ECON 2823 | Spring 2026 | Checkpoint 1

Due: Friday, March 27

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

Q1. Numerical Derivatives

In practice, data isn't always nicely organized. Measurements might be irregular. Suppose you have access to three values of a continuous function at unequally spaced points: $f(x+0.03)$, $f(x)$, and $f(x-0.02)$. *(Hint: think of ϵ as 0.01.)* What are three weights α_1, α_2, α_3 such that

$$\alpha_1 f(x+0.03) + \alpha_2 f(x) + \alpha_3 f(x-0.02) \tag{1}$$

gives a good approximation of the second derivative of f at the point x?

a) What is α_1 (the weight on $f(x+0.03)$)?

b) What is α_2 (the weight on $f(x)$)?

c) What is α_3 (the weight on $f(x-0.02)$)?

Q2. Root-Finding and Optimization

Consider the function:

$$g(x) = x^4 - 6.2x^3 + 1.61x^2 + 15.7x + 5.6 \tag{2}$$

a) Using `scipy.optimize.fsolve`, find all four roots of $g(x) = 0$. All solutions are in $[-3, 6]$. Try multiple starting points. Report from smallest to largest.

b) Find the global minimum of $g(x)$ on $[-3, 6]$. Report the x-value.

c) $g(x)$ also has a local minimum distinct from the global min. Report its x-value.

d) $g(x)$ also has a local maximum. Report its x-value.

Q3. Simulation

In soccer, goals in a match can be modeled as Poisson random variables. When Team i plays Team j, each team's goals are drawn independently:

$$\text{Goals}_i \sim \text{Poisson}\left(e^{\alpha_i - \delta_j}\right) \tag{3}$$

$$\text{Goals}_j \sim \text{Poisson}\left(e^{\alpha_j - \delta_i}\right) \tag{4}$$

where α_i measures Team i's attacking strength and δ_j measures Team j's defensive strength. A high α means a good offense. A high δ means a good defense. The exponential ensures the expected goals are always positive.

The file `TeamEstim.csv` contains these parameters for every team in the English Premier League (about which I, Taylor, know very little). You only need the `alpha` and `delta` columns. Liverpool and Arsenal are about to play a **two-leg tie**: the first match at Liverpool's home, the second at Arsenal's home. The winner is decided by aggregate goals (total across both matches). If the aggregate is tied, flip a coin. For example, if Liverpool wins 2-1 at home and Arsenal wins 1-0 at home, the aggregate is Liverpool 2, Arsenal 2, a tie, decided by coin flip.

a) Look up the α and δ values for Liverpool and Arsenal. Simulate a single match (Liverpool home vs Arsenal away) by drawing two Poisson random variables with the appropriate rates. What scoreline did you get?

b) Simulate 10,000 two-leg ties. For each tie, simulate both matches and determine the winner. What is the probability that Liverpool wins the tie?

c) Now introduce a common shock: before each two-leg tie, draw $\epsilon \sim N(0, 0.3)$ and add it to both teams' α parameters (so attacking strength varies from tie to tie, but both teams are affected equally). Re-run 10,000 simulations. What is the new probability that Liverpool wins? Is it higher, lower, or about the same as (b)?

d) Using your simulations from (c), plot a histogram of the aggregate goal difference (Liverpool total minus Arsenal total) across the 10,000 ties.

e) In 1-2 sentences: why does the common shock change (or not change) the probability? *(Hint: think about the election simulation from class, where adding a common shock across states changed the win probability.)*

Q4. Non-linear Least Squares

The file `MLB.csv` contains player-season observations with columns `Age`, `BA` (batting average), and `AB` (at-bats), among others. Load the data and filter to players with at least 200 at-bats (`AB >= 200`). Then count how many player-seasons appear at each age.

```
mlb = pd.read_csv(file_path + 'MLB.csv')
mlb = mlb[mlb.AB >= 200]
sns.histplot(mlb, x='Age')
```

The resulting histogram should look roughly bell-shaped: few very young or very old players qualify, with a peak in the mid-to-late 20s. Fit a Gaussian model to these counts:

$$f(\text{age}) = A \cdot \exp\left(-\frac{(\text{age} - \mu)^2}{2\sigma^2}\right) \tag{5}$$

where A is the peak count, μ is the peak age, and σ is the width.

a) Plot the count of qualified player-seasons by age. Based on the plot, propose starting values A_0, μ_0, and σ_0. Explain your reasoning in 1-2 sentences.

b) Estimate $\hat{A}$, $\hat{\mu}$, and $\hat{\sigma}$ by minimizing the sum of squared residuals using `scipy.optimize.minimize` with method `'Nelder-Mead'`.

```
# Get ages and counts
ages   = age_counts['Age'].values.astype(float)
counts = age_counts['count'].values.astype(float)

# Minimize
result = minimize(ssr, x0, method='Nelder-Mead')
```

c) Plot the fitted curve on top of the age-level counts.

d) According to your model, at what age are players most likely to be qualified hitters? How wide is the prime window ($\hat{\mu} \pm \hat{\sigma}$)? Where does the Gaussian fit well and where does it miss? *(2-3 sentences.)*