Homework 5

Due Thursday, October 8, in class

STAT5710 Applied Stochastic Processes, Fall 2026

Every February 2, a groundhog is produced before a crowd and its handlers announce one of two forecasts: more winter, or an early spring. Treat the sequence of annual announcements for a single animal as a realization of a time-homogeneous Markov chain $X_t \in \{W, S\}$, where

$$W = \text{more winter}, \qquad S = \text{early spring},$$

and t indexes consecutive years. In some years no forecast was recorded. Those years are gaps in the record rather than a third state, so a transition contributes to the counts only when both years are observed and consecutive. Both records used below are posted as a single file:

`https://tdruiz-stat5710.share.connect.posit.cloud/data/groundhogs.csv`

Where a part calls for computation, you may either attach or transcribe relevant output. You may include codes if you wish, but codes are not required.

1. **Punxsutawney Phil.** Phil's record runs from 1886 to 2026. Of those 141 years, 131 carry a recorded forecast; the ten gaps are 1886, 1889, 1891–1897, and 1899, leaving 127 usable transitions:

$$N = \begin{array}{c|cc|c} & W & S & \text{row total} \\ \hline W & 88 & 19 & 107 \\ S & 19 & 1 & 20 \end{array}$$

   (a) Compute the conditional MLE $\hat{P}$.

   (b) Construct two 95% confidence intervals for p_{WS}, one using the Wald approach and the other using the exact approach.

   (c) Are the Wald intervals valid for p_{SS} and p_{SW}? Explain.

   (d) Compute the Laplace-smoothed estimate $\tilde{P}$.

   (e) Recompute the estimate under a Dirichlet prior whose mean is the pooled column margin of N with strength $\alpha = 5$.

   (f) Put a symmetric Dirichlet(1) prior on row S and identify the marginal posterior distribution of p_{SS}. Report a 95% HPD interval by sampling from the posterior and taking the shortest interval containing 95% of the draws.

   (g) Phil called an early spring in 21 of the 131 recorded years, a rate of 0.160. Compare your estimate and interval for p_{SS} with that marginal rate. Does his record support the claim that he avoids calling an early spring two years running?

   (h) Phil called more winter in 2026, and you have \$20 riding on his 2027 call. Which way do you bet, and what are your odds of winning? How would your answer change if the bet were on the 2030 call?

2. **Staten Island Chuck.** Chuck has been making predictions in New York since 1982.

(a) Read in the data and construct the table of transition counts for Chuck's record, using the same two states. You should find 24 usable transitions. Report the table.

(b) Compute the conditional MLE. Which rows would you be willing to report as estimates, and which not? Smooth as necessary using a method of your choosing.

(c) Compare Chuck with Phil on p_{SS}, and again on p_{WS} with uncertainty quantification. Do their predictive behaviors differ?

(d) Chuck also called more winter in 2026. Would you bet differently on him, and would you be as confident?

3. *(Optional.)* **Is there any Markov structure here?** If a groundhog's forecast did not depend on the previous year's, the sequence of forecasts would be iid draws from a single distribution, and the chain would carry no memory at all.

(a) State that as a hypothesis about P, and find the MLE of the transition probabilities under it.

(b) Construct a test of the hypothesis from the likelihood, and determine its degrees of freedom.

(c) Carry out the test for Chuck and Phil and interpret the results.