Gov 2001: Problem Set 7

Spring 2026

Instructions:

- The Problem set is due on **April 7, 11:59 PM Eastern Time**.
- Please upload a PDF of your solutions to Gradescope. Make sure to assign to each question all the pages with your work on that question.
- **Do not use AI assistants (ChatGPT, Claude, Copilot, etc.) on this problem set.** Work with each other instead. The struggle is where learning happens.
- Remember: 70% of your grade comes from in-class exams. Use problem sets to *learn*, not just to get answers.

Short Questions

Note: In this problem set, we use a simple notation, $X_{1:n}$, to denote $X_1, X_2, ..., X_n$, the collection of sample R.V.s: $X_{1:n} = \{X_1, X_2, ..., X_n\}$.

1. Let $\phi(X_{1:n})$ be a binary sample statistic such that $\mathbb{P}(\phi(X_{1:n}) = 1) \leq 0.05$ when $\theta = \theta_0$. Explain in words how to use $\phi(X_{1:n})$ to perform a hypothesis testing on $\theta = \theta_0$ at the significance level of $\alpha = 0.05$. Make sure you state clearly H_0, H_1, and the criteria that we reject H_0.

Long Questions

For Q2 to Q5, we use the following common setting:

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$. We want to test $H_0 : \lambda = \lambda_0 = 2$ vs. $H_1 : \lambda > \lambda_0$ at significance level $\alpha = 0.05$. Let $\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the plug-in estimator of λ.

2. We first construct a test based on Chebyshev's inequality.

(a) Let $U(X_{1:n}) = \mathbf{1}\{\hat{\lambda} \geq 2\sqrt{5}\sigma(\bar{X}) + 2\}$. Prove $\mathbb{P}(U(X_{1:n}) = 1) \leq 0.05$ when $\lambda = 2$.

Hint: $\mathbb{P}(X - \mu \geq k) \leq \mathbb{P}(|X - \mu| \geq k)$.

(b) Explain why we can't use $U(X_{1:n})$ to test H_0 like what we did in Q1. Make a modification on $U(X_{1:n})$ so that we can use it and denote the modified statistics as $W(X_{1:n})$.

(c) Describe a situation where $W(X_{1:n})$ could be invalid as a hypothesis test for H_0. In other words, $\mathbb{P}(W = 1) > 0.05$ even when $\lambda = 2$.

Hint: Do estimators have estimation errors?

3. Next, we construct a test based on the exact distribution of $\hat{\lambda}$.

(a) What's the distribution of $\hat{\lambda}$ under H_0?

(b) Let $T(X_{1:n}) = \mathbf{1}\{\hat{\lambda} \geq k\}$. What is the value of k if we want to use $T(X_{1:n})$ to test H_0 at the significance level of $\alpha = 0.05$? You don't have to give a fully simplified answer, just state it as a solution to an equation.

(c) Is there any situation where T would be invalid as a hypothesis test of H_0 at the significance level of $\alpha = 0.05$?

4. Lastly, we construct a test based on approximate distribution of $\hat{\lambda}$.

(a) What's the approximate distribution of $\sqrt{n}\frac{(\hat{\lambda}-\lambda)}{\sqrt{\hat{\lambda}}}$? Make sure you clearly list the theorems used in each step.

(b) We want to construct $S(X_{1:n}) = \mathbf{1}\{\hat{\lambda} - 2 \geq c\}$ so that when $n \to \infty$, $\mathbb{P}(S(X_{1:n}) = 1) \leq 0.05$ under H_0. Use the result in (a) to express c with n, $\hat{\lambda}$, and $z_0.95$.

5. Now we have three different test statistics, $W(X_{1:n}), T(X_{1:n}), S(X_{1:n})$, all of which can be used to test H_0 at significance level $\alpha = 0.05$. However, they have different power, and we can use simulation to compare them. In this question, you are allowed to use AI to assist with coding tasks, but please provide your code and make sure you understand the logic.

For each given λ and n, run a simulation using the following process:

(1) Generate a random sample $x_1, x_2, \ldots, x_n \overset{iid}{\sim} \text{Poisson}(\lambda)$ and compute $\hat{\lambda} = \bar{x}$;

(2) Compute $W(x_{1:n}), T(x_{1:n}), S(x_{1:n})$ and decide whether each test rejects H_0 based on the results. Notice that you should first derive T's cutoff threshold k numerically;

(3) For a fixed λ, repeat steps (1)–(2) for 10,000 repetitions and record the results. You should have a $10{,}000 \times 3$ table if the code is correct.

(4) Calculate the rejection rate for each test. For example, for W,

$$\pi_W(\lambda) = \mathbb{P}(W = 1) = \frac{1}{10,000}\sum_{j=1}^{10,000} W_j.$$

Run the above simulation separately for $(n, \lambda) \in \{40, 400\} \times \{2, 2.5, 3\}$. You should get a 3×6 table (three tests by six combinations of n and λ) if the code is correct. Print the table and answer the following questions:

(a) Which entries in your table correspond to type I error rates? (Hint: use the columns with $\lambda = 2$.) Compare the three tests' sizes. Which one is more conservative?

(b) What are the power levels in this table? For fixed n, which test has more power?

(c) How does the difference of power between T and S change when n gets larger?