ECON 2823 | Spring 2026 | Checkpoint 2

Due: Friday, April 3

Work in groups. Use notebooks to compute answers. Learn something. Ask clarifying questions. Submit your answers to Gradescope. Have fun.

Q1. Analytical MLE (Exponential Distribution)

Customers arrive at a coffee shop. The time between arrivals follows an Exponential distribution with rate parameter λ:

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad x > 0 \tag{1}$$

You observe n iid arrival gaps $x_1, \ldots, x_n$.

a) Write the log-likelihood $\ell(\lambda)$ for n iid observations $x_1, \ldots, x_n$.

b) Derive the MLE $\hat{\lambda}$ analytically by taking the derivative of $\ell(\lambda)$, setting it to zero, and solving.

c) Compute the Fisher information $I(\lambda)$ using the Hessian method (i.e., $I(\lambda) = -\ell''(\lambda)$). What is the theoretical standard error of $\hat{\lambda}$?

d) Generate 200 draws from $\text{Exponential}(\lambda = 3)$ using `np.random.seed(42)` and `np.random.exponential(1/3, 200)`. Compute $\hat{\lambda}$ and its SE using your formulas from (b) and (c). Then compute the SE numerically using the score method with centered differences. Do they agree?

(Hint: the score for observation i is $s_i(\lambda) = \frac{\partial}{\partial \lambda} \log f(x_i|\lambda)$. Estimate the Fisher information as $\hat{I}(\hat{\lambda}) = \frac{1}{n} \sum_i s_i(\hat{\lambda})^2$, then $\text{SE} = 1/\sqrt{n \cdot \hat{I}}$.)

Q2. Poisson MLE for Soccer Goals

In Checkpoint 1, you simulated soccer matches where each team's goals depended on both teams' attack and defense parameters: $\text{Goals}_i \sim \text{Poisson}(e^{\alpha_i - \delta_j})$. Estimating that full model means jointly fitting parameters for all 20 teams, which we did in class but is a bigger problem than what we'll do here.

Instead, model all of Liverpool's home goals as iid draws from a single Poisson distribution, ignoring opponent identity:

$$\text{FTHG} \sim \text{Poisson}(\lambda_H) \tag{2}$$

This gives us Liverpool's average home scoring rate across all opponents. This is not the right model (it throws away information about who Liverpool is playing) but it's a good one-parameter MLE problem to practice with.

```
soccer = pd.read_csv(file_path + 'soccer/soccerData.csv')
soccer.head()
```

a) Filter to all matches where Liverpool is the home team. How many home matches does Liverpool play? What is the sample mean of their home goals (`FTHG`)? Under a Poisson model, this is the MLE of Liverpool's home scoring rate $\hat{\lambda}_H$.

b) Write a Python function `poisson_loglik(lam, goals)` that computes the Poisson log-likelihood: $\ell(\lambda) = \sum_i [k_i \ln(\lambda) - \lambda - \ln(k_i!)]$. Maximize it using `scipy.optimize.minimize_scalar` on the interval $[0.1, 10]$. Verify the answer matches (a).

c) Compute the standard error of $\hat{\lambda}_H$ using the Fisher information. Start from the Poisson log-likelihood for a single observation, $\log P(k|\lambda) = k \ln(\lambda) - \lambda - \ln(k!)$, take the second derivative with respect to λ, and use $I(\lambda) = -\mathbb{E}[\ell''(\lambda)]$ to derive the SE. *(Same idea as Q1(c) but for the Poisson.)*

d) Construct a 95% confidence interval for Liverpool's true home scoring rate. Based on this interval, is Liverpool's rate statistically different from 1.5 goals per match? *(1-2 sentences.)*

Q3. Monte Carlo Verification

Use $\hat{\lambda}_H$ from Q2 as the "true" parameter to verify your standard error and confidence interval results via simulation.

a) Simulate 10,000 Liverpool home seasons: for each, draw the same number of home matches as in Q2(a) from $\text{Poisson}(\hat{\lambda}_H)$ and compute the MLE (sample mean). Use `np.random.seed(0)`.

b) Plot a histogram of the 10,000 MLEs. Overlay the theoretical normal distribution $N(\hat{\lambda}_H,\ \text{SE}^2)$ using the SE from Q2(c).

c) What is the Monte Carlo standard deviation of the 10,000 estimates? How does it compare to the analytical SE from Q2(c)?

d) For each of the 10,000 simulations, compute a 95% CI using $\hat{\lambda} \pm 1.96 \cdot \text{SE}$ (where $\text{SE} = \sqrt{\hat{\lambda}/n}$ is computed from that simulation's estimate). What fraction of these CIs contain the true value $\hat{\lambda}_H$? Is it close to 95%?