541B Final Solutions[1]

1. Question 1

Suppose X is a binomial distributed random variable with parameters 2 and $\theta \in \{1/4, 3/4\}$. (That is, X is the number of heads that result from flipping two coins, where each coin has probability θ of landing heads.)
We want to test the hypothesis H_0 that $\theta = 1/4$ versus the hypothesis H_1 that $\theta = 3/4$.

Let $\mathcal{T}$ be the set of hypothesis tests with significance level at most $1/10$.
(Recall that the significance level of a hypothesis test $\phi\colon \mathbf{R} \to [0,1]$ is $\sup_{\theta\in\Theta_0} \mathbf{E}_\theta \phi(X)$.)

Find a uniformly most powerful (UMP) class $\mathcal{T}$ hypothesis test $\phi\colon \mathbf{R} \to [0,1]$.
Compute all constants that appear in the definition of ϕ. Justify your answer.

Solution. We first compute

$$\frac{f_{3/4}(0)}{f_{1/4}(0)} = \frac{1}{9}, \qquad \frac{f_{3/4}(1)}{f_{1/4}(1)} = 1, \qquad \frac{f_{3/4}(2)}{f_{1/4}(2)} = 9.$$

The Neyman-Pearson Lemma says that a UMP test for the class of tests with an upper bound on the significance level must be a likelihood ratio test with significance level equal to $1/10$. That is, there is some $k > 0$ and $\gamma \in [0,1]$ such that the following hypothesis test is UMP class $\mathcal{T}$.

$$\phi(x) := \begin{cases} 1 & \text{, if } f_{\theta_1}(x) > k f_{\theta_0}(x) \\ 0 & \text{, if } f_{\theta_1}(x) < k f_{\theta_0}(x) \\ \gamma & \text{, if } f_{\theta_1}(x) = k f_{\theta_0}(x). \end{cases}$$

After examining the likelihood ratios, we decide to choose $k = 1$, so that

$$\phi(x) := \begin{cases} 1 & \text{, if } f_{\theta_1}(x) > f_{\theta_0}(x) \\ 0 & \text{, if } f_{\theta_1}(x) < f_{\theta_0}(x) \\ \gamma & \text{, if } f_{\theta_1}(x) = f_{\theta_0}(x) \end{cases} \qquad = \begin{cases} 1 & \text{, if } x = 2 \\ 0 & \text{, if } x = 0 \\ \gamma & \text{, if } x = 1. \end{cases}$$

Then $\mathbf{E}_{\theta_0}\phi(X) = \gamma \mathbf{P}_{\theta_0}(X = 1) + \mathbf{P}_{\theta_0}(X = 2) = \gamma \mathbf{P}_{1/4}(X = 1) + \mathbf{P}_{1/4}(X = 2) = (3/8)\gamma + 1/16$. Since this quantity is equal to $1/10$ by assumption, we choose $\gamma := 1/10$. That is, our UMP test is

$$\phi(x) := \begin{cases} 1 & \text{, if } x = 2 \\ 0 & \text{, if } x = 0 \\ 1/10 & \text{, if } x = 1. \end{cases}$$

2. Question 2

Let $X_1, \ldots, X_n$ be a random sample from a Gaussian random variable with unknown mean $\mu \in \mathbf{R}$ variance one.

Fix $\mu_0 \in \mathbf{R}$. Suppose we want to test that hypothesis H_0 that $\mu = \mu_0$ versus the alternative H_1 that $\mu \neq \mu_0$.

- Explicitly describe the rejection region of the generalized likelihood ratio test for this hypothesis.
- Give an explicit formula for the p-value of this hypothesis test.

[1]December 9, 2023,

Solution. (This was Example 4.3 in the notes)

3. Question 3

Let P be the transition matrix of a finite Markov chain.
All eigenvectors and eigenvalues discussed below are left eigenvectors and left eigenvalues.

- Let $\lambda \in \mathbf{C}$ be an eigenvalue of P. Show that $|\lambda| \leq 1$.
- Given an example of a transition matrix P with at least one eigenvalue λ such that λ is not a real number.
- If P is irreducible and aperiodic, show that -1 is not an eigenvalue of P.
- If P is reversible, show that all eigenvalues of P are real.
  (Hint: show that $\langle f, Pg\rangle = \langle Pf, g\rangle$, where $\langle f, g\rangle := \sum_{x\in\Omega} f(x)g(x)\pi(x)$ for all $f, g\colon \Omega \to \mathbf{R}$, where π is stationary. To define Pf, we think of f as a column vector, so the matrix P applied to f is well-defined. You can then freely use the spectral theorem for self-adjoint matrices, which implies that all eigenvalues of a self adjoint matrix are real. Also, as a hint for the next part of the problem, P has an orthonormal basis of eigenvectors.)
- Let $\gamma := 1 - \max\{|\lambda| :\ \lambda \text{ is an eigenvalue of } P \text{ and } \lambda \neq 1\}$. Suppose P is irreducible and reversible with stationary distribution π. Show that, for all $n \geq 1$ and for all $f\colon \Omega \to \mathbf{R}$, we have

  $$\mathrm{Var}_\pi(P^n f) \leq (1-\gamma)^{2n}\mathrm{Var}_\pi(f).$$

  Here $\mathrm{Var}_\pi(f) := \mathbf{E}_\pi(f - \mathbf{E}_\pi f)^2$ and $\mathbf{E}_\pi f = \sum_{x\in\Omega} f(x)\pi(x)$.
  (You don't need to show this, but this inequality leads to a bound on the mixing time of a Markov Chain in terms of γ.)

Solution. Let $v \in \mathbf{R}^n \setminus \{0\}$ be a left eigenvector of P with eigenvalue $\lambda \in \mathbf{C}$. Denote $\|v\|_1 := \sum_{i=1}^n |v_i|$. Then

$$|\lambda|\,\|v\|_1 = \|vP\|_1 = \sum_{j=1}^n \left|\sum_{i=1}^n v_i P_{ij}\right| \leq \sum_{j=1}^n\sum_{i=1}^n |v_i|\,P_{ij}.$$

Using $\sum_{j=1}^n P_{ij} = 1$ (by definition of a transition matrix P) we conclude that $|\lambda|\,\|v\|_1 \leq \|v\|_1$. Since $v \neq 0$, we conclude that $|\lambda| \leq 1$.

Let $P := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$. Then P has eigenvalues λ satisfying $\lambda^3 = 1$, i.e. $e^{2\pi i/3} \notin \mathbf{R}$ is an eigenvalue of P.

If P is irreducible and aperiodic, then -1 cannot be an eigenvalue of P, since if -1 were an eigenvalue of P, it would contradict the Convergence Theorem 9.72. (If π is the unique stationary distribution, and if -1 is an eigenvalue with eigenvector v, then $vP^{2n} = v$, which does not converge to π as $n \to \infty$, since v is not a multiple of π, since they each have different eigenvalues.)

The hint implies that all eigenvalues of P are real when P is reversible.

Let $v_1, \ldots, v_n \in \mathbf{R}^n$ be an orthonormal set of eigenvectors of P with eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -1$. Let $f \in \mathbf{R}^n$. Then $f = \sum_{i=1}^n \langle f, v_i \rangle_\pi v_i$, and by definition of P we have

$$P^n f = \sum_{i=1}^n \langle f, v_i \rangle_\pi \lambda_i^n v_i.$$

$$\mathrm{Var}_\pi(P^n f) = \sum_{i=2}^n \langle f, v_i \rangle_\pi^2 \lambda_i^{2n}$$

$$\mathrm{Var}_\pi(f) = \sum_{i=2}^n \langle f, v_i \rangle_\pi^2 \qquad (*)$$

Therefore,

$$\mathrm{Var}_\pi(P^n f) = \sum_{i=2}^n \langle f, v_i \rangle_\pi^2 \lambda_i^{2n} \leq \max_{2 \leq j \leq n} \lambda_j^{2n} \sum_{i=2}^n \langle f, v_i \rangle_\pi^2 \stackrel{(*)}{=} (1-\gamma)^{2n} \mathrm{Var}_\pi(f).$$

4. Question 4

Suppose you can freely sample any number of i.i.d. Gaussian random variables with mean zero and variance one (on a computer).

Give an MCMC algorithm for estimating the integral

$$\int_{-\infty}^{\infty} e^{-x^2 - x^4 - x^8} e^x \frac{dx}{\alpha}, \qquad (*)$$

where $\alpha := \int_{-\infty}^{\infty} e^{-x^2-x^4-x^8} dx$ is an **unknown** quantity. (That is, you should not need to estimate α at all, in order to estimate the integral $(*)$.)

(Hint: Even though we didn't cover a continuous version of MCMC, all random variables on a computer are discrete, so the discrete versions of MCMC we dealt with should be sufficient to do this problem.)

(Hint: Consider approximating the integral by a Riemann sum, or something similar to that.)

Solution 1. Let $\Omega \subseteq \mathbf{R}$ be a finite, discrete set. For any $x \in \Omega$, let $\pi(x) := \frac{e^{-x^2-x^4-x^8}}{\sum_{y \in \Omega} e^{-y^2-y^4-y^8}}$. Then $\pi(x) > 0$ for all $x \in \Omega$ and $\sum_{x \in \Omega} \pi(x) = 1$. We are given the $|\Omega| \times |\Omega|$ transition matrix Q, which corresponds to sampling from a mean zero variance one Gaussian, i.e. $Q(x, y)$ is approximately $e^{-(x-y)^2/2}/z$, where z is an appropriate constant. The Metropolis Algorithm 9.80 in the notes then gives us a transition matrix P with stationary distribution π. Importantly, since P is a function of $\pi(x)/\pi(y)$, the definition of P does not require computing the constant $\sum_{y \in \Omega} e^{-y^2-y^4-y^8}$. Also, by definition of P, Q, we see that $P(x,y) > 0$ for all $x, y \in \Omega$, so that P is irreducible, and π is the unique stationary distribution of P. Likewise, P is aperiodic, so the convergence theorem implies that $\max_{x \in \Omega} \sum_{y \in \Omega} |P^n(x,y) - \pi(y)| \leq C\alpha^n$ for some $0 < \alpha < 1$.

Let $X_1, X_2, \ldots$ be samples from this Markov Chain. Define $Y_n := \frac{1}{n} \sum_{k=1}^n e^{X_k}$. Let T be the first time that $X_1, X_2, \ldots$ has a repeated value, and similarly let $T^{(j)}$ be the j^{th} time the Markov Chain has a repeated value. The Strong Markov Property implies that $\sum_{i=1}^{T-1} e^{X_i}, \sum_{i=T}^{T^{(2)}-1} e^{X_i}, \ldots$ are i.i.d. random variables, so the Strong Law of Large Numbers implies that their average converges to their expected value as time goes to infinity.

Solution 2. You could appeal directly to Corollary 6.28 or Corollary 6.26 or Theorem 6.25 in the Robert and Casella book, which considers MCMC on infinite state spaces (in this case, using the state space $\Omega = \mathbf{R}$). Theorem 6.25 requires checking that the Markov Chain is irreducible (on an infinite state space), which follows since the Gaussian density is positive everywhere. Similarly, Corollaries 6.26 and 6.28 require checking conditions about the conditional density of the Markov chain.

5. Question 5

Give an example of a Markov Chain that is **not** reversible.
Prove your assertion.

Solution. In the previous question, we showed that the transition matrix $P := \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ has at least one non-real eigenvalue. We also showed that reversible Markov chains have all real eigenvalues. It follows that P is not reversible.

6. Question 6

Give an example of a transition matrix P for a finite Markov Chain such that:

- The Markov Chain is reversible.
- There is a stationary distribution π that does **not** satisfy the detailed balance condition.

Prove your assertions.

Solution. Consider

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

We observe that $\pi' := (1,0,0,0)$ satisfies the detailed balance condition, so P is reversible. ($\pi(1)P(1,y) = 0$ unless $y = 1$, and $\pi(y)P(y,1) = 0$ unless $y = 1$, so $\pi(1)P(1,y) = \pi(y)P(y,1)$ for all $y \in \{1,2,3,4\}$. Also, if $x \neq 1$, then $\pi(x)P(x,y) = 0$ by definition of π, and $\pi(y)P(y,x) = 0$ since $P(1,x) = 0$ when $x \neq 1$ and $\pi(y) = 0$ when $y \neq 1$. So, in any case, $\pi(x)P(x,y) = \pi(y)P(y,x)$.)

However, $\pi = (0, 1/3, 1/3, 1/3)$ is also a stationary distribution that does not satisfy the detailed balance condition. (If it did, then the transition matrix from Question 5 would be reversible, but we showed it was not.)