CSE 312: Foundations of Computing II Summer 2024

Homework 6

Due: Monday, August 12, by 11:59pm

Instructions

See the instructions and FAQ for homeworks on the course website for important notes on the submission format!

Solutions submission. You must submit your solution via Gradescope. In particular:

- Submit a *single* PDF containing the solution to all of Tasks 1-8 to Gradescope under "**HW6 [Written]**".
- Each numbered task should be solved on its own page (or pages). Follow the prompt on Gradescope to link tasks to your pages.
- Do not write your name on the individual pages – Gradescope will handle that.

Task 0 – Collaborators [0 pts]

List the full names of anyone you collaborated with on this homework. If you did not collaborate with anyone, write "None" in this section.

Task 1 – Joining the Dots [15 pts]

A pet food company tracks the percentage of daily protein intake and carbohydrate intake in different pet food formulas. Let X be the protein intake and Y be the carbohydrate intake, both continuous random variables between 0 and 1. They had found a joint density function for X and Y, but lost all the data they used to come up with this density function! So, they need your help to analyze the relationship between X and Y. The joint density is: [1]

$$f_{X,Y}(x,y) = \begin{cases} 6x^2y & 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

(Observe that this is a probability density function since it is non-negative and we can use nested integrals to show that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(xy)\,\mathrm{d}y\,\mathrm{d}x = \int_0^1\int_0^1 6x^2y\,\mathrm{d}y\,\mathrm{d}x = \int_0^1 \left(3x^2y^2\Big|_{y=0}^{y=1}\right)\mathrm{d}x = \int_0^1 3x^2\,\mathrm{d}x = x^3\Big|_{x=0}^{x=1} = 1.)$$

Your answers below should **not** be evaluated unless otherwise specified. Your answers should usually be in terms of integrals or nested double integrals.

a) (5 points) The company is very doubtful of the formulas used in the pet food and wants to find the probability that the percentage intake of carbohydrates is more than twice the percentage intake of protein. Write an expression using nested integrals that we can evaluate to find $\mathbb{P}(Y \geqslant 2X)$.
Hint: draw the region of the joint density, and the desired region.

b) (5 points) Write an expression using a single integral that we can evaluate to find the marginal density $f_X(x)$. Be sure to specify the value of $f_X(x)$ for all $x \in \mathbb{R}$. Do the same for $f_Y(y)$.

c) (5 points) Are X and Y independent? Justify your answer. (You may need to evaluate an integral or two to do this.)

[1] These are completely made up probabilities, so ignore if it doesn't make sense in real life!

Example/Note

For questions with the law of total expectation, you MUST explicitly show what events you are partitioning into to use the law of total expectation.

Task 2 – Carrots and Crops [12 pts]

Peter Rabbit is in a large garden with 4 paths, and each path is equally likely to be chosen.

- The first path will exit the garden[2] after a number of hours which is Geometric with parameter $\frac{1}{6}$.
- The second path will exit the garden after a number of hours which is Poisson with parameter 2.
- The third path will exit the garden after a number of hours which is Exponential with parameter $\lambda = \frac{1}{7}$.
- The fourth path leads to a path which brings him back to where he started after 3 hours.

Use the law of total expectation to compute the expected number of hours until Peter Rabbit exits the garden.

Task 3 – Buggy Expectations [15 pts]

A software development team of 15 developers is working on a large codebase. They have a bug tracking system and the number of bugs that they find follows a Poisson random variable with mean 6. Each developer chooses one of the bugs to work on uniformly at random and independent of each other's choices (these developers are not great at communicating and may end up working on the same task). The developers simultaneously work on fixing the bugs. Each developer independently fixes their chosen bug with probability 0.8, use the law of total expectation to compute the expected number of bugs that are fixed (a bug is fixed if at least one developer fixes it).
Hint: Use law of total expectation, partitioning the number of bugs that are found.

Example/Note

In questions involving using tail bounds, you MUST explicitly show all the steps in manipulating the original expression into a form that the tail bound can be used for. Below is an example solution that would get full credit (also reference the section solutions!):

Let X be a random variable with mean 5 and variance 2. Use Chebyshev's inequality to bound $\mathbb{P}(X \leqslant 3)$.

$$
\begin{aligned}
\mathbb{P}(X \leqslant 3) &= \mathbb{P}(X - 5 \leqslant -2) && \text{Subtract } \mathbb{E}[X] = 5 \text{ from both sides} \\
&\leqslant \mathbb{P}(|X - 5| \leqslant -2) + \mathbb{P}(|X - 5| \geqslant 2) && \text{Adding the probability of the right tail} \\
&= \mathbb{P}(|X - 5| \geqslant 2) \\
&\leqslant 2/2^2 = 0.5 && \text{Applying Chebyshev's bound}
\end{aligned}
$$

[2] Peter Rabbit gets quite distracted stealing and eating the delicious carrots in the garden.

Task 4 – Frosting Requires Concentration [20 pts]

Chef Bob is slowly decorating 150 cupcakes for a big party. It takes an average of 3 minutes for Chef Bob to frost and decorate each cupcake, with a standard deviation of 0.5 minutes. The time to frost and decorate each cupcake is independent.

a) What is the expectation of the total time to frost and decorate all the cupcakes?

b) What is the variance of the total time to frost and decorate all the cupcakes?

c) Bob will have 10 hours (600 minutes) to frost and decorate all of the cupcakes between now and when his customer is coming to pick them up. Use Markov's Inequality to give a *lower bound* on the probability that Bob finishes the cupcakes before the customer comes to pick them up.

d) Can we improve the lower bound from **c)** using Chebyshev's inequality? What bound do you get?

Task 5 – Calling Call Centers [15 pts]

Suppose, there is a hotline that connects callers to one of 20 response stations. Over some time period, the hotline received 500 calls. Each call is equally likely to be routed to any of the 20 stations, and every call is successfully connected to a station, without any being left unanswered.

a) Use the Chernoff bound from class to bound the probability that Station 1 receives fewer than 11 calls. [7 points]

b) Is the probability of Station 1 having fewer than 11 calls independent of the probability that the student in the Station 2 has fewer than 11 calls? Briefly explain (you may give a formal derivation/calculation as an explanation or an informal one). [3 points]

c) We want to understand the probability of the calls being distributed very unevenly. Bound the probability that at least one station has fewer than 11 calls. Give the best bound you can from your answers in (a) and (b) (you must use a bound to compute this, rather than giving). [5 points]

The content in the last 3 questions will be covered by the end of Monday, August 5'th lecture.

Example/Note

For questions involve MLEs, you MUST explicitly show each of the steps: finding the likelihood function (make sure to justify how you get the likelihood function), finding the log-likelihood function, taking the derivative(s), and setting it to 0 and solving for the MLE(s).

Task 6 – For Hack-ademic Purposes [12 pts]

After taking CSE 484 (security), you are excited to hack IntoBit, a new business that has had a VERY profitable recommendation system. After some hacking, you have learned that they use the following probabilistic method: Each customer will independently be assigned one of the following *popular* items:

- An umbrella hat with probability θ_1
- A bowl of raw brussel sprouts with probability θ_2
- A used 1-inch pencil with probability $\frac{1}{2}$
- A rubber chicken handbag with probability $\frac{1}{2} - \theta_1 - \theta_2$

You use this recommendation system 100 times, and get recommended the umbrella hat 20 times, the bowl of brussel sprouts 25 times, the pencil 35 times, and the rubber chicken handbag 20 times. Find the maximum likelihood estimates for θ_1 and θ_2. Give exact answers as simplified fractions. You do not need to check second order conditions.

Task 7 – To be continued... [24 pts]

Given $\theta > 0$. Suppose that $x_1, \ldots, x_n$ are i.i.d. realizations (aka samples) from the model

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find the maximum likelihood estimate for θ. You do not need to check second order conditions.

Task 8 – Uniformly Unbiased? [10 pts]

Let $X_1, \ldots, X_n$ be independent samples from Unif$(0, \theta)$, the continuous uniform distribution on $[0, \theta]$. We stumble across a possible estimator for θ, where $\hat{\theta} = 2 \cdot X_1$.
Is this $\hat{\theta}$ an unbiased estimator of θ?