CS 70 Spring 2023 — Discrete Mathematics and Probability Theory — HW 14

Babak Ayazifar and Satish Rao

Due: Saturday 4/29, 4:00 PM
Grace period until Saturday 4/29, 6:00 PM

Sundry

Before you start writing your final homework submission, state briefly how you worked on it. Who else did you work with? List names and email addresses. (In case of homework party, you can just describe the group.)

1 It's Raining Fish

Note 19
Note 21

A hurricane just blew across the coast and flung a school of fish onto the road nearby the beach. The road starts at your house and is infinitely long. We will label a point on the road by its distance from your house (in miles). For each $n \in \mathbb{N}$, the number of fish that land on the segment of the road $[n, n+1]$ is independently Poisson(λ) and each fish that is flung into that segment of the road lands uniformly at random within the segment. Keep in mind that you can cite any result from lecture or discussion without proof.

(a) What is the distribution of the number of fish arriving in segment $[0, n]$ of the road, for some $n \in \mathbb{N}$?

(b) Let $[a, b]$ be an interval in $[0, 1]$. What is the distribution of the number of fish that lands in the segment $[a, b]$ of the road?

(c) Let $[a, b]$ be any interval such that $a \geq 0$. What is the distribution of the number of fish that land in $[a, b]$?

(d) Suppose you take a stroll down the road. What is the distribution of the distance you walk (in miles) until you encounter the first fish?

(e) Suppose you encounter a fish at distance x. What is the distribution of the distance you walk until you encounter the next fish?

2 Noisy Love

Note 21

Suppose you have confessed to your love interest on Valentine's Day and you are waiting to hear back. Your love interest is trying to send you a binary message: "0" means that your love interest is not interested in you, while "1" means that your love interest reciprocates your feelings. Let X be your love interest's message for you. Your current best guess of X has $\mathbb{P}(X=0)=0.7$ and $\mathbb{P}(X=1)=0.3$. Unfortunately, your love interest sends you the message through a noisy channel, and instead of receiving the message X, you receive the message $Y=X+\varepsilon$, where ε is independent Gaussian noise with mean 0 and variance 0.49.

(a) First, you decide upon the following rule: if you observe $Y>0.5$, then you will assume that your love interest loves you back, whereas if you observe $Y\leq 0.5$, then you will assume that your love interest is not interested in you. What is the probability that you are correct using this rule? (Express your answer in terms of the CDF of the standard Gaussian distribution $\Phi(z)=\mathbb{P}(\mathcal{N}(0,1)\leq z)$, and then evaluate your answer numerically.)

(b) Suppose you observe $Y=0.6$. What is the probability that your love interest loves you back? [*Hint*: This problem requires conditioning on an event of probability 0, namely, the event $\{Y=0.6\}$. To tackle this problem, think about conditioning on the event $\{Y\in[0.6,0.6+\delta]\}$, where $\delta>0$ is small, so that $f_Y(0.6)\cdot\delta\approx\mathbb{P}(Y\in[0.6,0.6+\delta])$, and then apply Bayes Rule.]

(c) Suppose you observe $Y=y$. For what values is it more likely than not that your love interest loves you back? [*Hint*: As before, instead of considering $\{Y=y\}$, you can consider the event $\{Y\in[y,y+\delta]\}$ for small $\delta>0$. So, when is $\mathbb{P}(X=1\mid Y\in[y,y+\delta])\geq\mathbb{P}(X=0\mid Y\in[y,y+\delta])$?]

(d) Your new rule is to assume that your love interest loves you back if (based on the value of Y that you observe) it is more likely than not that your love interest loves you back. Under this new rule, what is the probability that you are correct?

3 Chebyshev's Inequality vs. Central Limit Theorem

Note 17
Note 21

Let n be a positive integer. Let $X_1,X_2,\ldots,X_n$ be i.i.d. random variables with the following distribution:

$$\mathbb{P}[X_i=-1]=\frac{1}{12};\qquad \mathbb{P}[X_i=1]=\frac{9}{12};\qquad \mathbb{P}[X_i=2]=\frac{2}{12}.$$

(a) Calculate the expectations and variances of X_1, $\sum_{i=1}^n X_i$, $\sum_{i=1}^n (X_i-\mathbb{E}[X_i])$, and

$$Z_n=\frac{\sum_{i=1}^n (X_i-\mathbb{E}[X_i])}{\sqrt{n/2}}.$$

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n|\geq 2]$.

(c) Can you use b to bound $\mathbb{P}[Z_n\geq 2]$ and $\mathbb{P}[Z_n\leq -2]$?

(d) As $n\to\infty$, what is the distribution of Z_n?

(e) We know that if $Z\sim\mathcal{N}(0,1)$, then $\mathbb{P}[|Z|\leq 2]=\Phi(2)-\Phi(-2)\approx 0.9545$. As $n\to\infty$, can you provide approximations for $\mathbb{P}[Z_n\geq 2]$ and $\mathbb{P}[Z_n\leq -2]$?

4 Uniform Estimation

Note 17
Note 21

Let $U_1, \ldots, U_n \overset{\text{iid}}{\sim} \text{Uniform}(-\theta, \theta)$ for some unknown $\theta \in \mathbb{R}$, $\theta > 0$. We wish to estimate θ from the data $U_1, \ldots, U_n$.

(a) Why would using the sample mean $\overline{U} = \frac{1}{n}\sum_{i=1}^{n} U_i$ fail in this situation?

(b) Find the PDF of U_i^2 for $i \in \{1, \ldots, n\}$.

(c) Consider the following variance estimate:

$$V = \frac{1}{n}\sum_{i=1}^{n} U_i^2.$$

Show that for large n, the distribution of V is close to one of the famous ones, and provide its name and parameters.

(d) Use part (c) to construct an unbiased estimator for θ^2 that uses all the data.

(e) Let $\sigma^2 = \text{Var}[U_i^2]$. We wish to construct a confidence interval for θ^2 with a significance level of δ, where $0 < \delta < 1$.

(i) Without any assumption on the magnitude of n, construct a confidence interval for θ^2 with a significance level of δ using your estimator from part (d).

(ii) Suppose n is large. Construct an approximate confidence interval for θ^2 with a significance level of δ using your estimator from part (d). You may leave your answer in terms of Φ and Φ^{-1}, the normal CDF and its inverse.