4F7-STATISTICAL SIGNAL ANALYSIS

Extra Practise

Question 1: Let X and Y be continuous random variables. The probability density function of X is $p(x)$ and the conditional probability density function of Y given $X = x$ is $p(y \mid x)$. The random variable Y is observed to be $\bar{y}$.

(1) Let $q(x)$ be a probability density function. Given N independent and identically distributed samples, denoted $X^1, \ldots, X^N$, from $q(x)$, give the importance sampling estimates of $p(\bar{y})$, $\int h(x)p(x, \bar{y})dx$ and $\int h(x)p(x \mid \bar{y})dx$ where $h(x)$ is some function of interest.

(2) Find the variance of the importance sampling estimate of $p(\bar{y})$ and show that the variance is minimised when $q(x) = p(x \mid \bar{y})$.

Let $X_0, X_1, \ldots$ be a finite state Markov chain taking values in the set $S = \{1, 2, \ldots, n\}$. Let the matrix P with elements $P_{i,j}$ be the transition probability matrix of this Markov chain, i.e.

$$\Pr(X_{k+1} = j \mid X_k = i) = P_{i,j}.$$

Assume $X_0 = 1$.

Question 2: Find the probability of observing the sequence $X_1 = i_1, \ldots, X_k = i_k$.

Question 3: Let $Y_k = X_k + V_k$ for $k = 1, 2, \ldots$where $V_1, V_2, \ldots$ is a sequence of independent and identically distributed zero mean and unit variance Gaussian random variables. Find $p(i_1, \ldots, i_k \mid y_1, \ldots, y_k)$, which is the conditional probability mass function of $X_1, \ldots, X_k$ given $Y_1 = y_1, \ldots, Y_k = y_k$.

Question 4: What is the computational cost of calculating $p(i_1, \ldots, i_k \mid y_1, \ldots, y_k)$ exactly?

Question 5: Give an N-sample importance sampling estimate of $p(i_1, \ldots, i_k \mid y_1, \ldots, y_k)$ using the proposal probability mass function

$$q_k(i_1, \ldots, i_k) = P_{1,i_1} \cdots P_{i_{k-1},i_k}.$$

Question 6: Extend (sequentially) the importance sampling estimate of $p(i_1, \ldots, i_k \mid y_1, \ldots, y_k)$ to an importance sampling estimate of $p(i_1, \ldots, i_{k+1} \mid y_1, \ldots, y_k)$ and then to an estimate of $p(i_1, \ldots, i_{k+1} \mid y_1, \ldots, y_{k+1})$.

Question 7: Give the importance sampling estimates of

$$\sum_{i_1=1}^{n} h(i_1) p(i_1 \mid y_1, \ldots, y_{k+1})$$

and

$$\sum_{i_{k+1}=1}^{n} h(i_{k+1}) p(i_{k+1} \mid y_1, \ldots, y_{k+1})$$

where h is some real-valued function of interest.

Question 8: Comment on how effective this importance sampling estimate of $p(i_k \mid y_1, \ldots, y_k)$ is likely to be as k increases.

Question 9: Give the importance sampling estimate of $p(y_1, \ldots, y_k)$ and find the relative variance of this importance sampling estimate. Comment on how quickly the relative variance grows as a function of k? (The relative variance is defined to be the variance/mean^2.)

Solutions

Q1: Part 1. The importance sampling estimate of $\int h(x)p(x,\bar{y})dx$ is

$$\frac{1}{N}\sum_{i=1}^{N} h(X^i)w(X^i)$$

where $w(x) = p(x,\bar{y})/q(x)$. Set $h(x) = 1$ in the above expression to obtain the importance sampling estimate of $p(\bar{y}) = \int p(x,\bar{y})dx$. The importance sampling estimate of

$$\int h(x)p(x \mid \bar{y})dx = \frac{1}{p(\bar{y})}\int h(x)p(x,\bar{y})dx$$

is

$$\frac{\sum_{i=1}^{N} h(X^i)w(X^i)}{\sum_{i=1}^{N} w(X^i)}.$$

Part 2. The variance of the importance sampling estimate of $p(\bar{y})$, using the fact the samples are independent, is

$$\mathrm{Var}\left(\sum_{i=1}^{N}\frac{1}{N}w(X^i)\right) = \sum_{i=1}^{N}\mathrm{Var}\left(\frac{1}{N}w(X^i)\right) = \frac{1}{N}\mathrm{Var}\left(w(X^1)\right)$$

and

$$\begin{aligned}
\mathrm{Var}\left(w(X^1)\right) &= \mathbb{E}\left(w(X^1)^2\right) - \mathbb{E}\left(w(X^1)\right)^2 \\
&= \int \frac{p(x,\bar{y})}{q(x)}p(x,\bar{y})dx - p(\bar{y})^2 \\
&= p(\bar{y})^2\left(\int \frac{p(x \mid \bar{y})}{q(x)}p(x \mid \bar{y})dx - 1\right)
\end{aligned}$$

which is clearly zero when $q(x) = p(x \mid \bar{y})$. (We cannot do better than zero variance.)

Q2: The probability of observing the sequence $X_1 = i_1, \dots, X_k = i_k$ is

$$P_{1,i_1} P_{i_1,i_2} \cdots P_{i_{k-1},i_k}.$$

Q3: Using Bayes' law,

$$\begin{aligned} & p(i_1, \dots, i_k \mid y_1, \dots, y_k) \\ & \quad = \frac{p(i_1, \dots, i_k, y_1, \dots, y_k)}{p(y_1, \dots, y_k)} \\ & \quad = \frac{p(y_1, \dots, y_k \mid i_1, \dots, i_k) p(i_1, \dots, i_k)}{p(y_1, \dots, y_k)} \end{aligned}$$

The observations are conditionally independent give $X_1 = i_1, \dots, X_k = i_k$. Thus

$$p(y_1, \dots, y_k \mid i_1, \dots, i_k) = p(y_1 \mid i_1) \cdots p(y_k \mid i_k).$$

Using the fact that $p(y_1 \mid i_1) = \frac{1}{\sqrt{2\pi}} \exp\left(-0.5(y_1 - i_1)^2\right)$ we have

$$\begin{aligned} p(i_1, \dots, i_k \mid y_1, \dots, y_k) &= p(i_1, \dots, i_k, y_1, \dots, y_k) / p(y_1, \dots, y_k) \\ p(i_1, \dots, i_k, y_1, \dots, y_k) &= P_{1,i_1} P_{i_1,i_2} \cdots P_{i_{k-1},i_k} \left(\frac{1}{2\pi}\right)^{k/2} \\ & \qquad \times \exp\left(-0.5(y_1 - i_1)^2 - \cdots - 0.5(y_k - i_k)^2\right) \\ p(y_1, \dots, y_k) &= \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} p(i_1, \dots, i_k, y_1, \dots, y_k) \end{aligned}$$

Q4: To calculate $p(i_1, \dots, i_k \mid y_1, \dots, y_k)$ we need to calculate the value of this conditional probability mass function for all possible sequences $i_1, \dots, i_k$. There are n^k such terms. Clearly this will become too costly very quickly.

Q5: Let $X^j_{1:k} = (X^j_1, \dots, X^j_k)$ denote the j-th sample from $q_k(i_1, \dots, i_k)$. The weight is

$$w_k(X^j_{1:k}) = \frac{p(X^j_{1:k}, y_{1:k})}{q_k(X^j_{1;k})} = p(y_{1:k} \mid X^j_{1:k}) = p(y_1 \mid X^j_1) \cdots p(y_k \mid X^j_k)$$

since $q_k(X^j_{1;k}) = p(X^j_1, \dots, X^j_k)$. The importance sampling estimate of

$$\sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} H(i_{1:k}) p(i_{1:k} \mid y_{1:k})$$

for any function $H(i_1, \dots, i_k)$ is

$$\frac{\sum_{j=1}^{N} H(X^j_{1:k}) w_k(X^j_{1:k})}{\sum_{j=1}^{N} w_k(X^j_{1:k})}$$

Note that the N-sample importance sampling estimate now only stores $p(i_1, \dots, i_k \mid y_1, \dots, y_k)$ at N sequences.

Q6: To obtain an importance sampling estimate of $p(i_{1:k+1} \mid y_{1:k})$ extend each sample $X^j_{1:k}$ to $(X^j_{1:k}, X^j_{k+1})$ by sampling X^j_{k+1} from the transition probability matrix P. (If $X^j_k = m$ then row m of this matrix.) No change in weights.

Since $X^j_{1:k+1}$ is the j-th sample from $q_{k+1}(i_1, \dots, i_{k+1})$. The weight is

$$w_{k+1}(X^j_{1:k+1}) = \frac{p(X^j_{1:k+1}, y_{1:k+1})}{q_{k+1}(X^j_{1;k+1})} = p(y_{1:k+1} \mid X^j_{1:k+1}) = w_k(X^j_{1:k}) p(y_{k+1} \mid X^j_{k+1})$$

which is the previous weight multiplied by $p(y_{k+1} \mid X^j_{k+1})$. The importance sampling estimate of

$$\sum_{i_1=1}^{n} \cdots \sum_{i_{k+1}=1}^{n} H(i_{1:k+1}) p(i_{1:k+1} \mid y_{1:k+1})$$

for any function $H(i_1, \ldots, i_{k+1})$ is

$$\frac{\sum_{j=1}^{N} H(X_{1:k+1}^j) w_{k+1}(X_{1:k+1}^j)}{\sum_{j=1}^{N} w_{k+1}(X_{1:k+1}^j)}.$$

Q7: The importance sampling estimates of

$$\sum_{i_1=1}^{n} h(i_1) p(i_1 \mid y_1, \ldots, y_{k+1})$$

is

$$\frac{\sum_{j=1}^{N} h(X_1^j) w_{k+1}(X_{1:k+1}^j)}{\sum_{j=1}^{N} w_{k+1}(X_{1:k+1}^j)}.$$

The importance sampling estimate of $\sum_{i_{k+1}=1}^{n} h(i_{k+1}) p(i_{k+1} \mid y_1, \ldots, y_{k+1})$ is

$$\frac{\sum_{j=1}^{N} h(X_{k+1}^j) w_{k+1}(X_{1:k+1}^j)}{\sum_{j=1}^{N} w_{k+1}(X_{1:k+1}^j)}.$$

(The point here is to remind you that if you have an import8ant sampling estimate of $p(i_{1:k+1} \mid y_{1:k+1})$ then you can get an importance sampling estimate of $p(i_m \mid y_1, \ldots, y_{k+1})$ for any $m \leq k+1$.)

Q8: Note that $w_k(X_{1:k}^j)$ is a product of k terms. Eventually the weights will become unbalanced and only one particle will dominate when the weights are normalised. The full discussion of this issue was given in the lecture notes on page 46.

Q9: We can use the result in question 1 to calculate the variance of the importance sampling estimate of $p(y_1, \ldots, y_k)$. From question 1, the importance sampling estimate of $p(y_1, \ldots, y_k)$ is

$\frac{1}{N}\sum_{j=1}^{N} w_k(X_{1:k}^j)$ and the variance is

$$\begin{aligned}\frac{1}{N}\mathrm{Var}\left(w(X^1)\right) &= \frac{1}{N}p(y_1,\ldots,y_k)^2 \\ &\quad \times \left(\sum_{i_1=1}^{n}\cdots\sum_{i_k=1}^{n}\frac{p(i_1,\ldots,i_k \mid y_1,\ldots,y_k)}{q_k(i_1,\ldots,i_k)}p(i_1,\ldots,i_k \mid y_1,\ldots,y_k) - 1\right)\end{aligned}$$

Using the fact that $q_k(i_1,\ldots,i_k) = p_k(i_1,\ldots,i_k)$ to get

$$\begin{aligned}\frac{1}{N}\mathrm{Var}\left(w(X^1)\right) &= \frac{1}{N}p(y_1,\ldots,y_k)^2 \\ &\quad \times \left(\sum_{i_1=1}^{n}\cdots\sum_{i_k=1}^{n}\frac{p(y_1,\ldots,y_k \mid i_1,\ldots,i_k)}{p(y_1,\ldots,y_k)}p(i_1,\ldots,i_k \mid y_1,\ldots,y_k) - 1\right)\end{aligned}$$

The relative variance is variance/mean^2 or

$$\frac{1}{N}\left(\sum_{i_1=1}^{n}\cdots\sum_{i_k=1}^{n}\frac{p(y_1,\ldots,y_k \mid i_1,\ldots,i_k)}{p(y_1,\ldots,y_k)}p(i_1,\ldots,i_k \mid y_1,\ldots,y_k) - 1\right).$$

You should be able to reach this stage of the calculation. The final remark to be made (without proof) is that this expression grows exponentially in k.

S.S. Singh, Department of Engineering, University of Cambridge, Cambridge, CB1 7AT, UK

E-mail address: `sss40@cam.ac.uk`