PRACTICE MIDTERM EXAM – WITH CORRECTIONS

STAD91 WINTER 2026
University of Toronto Scarborough

Exam duration: 2H
No calculators will be allowed during the midterm exam.
A formula sheet with the usual families of distributions will be provided during the actual midterm.

Read the following instructions carefully:

1. Exam is closed book and internet. You can use an optional handwritten aid sheet (A4 double-sided).
2. If a question asks you to do some calculations, you must show your work for full credit.
3. Conceptual questions do not require long answers.
4. You will write your answers to each question in the space provided on the exam sheet.
5. After solving each question, you should write your answers immediately. Do not wait last minute to write them all at once.
6. Do not share the exam with anyone or in any platform!
7. Lastly, enjoy the problems!!!

1. Exponential Models and Bayes Estimators (30 pts)

Let $n \geqslant 3$. Consider the following Bayesian framework:

$$\theta \sim \Pi = \mathcal{E}(1)$$
$$\mathbf{X}|\theta \sim P_\theta^{\otimes n} = \mathcal{E}(\theta)^{\otimes n}$$

1. Determine the posterior distribution.

   Correction: We have the prior $\pi(\theta) = e^{-\theta}\mathbb{1}_{\theta>0}$ and the likelihood $L(\mathbf{X}|\theta) = \prod_{i=1}^n \theta e^{-\theta X_i} = \theta^n e^{-\theta \sum X_i} = \theta^n e^{-n\overline{X}_n\theta}$.

   The posterior density is proportional to:

   $$\pi(\theta|\mathbf{X}) \propto e^{-\theta} \cdot \theta^n e^{-n\overline{X}_n\theta}\mathbb{1}_{\theta>0} = \theta^n e^{-(n\overline{X}_n+1)\theta}\mathbb{1}_{\theta>0}$$

   We recognize the kernel of a Gamma distribution. Thus:

   $$\theta|\mathbf{X} \sim \Gamma(n+1, n\overline{X}_n + 1)$$

2. Consider the loss function $\ell(\theta, t) = e^\theta(t-\theta)^2$. Give the Bayes estimator $T^* = T^*(\mathbf{X})$ for Π and this loss function.

   Correction: The Bayes estimator minimizes the posterior expected loss.

   $$\rho(\Pi, T|\mathbf{X}) \propto \int_{\mathbb{R}} (T-\theta)^2 \underbrace{\theta^n e^{-\theta n\overline{X}_n}\mathbb{1}_{\theta>0}}_{\propto f_{\Gamma(n+1,n\overline{X}_n)}(\theta)} d\theta$$

   Calculating the expectations using the Gamma density properties (or by observing the modified kernels):

   $$T^* = \mathbb{E}[\Gamma(n+1, n\overline{X}_n)|\mathbf{X}] = \frac{n+1}{n\overline{X}_n}$$

   is the Bayes estimator.

3. Show that, under $P_\theta^{\otimes n}$, the variable $n\overline{X}_n$ follows a Gamma distribution and specify its parameters. Deduce $\mathbb{E}_\theta\left[\frac{1}{n\overline{X}_n}\right]$ and $\mathbb{E}_\theta\left[\frac{1}{(n\overline{X}_n)^2}\right]$.

   Correction: Under $P_\theta^{\otimes n}$, the X_i are i.i.d. $\mathcal{E}(\theta)$. The sum of n independent exponentials follows a Gamma distribution. Therefore, $n\overline{X}_n = \sum X_i \sim \Gamma(n, \theta)$.

   Using the moments of the inverse Gamma distribution (or direct integration):

   $$\mathbb{E}_\theta\left[\frac{1}{n\overline{X}_n}\right] = \frac{\theta}{n-1} \quad \text{and} \quad \mathbb{E}_\theta\left[\frac{1}{(n\overline{X}_n)^2}\right] = \frac{\theta^2}{(n-1)(n-2)}$$

4. Show that the point risk $R(\theta, T^*) = \mathbb{E}_\theta[\ell(\theta, T^*)]$ can be written in the form:

   $$R(\theta, T^*) = \frac{an+b}{(n-1)(n-2)}\theta^2 e^\theta$$

where a and b are two positive constants to be specified.

Correction:

$$\begin{aligned} R(\theta, T^*) &= \mathbb{E}_\theta\left[e^\theta (T^* - \theta)^2\right] = e^\theta \mathbb{E}_\theta\left[\left(\frac{n+1}{n\overline{X}_n} - \theta\right)^2\right] \\ &= e^\theta \mathbb{E}_\theta\left[\left(\frac{n+1}{n\overline{X}_n}\right)^2 - 2\theta\frac{n+1}{n\overline{X}_n} + \theta^2\right] \\ &= e^\theta\left((n+1)^2\mathbb{E}_\theta\left[\frac{1}{(n\overline{X}_n)^2}\right] - 2(n+1)\theta\mathbb{E}_\theta\left[\frac{1}{n\overline{X}_n}\right] + \theta^2\right) \end{aligned}$$

Substituting the moments from the previous question and simplifying:

$$R(\theta, T^*) = \frac{n+7}{(n-1)(n-2)}\theta^2 e^\theta$$

Thus, $a = 1$ and $b = 7$.

5. What is the Bayes risk $R_B(\Pi)$? Deduce the minimax risk R_M.

Correction:

$$R_B(\Pi) = \mathbb{E}[R(\theta, T^*)] = \int_0^{+\infty} R(\theta, T^*)\pi(\theta)d\theta = \int_0^{+\infty} \frac{n+7}{(n-1)(n-2)}\theta^2 e^\theta e^{-\theta} d\theta = +\infty$$

Since the Bayes risk is smaller than the minimax risk ($R_B \leqslant R_M$), we deduce that $R_M = +\infty$.

6. We wish to test

$$H_0 : \theta = 1 \quad \text{against} \quad H_1 : \theta = 2.$$

To do this, we now consider the prior $\Pi = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_2$ and the balanced loss function. Show that the Bayes test can be expressed in the form $\varphi^*(\mathbf{X}) = \mathbb{1}_{\{\overline{X}_n \leqslant c\}}$, for a certain constant $c \in \mathbb{R}$ that you will determine.

Correction: From the course, the Bayes test for balanced loss is given by:

$$\varphi^*(\mathbf{X}) = \mathbb{1}_{\{\Pi(\{1\}|\mathbf{X}) \leqslant \frac{1}{2}\}}$$

Or equivalently comparing the weighted likelihoods:

$$\begin{aligned} \Pi(\{1\}|\mathbf{X}) \leqslant \frac{1}{2} &\iff \frac{\Pi(\{1\})p_1(\mathbf{X})}{\Pi(\{1\})p_1(\mathbf{X}) + \Pi(\{2\})p_2(\mathbf{X})} \leqslant \frac{1}{2} \\ &\iff p_1(\mathbf{X}) \leqslant p_2(\mathbf{X}) \\ &\iff 1^n e^{-n\overline{X}_n} \leqslant 2^n e^{-2n\overline{X}_n} \\ &\iff e^{n\overline{X}_n} \leqslant 2^n \\ &\iff n\overline{X}_n \leqslant n\ln 2 \\ &\iff \overline{X}_n \leqslant \ln 2 \end{aligned}$$

Thus, $\varphi^*(\mathbf{X}) = \mathbb{1}_{\{\overline{X}_n \leqslant \ln(2)\}}$, so $c = \ln(2)$.

2. Gaussian Model and Empirical Bayes (25 pts)

Let $\lambda > 0$ and $n \geqslant 5$ be an integer. Consider the following Bayesian model:

$$\theta \sim \Pi = \mathcal{E}(\lambda)$$
$$\mathbf{X} = (X_1, \ldots, X_n)|\theta \sim \mathcal{N}\left(0, \frac{1}{\theta}\right)^{\otimes n}$$

1. Determine the posterior distribution $\Pi[\cdot|\mathbf{X}]$. We denote $S_{\mathbf{X}}^2 = \sum_{i=1}^n X_i^2$.

   Correction: For all $\theta > 0$:

   $$\pi(\theta|\mathbf{X}) \propto e^{-\lambda\theta}\theta^{n/2}e^{-\frac{1}{2}\theta\sum X_i^2}\mathbb{1}_{\theta>0} \propto \theta^{n/2}e^{-\theta(\lambda+\frac{1}{2}S_{\mathbf{X}}^2)}\mathbb{1}_{\theta>0}$$

   Therefore:

   $$\Pi[\cdot|\mathbf{X}] = \Gamma\left(1 + \frac{n}{2}, \lambda + \frac{1}{2}S_{\mathbf{X}}^2\right)$$

2. Give a Bayes estimator for the quadratic loss.

   Correction: A Bayes estimator for the quadratic loss is the posterior mean:

   $$T^*(\mathbf{X}) = \mathbb{E}[\theta|\mathbf{X}] = \frac{n/2+1}{\lambda + \frac{1}{2}S_{\mathbf{X}}^2} = \frac{n+2}{2\lambda + S_{\mathbf{X}}^2}$$

3. Consider the estimator $\widehat{\theta}_n(\mathbf{X}) = \frac{n}{S_{\mathbf{X}}^2}$. The objective of this question is to calculate its Bayes risk for Π and the quadratic loss.

   (a) Let $\theta > 0$. Given $\theta = \theta$, what distribution does $\theta S_{\mathbf{X}}^2$ follow?

   Correction: Given $\theta = \theta$, the variables $\sqrt{\theta}X_i$ are i.i.d standard Gaussians. Thus:

   $$\theta S_{\mathbf{X}}^2 = \sum_{i=1}^n (\sqrt{\theta}X_i)^2 \sim \chi_n^2$$

   (b) Deduce, for all $\theta > 0$, $\mathbb{E}_\theta[\frac{1}{S_{\mathbf{X}}^2}]$ and $\mathbb{E}_\theta[\frac{1}{S_{\mathbf{X}}^4}]$.

   Correction: Recall that if $Y \sim \chi_d^2$, it is identical to $\Gamma(\frac{d}{2}, \frac{1}{2})$. Using this:

   $$\mathbb{E}_\theta\left[\frac{1}{S_{\mathbf{X}}^2}\right] = \theta\mathbb{E}\left[\frac{1}{\theta S_{\mathbf{X}}^2}\right] = \frac{\theta}{n-2}$$
   $$\mathbb{E}_\theta\left[\frac{1}{S_{\mathbf{X}}^4}\right] = \theta^2\mathbb{E}\left[\frac{1}{(\theta S_{\mathbf{X}}^2)^2}\right] = \frac{\theta^2}{(n-2)(n-4)}$$

   (c) Deduce that for all $\theta > 0$, the point risk is $R(\theta, \widehat{\theta}_n(\mathbf{X})) = \frac{2(n+4)}{(n-2)(n-4)}\theta^2$.

   Correction:

   $$\begin{aligned}\mathbb{E}_\theta\left[\left(\theta - \frac{n}{S_{\mathbf{X}}^2}\right)^2\right] &= \theta^2 - 2n\theta\mathbb{E}_\theta\left[\frac{1}{S_{\mathbf{X}}^2}\right] + n^2\mathbb{E}_\theta\left[\frac{1}{S_{\mathbf{X}}^4}\right] \\ &= \theta^2\left(1 - \frac{2n}{n-2} + \frac{n^2}{(n-2)(n-4)}\right) \\ &= \frac{2(n+4)}{(n-2)(n-4)}\theta^2\end{aligned}$$

(d) Deduce the Bayes risk for Π of the estimator $\widehat{\theta}_n(\mathbf{X})$.

Correction: Since $\theta \sim \mathcal{E}(\lambda)$, we know $\mathbb{E}[\theta^2] = 2/\lambda^2$. Thus:

$$R_B(\Pi, \widehat{\theta}_n) = \frac{2(n+4)}{(n-2)(n-4)}\mathbb{E}[\theta^2] = \frac{4(n+4)}{(n-2)(n-4)} \times \frac{1}{\lambda^2}$$

4. Use the Empirical Bayes method to calibrate λ. Show that for all $\lambda > 0$, the marginal density $f_\lambda(\mathbf{X})$ satisfies:

$$f_\lambda(\mathbf{X}) \propto \frac{\lambda}{(\lambda + \frac{S_{\mathbf{X}}^2}{2})^{\frac{n}{2}+1}}$$

and determine the marginal maximum likelihood estimator of λ.

Correction: For all $\lambda > 0$, using the change of variable $\theta' = \theta(\lambda + \frac{1}{2}S_{\mathbf{X}}^2)$:

$$f_\lambda(\mathbf{X}) \propto \int_0^{+\infty} \lambda \theta^{n/2} e^{-\theta(\lambda + \frac{1}{2}S_{\mathbf{X}}^2)} d\theta \propto \frac{\lambda}{(\lambda + \frac{1}{2}S_{\mathbf{X}}^2)^{\frac{n}{2}+1}} \int_0^{+\infty} \theta'^{n/2} e^{-\theta'} d\theta' \propto \frac{\lambda}{(\lambda + \frac{1}{2}S_{\mathbf{X}}^2)^{\frac{n}{2}+1}}$$

Passing to the log-likelihood:

$$\ell_\lambda(\mathbf{X}) = c + \log(\lambda) - (n/2+1)\log(\lambda + \frac{1}{2}S_{\mathbf{X}}^2)$$

Differentiating with respect to λ:

$$\ell'_\lambda(\mathbf{X}) = \frac{1}{\lambda} - \frac{n/2+1}{\lambda + \frac{1}{2}S_{\mathbf{X}}^2}$$

Setting to 0 implies:

$$\lambda + \frac{1}{2}S_{\mathbf{X}}^2 = \lambda(n/2+1) \iff \frac{1}{2}S_{\mathbf{X}}^2 = \frac{n}{2}\lambda \iff \lambda = \frac{S_{\mathbf{X}}^2}{n}$$

The maximum is attained at $\widehat{\lambda}(\mathbf{X}) = \frac{1}{n}S_{\mathbf{X}}^2$.

5. Give the pseudo-posterior distribution obtained by the empirical Bayes method in this case. What do we find by considering the associated pseudo-posterior mean?

Correction: The pseudo-posterior obtained by empirical Bayes is $\Gamma(\frac{n+2}{2}, \frac{n+2}{2n}S_{\mathbf{X}}^2)$, which, given $\mathbf{X}$, has expectation $\widehat{\theta}_n(\mathbf{X}) = \frac{n}{S_{\mathbf{X}}^2}$.

3. Poisson Distribution and improper priors (25 pts)

Let X be a random variable following a Poisson distribution $\mathcal{P}(\theta)$ with $\theta \in \mathbb{R}_+^*$, and $x_1, \ldots, x_n$ a sample from this distribution.

1. Determine the Jeffreys prior measure $\pi^J(\theta)$.

   Correction: The Jeffreys prior is given by $\pi^J(\theta) \propto \sqrt{I(\theta)}$, where $I(\theta)$ is the Fisher information. For $X \sim \mathcal{P}(\theta)$, the likelihood is $f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}$. The log-likelihood is $\ell(\theta) = -\theta + x \ln \theta - \ln(x!)$. The second derivative is $\frac{\partial^2}{\partial \theta^2}\ell(\theta) = -\frac{x}{\theta^2}$. The Fisher information is $I(\theta) = -\mathbb{E}[-\frac{X}{\theta^2}] = \frac{\theta}{\theta^2} = \frac{1}{\theta}$. Thus, $\pi^J(\theta) \propto \sqrt{1/\theta} = \theta^{-1/2}$.

2. Evaluate, based on the existence of posterior distributions, which prior is more suitable between $\pi_0 \propto 1/\theta$ and π^J.

   Correction: The posterior density is proportional to likelihood $\times$ prior. The likelihood is proportional to $e^{-n\theta}\theta^{\sum x_i}$.

   - For $\pi_0 \propto \theta^{-1}$: The posterior is $\propto e^{-n\theta}\theta^{S_n - 1}$ where $S_n = \sum x_i$. After a change of variable $\theta' = n\theta$, this integrates to a quantity equal to $\Gamma(S_n)$ up to a multiplicative factor. If $S_n = 0$ (all $x_i = 0$), the integral diverges (Gamma(0) is undefined).
   - For $\pi^J \propto \theta^{-1/2}$: The posterior is $\propto e^{-n\theta}\theta^{S_n - 1/2}$. Since $S_n \geqslant 0$, the exponent $S_n - 1/2 \geqslant -0.5 > -1$, so the integral always converges to a proper Gamma distribution.

   Thus, π^J is more suitable as it yields a proper posterior even when $S_n = 0$.

3. Let $\pi_\alpha(\theta) \propto \theta^{-\alpha}$ with $\alpha \in \mathbb{R}^+$.

   (a) Show that the posterior distribution $\pi_\alpha(\theta|x_1, \ldots, x_n)$ is a Gamma distribution and specify its shape parameter A and rate parameter B.

   Correction:

   $$\pi_\alpha(\theta|\mathbf{x}) \propto \theta^{-\alpha} \prod \frac{e^{-\theta}\theta^{x_i}}{x_i!} \propto \theta^{-\alpha} e^{-n\theta}\theta^{S_n} = \theta^{(S_n - \alpha + 1) - 1} e^{-n\theta}$$

   This is the kernel of a Gamma distribution $\Gamma(A, B)$ with:

   $$A = S_n - \alpha + 1 \quad \text{and} \quad B = n$$

   (Note: Requires $A > 0$).

   (b) Write down the integral definition for the posterior predictive mass function $P(X_{new} = k|x_1, \ldots, x_n)$ using the posterior density you found in the previous step.

   Correction:

   $$P(X_{new} = k|\mathbf{x}) = \int_0^\infty P(X_{new} = k|\theta)p(\theta|\mathbf{x})d\theta = \int_0^\infty \frac{e^{-\theta}\theta^k}{k!} \frac{B^A}{\Gamma(A)} \theta^{A-1} e^{-B\theta} d\theta$$

   (c) Compute the integral to find the closed-form expression for $P(X_{new} = k|x_1, \ldots, x_n)$ in terms of Gamma functions. Identify the name of this known distribution.

Correction: Below, use the change-of-variable $\theta \to (B+1)\theta$:

$$\begin{aligned} P(X_{new} = k|\mathbf{x}) &= \frac{B^A}{k!\Gamma(A)} \int_0^\infty \theta^{k+A-1} e^{-(B+1)\theta} d\theta \\ &= \frac{B^A}{k!\Gamma(A)} \frac{\Gamma(k+A)}{(B+1)^{k+A}} \\ &= \frac{\Gamma(k+A)}{k!\Gamma(A)} \left(\frac{B}{B+1}\right)^A \left(\frac{1}{B+1}\right)^k \end{aligned}$$

This is the probability mass function of the **Negative Binomial** distribution $NB(r,p)$ with parameters $r = A$ and success probability $p = \frac{B}{B+1}$.

(d) Provide the expressions for its expectation and variance, and state their conditions for existence.

Correction: For $Y \sim NB(r,p)$:

$$\mathbb{E}[Y] = \frac{r(1-p)}{p} = \frac{A \cdot \frac{1}{B+1}}{\frac{B}{B+1}} = \frac{A}{B}$$

$$\mathrm{Var}(Y) = \frac{r(1-p)}{p^2} = \frac{A}{B^2}(B+1) = \frac{A(B+1)}{B^2}$$

These exist provided $A > 0$.

4. Bayesian Linear Regression (20 pts)

Consider a simple linear regression model where we assume the intercept is zero and the noise variance σ^2 is known. We observe data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$. The model is:

$$y_i = \beta x_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

We place a Gaussian prior on the slope parameter: $\beta \sim N(\mu_0, \tau_0^2)$.

1. Write down the likelihood $p(\mathbf{y}|\mathbf{x}, \beta)$. Express it as a single multivariate normal distribution.

   Correction: The likelihood is the product of independent Gaussian densities:

   $$p(\mathbf{y}|\mathbf{x}, \beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right)$$

   In vector notation, $\mathbf{y} \sim \mathcal{N}(\mathbf{x}\beta, \sigma^2 I_n)$.

2. Derive the posterior distribution $p(\beta|\mathbf{x}, \mathbf{y})$. Show that $\beta|\mathbf{y} \sim N(\mu_n, \tau_n^2)$ and find expressions for the posterior precision $(1/\tau_n^2)$ and the posterior mean (μ_n).

   Correction:

   $$\begin{aligned} p(\beta|\mathbf{y}) &\propto \pi(\beta) p(\mathbf{y}|\beta) \\ &\propto \exp\left(-\frac{(\beta - \mu_0)^2}{2\tau_0^2}\right) \exp\left(-\frac{\sum(y_i - \beta x_i)^2}{2\sigma^2}\right) \end{aligned}$$

   Expanding the terms in the exponent (focusing on β):

   $$-\frac{1}{2}\left[\beta^2\left(\frac{1}{\tau_0^2} + \frac{\sum x_i^2}{\sigma^2}\right) - 2\beta\left(\frac{\mu_0}{\tau_0^2} + \frac{\sum x_i y_i}{\sigma^2}\right) + C\right]$$

   This matches the exponent a Gaussian density $\mathcal{N}(\mu_n, \tau_n^2)$.

   Posterior Precision:

   $$\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{\sum x_i^2}{\sigma^2}$$

   Posterior Mean:

   $$\mu_n = \tau_n^2\left(\frac{\mu_0}{\tau_0^2} + \frac{\sum x_i y_i}{\sigma^2}\right)$$

3. Interpret the expression for the posterior precision. How does the "sample information" and the "prior information" combine to determine our certainty about β?

   Correction: The posterior precision is the sum of the prior precision $(1/\tau_0^2)$ coming from the prior belief and the data precision (Fisher information, $\sum x_i^2/\sigma^2$), provided by the data.

4. Suppose we use a "flat" prior by letting $\tau_0^2 \to \infty$. Show that the posterior mean μ_n converges to the ordinary least-square estimator $\hat{\beta}_{OLS} = \frac{\sum x_i y_i}{\sum x_i^2}$.

   Correction: As $\tau_0^2 \to \infty$, $1/\tau_0^2 \to 0$. Then $\frac{1}{\tau_n^2} \to \frac{\sum x_i^2}{\sigma^2}$, so $\tau_n^2 \to \frac{\sigma^2}{\sum x_i^2}$. The mean becomes:

   $$\mu_n \to \frac{\sigma^2}{\sum x_i^2}\left(0 + \frac{\sum x_i y_i}{\sigma^2}\right) = \frac{\sum x_i y_i}{\sum x_i^2}$$

   which is exactly $\hat{\beta}_{OLS}$.

5. What is the distribution of the posterior predictive $p(y^*|x^*, \mathbf{x}, \mathbf{y})$ for a new observation y^* at a new given point x^*?

   Correction: We have $y^* = \beta x^* + \epsilon$, with $\beta \sim \mathcal{N}(\mu_n, \tau_n^2)$ and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ independent. Thus y^* is a linear combination of independent Gaussians.

$$\mathbb{E}[y^*|\mathbf{y}] = x^* \mathbb{E}[\beta|\mathbf{y}] = x^* \mu_n$$

$$\mathrm{Var}(y^*|\mathbf{y}) = (x^*)^2 \mathrm{Var}(\beta|\mathbf{y}) + \mathrm{Var}(\epsilon) = (x^*)^2 \tau_n^2 + \sigma^2$$

   So, $y^*|\mathbf{y} \sim \mathcal{N}(x^* \mu_n, \sigma^2 + (x^*)^2 \tau_n^2)$.