Take-Home Quiz information
Each quiz will be a 47-hour open-book, open-note exam. We have designed this quiz to approximate about 1.5 hours of active work (before typesetting).

- You can submit multiple times; we will only grade the last submission you submit before 10:00 am (Pacific time) on Friday, June 4th. No late submissions can be accepted. When uploading, please assign pages to each question. You should upload your submission as a PDF to Gradescope. We provide a LaTeX template if you find it useful, but we will accept any legible submission.

- Course staff assistance will be limited to clarifying questions of the kind that might be allowed on a traditional, in-person exam. If you have questions during the exam, please ask them as private posts via our discussion forum. We will not have any office hours for answering quiz questions during the quiz, and we can’t answer any questions about course material while the quiz is out.

- For each problem, briefly explain/justify how you obtained your answer at a level such that a future CS109 student would be able to understand how to solve the problem. If it’s not fully clear how you arrived at your answer, you will not receive full credit. It is fine for your answers to be a well-defined mathematical expression including summations, products, factorials, exponents, combinations, and integrals, unless the question specifically asks for a numeric quantity or closed form. Where numeric answers are required, fractions are fine.

Honor Code Guidelines for Take-Home Quizzes
This exam must be completed individually. It is a violation of the Stanford Honor Code to communicate with any other humans about this exam (other than CS109 course staff), to solicit solutions to this exam, or to share your solutions with others.

The take-home exams are open-book: open lecture notes, handouts, textbooks, course lecture videos, and internet searches for conceptual information (e.g., Wikipedia). Consultation of other humans in any form or medium (e.g., communicating with classmates, asking questions on sites like Chegg or Stack Overflow) is prohibited. All work done with the assistance of any external material in any way (other than provided CS109 course materials) must include citation (e.g., “Referred to Wikipedia page on X for Question 2.”). Copying solutions is unacceptable, even with citation. If by chance you encounter solutions to the problem, navigate away from that page before you feel tempted to copy.

If you become aware of any Honor Code violations by any student in the class, your commitments under the Stanford Honor Code obligate you to inform course staff. Please remember that there is no reason to violate your conscience to complete a take-home exam in CS109.

I acknowledge and accept the letter and spirit of the Honor Code:

Name (typed or written): ____________________________
1 Burrow Smoke Detectors and Joint Probability Distributions [10 points]

Burrow Labs has taken on other startups in the home safety and security space and has recently started marketing a new smoke detector. Burrow’s smoke detectors rely on CO₂ sensors that eventually fail, and that failure time dictates the average product lifetime of the smoke detector. Burrow manufactures three quarters of its smoke detectors in central Idaho, and the rest are manufactured in suburban Maine. Any single smoke detector’s product lifetime can be modeled as an Exponential random variable.

Each of the two locations sources its CO₂ sensors from different suppliers, so the smoke detectors manufactured in Maine have an average product lifetime of 7 years and the smoke detectors manufactured in Idaho have an average product lifetime of 6 years. All smoke detectors are sold online, so aside from the fact that a smoke detector is three times more likely to ship from the Idaho facility, you can’t tell by looking at a single smoke detector where it was manufactured.

Let $T$ model the amount of time that passes until the CO₂ sensor (and therefore the smoke detector) fails, and let $M$ be a discrete random variable that takes on the value of 1 for a smoke detector manufactured in Maine, and 0 otherwise.

a. (6 points) Present the cumulative distribution and probability density functions for the random variable $T$. Both your CDF and your PDF should be analytic functions on $t$.

**Answer.**

b. (4 points) Compute the probability that a smoke detector was manufactured in Maine, given that it lasts more than 15 years. If needed, you can keep your answer in terms of $F_T(15)$ or $f_T(15)$ from part (a). However, any conditional expression of the form $P(\cdot | \cdot)$ or $f(\cdot | \cdot)$ must be evaluated.

**Answer.**
2 Desalination Plants, Water Supply, and the CLT [8 points]

Desalination is the process of removing salt and other minerals from seawater to produce water that is drinkable. As potable drinking water sources dwindle, many countries are exploring sustainable ways to power desalination plants to supplement existing potable water supplies. However, some countries restrict the use of desalination plants to small, isolated communities that have no other access to fresh drinking water. In some cases, desalination plant are so small scale that they produce only 5,000 gallons of drinkable water each day. That’s enough potable drinking water for about 500 households.

Assume that the water consumption of a community on any particular day can be modeled by a continuous random variable, and the consumption between any two days are independent. The community consumes an average of 5,000 gallons of potable drinking water each day, but that the community’s per-day drinking water consumption varies with a standard deviation of 1,000 gallons. The community’s potable water reservoir initially contains 70,000 gallons of desalinated water and it’s reliably supplanted with 5,000 gallons of additional desalinated water per day.

What’s the approximate probability that, after 60 days, the community reservoir stores less than 60,000 gallons of drinkable water? You shouldn’t subdivide days into smaller time units or worry about the remote possibility that the reservoir is ever fully depleted. Your final answer can be left in terms of the $\Phi$ function used to compute probabilities related to the Standard Normal distribution.

Answer.
3 Parameter Estimation and Wealth Distribution [12 points]

The broader field of economics also relies on likelihood estimation and parameter estimation. One continuous probability distribution—one with a long tail as $x$ approaches infinity—is used to model wealth inequality and the socioeconomic problems that stem from it. This probability distribution is given as:

$$f(x|\omega) = \frac{\omega x^{\omega}}{x^{\omega+1}}, \text{ where } x \geq 3, \omega > 1$$

Assume that you’ve observed a sample of i.i.d. random variables $(X_1, X_2, X_3, ..., X_n)$, where each of the $X_i$ is modeled according to the above probability distribution function.

a. (6 points) What is the log-likelihood function $LL(\omega)$ of the sample $(X_1, X_2, X_3, ..., X_n)$? Simplify using properties of logarithms wherever possible.

Answer.

b. (6 points) Set up the equation that would need to be solved in order to compute $\hat{\omega}_{MLE}$. Once you arrive at the equation and have worked through any calculus, you can stop and simply present the equation that can be solved via simple algebraic manipulation.

Answer.
4 Naive Bayes [10 points]

Below is a table relating two input variables $X_1$ and $X_2$ with one output value $Y$. Assume that $X_1$ is conditionally independent of $X_2$ given $Y$. Assume further that $\text{range}(X_1) = \{0, 1\}$ and $\text{range}(X_2) = \{0, 1, 2\}$

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

A Categorical random variable is a generalization of a Bernoulli random variable. Whereas a Bernoulli has two possible outcomes, a Categorical has $k$ possible outcomes and is parameterized by $p_1 \ldots p_k$ subject to the constraint $\sum_i^k p_i = 1$. When $k = 2$, it is equivalent to the Bernoulli. For example, $W \sim \text{Categorical}(0.3, 0.7)$ is the same as $W \sim \text{Bernoulli}(0.7)$.

a. (3 points)
Assume $X_1$ conditioned on $Y$ has the following distributions. Note that the superscripts (e.g. the "a" in $p_1^{(a)}$) do not indicate an exponent, but are just used to distinguish the value of $Y$ to which the parameter corresponds.

(\begin{align*}
(a) & \quad (X_1|Y = 0) \sim \text{Categorical}(p_1^{(a)}, p_2^{(a)}) \\
(b) & \quad (X_1|Y = 1) \sim \text{Categorical}(p_1^{(b)}, p_2^{(b)}) \\
(c) & \quad (X_1|Y = 2) \sim \text{Categorical}(p_1^{(c)}, p_2^{(c)})
\end{align*})

Compute the values for $p_2^{(a)}, p_2^{(b)}, p_2^{(c)}$ using Laplace add-one smoothing.

Answer.

b. (3 points)
Assume $X_2$ conditioned on $Y$ has the following distributions.

(\begin{align*}
(d) & \quad (X_2|Y = 0) \sim \text{Categorical}(p_1^{(d)}, p_2^{(d)}, p_3^{(d)}) \\
(e) & \quad (X_2|Y = 1) \sim \text{Categorical}(p_1^{(e)}, p_2^{(e)}, p_3^{(e)}) \\
(f) & \quad (X_2|Y = 2) \sim \text{Categorical}(p_1^{(f)}, p_2^{(f)}, p_3^{(f)})
\end{align*})

Compute the values for $p_3^{(d)}, p_3^{(e)}, p_3^{(f)}$ using Laplace add-one smoothing.

Answer.
c. (4 points)
Compute \( \arg \max_y P(Y = y|X_1 = 0, X_2 = 2) \) using Laplace add-one smoothing. As part of your calculation, we provide here the values of the following expressions to reduce the algebra you require.

\[
P(Y = 1, X_1 = 0, X_2 = 2) = \frac{15}{224}
\]
\[
P(Y = 2, X_1 = 0, X_2 = 2) = \frac{1}{120}
\]

Answer.