of random strings, of total length n and made up of 26 distinct symbols (an "alphabet"), drawing random numbers so that the probability of picking the ith letter of the English alphabet was

$$p_i \propto i^{-\alpha}.$$

I used "gzip," a universal lossless compression tool, to compress these random strings ("messages"), and showed a plot of the ratio of the length of the original string to the compressed string (on average) on the y-axis vs. the parameter α on the x-axis. I showed that in the limit that n got large (but where, in practice, it didn't have to be that large), the properly normalized compression ratio approached the Shannon entropy associated with the probability distribution of the alphabet, $H(\alpha)$.

Conduct your own numerical experiment!

Choose a discrete alphabet made up of some number of characters, and think of a one-parameter family of probability distributions by which you could choose letters in your randomly generated messages:

$$p_i \propto F(i, \alpha)$$

for some function F (as in the above, but choose a different distribution). Try to compress[1] randomly generated messages for different α and message length n, and show that, after proper normalization, your compression efficiency is bounded by the Shannon entropy.

To get full credit on this problem, you should turn in (a) whatever code you wrote, (b) a plot showing some measure of "compression ratio" and Shannon

[1] Python has a gzip.py module, Mathematica has ExportString[*string*,"GZIP"] commands, etc.

entropy vs α for various n, and (c) an explanation of how you had to define and normalize the "compression ratio" (length of alphabet? choice of ln vs $\log_2$? etc) so that compression approaches the bound set by the Shannon entropy in the $n \to \infty$ limit.

If you're interested in learning more about how these ideas are being used, here are two papers from 2019:

1. Universal and Accessible Entropy Estimation Using a Compression Algorithm
2. Quantifying Hidden Order out of Equilibrium

Question: (0 points): Measurement of homework difficulty

How much time did you spend on this homework? Feel free to answer either in absolute terms (i.e., number of hours worked) or in qualitative terms relative to the average homework from last semester. Thanks!

Question: (0 points): Lecture survey

Please fill out the (tiny, extremely short, and painless) survey from the course webpage about the lectures. Much appreciated!