Stat 516

Homework 2

Alden Timme and Marco Ribeiro

Due Date: Tuesday, October 14

1. Conditional independence

(a) If (X, Y, Z) is multivariate normal with $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y|Z$ then we have that the covariance matrix Σ is has the constraints $\Sigma_{12} = 0$ and $\Lambda_{12} = 0$, where $\Lambda = \Sigma^{-1}$ is the precision matrix. Using these constraints and finding the $(1, 2)$ element of the Σ matrix, we have

$$\frac{1}{\det(\Sigma)} (\sigma_{xz}\sigma_{yz}) = 0$$

which implies that $\sigma_{xz} = 0$ or $\sigma_{yz} = 0$.
In the first case ($\sigma_{xz} = 0$), we have $X \perp\!\!\!\perp Z$. This implies that $X \perp\!\!\!\perp (Y, Z)$, since

$$\begin{aligned}
\Pr(X, Y, Z) &= \Pr(X, Y|Z)\Pr(Z) & \\
&= \Pr(X|Z)\Pr(Y|Z)\Pr(Z) & X \perp\!\!\!\perp Y|Z \\
&= \Pr(X)\Pr(Y|Z)\Pr(Z) & X \perp\!\!\!\perp Z \\
&= \Pr(X)\Pr(Y, Z) &
\end{aligned}$$

In the second case ($\sigma_{yz} = 0$), we have $Y \perp\!\!\!\perp Z$. This implies that $(X, Z) \perp\!\!\!\perp Y$, since

$$\begin{aligned}
\Pr(X, Y, Z) &= \Pr(X, Z|Y)\Pr(Y) & \\
&= \Pr(X|Y)\Pr(Z|Y)\Pr(Y) & X \perp\!\!\!\perp Y|Z \\
&= \Pr(X)\Pr(Y)\Pr(Z) & Y \perp\!\!\!\perp Z \\
&= \Pr(X, Z)\Pr(Y) &
\end{aligned}$$

Therefore, we have that if (X, Y, Z) is multivariate normal with $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y|Z$, then either $X \perp\!\!\!\perp (Y, Z)$ or $(X, Z) \perp\!\!\!\perp Y$.

(b) Take the following joint distribution on (X, Y, Z):

X	Y	Z	$\Pr(X, Y, Z)$
0	0	0	0.2
0	0	1	0.1
0	1	0	0.15
0	1	1	0.05
1	0	0	0.25
1	0	1	0.05
1	1	0	0.1
1	1	1	0.1

With this joint distribution, we have

$$\Pr(X = 1) = \Pr(X = 1|Y = 1) = \Pr(X = 1|Y = 0) = 0.5$$
$$\Pr(X = 0) = \Pr(X = 0|Y = 1) = \Pr(X = 0|Y = 0) = 0.5$$

so $X \perp\!\!\!\perp Y$, and

$$\Pr(X=1) = \Pr(X=1|Z=1) = \Pr(X=1|Z=0) = 0.5$$
$$\Pr(X=0) = \Pr(X=0|Z=1) = \Pr(X=0|Z=0) = 0.5$$

so $X \perp\!\!\!\perp Z$. However,

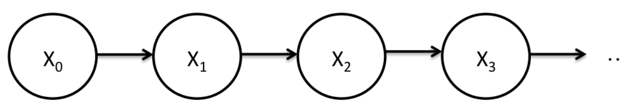
$$\Pr(X=1|Y=1, Z=1) = 2/3 \neq \Pr(X=1)$$

so we see that X is not independent of the pair (Y, Z).

(c) No, there does not exist a binary random variable Z such that $X \perp\!\!\!\perp Y|Z$.
Note first that if two discrete random variables are independent, then the matrix of probabilities p_{ij} must be of rank at most 1. This is because $p_{ij} = p_{i\cdot}p_{\cdot j}$ – i.e. the matrix of probabilities is the outer product of the probabilities of two random variables.
As a direct consequence, if two discrete random variables are independent given a third discrete random variable, then all matrices of conditional probabilities of the first two random variables given the third random variable must also have rank at most 1.
Now, in this case, notice that the matrix P has rank 3. We know, as stated above, that conditional independence of X and Y given Z requires that each matrix $P(X, Y|Z = z_0)$ and $P(X, Y|Z = z_1)$ must have rank at most one. However, these constitute only two matrices, and we have $P(X, Y) = P(X, Y|Z = z_0)P(Z = z_0) + P(X, Y|Z = z_1)P(Z = z_1)$, which is just a weighted sum of two rank-1 matrices. But the sum of two rank-1 matrices cannot have rank more than 2. Therefore, there does not exist a binary random variable Z such that $X \perp\!\!\!\perp Y|Z$.

2. Conditional independence and graph separation

The class definition of a Markov Chain defines a distribution that clearly factorizes according to the following graph:

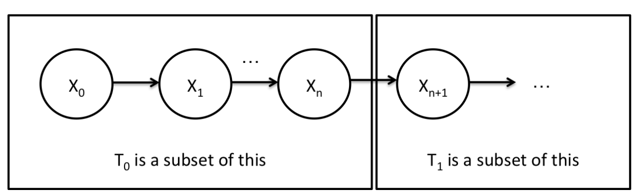


From the definition, $X_a \perp\!\!\!\perp X_b|X_c$ if there is no path from X_a to X_b that doesn't pass through X_c in the graph. Furthermore, by the definitions of conditional independence,

(1) $X \perp\!\!\!\perp Y|Z \Rightarrow P(X|Y, Z) = P(X|Z)$

(2) $X \perp\!\!\!\perp Y|Z \Rightarrow P(X, Y|Z) = P(X|Z)P(Y|Z)$

(a) In this statement, the sets are pictured below.

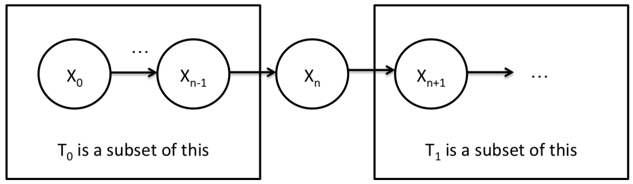


It is clear from the picture that the path betweeen any X_k in T_1 will have to go through t_0 to reach any X_l in X_{t_0}, and thus $X_k \perp\!\!\!\perp X_l|X_{t_0}$, which by definition (1) above means

$$P(X_k = i_k \forall k \in T_1 | X_l = i_l \forall l \in T_0 \text{ (which includes } X_{t_0})) = P(X_k = i_k \forall k \in T1 | X_{t_0} = i_{t_0})$$

We see then that the original graph also describes this statement.

(b) A visualization of the chain and sets is provided below:



From the picture above, it is clear that any path from X_k in T_1 to any X_l in T_0 must go through X_n, which means that $X_k \perp\!\!\!\perp X_l | X_n$ for all k, l. Applying definition (2) of conditional independence from above with $X = X_k$ for any $k \in T_1$, $Y = X_l$ for any $l \in T_0$, and $Z = X_n$ gives the statement. Once again, the same graph factorizes according to the distribution given by this statement.

3. Recognizing Markov chains

(a) To show that X_n is a Markov chain, we need only show that $P(X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1}$. That is, the distribution of X_n depends only on X_{n-1}. To do this, we need only specify the transition probabilities p_{ij}. Note that the state space is $\{1, 2, 3, 4, 5, 6\}$. From this, we can specify the transition matrix P

$$P = \begin{pmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 1/3 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 1/2 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 2/3 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

or

$$p_{ij} = \begin{cases} 0 & j < i \\ j/6 & j = i \\ 1/6 & j > i \end{cases}$$

and

$$P^{(n)} = \begin{pmatrix} 1 & 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{4} & 1/\sqrt{5} & 1/\sqrt{6} \\ 0 & 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{4} & 1/\sqrt{5} & 1/\sqrt{6} \\ 0 & 0 & 1/\sqrt{3} & 1/\sqrt{4} & 1/\sqrt{5} & 1/\sqrt{6} \\ 0 & 0 & 0 & 1/\sqrt{4} & 1/\sqrt{5} & 1/\sqrt{6} \\ 0 & 0 & 0 & 0 & 1/\sqrt{5} & 1/\sqrt{6} \\ 0 & 0 & 0 & 0 & 0 & 1/\sqrt{6} \end{pmatrix}$$

$$\cdot \begin{pmatrix} (1/6)^n & 0 & 0 & 0 & 0 & 0 \\ 0 & (1/3)^n & 0 & 0 & 0 & 0 \\ 0 & 0 & (1/2)^n & 0 & 0 & 0 \\ 0 & 0 & 0 & (2/3)^n & 0 & 0 \\ 0 & 0 & 0 & 0 & (5/6)^n & 0 \\ 0 & 0 & 0 & 0 & 0 & 1^n \end{pmatrix}$$

$$\cdot \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{2} & -\sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & \sqrt{3} & -\sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{4} & -\sqrt{4} & 0 \\ 0 & 0 & 0 & 0 & \sqrt{5} & -\sqrt{5} \\ 0 & 0 & 0 & 0 & 0 & \sqrt{6} \end{pmatrix}$$

$$= \begin{pmatrix} \left(\frac{1}{6}\right)^n & \left(\frac{1}{3}\right)^n - \left(\frac{1}{6}\right)^n & \left(\frac{1}{2}\right)^n - \left(\frac{1}{3}\right)^n & \left(\frac{2}{3}\right)^n - \left(\frac{1}{2}\right)^n & \left(\frac{5}{6}\right)^n - \left(\frac{2}{3}\right)^n & 1 - \left(\frac{5}{6}\right)^n \\ 0 & \left(\frac{1}{3}\right)^n & \left(\frac{1}{2}\right)^n - \left(\frac{1}{3}\right)^n & \left(\frac{2}{3}\right)^n - \left(\frac{1}{2}\right)^n & \left(\frac{5}{6}\right)^n - \left(\frac{2}{3}\right)^n & 1 - \left(\frac{5}{6}\right)^n \\ 0 & 0 & \left(\frac{1}{2}\right)^n & \left(\frac{2}{3}\right)^n - \left(\frac{1}{2}\right)^n & \left(\frac{5}{6}\right)^n - \left(\frac{2}{3}\right)^n & 1 - \left(\frac{5}{6}\right)^n \\ 0 & 0 & 0 & \left(\frac{2}{3}\right)^n & \left(\frac{5}{6}\right)^n - \left(\frac{2}{3}\right)^n & 1 - \left(\frac{5}{6}\right)^n \\ 0 & 0 & 0 & 0 & \left(\frac{5}{6}\right)^n & 1 - \left(\frac{5}{6}\right)^n \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

or

$$p_{ij}^{(n)} = \begin{cases} 0 & j < i \\ \left(\frac{i}{6}\right)^n & j = i \\ \left(\frac{j}{6}\right)^n - \left(\frac{j-1}{6}\right)^n & j > i \end{cases}$$

(b) No, Z_n is not a Markov chain. Intuitively, if we're trying to calculate $P(Z_4 = 2|Z_3 = 2)$, the information we have is that in the first three rolls:

- one of the rolls was a 2,
- one of the rolls was 2 or 1, and
- one of the rolls was a number $X \geq 2$.

Now, let's call the 4^{th} roll Y. For $Z_4 = 2|Z_3 = 2$, we need either $X > 2$ and $Y \leq 2$ or $X = 2$.
If we calculate $P(Z_4 = 2|Z_3 = 2, Z_2 = 1)$, we have the following information:

- one of the rolls was a 2,
- one of the rolls (the first or second one) was a 1, and
- one of the rolls was a number $X \geq 2$.

Now, for $Z_4 = 2|Z_3 = 2, Z_2 = 1$ we need $X > 2$ and $Y \geq 2$ or $X = 2$.
From the description of the three rolls in the two scenarios, it is clear that the distribution of $Z_4|Z_3 = 2, Z_2 = 1$ is different from $Z_4|Z_3 = 2$.
In order to prove this, one just has to calculate these values:

$$P(Z_4 = 2|Z_3 = 2) = 0.4$$
$$P(Z_4 = 2|Z_3 = 2, Z_2 = 1) = 0.4074$$

4. Markov chain with two states

To find a_n and b_n, we see that

$$\begin{aligned} a_n &= (1-a_{n-1})a + a_{n-1}(1-b) \\ &= a + (1-a-b)a_{n-1} \\ b_n &= (1-b_{n-1})b + b_{n-1}(1-a) \\ &= b + (1-a-b)b_{n-1} \end{aligned}$$

Looking at a_n specifically, we see

$$\begin{aligned} a_1 &= a \\ a_2 &= a + a(1-a-b) = a\left[1 + (1-a-b)\right] \\ a_3 &= a + a\left[1 + (1-a-b)\right](1-a-b) \\ &= a\left[1 + (1-a-b) + (1-a-b)^2\right] \\ a_4 &= a + a\left[1 + (1-a-b) + (1-a-b)^2\right](1-a-b) \\ &= a\left[1 + (1-a-b) + (1-a-b)^2 + (1-a-b)^3\right] \\ &\vdots \\ a_n &= a\sum_{j=0}^{n-1}(1-a-b)^j \end{aligned}$$

and similarly,

$$b_n = b\sum_{j=0}^{n-1}(1-a-b)^j$$

These can also be expressed

$$\begin{aligned} a_n &= a\frac{1-(1-a-b)^n}{a+b} \\ b_n &= b\frac{1-(1-a-b)^n}{a+b} \end{aligned}$$

As $n \to \infty$, the limit of P^n only converges when a and b are either both not 0 or both not 1. If both a and b are not 0 or 1, then

$$\begin{aligned} a_n &\to_{n\to\infty} \frac{a}{a+b} \\ b_n &\to_{n\to\infty} \frac{b}{a+b} \end{aligned}$$

yielding

$$P^n \to_{n\to\infty} \begin{pmatrix} \frac{b}{a+b} & \frac{a}{a+b} \\ \frac{b}{a+b} & \frac{a}{a+b} \end{pmatrix}$$

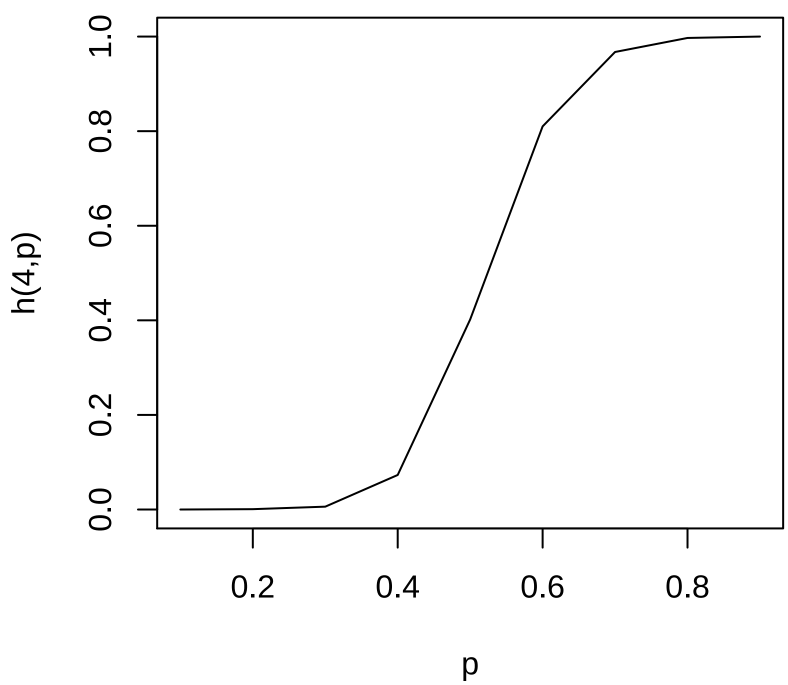
5. Simulating gambler's ruin

Code for parts (a) and (b) can be found in the appendix.

(a) The output of the 20 runs with $N = 10$, $i = 3$, and $p = 0.32$ is

```
## 3 4 3 2 1 2 1 0
## 3 2 1 2 3 4 3 2 1 0
## 3 2 1 0
## 3 2 1 2 1 2 3 2 3 2 1 0
## 3 2 1 0
## 3 4 3 2 1 2 1 0
## 3 2 1 0
## 3 2 1 0
## 3 4 3 2 1 0
## 3 2 1 2 3 4 5 4 3 2 1 2 1 0
## 3 2 3 2 1 2 1 0
## 3 2 1 0
## 3 2 1 0
## 3 2 1 0
## 3 2 1 0
## 3 2 1 2 1 2 3 4 5 6 5 4 3 2 1 0
## 3 2 1 0
## 3 2 3 2 1 2 3 2 1 0
## 3 2 1 0
## 3 2 3 4 3 4 3 2 1 0
```

(b) Estimating $h(4, p)$ against p and plotting, we find



Code

5(a)

```
set.seed(1)
GamblersRuin <- function(i, N, p) {
  states <- c(i)
  while (i != 0 && i != N) {
    if (runif(1) < p) {
      i <- i + 1
    } else {
      i <- i - 1
    }
    states <- c(states, i)
  }
  return(states)
}
all.runs <- list()
for (i in 1:20) {
  all.runs[[i]] <- GamblersRuin(3, 10, 0.32)
}
```

5(b)

```
num.runs <- 10000
probs <- (1:9)/10
N <- 10
i <- 4
h <- rep(0, length(probs))
for (j in 1:length(probs)) {
  p <- probs[j]
  num.absorbed <- 0
  for (n in 1:num.runs) {
    states <- GamblersRuin(i, N, p)
    if (states[length(states)] == N) {
      num.absorbed <- num.absorbed + 1
    }
  }
  h[j] <- num.absorbed / num.runs
}
```

```
plot(probs, h, type='l', xlab='p', ylab='h(4,p)')
```