# Math 341 / 641 Fall 2025
# Final Examination Solutions

Professor Adam Kapelner

December 16, 2025

Full Name ________________________________________

## Code of Academic Integrity

Since the college is an academic community, its fundamental purpose is the pursuit of knowledge. Essential to the success of this educational mission is a commitment to the principles of academic integrity. Every member of the college community is responsible for upholding the highest standards of honesty at all times. Students, as members of the community, are also responsible for adhering to the principles and spirit of the following Code of Academic Integrity.

Activities that have the effect or intention of interfering with education, pursuit of knowledge, or fair evaluation of a student's performance are prohibited. Examples of such activities include but are not limited to the following definitions:

**Cheating** Using or attempting to use unauthorized assistance, material, or study aids in examinations or other academic work or preventing, or attempting to prevent, another from using authorized assistance, material, or study aids. Example: using an unauthorized cheat sheet in a quiz or exam, altering a graded exam and resubmitting it for a better grade, etc.

I acknowledge and agree to uphold this Code of Academic Integrity.

________________________________________ ____________________

signature date

## Instructions

This exam is 120 minutes (variable time per question) and closed-book. You are allowed **three** pages (front and back) of a "cheat sheet", blank scrap paper (provided by the proctor) and a graphing calculator (which is not your smartphone). Please read the questions carefully. Within each problem, I recommend considering the questions that are easy first and then circling back to evaluate the harder ones. No food is allowed, only drinks.

**Problem 1** Consider the iid Poisson DGP where instead of being parameterized by the usual $\lambda > 0$, we parameterize it by $\theta = g(\lambda) = \ln(\lambda)$ so that $\lambda = e^\theta$ hence,

$$X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\theta) := \frac{e^{-e^\theta}\left(e^\theta\right)^x}{x!}\mathbb{1}_{x\in\mathbb{N}_0}.$$

(a) [2 pt / 2 pts] What is $\Theta$?

$$\lambda \in (0,\infty) \quad \Rightarrow \quad \ln(\lambda) = \theta \in \mathbb{R} = \Theta$$

(b) [4 pt / 6 pts] Show the following and justify each step:

$$\mathcal{L}(\theta; \boldsymbol{X}) = e^{-ne^\theta} e^{n\bar{X}\theta} \prod_{i=1}^n \frac{\mathbb{1}_{X_i\in\mathbb{N}_0}}{X_i!}$$

$$\mathcal{L}(\theta; \boldsymbol{X}) = \prod_{i=1}^n \frac{e^{-e^\theta}\left(e^\theta\right)^{X_i}}{X_i!}\mathbb{1}_{X_i\in\mathbb{N}_0} = e^{-ne^\theta}\left(e^\theta\right)^{\sum_{i=1}^n X_i} \prod_{i=1}^n \frac{\mathbb{1}_{X_i\in\mathbb{N}_0}}{X_i!} = e^{-ne^\theta} e^{n\bar{X}\theta} \prod_{i=1}^n \frac{\mathbb{1}_{X_i\in\mathbb{N}_0}}{X_i!}$$

(c) [3 pt / 9 pts] Find $\ell(\theta; \boldsymbol{X})$.

$$\ell(\theta; \boldsymbol{X}) = -ne^\theta + n\bar{X}\theta + \sum_{i=1}^n \ln\left(\frac{\mathbb{1}_{X_i\in\mathbb{N}_0}}{X_i!}\right)$$

(d) [2 pt / 11 pts] Find $\ell'(\theta; \boldsymbol{X})$.

$$\ell'(\theta; \boldsymbol{X}) = -ne^{\theta} + n\bar{X}$$

(e) [4 pt / 15 pts] Find $\hat{\theta}^{\text{MLE}}$. Simplify.

$$\ell'(\theta; \boldsymbol{X}) \overset{set}{=} 0 \quad \Rightarrow \quad -ne^{\theta} + n\bar{X} = 0 \quad \Rightarrow \quad e^{\theta} = \bar{X} \quad \Rightarrow \quad \hat{\hat{\theta}}^{\text{MLE}} = \ln\left(\bar{X}\right)$$

(f) [2 pt / 17 pts] Find $I_n(\theta)$.

$$I_n(\theta) = \mathbb{E}\left[-\ell''(\theta; \boldsymbol{X})\right] = \mathbb{E}\left[-\left(-ne^{\theta}\right)\right] = ne^{\theta}$$

(g) [2 pt / 19 pts] Find $I(\theta)$.

$$I_n(\theta) = nI(\theta) \quad \Rightarrow \quad ne^{\theta} = nI(\theta) \quad \Rightarrow \quad I(\theta) = e^{\theta}$$

(h) [3 pt / 22 pts] Given your answer in (a), what is the Laplace prior on $\theta$?

Since the parameter space is all $\mathbb{R}$, $f(\theta) \propto 1$.

(i) [1 pt / 23 pts] Is Laplace's prior proper? Yes / no.

No

(j) [3 pt / 26 pts] Show that the kernel of Jeffrey's prior on $\theta$ is $e^{\theta/2}$.

$$f(\theta) \propto \sqrt{I(\theta)} = \sqrt{e^\theta} = \left(e^\theta\right)^{\frac{1}{2}} = e^{\theta/2}$$

(k) [3 pt / 29 pts] Is Jeffrey's prior proper? Yes / no and justify your answer.

No, because $\int_\Theta e^{\theta/2} d\theta = \infty$ and thus Humpty Dumpty cannot be satisfied.

For the rest of this question, you will need to know about the following new rv:

$$Y \sim \text{LogGamma}(a, b) := \frac{b^a}{\Gamma(a)} e^{ay - be^y}, \ a, b > 0, \ \mathbb{E}[Y] = \psi(a) - \ln(b), \ \text{Mode}[Y] = \ln\left(\frac{a}{b}\right)$$

where $\psi$ is called the "digamma" function.

(l) [6 pt / 35 pts] Using Laplace's prior, derive the posterior. If you did not figure out Laplace's prior, let $f(\theta) \propto 1$ going forward. Show the posterior is a LogGamma distribution and find the posterior parameters.

$$\begin{aligned} f(\theta \mid \boldsymbol{x}) &= \frac{f(\boldsymbol{x} \mid \theta) f(\theta)}{f(\boldsymbol{x})} \propto f(\boldsymbol{x} \mid \theta) f(\theta) \propto \mathcal{L}(\theta; \boldsymbol{x}) = \frac{e^{-ne^\theta} \left(e^\theta\right)^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \prod_{i=1}^n \mathbb{1}_{x_i \in \mathbb{N}_0} \\ &\propto e^{-ne^\theta} \left(e^\theta\right)^{\sum_{i=1}^n x_i} \\ &= e^{n\bar{x}\theta - ne^\theta} \\ &\propto \text{LogGamma}(n\bar{x}, n) \end{aligned}$$

(m) [3 pt / 38 pts] Is this posterior always proper? Yes / no and justify your answer.

No. If $\bar{x} = 0$ (which is possible for a Poisson rv), then $b = 0$ and the posterior is improper.

(n) [2 pt / 40 pts] Write Jeffrey's prior as a LogGamma (whether proper or improper).

$$e^{\theta/2} \propto \text{LogGamma}(1/2, 0)$$

(o) [6 pt / 46 pts] Now let $f(\theta) = \text{LogGamma}(a, b)$, the general rv where $a, b > 0$, and rederive the posterior.

Note $f(\theta) \propto e^{a\theta - be^{\theta}}$. We can piggyback from part (l) and start here:

$$\begin{aligned} f(\theta \mid \boldsymbol{x}) &\propto e^{n\bar{x}\theta - ne^{\theta}} f(\theta) \\ &\propto e^{n\bar{x}\theta - ne^{\theta}} \left( e^{a\theta - be^{\theta}} \right) \\ &= e^{(n\bar{x}+a)\theta - (n+b)e^{\theta}} \\ &\propto \text{LogGamma}(n\bar{x} + a, n + b) \end{aligned}$$

(p) [3 pt / 49 pts] Is the LogGamma the conjugate prior for the iid Poisson with our parameterization of $\theta = \ln(\lambda)$? Yes / no and justify your answer.

Yes. The prior is LogGamma and the posterior is LogGamma hence we satisfied the definition of conjugacy.

(q) [3 pt / 52 pts] Interpret the likely significance of the hyperparemeter $b$ in the context of the pseudoobservations.

Given that it adds to $n$, we can surmise that $b = n_0$, the number of pseudoobservations.

(r) [4 pt / 56 pts] Interpret the likely significance of the hyperparemeter $a$ in the context of the pseudoobservations.

Given that it adds to $n\bar{x} = \sum_{i=1}^{n} x_i$, we can surmise that $a = n_0\mu_0$, where $\mu_0$ is the average of the $n_0$ pseudoobservations. Or alternatively, $a$ is the sum of the $n_0$ pseudoobservations which in the context of the Poisson DGP is the number of pseudosuccesses.

(s) [1 pt / 57 pts] Are the interpretations of $a, b$ the same as the interpretation of $\alpha, \beta$ in the Gamma $(\alpha, \beta)$ prior when we parameterize the Poisson with $\lambda$, the canonical parameterization we used in class? Yes / no.

Yes.

(t) [3 pt / 60 pts] Is Haldane's prior the same as Laplace's prior in this model? Yes / no and justify your answer.

Yes. Haldane's prior is setting $n_0 = 0$. When doing so, we get the same posterior as Laplace's prior, therefore the prior must be identical.

(u) [4 pt / 64 pts] Consider the setting with a sample size of 17 observations. Provide an example of an informative conjugate prior.

All we need to do is let $n_0 >> 17$ and $\mu_0$ doesn't matter. So the following qualifies: $\text{LogGamma}(\cdot, 1000)$.

For the remainder of this problem, let

$$f(\theta \mid \boldsymbol{x}) = \text{LogGamma}(n\bar{x} + n_0\mu_0, n + n_0)$$

where we have hyperparameters $n_0 \geq 0, \mu_0 \geq 0$.

(v) [2 pt / 66 pts] Find $\hat{\hat{\theta}}^{\text{MMSE}}$ explicitly as a function of $n, \bar{x}, n_0, \mu_0$ or write a mathematical expression that will compute it.

$$\hat{\hat{\theta}}^{\text{MMSE}} := \mathbb{E}\left[\theta \mid \boldsymbol{x}\right] = \psi(n\bar{x} + n_0\mu_0) - \ln\left(n + n_0\right)$$

(w) [4 pt / 70 pts] Find $\hat{\hat{\theta}}^{\text{MMAE}}$ explicitly as a function of $n, \bar{x}, n_0, \mu_0$ or write a mathematical expression that will compute it.

$$\hat{\hat{\theta}}^{\text{MMAE}} := \text{Med}\left[\theta \mid \boldsymbol{x}\right] = \left\{\theta : \int_{-\infty}^{\theta} \frac{(n+n_0)^{n\bar{x}+n_0\mu_0}}{\Gamma(n\bar{x}+n_0\mu_0)} e^{(n\bar{x}+n_0\mu_0)u-(n+n_0)e^u} du = 0.5\right\}$$

(x) [2 pt / 72 pts] Find $\hat{\hat{\theta}}^{\text{MAP}}$ explicitly as a function of $n, \bar{x}, n_0, \mu_0$.

$$\hat{\hat{\theta}}^{\text{MAP}} := \text{Mode}\left[\theta \mid \boldsymbol{x}\right] = \ln\left(\frac{n\bar{x} + n_0\mu_0}{n + n_0}\right)$$

For the remainder of this problem, let `qlgamma`($q$, $a$, $b$) be the function that returns the $q$th quantile of the rv LogGamma($a$, $b$) and `plgamma`($x$, $a$, $b$) be the CDF $F(x)$ of the rv LogGamma($a$, $b$). To be clear: you cannot use these functions to answer questions before this point in the exam.

(y) [5 pt / 77 pts] Write an expression that will return a 95% credible region for $\theta$ after seeing the data $\boldsymbol{x}$.

$$CR_{\theta,95\%} = \left[\texttt{qlgamma}\left(.025, n\bar{x} + n_0\mu_0, n + n_0\right), \texttt{qlgamma}\left(.975, n\bar{x} + n_0\mu_0, n + n_0\right)\right]$$

(z) [7 pt / 84 pts] Note: we are switching to Frequentist inference for this one question. Find an approximate 95% confidence interval for $\theta$ after seeing the data $\boldsymbol{x}$. You will need to use the fact that if $Y \sim \text{Poisson}\left(\lambda\right)$ then $\mathbb{V}\text{ar}\left[Y\right] = \lambda$.

For this we need the delta method. In a previous question we found that $\hat{\hat{\theta}}^{\text{MLE}} = \ln\left(\bar{x}\right)$. We have $\theta = g(\lambda) = \ln\left(\lambda\right)$ and thus $g'(\lambda) = \lambda^{-1}$. You either have proved or can prove quickly that $\hat{\hat{\lambda}}^{MLE} = \bar{x}$. Hence $\mathbb{SD}\left[\hat{\lambda}^{MLE}\right] = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{\lambda}{n}}$ and thus $\hat{\hat{\mathbb{SD}}}\left[\hat{\lambda}^{MLE}\right] = \sqrt{\frac{\bar{x}}{n}}$. We now have everything we need for the formula:

$$\begin{aligned} CI_{\theta,95\%} &\approx \left[\hat{\hat{\theta}}^{\text{MLE}} \pm 1.96\, g'\left(\hat{\hat{\lambda}}^{MLE}\right) \hat{\hat{\mathbb{SD}}}\left[\hat{\lambda}^{MLE}\right]\right] \\ &= \left[\ln\left(\bar{x}\right) \pm 1.96 \left(\frac{1}{\bar{x}}\right) \sqrt{\frac{\bar{x}}{n}}\right] \\ &= \left[\ln\left(\bar{x}\right) \pm 1.96 \frac{1}{\sqrt{n\bar{x}}}\right] \end{aligned}$$

(aa) [1 pt / 85 pts] If $n$ is large, do you expect $CR_{\theta,95\%} \approx CI_{\theta,95\%}$? Yes / no.

Yes.

(ab) [5 pt / 90 pts] Write an expression that will return a the Bayesian p-value for the test $H_0 : \theta = \theta_0$ with a margin of equivalence $\delta$ after seeing the data $\boldsymbol{x}$.

$$p_{val} = \texttt{plgamma}\left(\theta_0 + \delta, n\bar{x} + n_0\mu_0, n + n_0\right) - \texttt{plgamma}\left(\theta_0 - \delta, n\bar{x} + n_0\mu_0, n + n_0\right)$$

(ac) [10 pt / 100 pts] Derive the distribution of $X_\star$, the next future observation, given the data $\boldsymbol{x}$. Don't forget the $\mathbb{1}_{x_\star \in S_{X_\star}}$ term. If the answer is a brand name rv, indicate it as so and find its parameters.

$$\begin{aligned}
p_{X_\star|\boldsymbol{x}}(X_\star \mid \boldsymbol{x}) &= \int_{\mathbb{R}} p_{X_\star|\theta}(X_\star|\theta) f_{\theta|\boldsymbol{x}}(\theta \mid \boldsymbol{x}) d\theta \\
&= \int_{\mathbb{R}} \left( \frac{e^{-e^\theta} \left(e^\theta\right)^{x_\star}}{x_\star!} \mathbb{1}_{x_\star \in \mathbb{N}_0} \right) \left( \frac{(n+n_0)^{n\bar{x}+n_0\mu_0}}{\Gamma(n\bar{x}+n_0\mu_0)} e^{(n\bar{x}+n_0\mu_0)\theta - (n+n_0)e^\theta} \right) d\theta \\
&\propto \frac{\mathbb{1}_{x_\star \in \mathbb{N}_0}}{x_\star!} \int_{\mathbb{R}} e^{(n\bar{x}+n_0\mu_0+x_\star)\theta - (n+n_0-1)e^\theta} d\theta \\
&= \frac{\Gamma(n\bar{x}+n_0\mu_0+x_\star)}{(n+n_0+1)^{n\bar{x}+n_0\mu_0+x_\star} x_\star!} \mathbb{1}_{x_\star \in \mathbb{N}_0} \\
&\propto \frac{\Gamma(n\bar{x}+n_0\mu_0+x_\star)}{x_\star!} \left( \frac{1}{n+n_0+1} \right)^{x_\star} \mathbb{1}_{x_\star \in \mathbb{N}_0} \\
&\propto \text{ExtNegBin}\left( n\bar{x}+n_0\mu_0, \frac{n+n_0}{n+n_0+1} \right)
\end{aligned}$$

Note: this is the same posterior predictive distribution we derived in class using the hyperparemeters specified by pseudodata, i.e., $f(\theta) = \text{Gamma}\,(\alpha = n_0\mu_0,\ \beta = n_0)$. It makes sense that the posterior predictive distribution will be the same regardless of the parameterization as the parameter itself has been margined out when computing $p_{X_\star|\boldsymbol{x}}(X_\star \mid \boldsymbol{x})$. (It shouldn't matter which scale the parameter is measured).