Bayesian Dynamic Modeling in M&F **Hedibert F. Lopes**
Undergraduate in Economics **Insper**
Homework 1 **Due: 9:45, March, 5th 2026**

Individual Work & Oral Assessment: This homework assignment must be completed individually. To ensure a comprehensive understanding of the material, students will be selected at random for a brief oral examination to discuss their solutions and the underlying methodology.

Probem 1

A few years ago, researchers in the United Kingdom conducted a meta-analysis of published studies regarding commercial antigen tests. The analysis comprised 24,087 samples, 7,415 of which were confirmed positive for SARS-CoV-2. The study revealed that for symptomatic individuals, these tests demonstrated an average sensitivity (true positive rate) of 72%. This indicates a 28% probability of an infected patient receiving a false negative result. Conversely, false positives were found to be exceptionally rare. The researchers concluded that the tests maintain a Positive Predictive Value (PPV) of 99.6%, meaning that when a test confirms an infection, the result is accurate in nearly all cases. Based on this research, we can define the following probabilities:

$$\begin{aligned}
P(X=1|\theta=0) &= 1\% \\
P(X=0|\theta=1) &= 28\% \\
P(X=1,Y=1|\theta=0) &= 0.01\% \\
P(X=1,Y=1|\theta=1) &= 51.84\%
\end{aligned}$$

for two tests X and Y, conditionally independent given θ. The prevalence of Covid in my mother's age group is 15.71%, i.e., the accumulated scientific knowledge is that $P(\theta=1)=0.1571$. What is the posterior probability of her having Covid after the two independent tests, X and Y, come back positive, $P(\theta=1|X=1,Y=1)$?

Problem 2

Our choice of words often serves as a subtle map of where we live. Imagine watching an interview with someone in the United States: even without any personal background on the speaker, we can rely on U.S. Census data to establish a prior probability of their geographic origin. These figures allow us to weigh the likelihood that they reside in one of the four primary regions: the Midwest (M), Northeast (N), South (S), or West (W).

$$\begin{aligned} Pr(\text{region} = M) &= 0.21 \\ Pr(\text{region} = N) &= 0.17 \\ Pr(\text{region} = S) &= 0.38 \\ Pr(\text{region} = W) &= 0.24 \end{aligned}$$

The South stands as the most populous region, while the Northeast is the least. Relying solely on these population statistics, we establish a prior probability of 38% that the interviewee resides in the South. However, new evidence emerges: the speaker points to a carbonated beverage and says, “Please pass my pop.” While the nation may be united in its affinity for fizzy drinks, it is deeply divided by the terminology used to describe them: “pop”,“soda,” or “coke.” This linguistic choice provides the data (y) necessary to update our initial beliefs. To evaluate the strength of this evidence, we can utilize the ”Pop vs. Soda” dataset, which comprises 374,250 responses from a large-scale survey at popvssoda.com.

$$\begin{aligned} Pr(\text{say} = \text{pop} \mid \text{region} = M) &= 0.6447 \\ Pr(\text{say} = \text{pop} \mid \text{region} = N) &= 0.2734 \\ Pr(\text{say} = \text{pop} \mid \text{region} = S) &= 0.0792 \\ Pr(\text{say} = \text{pop} \mid \text{region} = W) &= 0.2943 \end{aligned}$$

2.a

Show that there is a 28.26% chance that a person in the U.S. uses the word “pop”.

2.b

Considering the fact that 38% of people live in the South but that “pop” is relatively rare to that region, what’s the posterior probability that the interviewee lives in the South?

Problem 3

According to a 2017 Pew Research survey, 10.2% of LGBT adults in the U.S. were married to a same-sex spouse. Fast-forwarding to 2024, Pamela anticipates that this proportion, represented by the parameter π, has shifted. She hypothesizes that π has most likely increased to approximately 15%, though she believes it could reasonably range anywhere from 10% to 25%.

4.a Identify and plot a Beta prior model, $Beta(\alpha, \beta)$, that reflects Pamela's prior beliefs about π.

4.b To update her beliefs, Pamela collects data from a random sample of 90 LGBT adults in the U.S. and finds that 30 are married to a same-sex partner. Using this data, derive the posterior distribution of π.

4.c Calculate the posterior mean, mode, and standard deviation of π based on your results from part (4.b).

4.d Does the resulting posterior distribution more closely reflect Pamela's prior information or the observed data? Justify your reasoning by comparing the prior and posterior parameters.

Problem 4

Suppose you have the data $x_1, \ldots, x_n$ and that you want to entertain the adherence of the data to the following configuration of model and prior: The data $(x_1, \ldots, x_n)$ are conditionally independent and identically distributed, given μ, as a $N(\mu, \sigma^2)$, while the of μ is $N(0, 1)$, for $\sigma = 0.25$ a known quantity. You have a small sample with $n = 15$ observations such that $\sum_{i=1}^{n} x_i = 8.31$. The posterior distribution of μ is also normal. (i) Find its mean and variance, and (ii) plot $p(\mu)$ (the prior) along with $p(\mu | x_1, \ldots, x_{15})$ (the posterior).