Problem Set 11
Due: Never

1. Alice has two coins. The probability of heads for the first coin is 1/3; the probability of heads for the second coin is 2/3. Other than this difference in their bias, the coins are indistinguishable through any measurement known to man. Alice chooses one of the coins randomly and sends it to Bob. Let p be the probability that Alice chose the first coin. Bob tries to guess which of the two coins he received by flipping it 3 times in a row and observing the outcome. Assume that all coin flips are independent. Let Y be the number of heads Bob observed.

   (a) Given that Bob observed k heads, what is the probability that he received the first coin?

   (b) Find values of k for which the probability that Alice sent the first coin increases after Bob observes k heads out of 3 tosses. In other words, for what values of k is the probability that Alice sent the first coin given that Bob observed k heads greater than p? If we increase p, how does your answer change (goes up, goes down, or stays unchanged)?

   (c) Help Bob develop the rule for deciding which coin he received based on the number of heads k he observed in 3 tosses if his goal is to minimize the probability of error.

   (d) For this part, assume $p = 2/3$.

      i. Find the probability that Bob will guess the coin correctly using the rule above.

      ii. How does this compare to the probability of guessing correctly if Bob tried to guess which coin he received before flipping it?

   (e) If we increase p, how does that affect the decision rule?

   (f) Find the values of p for which Bob will never guess he received the first coin, regardless of the outcome of the tosses.

   (g) Find the values of p for which Bob will always guess he received the first coin, regardless of the outcome of the tosses.

2. Caleb builds a particle detector and uses it to measure radiation from far stars. On any given day, the number of particles Y that hit the detector is conditionally distributed according to a Poisson distribution conditioned on parameter x. The parameter x is unknown and is modeled as the value of a random variable X, exponentially distributed with parameter μ as follows.

$$f_X(x) = \begin{cases} \mu e^{-\mu x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

   Then, the conditional PDF of the number of particles hitting the detector is,

$$p_{Y|X}(y \mid x) = \begin{cases} \frac{e^{-x}x^y}{y!} & \text{y = 0, 1, 2, ...} \\ 0 & \text{otherwise} \end{cases}$$

   (a) Find the MAP estimate of X from the observed particle count y.

(b) Our goal is to find the conditional expectation estimator for X from the observed particle count y.

i. Show that the posterior probability distribution for X given Y is of the form

$$f_{X|Y}(x \mid y) = \frac{\lambda^{y+1}}{y!} x^y e^{-\lambda x}, \quad x > 0$$

and find the parameter λ. You may find the following equality useful (it is obviously true if the equation above describes a true PDF):

$$\int_0^\infty a^{y+1} x^y e^{-ax}\, dx = y! \quad \text{for any } a > 0$$

ii. Find the conditional expectation estimate of X from the observed particle count y. *Hint:* you might want to express $x f_{X|Y}(x \mid y)$ in terms of $f_{X|Y}(x \mid y+1)$.

(c) Compare the two estimators you constructed in part (a) and part (b).

3. The joint PDF of random variables X and Θ is of the form

$$f_{X,\Theta}(x,\theta) = \begin{cases} c, & \text{if } (x,\theta) \in S \\ 0, & \text{otherwise,} \end{cases}$$

where c is a constant and S is the set

$$S = \{(x,\theta) | 0 \le x \le 2, 0 \le \theta \le 2, x-1 \le \theta \le x\}.$$

We want to estimate Θ based on X.

(a) Find the LMS estimator $g(X)$ of Θ. Sketch the region where $f_{X,\Theta}(x,\theta)$ is positive and plot the LMS estimator $g(X)$.

(b) Calculate $\mathbf{E}[(\Theta - g(X))^2 | X = x]$, $\mathbf{E}[g(X)]$, and $\text{var}(g(X))$.

(c) Calculate the mean squared error $\mathbf{E}[(\Theta - g(X))^2]$. Is it the same as $\mathbf{E}[\text{var}(\Theta | X)]$?

(d) Calculate $\text{var}(\Theta)$ using the law of total variance.

(e) Derive the linear LMS estimator of Θ based on X, and calculate its mean squared error.

(f) Sketch the linear LMS estimator $\hat{\Theta}_{\text{lin}}(X)$ on the plot you drew for part a).

4. The joint PDF of X and Y is defined as follows:

$$f_{X,Y}(x,y) = \begin{cases} cxy & \text{if } 0 < x \le 1,\ 0 < y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the normalization constant c.

(b) Compute the conditional expectation estimator of X based on the observed value $Y = y$.

(c) Is this estimate different from what you would have guessed before you saw the value $Y = y$? Explain.

(d) Repeat (b) and (c) for the MAP estimator.

5. Consider a noisy channel over which you send messages consisting of 0s and 1s to your friend. It is known that the channel independently flips each bit sent with some fixed probability p; however the value of p is unknown. You decide to conduct some experiments to estimate p and seek your friend's help. Your friend, cheeky as she is, insists that you send her messages consisting of three bits each (which you will both agree upon in advance); for each message, she will only tell you the total number of bits in that message that were flipped. Let X denote the number of bits flipped in a particular three-bit message.

   (a) Find the PMF of X.

   (b) Derive the ML estimator for p based on $X_1, \ldots, X_n$, the numbers of bits flipped in the first n three-bit messages.

   (c) Is the ML estimator unbiased ?

   (d) Is the estimator consistent?

6. A body at temperature Θ radiates photons at a given wavelength. This problem will have you estimate Θ, which is fixed but unknown. The PMF for the number of photons K in a given wavelength range and a fixed time interval of one second is given by,

$$p_K(k;\theta) = \frac{1}{Z(\theta)} e^{-\frac{k}{\theta}}, k = 0, 1, 2, \ldots$$

   $Z(\theta)$ is a normalization factor for the probability distribution (The physicists call it the partition function). You are given the task of determining the temperature of the body to two significant digits by photon counting in non-overlapping time intervals of duration one second. The photon emissions in non-overlapping time intervals are statistically independent from each other.

   (a) Determine the normalization factor $Z(\theta)$.

   (b) Compute the expected value of the photon number measured in any 1 second time interval, $\mu_K = \mathbf{E}_\theta[K]$ and its variance, $\text{var}_\theta(K) = \sigma_K^2$.

   (c) You count the number k_i of photons detected in n non-overlapping 1 second time intervals. Find the maximum likelihood estimator, $\hat{\Theta}_n$, for temperature Θ. Note, it might be useful to introduce the average photon number $s_n = \frac{1}{n}\sum_{i=1}^{n} k_i$. And in order to keep the analysis simple we assume that the body is hot, i.e. $\Theta \gg 1$.

   In the following questions we wish to estimate the mean of the photon count in a one second time interval using the estimator $\hat{K}$, which is given by,

$$\hat{K} = \frac{1}{n}\sum_{i=1}^{n} K_i.$$

   (d) Find the number of samples n for which the noise to signal ratio for $\hat{K}$, (i.e.,$\frac{\sigma_{\hat{K}}}{\mu_{\hat{K}}}$), is 0.01.

   (e) Find the 95% confidence interval for the mean photon count estimate for the situation in part (d). (You may use the central limit theorem.)

7. Suppose that we have two populations, population a and population b, where individuals in population a (population b) are modeled as iid Gaussian random variables with unknown mean δ_a and known variance σ_a^2 (δ_b, σ_b^2). We'd like to estimate $\delta = \delta_a + \delta_b$, and also determine a 95% confidence interval for δ. We use the natural estimator: getting a sample mean from each distribution, and then adding them together.

$$\hat{\Delta}_a = \frac{X_1^a + \cdots + X_{n_1}^a}{n_1}$$
$$\hat{\Delta}_b = \frac{X_1^b + \cdots + X_{n_2}^b}{n_2}$$
$$\hat{\Delta} = \hat{\Delta}_a + \hat{\Delta}_b$$

We'd like to compare two ways for developing a confidence interval.

(a) Find the variance of the estimator $\hat{\Delta}$ and use it to construct an ordinary confidence interval for $\delta = \delta_a + \delta_b$.

(b) Get separate 97.5% confidence intervals for each δ_a and δ_b. 'Add' the confidence intervals using the union bound:

$$P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$$

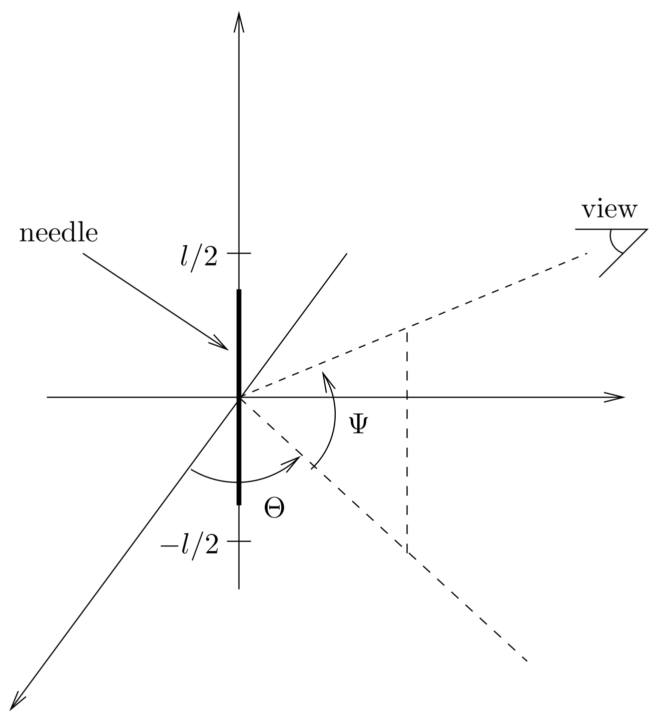
(c) Which of these confidence intervals do you expect to be tighter ? Explain.

(d) Suppose that you have a fixed budget of n samples and so you want to choose $n = n_1 + n_2$ such that the confidence interval is as tight as possible. Which of the methods above would you use ? What value will you choose for n_1 (as a function of n).

(e) How would you estimate the mean and determine the confidence interval that minimizes $n = n_1 + n_2$ if the random variables were iid Bernoulli with unknown parameters θ_a and θ_b ?

G1†. **Swallowed Buffon's Needle:** A doctor is treating a patient who has accidentally swallowed a needle. The key factor in whether to operate on the patient is the true length of the needle, which is not known exactly but is assumed to be uniformly distributed between 0 and $\ell > 0$. While the needle may show up on an X-ray, the doctor recognizes that the random orientation of the needle within the patient's stomach implies the needle's length on film could be misleading. The doctor has asked you to analyze this scenario and form an estimate of the needle's true length based on its projected length in the X-ray.

We attach a 3-dimensional coordinate system to the problem such that the origin is at the midpoint of the needle, and the needle lies parallel to the vertical-axis. Our view, then, is described by the random pair of angles (Θ, Ψ), where $\Theta \in [0, 2\pi)$ is the azimuth angle (it describes the orientation in the horizontal-plane) and $\Psi \in [-\pi/2, \pi/2]$ is the elevation angle (it describes the orientation out of the horizontal-plane).



Let X be the true length of the needle, Y be the projected length of the needle, Θ be the azimuth angle, and Ψ be the elevation angle. We see that the azimuth angle does not affect the length of the projection and that the projected length is given by

$$Y = X \cos \Psi.$$

It is evident that the problem is symmetric about the xy-plane. Hence an elevation angle of Ψ is equivalent to an elevation angle of $-\Psi$. Thus, we can define W to be $W = |\Psi|$, so W is uniformly-distributed between 0 and $\pi/2$.

(a) Determine the *least-squares estimate* of the needle's true length given the value of the needle's projected length in a single X-ray. In other words, if X were to denote the needle's true length and Y its length as measured from the resulting X-ray, determine the mathematical formula for $\mathbf{E}[X|Y]$. Proceed in the order implied by the steps below:

i. Derive $f_{Y|X}(y \mid x)$, the conditional PDF of Y given $X = x$.
ii. Determine $f_Y(y)$, the unconditional PDF of Y.
iii. Determine $f_{X|Y}(x \mid y)$, the conditional PDF of X given $Y = y$, and then compute $\mathbf{E}[X|Y = y]$ directly.
iv. Plot $\mathbf{E}[X|Y = y]$ as a function of y.

Hint: You may find the following integral formulas useful:

$$\int_a^b \frac{1}{\sqrt{\alpha^2 - c^2}} d\alpha = \ln\left(\alpha + \sqrt{\alpha^2 - c^2}\right)\Big|_a^b \qquad \int_a^b \frac{\alpha}{\sqrt{\alpha^2 - c^2}} d\alpha = \sqrt{\alpha^2 - c^2}\Big|_a^b$$

(b) i. Determine the *linear least-squares estimate* for X, given Y. Also compute the resulting estimation error. *Hint:* Consider using the Law of Iterated Expectations to simplify the necessary calculations.

ii. Plot your result on the same graph as part (a)(iv).

(c) Consider a unit sphere centered at the origin, and the viewer is viewing from the North pole, (0,0,1). Assume that the needle has a dull end and a sharp end, and $\ell < 1$ (so the needle lies entirely within the sphere).

One criterion for the orientation of the needle to be truly random is for the intersection of the sphere with the straight line extended from the sharp end to have a uniform distribution on the sphere.

Is the distribution of angles in this problem such that the intersection point would be uniformly distributed on the sphere? If not, can you change the distribution of angles such that it would be?