University of Technology Sydney
School of Mathematical and Physical Sciences

Probability and Random Variables (37161) – Tutorial/Laboratory 9
TO BE HANDED IN FOR ASSESSMENT

Please hand in your answers, showing all relevant working in the spaces provided. You may use additional sheets for rough working, but only these worksheets should be submitted for assessment.

1. For each of the following variables, state whether or not the sequence could reasonably be assumed to form a Markov Chain. Briefly explain each of your answers.

    i) The number of minutes past the hour (e.g. 1:34pm recorded as 34) when recorded at timepoints exactly one minute apart.

    > This is a Markov Chain, since we know that each observation will be one larger than the previous one, unless it is 59 minutes past the hour, in which case it will certainly be 0 next.

    ii) The day of the month (between 1 and 31) when recorded at timepoints exactly one day apart.

    > This is not a Markov Chain, since we do not know whether the observation after 30 will be 31 or 1 without additional information.

    iii) When rolling a regular fair six-sided die repeatedly, an indicator variable which records 1 if the outcome is equal to the previous roll and 0 otherwise.

    > This is a Markov Chain, since any consecutive pair of rolls are equal with probability 1/6.

    iv) When rolling a regular fair six-sided die repeatedly, an indicator variable which records 1 if the outcome is equal to the sum of the previous two rolls and 0 otherwise.

    > This does not form a Markov Chain, since you need to know both of the previous two rolls in order to assess if the next roll is equal.

(8 Marks)

2. A standard (European-style) roulette wheel contains 18 red numbers, 18 black and 1 green. A gambler walks into a casino and bets on red repeatedly. Each time he/she wins, he/she gains an amount equal to his/her stake. Each time he/she loses, he/she loses his/her stake.

The gambler keeps placing bets until he/she either has \$80 and leaves the casino in profit, or runs out of money i.e. has \$0.
Let W_k be the probability that the gambler eventually wins \$80, given that he/she has \$$k$ at a given time.

i) The gambler starts with \$50 and bets \$1 each time. By conditioning on each possible event, show that W_k satisfies the difference equation $37W_k = 18W_{k+1} + 19W_{k-1}$.

If the gambler has \$$k$, then he/she either loses the next bet and has $\$(k-1)$ after the next bet, or wins and has $\$(k+1)$ after the next bet. These occur with probabilities 19/37 and 18/37 respectively.

Hence $W_k = \frac{18}{37}W_{k+1} + \frac{19}{37}W_{k-1}$ or $37W_k = 18W_{k+1} + 19W_{k-1}$.

ii) What are the boundary conditions, W_0 and W_{80}? Briefly explain your answers.

If he has \$0 he/she has definitely lost. If he has \$80, he/she definitely leaves in profit, hence $W_0 = 0$ and $W_{80} = 1$.

iii) Solve the difference equation to show that the probability the gambler leaves the casino in profit is approximately 10.3%

Solving $37W_k = 18W_{k+1} + 19W_{k-1}$, we seek a solution of the form $W_k = CM^k$, then we get the auxiliary equation $37M = 18M^2 + 19 = 0$ or $(M-1)(18M-19) = 0$. so the roots of this are $M = 1$ or 19/18.

$$W_k = C_1 1^k + C_2 \left(\frac{19}{18}\right)^k$$

The boundary conditions give $C_1 = \dfrac{1}{1-\left(\frac{19}{18}\right)^{80}}, C_2 = \dfrac{-1}{1-\left(\frac{19}{18}\right)^{80}}$ that

so $W_k = \dfrac{1-\left(\frac{19}{18}\right)^k}{1-\left(\frac{19}{18}\right)^{80}}$ .This then gives $W_{40} = \dfrac{1-\left(\frac{19}{18}\right)^{40}}{1-\left(\frac{19}{18}\right)^{80}} \approx 10.3\%$.

iv) If he/she bets in multiples of $10 instead of $1, does his/her probability of leaving in profit increase, decrease or stay the same? Justify your answer.

Changing the bet size from $1 to $10 is effectively the same as starting with $4 and betting $1 increments until reaching $0 or $8.

The resulting probability is therefore $W_{40} = \dfrac{1-\left(\frac{19}{18}\right)^4}{1-\left(\frac{19}{18}\right)^8} \approx 44.6\%$,

which is an increased probability of profit.

v) Starting with $40 and playing until first reaching $80 or $0, which bet size maximises the player's chance of leaving in profit? Justify your answer.

The chance of winning is maximised with one single bet of $40. This gives a probability of profit of 18/37 i.e. around 48.6%.

(9 Marks)

3.

i) Show that the generating function of $X \sim Geo(p)$ is

$$g_X(z) = \frac{pz}{1-(1-p)z}.$$

$$g_X(z) = E(z^X) = \sum_{k=0}^{\infty} z^k p(1-p)^{k-1}$$
$$= pz + pz(1-p)z + pz(1-p)^2 z^2 + pz(1-p)^3 z^3 + \ldots$$

This is a geometric series, first term pz, common ratio $(1-p)z$.

Hence (assuming $-1 < (1-p)z < 1$) $g_X(z) = \dfrac{pz}{1-(1-p)z}$.

The random variable Y describes how many independent $Bern(p)$ variables must be counted until the nth $(n \in \mathbb{Z}^+)$ 1 is seen.
This describes a negative binomial variable, $Y \sim NegBin(n,p)$.

(For example, if the sequence of $Bern(p)$ variables were 0, 0, 1, 0, 0, 1,… then the corresponding $NegBin(2,p)$ variable would take the value 6 since the 2nd 1 is seen on the 6th $Bern(p)$ variable)

i) What is the range of Y? Justify your answer.

The nth 1 cannot be observed until at least n $Bern(p)$ variables have been observed, but there is no theoretical maximum value hence the range of Y is $\{n, n+1, n+2, n+3, \ldots\}$

ii) Calculate the generating function of Y.

As Y is the sum of n independent $Geo(p)$ variables, its generating function is equal to the product of n $Geo(p)$ variables' generating functions. This gives $g_Y(z) = \left(\dfrac{pz}{1-(1-p)z}\right)^n$

iii) Hence or otherwise, find $E(Y)$. Justify your answer.

We could find $E(Y)$ by differentiating $g_Y(z) = \left(\dfrac{pz}{1-(1-p)z}\right)^n$.

A simpler method is to reason that as, for $X \sim Geo(p)$, $E(X) = \dfrac{1}{p}$

then summing n of these variables gives $E(Y) = \dfrac{n}{p}$.

(7 Marks)

4.

If $X \sim Bin(n, p_1)$ then $g_X(z) = E(z^X) = \left[(1-p_1) + p_1 z\right]^n$

Let $T_1, T_2, \ldots, T_X$ each be independent variables such that each $T_i \sim Bern(p_2)$ and where $X \sim Bin(n, p_1)$. Let $S = \sum_{i=0}^{X} T_i$.

i) Show that S is a binomial random variable and find both its parameters.

$$g_S(z) = g_X(g_{T_i}(z)) = \left[(1-p_1) + p_1\left[(1-p_2) + p_2 z\right]\right]^n$$
$$= \left[(1-p_1 p_2) + p_1 p_2 z\right]^n.$$

Hence $S \sim Bin(n, p_1 p_2)$.

Hints: You may use without proof the result that $g_S(z) = g_X(g_{T_i}(z))$.

Recall that if $T \sim Bern(p)$ then $T \sim Bin(1, p)$.

ii) If $X \sim Bin(n, p_1)$, $T \sim Bern(p_2)$, $U \sim Bern(p_3)$ and $V \sim Bern(p_4)$, write down the distribution of C *where*

$$S = \sum_{i=0}^{X} T_i \quad D = \sum_{i=0}^{S} U_i \text{ and } C = \sum_{i=0}^{D} V_i.$$

Justify your answer.

$S \sim Bin(n, p_1 p_2)$.

By the same logic, $D \sim Bin(n, p_1 p_2 p_3)$.

Hence, $C \sim Bin(n, p_1 p_2 p_3 p_4)$.

This can be seen by repeatedly calculating generating functions of generating functions i.e. $g_S(z) = g_X(g_{T_i}(z))$ so

$$g_C(z) = g_D(g_S(g_X(g_{T_i}(z)))).$$

(6 Marks)

5. Written task. Please see the marking rubric on UTSOnline before attempting this problem.

In your own words, clearly describe a situation which would give rise to a sequence of observations which could reasonably be modelled by a Markov Chain with four possible states. Explain why you believe the Markov property holds for this sequence.

In your own words, clearly describe a second situation which would give rise to a sequence of observations with four possible states which could not reasonably be modelled by a Markov Chain. Explain why you believe the Markov property does not hold for this sequence.

Explain your example in everyday language, not mathematical notation.

Consider two random experiments, each consisting of repeatedly drawing numbered balls from a bag and noting the selected number. Initially there are 4000 balls in the bag. 1000 are numbered one, 1000 are numbered two, 1000 are numbered three and 1000 are numbered four.
If the number of the ball is noted each time, then we have a sequence of observations with four states (one, two, three and four.)

If balls are replaced after observation, then we have a Markov Chain with four states since each draw would be independent and hence no prior information is useful for predicting future outcomes.

If balls are not replaced after observation, then this would not be a Markov Chain. For example, knowing that the first 1000 balls drawn were all numbered three, we would be certain that no future ball drawn would be a three. Knowing simply one observation (i.e. that the last ball drawn was a three) would not tell us as much as this and hence our prediction based on only one observation would not be the same as our prediction based on numerous past observations.

(5 marks, including assessment for English Language)