STAT 460/560: Statistical Inference 1 **Winter Term 1 (Fall) 2026**

Assignment 8

Due date: Nov. 27, 11:59 p.m. *Instructor: Daniel McDonald and Benjamin Bloem-Reddy*

Instructions

Academic integrity policy: I encourage you to discuss verbally with other students about the assignment. However, you **must** write your answers by yourself. If you discuss the assignment with anyone (a classmate or anyone else), you must say so at the top of your solutions. Use of AI tools to produce full or partial answers to these questions or to related questions is strictly prohibited. This includes tools like Co-pilot, auto-complete, etc. Also, refrain from looking at answer keys from other schools, previous years, or the Math Stack Exchange. For more information, see: the syllabus or contact the instructor.

1. Solutions may be handwritten or typeset with LaTeX (see the course website for a template). It must be legible. Difficult-to-read solutions will be marked down or ignored.
2. *Justify your answers formally. (Err on the side of too much detail so that the teaching team knows exactly what you're doing and why.)*
3. Submit your work by 11:59 pm (Vancouver time) on the due date to Canvas (.pdf).
4. For questions that require computer simulations, you may use any language from the list of `(R, Julia, Python)`, with level of support provided by the teaching team decreasing further in the list. Please include in your PDF submission the relevant plots generated by your simulation code, and append your code (in PDF) to your submission. Please retain the code you used, as you may be asked for it if required.

Questions

1 EM algorithm for censored data

Suppose we have i.i.d. $X^n := (X_1, \dots, X_n)$ from some unknown distribution with PDF $f(x; \lambda)$. Let T be a fixed censoring time (e.g., the end of a drug trial's observation window). Define

$$U_i = \min\{X_i, T\} \,, \quad V_i = \mathbb{I}\{X_i \leq T\} \,. \tag{1}$$

We observe (U_i, V_i).

1. What is the "complete data" log-likelihood, $\ln f(x^n, u^n, v^n; \lambda)$?
2. Hence, what is the marginal *likelihood* (marginalized over the X_i's for which $V_i = 0$)? Where have we seen this before?

3. Suppose that $f(x;\lambda)$ is the exponential PDF with rate parameter λ (i.e., mean $1/\lambda$). The EM algorithm for this problem will require that for the E-step we calculate $\mathbb{E}_{\lambda^{\text{old}}}[\ell_n(X^n, U^n, V^n; \lambda)|u^n, v^n]$. Show that

$$\mathbb{E}_{\lambda^{\text{old}}}[\ell_n(X^n, U^n, V^n; \lambda)|u^n, v^n] = \sum_{i=1}^{n} v_i(\log\lambda - \lambda u_i) + (1-v_i)(\log\lambda - \lambda\mathbb{E}_{\lambda^{\text{old}}}[X_i|u_i, v_i]) \quad (2)$$

$$= \sum_{i=1}^{n} v_i(\log\lambda - \lambda u_i) + (1-v_i)(\log\lambda - \lambda(T + \frac{1}{\lambda^{\text{old}}})) \; . \quad (3)$$

4. Write the M-step.

5. **(Optional - bonus)** Implement this EM algorithm on simulated data with $n = 50$ observations, for $\lambda \in \{0.1, 0.25, 0.5, 1, 5\}$, with $T = 1$. For each parameter value, use $B = 100$ simulations and plot the average number of steps needed for convergence. (You can set the maximum number of steps to 1000 if necessary.) Plot the estimated risk under squared error loss $L(\hat{\lambda}, \lambda) = (\hat{\lambda} - \lambda)^2$.

2 Optional: EM for Gaussian Mixture Models

Do not turn in; not graded.

Read Chapter 9 (focus on 9.2) of Bishop, *Pattern Recognition and Machine Learning* (https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf) and do Exercises 9.6–9.9 (p. 456).