Data 102 Fall 2026 **Midterm 2** **A. Strang**

NAME (FIRST LAST): _____ **SID:** _____

Room: ◯ Mulford, ◯ Gateway, ◯ Pimentel

Seat Number: _____

TIME: 110 minutes

Conditions: closed book/internet; no reference sheet; no calculator/computer

GRADING

- There are 10 questions and 80 points total.
- The exam will be graded for accuracy. Partial credit will be granted.
- Please mark all selection bubbles clearly. Fill in the bubble fully. Do not use x's or check marks.

YOUR WRITTEN WORK

- There is a space for your full name and SID number on each page. **You must fill this in.** It will ensure that we can identify your work during the scanning process.
- If you need scratch paper, please use the backs of the pages of the exam. Note: **only the front pages will be scanned and graded.**
- When working on short-answer problems, clearly indicate your answers, and justify your answers. We will give partial credit based on demonstrated knowledge. Make sure your work is organized and legible. We will not give partial credit for work we cannot follow.
- When working on short-answer problems from practice problems you must justify your answers as when solving the problems from scratch. **Correct answers without justification will not receive full credit.**
- Numerical answers may be left as mathematical expressions (need not be fully simplified) unless we explicitly ask for simplification. For example, $(6 \times 10 + 15 \times 3)/3$.
- When writing an integral involving a density function, only integrate over regions where the density is nonzero. We will check the bounds of your integrals.

HONOR CODE

Exams are individual work. By signing on the line below you verify that all answers are your own and that you will not share the content of exams until answers are released. Unsigned exams may not be graded. **Please abide by UC Berkeley's Honor Code:**

"As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others."

Your signature: _____

Full Name and SID: _____

Part 1: *Answer all parts of questions 1 - 6.*

Question 1 [11 points]. Suppose that Θ, X are discrete, jointly random quantities where X is observable. Suppose that Θ is drawn from a distribution with fixed, but unknown parameters α.

(a) Let $\{X_j = x_j\}_{j=1}^n$ denote a string of i.i.d. observations drawn with the same underlying Θ. In the space below, draw a graphical model representing the relationships between Θ and the collection of observations $\{X_j = x_j\}_{j=1}^n$. Follow all graphical conventions established in class. For space, draw your model working left to right instead of top to bottom.

(b) Suppose we only have a single observation $X_1 = x_1$. Then, consider the distributions:

Marginals:	**(a)** $p_\Theta(\theta) = \Pr(\Theta = \theta)$, **(b)** $p_{X_j}(x) = \Pr(X_j = x)$
Joint:	**(c)** $p_{\Theta, X_j}(\theta, x) = \Pr(\Theta = \theta, X_j = x)$
Conditionals:	**(d)** $p_{X_j \mid \Theta=\theta}(x) = \Pr(X_j = x \mid \Theta = \theta)$, **(e)** $p_{\Theta \mid X_j = x}(\theta) = \Pr(\Theta = \theta \mid X_j = x)$
	(f) $p_{X_{n+1} \mid \{X_j = x_j\}_{j=1}^n}(x) = \Pr(X_{n+1} = x \mid \{X_j = x_j\}_{j=1}^n)$

Given only the observation $X_1 = x_1$ (that is, for $j = 1$), define each function below by filling in the first blank with a letter from (**(a)** - **(f)**), and, when asked, specifying where the function is evaluated.

1. The **prior** distribution is _____.
2. The **likelihood** function is _____ evaluated at _____.
3. The **evidence** is _____ evaluated at _____.
4. The **posterior distribution** is _____.
5. The **prior predictive distribution** is _____.
6. The **posterior predictive distribution** is _____.

(c) Use Bayes rule to provide an **exact** formula for the posterior distribution. Use the notation provided in part (b). Then, express the posterior up to proportionality.

Posterior: _____ = (_____)/(_____)

$\propto$ _____

Full Name and SID: _____

Question 2 [6 points]. Use the same assumptions and notation as in question 1.

(a) Answer the questions below regarding the prior predictive distribution:

(i) The prior predictive distribution is the distribution of (◯ Θ, ◯ X) formed by marginalizing:

(◯ **(a)**, ◯ **(b)**, ◯ **(c)**, ◯ **(d)**, ◯ **(e)**, ◯ **(f)**)

over (◯ Θ, ◯ X) drawn from:

(◯ **(a)**, ◯ **(b)**, ◯ **(c)**, ◯ **(d)**, ◯ **(e)**, ◯ **(f)**)

(ii) A colleague suggests using empirical Bayes to estimate the unknown parameters α. They suggest you select the parameters α which maximize the evidence. Their suggestion is equivalent to:

◯ MLE of α using the prior predictive distribution as the likelihood.
◯ MAP estimation of α using the prior predictive distribution as the prior.

(iii) Another colleague suggests you use a hierarchical Bayesian procedure to estimate α. What would you need to change about the model provided in order to apply a hierarchical Bayesian procedure? You may answer by drawing an updated graphical model below, or by describing the necessary change.

(b) Suppose that, to check your model, you decide to run a posterior predictive check. To run a posterior predictive check you decide to generate samples from the posterior predictive distribution. In the space below, complete the pseudocode for sampling from the posterior predictive distribution.

Algorithm 1 Two-stage sampling from the posterior predictive distribution

Input: Given a prior, likelihood, observed data, ensuing posterior, and desired sample size m:
Output: Return $\{X_j\}_{j=1}^m$ drawn i.i.d. from the posterior predictive distribution

1: **for** $j = 1$ **to** m **do**
2: Sample (◯ Θ_j, ◯ X_j, ◯ Θ_j, X_j) from:

(◯ **(a)**, ◯ **(b)**, ◯ **(c)**, ◯ **(d)**, ◯ **(e)**, ◯ **(f)**)

3: Sample (◯ Θ_j, ◯ X_j, ◯ Θ_j, X_j) from:

(◯ **(a)**, ◯ **(b)**, ◯ **(c)**, ◯ **(d)**, ◯ **(e)**, ◯ **(f)**)

setting (◯ $\theta = \Theta_j$, ◯ $x = X_j$, ◯ $\theta = \Theta_j, x = X_j$)
4: **end for**

Full Name and SID: _____

Question 3 [3 points]. Suppose that $p \in \mathcal{P}$ is a prior distribution in the parametric family of distributions $\mathcal{P}$. Suppose that $\ell \in \mathcal{L}$ is a likelihood function in the parametric family of likelihoods $\mathcal{L}$.

Complete the definition: "*$\mathcal{P}$ is a conjugate for $\mathcal{L}$ if, for any observation, the posterior distribution is*":

(◯ in $\mathcal{L}$, ◯ in $\mathcal{P}$, ◯ in an exponential family)

Question 4 [6 points]. Suppose that Θ, X are discrete, jointly random quantities, Θ is unobservable, and X is observable. Suppose we observe $X = x$.

(a) Complete each sentence below given:

(a) the maximum likelihood estimator, **(b)** the maximum a posteriori estimator,

- The most likely value of Θ given $X = x$ is (◯ **(a)**, ◯ **(b)**)
- The value of Θ that would make the observation $X = x$ most likely is: (◯ **(a)**, ◯ **(b)**)

(b) The posterior mean estimator, $\hat{\theta}(\cdot)$ minimizes the expected:

(◯ absolute error $|\hat{\theta}(X) - \Theta|$, ◯ square error $(\hat{\theta}(X) - \Theta)^2$, ◯ 0-1 loss $\{1$ if $\hat{\theta}(X) \neq \Theta, 0$ else $\}$)

averaging over (select all that are true):

(☐ $X|\Theta = \theta$ for any θ, ☐ $\Theta|X = x$ for any x, ☐ Θ and X jointly)

(c) The posterior median estimator, $\hat{\theta}(\cdot)$ minimizes the expected:

(◯ absolute error $|\hat{\theta}(X) - \Theta|$, ◯ square error $(\hat{\theta}(X) - \Theta)^2$, ◯ 0-1 loss $\{1$ if $\hat{\theta}(X) \neq \Theta, 0$ else $\}$)

averaging over (select all that are true):

(☐ $X|\Theta = \theta$ for any θ, ☐ $\Theta|X = x$ for any x, ☐ Θ and X jointly)

Full Name and SID: _____

Question 5 [7 points].

(a) **Gibbs Sampling:** Suppose that we want to sample $\vec{X} = [X_1, X_2, ..., X_n] \sim p_{\vec{X}}$ but we cannot sample from p jointly. Suppose that we can sample from the conditionals $p_{X_j|\{X_i=x_i\}_{i\neq j}}$. Provide a pseudocode for a Gibbs sampling procedure that approximately samples $\vec{X} \sim p_{\vec{X}}$. You do not need to specify the burn-in or stopping criteria.

(b) **Metropolis-Hastings Sampling:** Suppose that we want to sample $X \sim p$ but we only know p up to proportionality: $p(x) = u(x)/Z$ for some unknown normalization constant Z. Suppose that we can sample from a transition kernel q where $q(x \to x')$ is the probability that a random walk, governed by q, jumps to x', if starting from x. Provide pseudocode for a Metropolis-Hastings procedure that approximately samples $X \sim p$ using only u and q. You may assume that the transition kernel q is symmetric. You do not need to specify the burn-in or stopping criteria.

(c) Suppose that $X(t)$ obeys a discrete time Markov chain with transition probabilities $\Pr(X(t+1) = x'|X(t) = x) = p_{x\to x'}$. In the space below, state the detailed balance condition that, if satisfied, would ensure that the steady state distribution of $X(t)$ is proportional to u.

Full Name and SID: _____

Question 6 [7 points].

(a) A statistical model with features x, regression coefficients β, and prediction parameter $\hat{y}$ is a generalized linear model if (select all that must be true):

- ☐ $\hat{y}(\beta, x)$ is linear in x
- ☐ $\hat{y}(\beta, x)$ is linear in β
- ☐ $g(\hat{y}(\beta, x))$ is linear in x for some link function g
- ☐ $g(\hat{y}(\beta, x))$ is linear in β for some link function g

(b) In the space below, fully define a logistic regression model. Let $x = [1, x_1, x_2, ..., x_d]$ denote the vector of features/predictors, and $\beta = [\beta_0, \beta_1, \beta_2, ..., \beta_d]$ denote the vector of regression coefficients. Let Y denote the random variable you are trying to predict.

Your answer must:

1. Clearly state how β and x are combined to produce a single variable, z.
2. Clearly define both the inverse-link $\hat{y} = g^{-1}(z)$, and link functions, $g(\hat{y})$. You must provide a definition in terms of simple functions (e.g. sums, products, ratios, exponentials, or logarithms).
3. State the distributional model used for Y, and provide an expression for the parameter that depends on x in terms of x, β, and the inverse-link.

Full Name and SID: _____

Part 2: *Complete 2 of the 4 problems. Indicate which problems you want graded below.*

Please Grade: ◯ Question 7, ◯ Question 8, ◯ Question 9, ◯ Question 10

If you fill in more than two bubbles, we will grade the first two problems you mark for grading.

Question 7 [20 points]. Cindy observes one beetle everyday and keeps track of their lengths. The lengths of the beetles come from a uniform distribution on $[0, w]$ where w is an unknown parameter corresponding to the size of the largest possible beetle. Cindy wants to estimate w. She observes lengths of n beetles, and calls their lengths $x_1, \ldots, x_n$.

(a) What is the likelihood function of the observations $x_1, \ldots, x_n$? Express your answer as a function of the parameter w. Your answer must be valid for all $w > 0$, so should include an indicator function. Use $\mathbf{1}(E)$ to denote the indicator function that returns 1 if the event E occurs and zero otherwise.

(b) Use your answer from Part (a) to explain why the maximum likelihood estimate (MLE) for w is the maximum of the observed lengths, that is,

$$\hat{w}_{\text{MLE}} = \max\{x_1, x_2, \ldots, x_n\}$$

Full Name and SID: _____

Question 7 continued.

(c) Cindy decides to instead use a Bayesian approach. She has a prior belief that w follows a *Pareto distribution* with parameters $\alpha, \beta > 0$. If $w \sim \text{Pareto}(\alpha, \beta)$ then the density function of w is:

$$f_W(w) = \frac{\alpha\beta^\alpha}{w^{\alpha+1}} 1(w \geq \beta)$$

Show that the posterior distribution for w is also a Pareto distribution, and compute the parameters as a function of α, β, and the observations $x_1, \ldots, x_n$.

(d) Match the prior parameters α and β, to a valid interpretation selected from the menu below. Your answers should be consistent with the update rule for the posterior parameters.

(a) the minimum length observed in a past sample
(b) the average length observed in a past sample
(c) the maximum length observed in a past sample
(d) the variance in lengths observed in a past sample
(e) the number of observations in a past sample

The parameter α is analogous to: (◯ **(a)**, ◯ **(b)**, ◯ **(c)**, ◯ **(d)**, ◯ **(e)**) and the parameter β is analogous to: (◯ **(a)**, ◯ **(b)**, ◯ **(c)**, ◯ **(d)**, ◯ **(e)**).

Full Name and SID: _____

Question 8 [20 points] Suppose that:

$$\Theta \sim \text{Normal}(\mu_\theta, \sigma_\theta^2), \qquad \{X_i\}_{i=1}^n | \Theta = \theta \underset{\text{i.i.d}}{\sim} \text{Normal}(\theta, \sigma_x^2).$$

You can observe $\{X_i\}_{i=1}^n$ and want to estimate Θ. Then the MLE is the sample average $\hat{\theta}_{\text{MLE}}(\{x_i\}_{i=1}^n) = \frac{1}{n}\sum_{i=1}^n x_i$ and:

$$\Theta | \{X_i = x_i\}_{i=1}^n \sim \text{Normal}\,(a, b)$$

where:

$$a = (1-\lambda)\hat{\theta}_{\text{MLE}}(\{x_i\}_{i=1}^n) + \lambda\mu_\theta, \quad \lambda = \frac{1/\sigma_\theta^2}{n/\sigma_x^2 + 1/\sigma_\theta^2}, \quad \text{and} \quad b = \frac{1}{n/\sigma_x^2 + 1/\sigma_\theta^2}$$

Let $\hat{\theta}_{\text{mean}}(\{x_i\}_{i=1}^n) = \mathbb{E}[\Theta | \{X_i = x_i\}_{i=1}^n] = a$ denote the posterior mean estimator.

(a) Suppose that $\hat{\theta}_{\text{MLE}} \neq \mu_\theta$ is fixed. Then:

- $\lim_{\sigma_\theta \to \infty} \hat{\theta}_{\text{mean}} = (\bigcirc\ \hat{\theta}_{\text{MLE}}, \bigcirc\ \mu_\theta)$
- $\lim_{\sigma_\theta \to 0} \hat{\theta}_{\text{mean}} = (\bigcirc\ \hat{\theta}_{\text{MLE}}, \bigcirc\ \mu_\theta)$
- $\lim_{n \to \infty} \hat{\theta}_{\text{mean}} = (\bigcirc\ \hat{\theta}_{\text{MLE}}, \bigcirc\ \mu_\theta)$

(b) Find the Frequentist bias, $\mathbb{E}_{\{X_i\}_{i=1}^n | \Theta=\theta}[\hat{\theta}_{\text{mean}}(\{X_i\}_{i=1}^n) - \Theta]$, in the posterior mean estimator.

(c) Find the Bayesian bias, $\mathbb{E}_{\Theta, \{X_i\}_{i=1}^n}[\hat{\theta}_{\text{mean}}(\{X_i\}_{i=1}^n) - \Theta]$, in the posterior mean estimator.

Full Name and SID: _____

Question 8 continued.

(d) Find the Frequentist variance in the errors in both estimators:

$$\text{Var}_{\{X_i\}_{i=1}^n|\Theta=\theta}[\hat{\theta}_{\text{MLE}}(\{X_i\}_{i=1}^n) - \Theta] = ?, \qquad \text{Var}_{\{X_i\}_{i=1}^n|\Theta=\theta}[\hat{\theta}_{\text{mean}}(\{X_i\}_{i=1}^n) - \Theta] = ?$$

Based on your answer, which estimator is more stable (i.e. less variable)? (◯ MLE, ◯ Post. Mean)

(e) Using the law of total variance, find the Bayesian variance in the errors for the posterior mean estimator:

$$\text{Var}_{\Theta,\{X_i\}_{i=1}^n}[\hat{\theta}_{\text{mean}}(\{X_i\}_{i=1}^n) - \Theta] = ?$$

Full Name and SID: _____

Question 9 [20 points] Suppose that q is an unnormalized target distribution. Let

$$Z = \int_{\text{all x}} q(x)dx$$

denote the (unknown) normalizing constant such that $p(x) = q(x)/Z$ is the normalized target.

To sample from q, adopt a rejection sampling scheme:

Algorithm 2 Rejection Sampling

Input: Unnormalized target distribution $q(x)$, proposal distribution $f(x)$ chosen such that the support of f includes the support of q, and a constant M such that $Mq(x) \leq f(x)$ for all x.

1: **repeat**
2: Draw sample Y from $f(x)$
3: Compute the ratio $R = \frac{Mq(Y)}{f(Y)}$
4: Draw sample U from Uniform$[0, 1]$
5: **until** $U \leq R$
6: **return** $X = Y$

(a) The unconditional acceptance probability is the chance we accept a sample before we have proposed a specific sample value $Y \sim f$. Under the rejection sampling scheme, what is the unconditional acceptance probability? Simplify fully.

(b) Based on the stated algorithm (or your answer to part a), complete the sentence: "*To maximize the unconditional acceptance probability we should:*

◯ *choose M as large as possible.*"

◯ *choose M as small as possible.*"

Full Name and SID: _____

Question 9 continued.

(c) Given f and q, find, by optimizing over M, the largest possible unconditional acceptance probability. Simplify your answer as far as possible. Your final answer should depend only on p and f.

(d) Suppose you were free to choose any proposal density f (subject to normalization). What choice of f would make rejection sampling as efficient as possible? For that choice of f, what is the unconditional acceptance probability?

(e) Based on your answers to parts (c) and (d), complete the sentence, "*to maximize the efficiency of a rejection sampler, we should pick a proposal distribution that is:*

- ◯ *less variable than the target.*"
- ◯ *as similar to the target as possible.*"
- ◯ *as near to uniform as possible.*"
- ◯ *more variable than the target.*"

Full Name and SID: _____

Question 10 [20 points] You want to estimate the unknown concentration ρ_0 of an infectious microbe in a solution. Consider the following procedure: progressively dilute the original solution, and after each dilution, pour out some small volume v onto a sterile plate. If zero microbes land on the plate, it will remain sterile, but if any microbes land on a plate, they will infect it. After waiting for microbe growth, you can observe whether a plate was infected ($Y_t = 1$) or is sterile ($Y_t = 0$).

Let ρ_t denote the concentration at dilution t. Each time, dilute the solution to half its concentration. Then:

$$\rho_t = \frac{\rho_0}{2^t} \text{ for } t = 0, 1, 2 \ldots$$

(a) At dilution t, the data $Y_t \in \{0, 1\}$ indicates whether or not the plate is infected. The chance that a plate gets infected is denoted by $\mu(t) := \mathbb{E}[Y_t]$. How is Y_t distributed given $\mu(t)$?

$$Y_t \sim _____$$

(b) At dilution t, you pour out volume v onto a plate, so the expected number of microbes on the plate is $\rho_t v$. Let X_t denote the number of microbes on the plate at dilution t. The actual number of microbes is distributed as a Poisson random variable with mean $\rho_t v$:

$$\Pr(X_t = n) = \frac{(\rho_t v)^n}{n!} e^{-\rho_t v}$$

Write out an expression for $\mu(t) := \mathbb{E}[Y_t]$ in terms of ρ_0, t, and other constants.

Full Name and SID: _____

Question 10 continued.

(c) Find a link function g such that

$$g(\mu(t)) = \beta_0 + \beta_1 t \quad , \quad \mu(t) = g^{-1}(\beta_0 + \beta_1 t)$$

for some constants β_0 and β_1. Your answer should provide the link function, and explicit formula for β_0 and β_1 in terms of ρ_0 and v.

(d) Suppose you've estimated β_0 and β_1 using maximum-likelihood estimation. Call your estimates $\hat{\beta}_0$ and $\hat{\beta}_1$. Write down an estimate of ρ_0 in terms of $\hat{\beta}_0$ and $\hat{\beta}_1$.