Assignment: You may discuss the problems with other students, but all work must be your own. If you consult a book or online resource, cite it.

Each problem below refers to the following setup. We are conducting a study of people who have epilepsy. We record the number X_i of seizures person i experiences in one week, for $i = 1, \ldots, n$. We observe $\mathbf{X} = (X_1, \ldots, X_n)$, where we model $X_i \sim \text{Poisson}(\lambda)$ independently. Our prior beliefs about the seizure rate λ (before seeing the data $\mathbf{X}$) are summarized by $\lambda \sim \text{Exponential}(\alpha)$, where $\alpha > 0$ is fixed and known. Call the prior density $p(\lambda|\alpha)$. In Bayesian statistics, our new belief about λ (after seeing the data) is embodied by the *posterior distribution* of the unknown parameter λ:

$$p(\lambda|\mathbf{X}, \alpha) = \frac{L(\lambda; \mathbf{X})p(\lambda|\alpha)}{\int_0^\infty L(\lambda; \mathbf{X})p(\lambda|\alpha)\ \mathrm{d}\lambda}$$

where $L(\lambda; \mathbf{X})$ is the likelihood.

Problem 1: The posterior distribution

a. Write the likelihood and an expression for the posterior distribution $p(\lambda|\mathbf{X}, \alpha)$.

b. Find the maximum likelihood estimate $\hat{\lambda}_{\text{MLE}} = \text{argmax}_\lambda\ L(\lambda; \mathbf{X})$ analytically. Derive an expression for its asymptotic standard error.

c. Find the maximum *a posteriori* (MAP) estimate $\hat{\lambda}_{\text{MAP}} = \text{argmax}_\lambda\ p(\lambda|\mathbf{X}, \alpha)$ analytically. How is the MAP estimate different from the MLE in part (a)? Which one is bigger? What happens as $n \to \infty$? Why is it useful to have prior beliefs about λ, and what is the role of α in the MAP estimate?

Problem 2: The mean and variance of the posterior

a. Is it easy or difficult to find an analytic formula for $\mathbb{E}[\lambda|\mathbf{X}, \alpha]$? Why?

b. Derive the normal approximation $g(\lambda)$ to the posterior distribution $p(\lambda|\mathbf{X})$ at $\lambda = \hat{\lambda}_{\text{MAP}}$. Remember that $\lambda > 0$, and specify your importance distribution accordingly. State explicitly the distribution or density for your approximation.

c. Let $\alpha = 0.7$ and write a computer program to draw the true value of $\lambda \sim \text{Exponential}(\alpha)$. Then, conditional on this value of λ, generage $n = 13$ independent Poisson(λ) variables. Call these values $X_1, \ldots, X_n$.

d. Derive an importance sampling estimator to find the mean and variance of the posterior $p(\lambda|\mathbf{X}, \alpha)$. For the importance distribution $g(y)$ use your normal approximation. Since the normalizing constant of $p(\lambda|\mathbf{X}, \alpha)$ is unknown, you may use the following fact:

$$\frac{\frac{1}{m}\sum_{j=1}^{m} \frac{\lambda_j p(\lambda_j|\mathbf{X},\alpha)}{g(\lambda_j)}}{\frac{1}{m}\sum_{k=1}^{m} \frac{p(\lambda_k|\mathbf{X},\alpha)}{g(\lambda_k)}} \to \mathbb{E}[\lambda|\mathbf{X}, \alpha]$$

for $\lambda_1, \ldots, \lambda_m \sim g$, as $m \to \infty$.

e. Apply your importance sampling estimator to numerically compare the posterior mean and standard deviation to the MLE and its asymptotic SE, and the MAP estimate. How do they differ?

Problem 3: MCMC for the posterior distribution of λ

We will derive an MCMC algorithm for learning about the posterior distribution $p(\lambda|\mathbf{X}, \alpha)$.

a. Consider the proposal density

$$g(\lambda^*|\lambda) = \frac{\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2\sigma^2}(\lambda^* - \lambda)^2]}{\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2\sigma^2}(\lambda' - \lambda)^2]\, \mathrm{d}\lambda'}$$

Why is this a reasonable proposal function for λ? Is it symmetric?

b. Derive the Metropolis-Hastings acceptance probability $\rho(\lambda, \lambda^*)$.

c. Using the values of $\mathbf{X}$ and α from the last part, implement a Metropolis-Hastings algorithm to draw $N = 10,000$ samples $\lambda_1, \ldots, \lambda_N \sim p(\lambda|\mathbf{X}, \alpha)$. Use the proposal density $g(\lambda^*|\lambda)$ given above. Explain how you chose σ^2.

d. Plot the samples of λ from your algorithm by iteration and as a histogram. Overlay your previous estimates on this histogram. What do you see?

e. Compute an estimate of the posterior mean and variance of $p(\lambda|\mathbf{X}, \alpha)$ using your Metropolis-Hastings samples and compare to your importance sampling estimates and the MAP estimate and the MLE.