Tutorial and Lab Problems # 6
MATH3871/MATH5970

1. **Markov Chain Simulation.** Simulate the Markov chain in Example 1.1 in the lecture notes in R/Matlab. We know that the one-step transition density $\kappa(\boldsymbol{x} \,|\, \boldsymbol{y})$ is given by $\boldsymbol{y}^\top \mathbf{P}$, where $\boldsymbol{y}^\top$ is a row vector indicating the initial state (or distribution) of the Markov chain (e.g., $(0, 1, 0, 0)^\top$ indicates we have started in state two and $(1, 1, 1, 1)^\top/4$ indicates that the initial state was chosen at random). It follows that the the t-step transition density $\kappa_t(\boldsymbol{x} \,|\, \boldsymbol{y})$ is given by the matrix power:

$$\boldsymbol{y}^\top \mathbf{P}^t = \boldsymbol{y}^\top \underbrace{\mathbf{P} \cdots \mathbf{P}}_{t} \,.$$

   Compute $\kappa_{10}(\boldsymbol{x} \,|\, \boldsymbol{y})$ and $\kappa_{100}(\boldsymbol{x} \,|\, \boldsymbol{y})$ for $\boldsymbol{y} = (1, 0, 0, 0)^\top$ and $\boldsymbol{y} = (0, 1, 0, 0)^\top$. What do you see? How long does it take to "forget" the initial state?

2. **Effective Sample Size.** The effective sample size using weights/likelihood ratios $\{W_k\}_{k=1}^t$ is $(\sum_{k=1}^t W_k)^2 / \sum_{k=1}^t W_k^2$. Show that it ranges between 1 and t.

3. **Lowest-variance importance sampling pdf.** Suppose we wish to estimate

$$\alpha := \int \pi(x)\phi(x)\mathrm{d}x$$

   via importance sampling. Show that the minimum-variance importance sampling density is:

$$g^*(x) = \frac{\pi(x)|\phi(x)|}{\int \pi(x)|\phi(x)|\mathrm{d}x} \,.$$

   What is the value of the variance when $\phi \geq 0$ for all x? Explain what prevents us from using this optimal importance sampling density.

Answers:

1. We observe that whatever the initial state

$$\kappa_t(\boldsymbol{x} \mid \boldsymbol{y}) = (0.1733\ldots, 0.1467\ldots, 0.1600\ldots, 0.5200\ldots)$$

   for $t > 100$. Thus, the initial state seems to be forgotten after approximately 100 steps. We obtain that $\mathbb{P}[X_{100} = 1 \mid Y = y] = 0.1733\ldots$ etc.

2. The $\leq t$ result follows by rearranging the obvious variance inequality: $1/t \sum_k W_k^2 \geq (1/t \sum_k W_k)^2$. The second one follows from the fact that

$$\max\{\textstyle\sum_k p_k^2 : p_k \geq 0, \sum_k p_k = 1\} = 1.$$

3. The variance of the importance sampling estimator is:

$$\frac{\mathbb{E}_g(\pi\phi/g)^2 - \alpha^2}{t}.$$

   This is minimized when $\mathbb{E}_g(\pi\phi/g)^2$ is minimal. We also have the variance inequality:

$$\mathbb{E}_g(\pi(X)\phi(X)/g(X))^2 = \mathbb{E}_g(\pi(X)|\phi(X)|/g(X))^2 \geq \left(\int \pi(x)|\phi(x)|\mathrm{d}x\right)^2 .$$

   Thus, the smallest $\mathbb{E}_g(\pi\phi/g)^2$ it can ever be (for any conceivable g) is $\left(\int \pi(x)|\phi(x)|\mathrm{d}x\right)^2$ and it is easy to check that g^* achieves this lower bound. Variance is zero when $\phi \geq 0$. Sadly, this importance sampling density requires knowledge of the ery thing we wish to estimate and so cannot be used in practice. For maths students: proof similar to establishing Cramer-Rao lower bound for MLE.