Contrasts and Bonferroni method.

Motivation: suppose we have g groups and would like to test $H_0: \mu_1 + \dots + \mu_{g-1} = g\mu_g$ (the mean in the last group is the average of the means of the remaining) or

$$H_0: \frac{\mu_1 + \mu_g}{2} = \frac{\mu_2 + \dots + \mu_{g-1}}{2}.$$

Def-n: a contrast is a linear combination of population means $C = \sum_{i=1}^{g} a_i \mu_i$, s.t. $\sum_{i=1}^{g} a_i = 0$.

Question: how can we test $H_0: C = 0$?

Rmk (partial answer): consider the case $a_1 = 1, a_2 = -1, a_i = 0, i \neq 1,2$, then we can use Student's t-test (see page 15)

Our contrast c has an estimator $\tilde{c} = \sum_{i=1}^{g} a_i \bar{X}_i$ and we will use $S_c^2 = \frac{1}{N-g} \sum_{i=1}^{g} (N_i - 1) S_i^2 \cdot \sum_{i=1}^{g} \frac{a_i^2}{N_i} =$

$= \frac{SSE}{N-g} \sum_{i=1}^{g} \frac{a_i^2}{N_i}$ for the variance of c estimator.

<u>Thm.</u> To test $H_0: c=0$ against $H_1: c \neq 0$ at the α level of significance, use the test statistic $t = \frac{\tilde{c}}{S_c}$ and reject H_0 provided $|t| > t_{\alpha/2, N-g}$.

<u>Example (p. 15, problem 3 in Handout file).</u>

Below are the numbers of students in classes given in the Spring, Summer and Fall semesters:

Spring	40, 44, 36, 33, 15, 32, 13, 41, 31
Summer	21, 47, 42, 35, 25, 30
Fall	66, 36, 27, 36, 46, 22, 28, 32, 19, 43, 17, 35.

Test at the 5% level of significance if the mean in Summer is different from the average of the remaining two.

We will use the calculator. First, the setup:

$$C = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_3 - \mu_2, \text{ so } \tilde{C} = \sum_{i=1}^{3} a_i \bar{x}_i = \frac{1}{2}\cdot 31.3 + \frac{1}{2}\cdot 33.9 - 33.3$$

$$= -0.7$$

Use 'STAT' → 'CALC' → '1VAR STATS' to find $\bar{x}_i$'s.

$n_1 = 9$, $n_2 = 6$, $n_3 = 12$, $df = N - g = 27 - 3 = 24$

$$\sum_{i=1}^{3} \frac{a_i^2}{n_i} = \frac{1}{4\cdot 9} + \frac{1}{4\cdot 12} + \frac{1}{6} = 0.215$$

If we run the 'ANOVA' test on the calculator, the output will contain a value Sxp, which is equal to $\sqrt{\frac{SSE}{N-g}}$. In our case $\sqrt{\frac{SSE}{N-g}} = Sxp = 11.89$

Finally, $t = \frac{\tilde{C}}{S_C} = \frac{-0.7}{11.89\cdot\sqrt{0.215}} = -\frac{0.7}{5.513} = -0.127$.

The critical value is $t_{0.025,24} = 2.064$.

As $|t| < t_{0.025,24}$, we accept H_0.

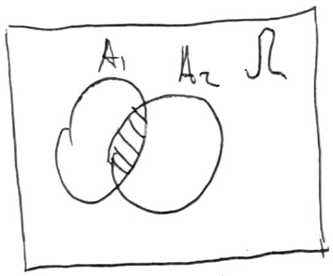
Bonferroni's Method.

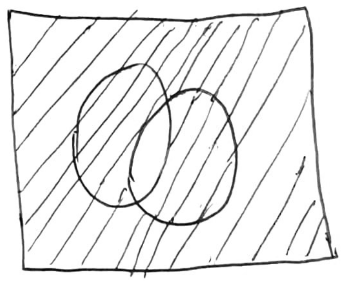
Suppose we have a collection of contrasts $C_1, C_2, \ldots, C_k$ and would like to test the hypothesis H_0 that every $C_i = 0$ at significance level α.

Lemma (Bonferroni). Let $A_1, \dots, A_g$ be a collection of events in a probability space (Ω, P) and $\bar{A}_i$ stand for the complement of A_i. Then the following inequality holds:

$$P\left(\bigcap_{i=1}^{g} A_i\right) \ge 1 - \sum_{i=1}^{g} P(\bar{A}_i) \qquad (*)$$

Example. $g=2$.





$\bar{A}_1 \cup \bar{A}_2$

$$P(\bar{A}_1 \cup \bar{A}_2) \le P(\bar{A}_1) + P(\bar{A}_2)$$

Since $\overline{A_1 \cap A_2} = \bar{A}_1 \cup \bar{A}_2$ (De Morgan's law) and $P(\overline{A_1 \cap A_2}) = 1 - P(A_1 \cap A_2)$, we have $P(\bar{A}_1 \cup \bar{A}_2) = 1 - P(A_1 \cap A_2) \le$

$\le P(\bar{A}_1) + P(\bar{A}_2)$ or $P(A_1 \cap A_2) \ge 1 - P(\bar{A}_1) - P(\bar{A}_2)$.

Exercise. Prove the lemma for general g.

Now let A_i be the event that the confidence interval for the contrast C_i contains the true value of that contrast, then the l.h.s. of inequality (*) is the probability that all confidence intervals simultaneously cover the true values of the respective contrasts.

Therefore, if each interval has confidence coeff. $1-\alpha/k$, the Bonferroni's inequality implies that the overall confidence coeff. does not go below $1-\alpha$.

Thm 2. Let $C_1, \dots, C_k$ be a collection of contrasts. We reject H_0 that all $C_i = 0$ (simultaneously) at signif. level α provided there exists C_i, s.t. $t_i = \frac{\hat{C}_i}{S_{C_i}}$ has $|t_i| > t_{\alpha/2k, N-g}$.

Linear regression.

Suppose we have collected bivariate data (x_i, y_i) $i=1,\dots,n$ and would like to model the relation between x and y by finding ~~the~~ a f-n $y=f(x)$ that is a good fit for the data. We will assume that x_i's are not random and y_i's are f-ns of x_i's plus some random noise.

Def-n. With the above assumptions x is called a **predictor (independent) variable** and y a **response (dependent) variable**.

Example. Suppose we have n pairs of fathers and adult sons. Let x_i and y_i be the heights of the i^{th} father and his son, respectively.

Def-n: the **best fit line** for the bivariate data $(x_i, y_i)_{i\in\{1,\dots,n\}}$ is the line $y=f(x)$ for which the sum of the squared errors is minimal, i.e. $\sum_{i=1}^{n}(f(x_i)-y_i)^2$ is minimized.

Rmk: the same applies for $y=f(x)$ a polynomial of degree $d>0$.

How can we find the eq-n of the best fit line?

1. The best fit line is a line, so $y = \beta_0 + \beta_1 x$.
We need to determine β_0 and β_1.

2. $y_i = \beta_1 x_i + \beta_0 + \varepsilon_i$ (here ε_i is some error that we want to minimize the square of).

$$\varepsilon_i = y_i - \beta_1 x_i - \beta_0.$$

The method of least squares finds the parameters $\hat{\beta}_1$ and $\hat{\beta}_0$ for which the sum of the squared errors $S(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_1 x_i - \beta_0)^2$ is minimal.

For a point $(\hat{\beta}_0, \hat{\beta}_1)$ to be a min of $S(\beta_0, \beta_1)$ it must be a critical point, i.e.

$$\begin{cases} \frac{\partial S}{\partial \beta_0}(\hat{\beta}_0, \hat{\beta}_1) = 0 \\ \frac{\partial S}{\partial \beta_1}(\hat{\beta}_0, \hat{\beta}_1) = 0. \end{cases}$$

Using some algebra, we find

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \text{ and } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \text{ where}$$

$$S_{xx} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad S_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

S_{xx} is called the <u>sample variance</u> and S_{xy} the <u>sample covariance</u>.

<u>Example.</u> Fit a line to the following data: $(0,1), (2,1), (3,4)$.

<u>Answer:</u> $(x_1, y_1) = (0,1)$

$(x_2, y_2) = (2,1)$

$(x_3, y_3) = (3,4)$

$\bar{x} = \frac{5}{3}$, $\bar{y} = 2$, $S_{xx} = \frac{1}{3}\left[\left(0-\frac{5}{3}\right)^2 + \left(2-\frac{5}{3}\right)^2 + \left(3-\frac{5}{3}\right)^2\right] =$

$= \frac{14}{9}$, $S_{xy} = \frac{1}{3}\left[\left(0-\frac{5}{3}\right)(1-2) + \left(2-\frac{5}{3}\right)(1-2) + \left(3-\frac{5}{3}\right)(4-2)\right] =$

$= \frac{4}{3}$

$\hat{\beta}_1 = \frac{6}{7}$

$\hat{\beta}_0 = \frac{4}{7}$

The best fit line is $y = \frac{6}{7}x + \frac{4}{7}$.

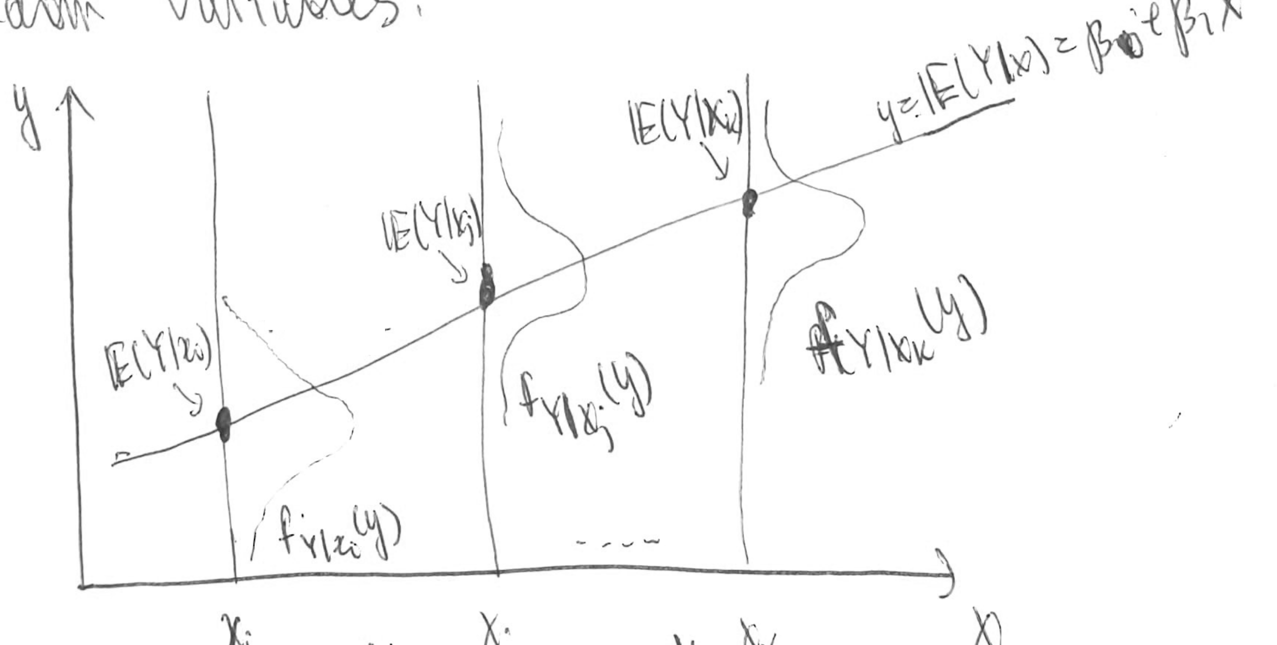
<u>Exercise.</u> Construct a quadratic polynomial $y = ax^2 + bx + c$ that passes through the three given points (x_i, y_i), $i = 1,2,3$.

Now we will formalize the discussion above.

<u>Def-n.</u> Let $f_{Y|x}(y)$ denote the pdf of the random variable Y for a given value x, and let $\mathbb{E}(Y|x)$ denote the corresponding expected value. Then the f-n $y = \mathbb{E}(Y|x)$ is called the <u>regression curve of Y on x</u>.

We would like to consider the special case when $y = \mathbb{E}(Y|x)$ a <u>simple linear model</u>. This requires the following assumptions:

1. $f_{Y|x}(y)$ is a normal pdf for all x.
2. The standard deviation σ associated with $f_{Y|x}(y)$ is the same for all x.
3. The means of the conditional Y distr-ns are collinear, i.e. $y = \mathbb{E}(Y|x) = \beta_0 + \beta_1 x$.
4. All of the conditional distr-ns represent indep. random variables.



Thm. Let $(x_1, y_1), \dots, (x_n, y_n)$ be a set of points that satisfy the assumptions above, and let

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 .$$

Then

$$T_{n-2} = \frac{\hat{\beta}_1 - \beta_1}{S \Big/ \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

has a Student t-distribution with $n-2$ degrees of freedom.

Moreover, to test $H_0: \beta_1 = \beta_1'$ at the α level of signif. versus

(a) $H_1: \beta_1 > \beta_1'$, reject if $t \geq t_{\alpha, n-2}$;

(b) $H_1: \beta_1 < \beta_1'$, reject if $t \leq t_{\alpha, n-2}$;

(c) $H_1: \beta_1 \neq \beta_1'$, reject if $|t| \geq t_{\alpha/2, n-2}$, where

$$t = \frac{\hat{\beta}_1 - \beta_1'}{S \Big/ \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}.$$

Rmk: One of the most common applications of the thm above is to test if the slope is zero ($\beta_1' = 0$). If the null hypothesis is rejected it can be concluded that (at the level of signif. α) the $\mathbb{E}(Y)$ changes with x. On the other hand the ~~rejection~~ acceptance of H_0 implies we cannot rule out that variation in Y can be unaffected by x.

Example. Consider the following bivariate data

x	20	22	23	27	28	30	32
y	35	30	32	31	32	28	27

Test at the 5% level of signif. to see if the slope of regression line is 0. Also give the equation of regression line and the value of S.

We will use the TI-84 calculator.

1. Press 'STAT' $\to$ 'EDIT' and store the x-values in the column L_1 and y-values in L_2.
2. Go to 'STAT' $\to$ 'TESTS' $\to$ 'F: LINREGTTEST' (store the fn RegEq Y_1)

The answer given by the calculator is $y = 43.27 - 0.49x$ with the p-value $p = 0.03$. Since $0.03 < 0.05$ we reject H_0. Alternatively, we can use that $|t| = 2.94 > 2.571 = t_{0.025,5}$. Also, $S = 1.783$.

<u>Thm 2.</u> Same setup as in Thm 1. The $100(1-\alpha)\%$ confidence interval for β_1 is

$$\left[\hat{\beta}_1 - t_{\alpha/2, n-2} \cdot \frac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} ; \hat{\beta}_1 + t_{\alpha/2, n-2} \cdot \frac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}\right]$$

Thm 3. Same setup as before. A $100(1-\alpha)\%$ confidence interval for $\mathbb{E}(Y|x) = \beta_0 + \beta_1 x$ is given by $(\hat{y} - w, \hat{y} + w)$, where $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ and

$$w = t_{\alpha/2,\, n-2} \cdot S \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

The predicted value of y, for a given x, has the $100(1-\alpha)\%$ confidence interval given by

$(\hat{y} - \tilde{w}, \hat{y} + \tilde{w})$, where $\tilde{w} = t_{\alpha/2,\, n-2} \cdot S \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

Example. Same setup as in the previous example.

(a) Find a 95% confidence interval for the mean value of y when $x = 20$.

(b) Find a 95% conf. interval for the predicted value at $x = 21$.

Answers:

(a) $\hat{y}(20) = 33.61$, so we get the interval

$$CI = 33.61 \pm 2.571 \cdot 1.783 \cdot \sqrt{\frac{1}{7} + \frac{(20-26)^2}{118}} = 33.61 \pm 3.068$$

(b) $\hat{y}(21) = 33.13$ ($Y_1(21)$ on calculator).

$$CI = 33.13 \pm 2.571 \cdot 1.783 \sqrt{\frac{8}{7} + \frac{(20-26)^2}{118}} = 33.13 \pm 5.52$$

Rmk: notice that the intervals in Thm 3 are smaller for the values of x_0 closer to $\bar{x}$ and get larger when $|x_0 - \bar{x}|$ grows, i.e. we can predict the location of the regression line for x_0-values closer to $\bar{x}$ better.

Rmk: We can calculate $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ on the calculator. For this go to 'STAT' → 'CALC' → '2-Var Stats'. The resulting value of σ_x is ~~what~~ what we are looking for.

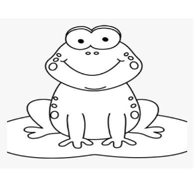
Markov Chains.

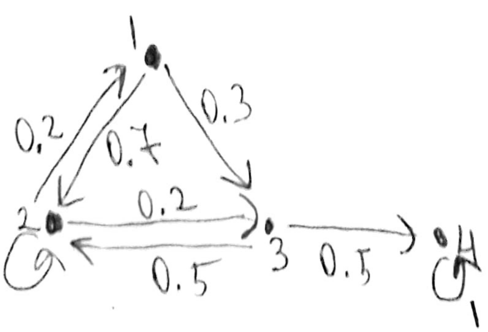
Suppose we have a set of states $S = \{s_1, \dots, s_k\}$ and a process that starts in one of these states and moves successively from one state to the other. Each such move is called a step. Let the probability of moving from state i to j be denoted by p_{ij}.
Then $\sum_{j=1}^{k} p_{ij} = 1$ for any i, let

$$P = \left(p_{ij} \right)$$ be the corresponding matrix.

<u>Def-n.</u> The probabilities p_{ij} are called <u>transition probabilities</u> and the matrix P is called <u>transition matrix</u>. The pair (S, P) is referred to as <u>Markov chain</u>.

<u>Example.</u> A frog hops about on 4 lily pads with tr. probabilities assigned to the arrows on the graph below.





$$P = \begin{pmatrix} 0 & 0.7 & 0.3 & 0 \\ 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Reasonable questions:

(1) If the frog is at state 1, what is the probability that it is at state 3 after 3 steps?

(2) What is the probability that it eventually reaches state 4? State 2?

(3) Starting at state 2, what is the average number of steps before the frog jumps to pad 3 for the first time?

We will give an answer to (1) now, while (2) and (3) will be answered in the due course.

It is not hard to see that the answer to (1) is given by $\underbrace{P^3_{13}}_{\text{the value of}}$. Indeed, notice first that

$$P^2_{13} = \sum_{i=1}^{4} p_{1i} \cdot p_{i3}$$

is the probability of jumping from the first to the third pad in 2 steps. Similarly,

$$P^3_{13} = \sum_{i=1}^{4} \sum_{j=1}^{4} p_{1i} \cdot p_{ij} \cdot p_{j3}$$

gives the probability from (1)

We compute $P^3_{13} = 0.156$

Thm: More generally, P^n_{ij} gives the probability of getting from state i to state j in n steps.

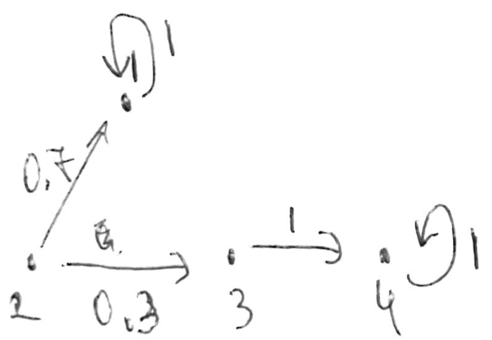
Absorbing Markov Chains (MC)

Def-n. A state s_i of a MC (S,P) is called absorbing if it is impossible to leave it, i.e. $p_{ii}=1$ $p_{ij}\neq 0$ for $j\neq i$. A MC is called absorbing if it has at least one absorbing state and it is possible to reach an absorbing state from any other state (not necessarily in a single step).

Examples (1) In the 'Frog' example on page state 4 is absorbing and the MC is absorbing as well.

(2)

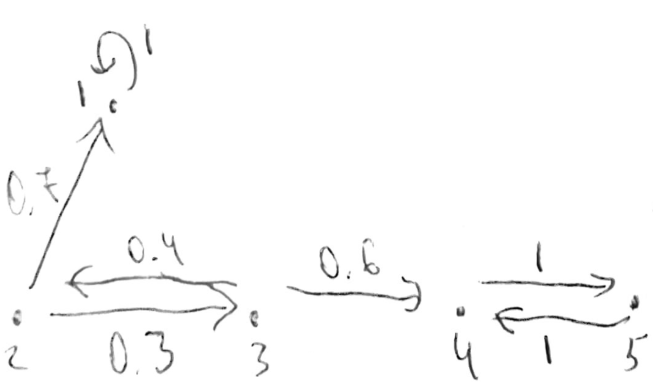
MC =



Absorbing states: S_1, S_4
MC is absorbing

(3)

MC =



Absorbing states: S_1
MC is not absorbing, since we can not reach S_1 from S_4 and S_5.

Interesting questions:

(1) What is the probability that the process will end up in a given absorbing state?

(2) What is the average number of steps prior to absorption?

(3) How many times (on average) will a transient (not absorbing) be visited prior to absorption?

Canonical form of P.

Consider an absorbing MC (S, P) with a absorbing and t transient states.

Def-n: the canonical form of the transient matrix P is the 'block form'

$$P = \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array}\right), \text{ where}$$

Q is a t×t matrix of transition probabilities between transient states,

R is an a×t matrix of transition probabilities between transient and absorbing states and

I is the identity matrix of size a×a.

$= \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$

Rmk: for P in canonical form, we have

$$P^n = \left(\begin{array}{c|c} Q^n & * \\ \hline 0 & I \end{array}\right).$$

Thm 2. In an absorbing MC we have $Q^n \to 0$ as $n \to \infty$. In other words, the process is absorbed with probability 1.

Proof. According to the def-n of an absorbing MC, it is possible to reach an absorbing state starting from any transient state s_j. Let m_j be the minimal number of steps required to reach an absorbing state from s_j. Then the probability that we do not reach an absorbing state starting from s_j in m_j steps is strictly less than 1. ~~We will denote this probability by p_j.~~

Let $m := \max\limits_{s_j \in \text{transient states}} m_j$ be the maximal number of steps required to reach an absorbing state from a transient in the minimal number of steps. Then, regardless which transient state we start in, with probability $p < 1$ an absorbing state is not reached after m steps. Similarly (due to independence of the second group of m steps from the first), the probability of not reaching an absorbing state after $2m$ steps does not exceed p^2, etc.

As the entries of Q^{nm} are bounded above by p^n and $\lim_{n\to\infty} p^n = 0$, the result follows. $\square$

<u>Examples</u> (1) What is the canonical form of the matrix
$$P = \begin{pmatrix} 0 & 0.7 & 0.3 & 0 \\ 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$
from example with the frog and lily pads?

Answer: P is in a canonical form.

(2) What about the matrix $P = \begin{pmatrix} 1 & 0 & 0 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0 & 0.8 \end{pmatrix}$?

Answer: switching rows I and III, we get
$$\begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0 & 0.8 \\ 1 & 0 & 0 \end{pmatrix}.$$

Next switching columns I and III, obtain the answer
$$\left(\begin{array}{cc|c} 0.3 & 0.2 & 0.5 \\ 0 & 0.8 & 0.2 \\ \hline 0 & 0 & 1 \end{array}\right) \quad Q = \begin{pmatrix} 0.3 & 0.2 \\ 0 & 0.8 \end{pmatrix},\ R = \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix}$$

The Fundamental Matrix.

Def-n: for an absorbing MC the matrix $N = (I-Q)^{-1} = I + Q + Q^2 + \ldots$ is called the fundamental matrix.

Thm 3. The ij-entry n_{ij} of N is the expected number of times the chain 'visits' state j prior to absorption, provided it starts in state i (if $i=j$ the initial state is counted).

Rmk: this gives the answer to question (3) (page 14).

Proof: recall that a $k \times k$ matrix M is invertible provided it has no kernel, i.e. no vector v, s.t. $Mv = 0$.
We start by showing that $I - Q$ is invertible.
Suppose there was a v with $(I-Q)v = 0$, equivalently, $v = Qv$. But this would imply $v = Qv = Q^2 v = \ldots$
As we established in Thm 2, $\lim_{n\to\infty} Q^n v = 0$, which gives a contradiction.

Next we check that $N = I + Q + Q^2 + \ldots$ is the inverse of $I - Q$.

Rmk: in case $k=1$, we get the formula for geometric progression $\frac{1}{1-q} = 1 + q + q^2 + \ldots$

$(I-Q)(I+Q+Q^2+\dots)=I+(Q-Q)+(Q^2-Q^2)+\dots=I.$

Let $X_{ij}^{(n)}$ be a Bernoulli random variable with

$$\begin{cases} X_{ij}^{(n)}=1, & \text{if we end up in state } s_j \text{ after } n \text{ steps (starting in state } s_i) \\ X_{ij}^{(n)}=0, & \text{otherwise.} \end{cases}$$

It follows from Thm 1 that the probability $P(X_{ij}^{(n)}=1)$ is equal to Q_{ij}^{n}, the ij entry of Q^n.

This allows to find the expected number of times the chain visits state s_j if it starts in s_i!

$$\mathbb{E}(X_{ij}^{(0)}+X_{ij}^{(1)}+\dots)=\sum_{n=0}^{\infty} Q_{ij}^{(n)}=n_{ij}.$$

Example. What is the expected number of times we visit state 2 starting in state 1 before absorption for the MC with transient matrix

$$P=\left(\begin{array}{cc|c} 0.3 & 0.2 & 0.5 \\ 0 & 0.8 & 0.2 \\ \hline 0 & 0 & 1 \end{array}\right) \; ?$$

Answer: $Q=\begin{pmatrix} 0.3 & 0.2 \\ 0 & 0.8 \end{pmatrix} \rightsquigarrow I-Q=\begin{pmatrix} 0.7 & -0.2 \\ 0 & 0.2 \end{pmatrix}$

$N=(I-Q)^{-1}=\frac{100}{14}\begin{pmatrix} 0.2 & 0.2 \\ 0 & 0.7 \end{pmatrix}$, so $n_{12}=\frac{100}{14}\cdot 0.2=\frac{10}{7}.$

Time to Absorption.

Next we will give an answer to question (2): What is the average number of steps prior to absorption (if we start in state s_i)?

Thm 4. Let t_i denote the answer to the question above. Then we have the equality $\begin{pmatrix} t_1 \\ \vdots \\ t_k \end{pmatrix} = N \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ (*)

Proof. Notice that (*) implies $t_i = \sum_{j=1}^{k} n_{ij}$ (the sum of elements in the i^{th} row of N). Since the element n_{ij} equals to the average number of times state s_j is visited prior to absorption (Thm 3), the result follows.

Thm 5. Let b_{ij} be the probability that an absorbing MC will be absorbed in state s_j if it starts in transient state s_i. Then b_{ij} is the ij entry of the matrix $B = N \cdot R$.

Proof. By def-n $b_{ij} = \sum_{n=0}^{\infty} \sum_{k=1}^{t} q_{ik}^{n} \cdot r_{kj} = \sum_{k=1}^{t} \sum_{n=0}^{\infty} q_{ik}^{n} \cdot r_{kj}$

$= (NR)_{ij}$.

Problem Consider the game of tennis when deuce is reached. If a player wins the next point, he has advantage. On the following point, he either wins the game or the game returns to deuce. Assume that for any point, player A has probability .6 of winning the point and player B has probability .4 of winning the point.

(a) Set this up as a Markov chain with state 1: advantage A; 2: deuce; 3: advantage B; 4: A wins; 5: B wins. Write the transition matrix P.

(b) Find the absorption probabilities.

(c) At deuce, find the expected duration of the game and the probability that B will win.

Solution:

(a) Set this up as a Markov chain with state 1: advantage A; 2: deuce; 3: advantage B; 4: A wins; 5: B wins. Write the transition matrix P.

$$P = \begin{pmatrix} 0 & 0.4 & 0 & 0.6 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0 & 0.4 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(b) $Q = \begin{pmatrix} 0 & 0.4 & 0 \\ 0.6 & 0 & 0.4 \\ 0 & 0.6 & 0 \end{pmatrix}$

$$N = (I - Q)^{-1} = \begin{pmatrix} 1 & -0.4 & 0 \\ -0.6 & 1 & -0.4 \\ 0 & -0.6 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{19}{13} & \frac{10}{13} & \frac{4}{13} \\ \frac{15}{13} & \frac{25}{13} & \frac{10}{13} \\ \frac{9}{13} & \frac{15}{13} & \frac{19}{13} \end{pmatrix}.$$

$$R = \begin{pmatrix} 0.6 & 0 \\ 0 & 0 \\ 0 & 0.4 \end{pmatrix}$$

The absorption probabilities are given by the matrix $NR = \begin{pmatrix} \frac{57}{65} & \frac{8}{65} \\ \frac{45}{65} & \frac{20}{65} \\ \frac{27}{65} & \frac{38}{65} \end{pmatrix}$ (see Thm 5).

Remark. We can find the inverse of a matrix using the calculator TI-84: press

$$2^{nd} \to \text{MATRIX} \to EDIT$$

and enter the matrix as A. Then press

$$2^{nd} \to \text{QUIT}$$

to get to the main screen and plug in A^{-1} ($2^{nd} \to$ MATRIX and choose A). Similarly, one computes the product NR as $A^{-1} \times B$, where $B = R$. In order to convert the decimal entries to fractions, press

$$MATH \to Frac.$$

(c) The expected duration of the game at deuce is equal to the sum of expected number of occurences in each transient state ($1, 2$ and 3) provided we started in state 2 (deuce). By Thm 4 this is equal to the sum of the elements in the second row of fundamental matrix N, i.e.

$$\mathbb{E}(\text{duration}) = \frac{15}{13} + \frac{25}{13} + \frac{10}{13} = \frac{50}{13},$$

probability that B wins $= NR_{22} = \frac{20}{65}$ (see Thm 5).

Ergodic MC.

Def-n. A MC is ergodic (irreducible) if it is possible to get from every state to every state (not necessarily in one move).

Thm 6. For an ergodic MC there is a unique probability vector $w = (w_1, \dots, w_k)$ $(\sum_{i=1}^{k} w_i = 1)$ such that $wP = w$.

Recall one of the main results of Probability Theory, the Law of Large numbers. It states that in case $X_1, X_2, \dots$ is an infinite sequence of i.i.d. random variables with the same expected values $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \dots = \mu$, than the sample average $\frac{1}{n}(X_1 + \dots + X_n) =: \overline{X}_n$ converges to the expected value:

$$X_n \to \mu \text{ as } n \to \infty.$$

Here by 'convergence' we understand that for any $\varepsilon > 0$

$$P(|\overline{X}_n - \mu| \geq \varepsilon) \to 0 \text{ as } n \to \infty.$$

Thm 6 allows* to obtain an analagous statement for ergodic MC, which we now describe.

* with quite a bit of extra work.

Thm 7. Let P be the transition matrix for an ergodic MC. Let $A_n := \frac{I + P + \dots + P^n}{n+1}$. Then $A_n \to W = \begin{pmatrix} w_1 & \dots & w_k \\ w_1 & \dots & w_k \\ w_1 & \dots & w_k \end{pmatrix}$, where

w is the fixed vector for P.

Consider an ergodic MC that starts in state s_i. Let $X^{(m)}$ be a random variable, s.t. $X^{(m)}=1$ if we are at s_j on step m and $X^{(m)}=0$ otherwise. Then the expression

$$H^{(n)}=\frac{X^{(0)}+X^{(1)}+\dots+X^{(n)}}{n+1}$$

gives the average number of times s_j was visited on the first n steps.

Recall that due to Thm1 $P(X^{(m)}=1)=P_{ij}^{m}$, hence, $\mathbb{E}(X^{m})=P_{ij}^{m}$, so $\mathbb{E}(H^{(n)})=ij$-entry of A_n.

<u>Thm 8.</u> Let $H_j^{(n)}$ be the proportion of times in n steps that an ergodic MC is in state s_j. Then for any $\varepsilon>0$,

$$P(|H_j^{(n)}-w_j|\geq\varepsilon)\xrightarrow{n\to\infty}0,$$

independent of the starting state s_i, i.e. the vector w is the analogue of the mean μ in the 'usual' law of large numbers.

Thm 8 is called 'Law of Large Numbers for Ergodic MC'.

We conclude the discussion of MC with the following results.

<u>Def-n:</u> (1) If an ergodic MC starts in s_i, the expected number of steps to reach s_j for the first time is called <u>mean first passage time from s_i to s_j</u>. It is denoted by m_{ij} with the convention $m_{ii}=0$.

(2) The expected number of steps to return to s_i for the first time is called the <u>mean recurrence time for s_i</u> and is denoted by r_i.

Recall that an ergodic MC has no absorbing states. To compute m_{ij} we adhere to the following procedure:

<u>Step 1.</u> Make j an absorbing state, i.e. produce the chain MC' with the same collection of states S and transient matrix P' with $P'_{ab}=P_{ab}$ if $a\neq j$ and
$$\begin{cases} P'_{js}=0, & s\neq j \\ P'_{jj}=1. \end{cases}$$

<u>Step 2.</u> Write P' in the canonical form, compute $N=(I-Q)^{-1}$ and $N\cdot\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}_{ij}$ will be the ~~answer~~ value of m_{ij}.

In order to find r_j, notice that after the first step is made 'the problem reduces to the previous one', more precisely,

$$r_j = 1 + \sum_{i \neq j}^{k} p_{ji} \cdot m_{ij}$$

<u>Example.</u> A die with 4 sides is rolled repeatedly. Find the mean time between occurrences of a given number.

Let us choose '4'.

$$P = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \qquad P' = \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$Q = \begin{pmatrix} 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 \end{pmatrix} \qquad N = (I-Q)^{-1} = \begin{pmatrix} 3/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & 3/4 \end{pmatrix}^{-1} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

$$N \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix} \quad \text{and} \quad r_4 = 1 + \sum_{i=1}^{3} \frac{1}{4} \cdot 4 = 4.$$