Practice Questions (week 8)

Semester 2, 2019

The first two questions are about principal component analysis (PCA). The remainder are about probability – permutations and combinations, discrete random variables, and discrete probability distributions.

1. Consider a principal component analysis (PCA) of a set of data $X \in \mathbb{R}^{n\times p}$ (with $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} X_{ij}$, $X' = X - \mathbf{1}_{n\times 1}\bar{\mathbf{x}}$ and $C = \frac{1}{n-1}(X')^T X'$ as usual) Let $a \in \mathbb{R}$ be non-zero.

   (a) How do the principal components of aX compare to those of X?

   (b) Let $A = a\mathbf{1}_{n\times p}$, how do the principal components of $A + X$ compare to those of X?

Solution:

(a) The mean of aX is $a\bar{\mathbf{x}}$. The difference of aX with its mean is therefore $aX - a\mathbf{1}_{n\times 1}\bar{\mathbf{x}} = a(X - \mathbf{1}_{n\times 1}\bar{\mathbf{x}}) = aX'$. The corresponding covariance matrix is then $a^2 C$. If λ, v is an eigen-pair of C then observe that

$$(a^2 C)v = a^2(Cv) = a^2(\lambda v) = (a^2\lambda)v\,.$$

Thus the principle components of aX are the same as those of X (but with their eigenvalues scaled by a^2).

(b) The mean of $A + X$ is $a\mathbf{1}_{1\times p} + \bar{\mathbf{x}}$. The difference between $A + X$ and its mean is then

$$(A+X)-\mathbf{1}_{n\times 1}(a\mathbf{1}_{1\times p}+\bar{\mathbf{x}}) = (A+X)-(A+\mathbf{1}_{n\times 1}\bar{\mathbf{x}}) = X-\mathbf{1}_{n\times 1}\bar{\mathbf{x}} = X'\,.$$

Since the difference is unchanged the principle components (and their eigenvalues) remain the same.

---

2. Determine the principal components of the following data

$$X = \begin{bmatrix} 3 & 4 \\ 1 & 0 \\ 4 & 3 \\ 0 & 1 \\ 2 & 2 \end{bmatrix}$$

Solution: The mean is $\bar{\mathbf{x}} = (2, 2)$ and therefore we have the differences

$$X' = \begin{bmatrix} 1 & 2 \\ -1 & -2 \\ 2 & 1 \\ -2 & -1 \\ 0 & 0 \end{bmatrix}.$$

The covariance matrix is then

$$C = \frac{1}{5-1} \begin{bmatrix} 1 & -1 & 2 & -2 & 0 \\ 2 & -2 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & -2 \\ 2 & 1 \\ -2 & -1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2.5 & 2 \\ 2 & 2.5 \end{bmatrix}$$

The eigenvalues are then given by the solution of $(2.5 - \lambda)^2 - 4 = 0$. Upon expansion we have $2.25 - 5\lambda + \lambda^2 = 0$ and the formula for the general solution of a quadratic polynomial yields

$$\lambda = \frac{5 \pm \sqrt{25 - 9}}{2} = \frac{5 \pm 4}{2} = \frac{9}{2}, \frac{1}{2}.$$

We now want to find the eigenvectors (i.e. principal components). For $\lambda = 1/2$ we have

$$\frac{1}{2} \begin{bmatrix} a \\ b \end{bmatrix} = C \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} (5/2)a + 2b \\ 2a + (5/2)b \end{bmatrix}$$

and subtracting the LHS from the RHS leads to the equation $0 = 2a + 2b$, or equivalently $a = -b$, and thus the corresponding eigenvector is $a[1, -1]$ for any $a \neq 0$. Similarly for $\lambda = 9/2$ we have

$$\frac{9}{2} \begin{bmatrix} a \\ b \end{bmatrix} = C \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} (5/2)a + 2b \\ 2a + (5/2)b \end{bmatrix}$$

and subtracting the RHS from the LHS leads to the equation $2a - 2b = 0$, or equivalently $a = b$, and thus the corresponding eigenvector is $a[1, 1]$ for any $a \neq 0$. The two principal components are therefore $[1, 1]$ and $[1, -1]$ (up to constant factors).

---

3. The university assigns IDs to students and staff using 7 digit numbers.
   (a) How many IDs are there assuming that the first digit cannot be zero?
   (b) How many IDs are there in which no digit is repeated twice (allowing a zero as the first digit)?

(c) How many IDs are there in which no digit is repeated twice if the first digit cannot be zero?

Solution:

(a) The first digit can be any of $1, 2, \ldots, 9$ while the remaining six can be any of $0, 1, 2, \ldots, 9$, therefore there are $9^1 10^6 = 9000000$ possible IDs.

(b) The first digit can be any of $0, 1, 2, \ldots, 9$, the second can only be one of the remaining 9 numbers, the third can be any of the remaining 8 numbers, and so on. Therefore there are $10!/(10 - 7)! = 10!/3! = 604800$ IDs in which no digit is repeated twice.

(c) The first digit can be any of $1, 2, \ldots, 9$. The remaining digits must all then be one of $0, 1, 2, \ldots, 9$ excluding whichever digit the first is. It follows that regardless of the choice of first digit there are $9!/(9 - 6)! = 9!/3! = 60480$ possiblities for the last six digits. Therefore, the total number of possibilities is $9 \times 60480 = 544320$ (by the multiplication principle).

---

4. Suppose a tutorial of 30 students is broken up into 5 working groups of 6 students to work together on problems.

(a) Taking a single working group of 6 students, how many possible outcomes are there?

(b) Suppose the 30 students consist of 15 distinct pairs of friends who always work together, how many possible outcomes are there now for a single working group of 6 students?

Solution:

(a) A group of 6 students out of 30 is an unordered combination without repetition and therefore has $\binom{30}{6} = \frac{30!}{(30-6)!6!} = 593775$ possible outcomes.

(b) A group of 6 now consists of 3 pairs out of 15 (again an unordered combination without repitition) and therefore has $\binom{15}{3} = \frac{15!}{(15-3)!3!} = 455$ possible outcomes.

---

5. You've just sat down to tackle these practice questions but only have time to do five of them.

(a) How many possible combinations of the practice questions could you answer?

(b) How many possible combinations of the practice questions could you answer assuming you are completing this one?

(c) Suppose you answer another four questions tomorrow, how many combinations are there of the five questions you do today and four you do tomorrow?
(Note: this question does not have to be one of the nine, and we implicitly distinguish the questions done on each day)

Solution:

(a) Since there are 15 questions, and choosing five forms an unordered permutation without repetition there are $\binom{15}{5} = 3003$ possibilities.

(b) If this question must be one of the 5 then the problem is reduced to choosing four from 14, thus there are $\binom{14}{4} = 1001$ possibilities.

(c) We already know there are $\binom{15}{5} = 3003$ possibilities for todays five questions, in each case that leaves $x - 5$ questions from which to choose four tomorrow. Therefore there are $\binom{15-5}{4} = \binom{10}{4} = 210$ possiblities tomorrow regardless of those chosen today. Thus the total number of possibilities over the two days is $3003 \times 210 = 630630$.
(Note: the question implied the day in which questions are done is a distinguishing factor, if this were not the case and we were only interested in the possibilities of the total nine questions then obviously we would get $\binom{15}{9} = 5005$. As an aside, we could then distinguish over the two days by choosing which five of the nine we do on the first day and then obtain $\binom{15}{9}\binom{9}{5} = 5005 \times 126 = 630630$ again for the possible outcomes across the two days.)

---

6. How many different letter arrangements can be made from the letters in the words

(a) MODEL;

(b) DATA;

(c) ADELAIDE;

(d) KARRAWIRRA?

Solution:

(a) There are five distinct letters, therefore there are $5! = 120$ possible letter arrangements of MODEL.

(b) There are four letters, but the A is repeated twice, therefore there are $4!/2! = 12$ possible letter arrangments of DATA.

(c) There are eight letters, but A, D and E are each repeated twice, therefore there are $8!/(2!2!2!) = 40320/8 = 5040$ possible letter arrangements of ADELAIDE.

(d) There are ten letters, but A is repeated thrice (three times) and R is repeated quarce (four times), therefore there are $10!/(3!4!) = 25200$ possible letter arrangements of KARRAWIRRA (Karrawirra Parri is the native Kaurna name of the River Torrens on the north side of campus).

---

7. In Australian rules football (AFL) there are eighteen teams. Eight of these teams will make it into the finals (playoffs). Of these eight, only two will participate in the grand final. Determine the number of possible ways the final eight and final two can be formed. That is, how many ways are there to form a list of 10 teams from 18 where the first 8 are those that make it to the finals and the last 2 are those that make it to the grand final (and thus appear twice in the list).

Solution: The number of combinations for the eight teams in the finals is $\binom{18}{8} = 43758$. Given these eight are in the finals then the possible combinations of teams in the grand final is $\binom{8}{2} = 28$. Thus the total number of combinations is $\binom{18}{8}\binom{8}{2} = 1225224$.
Alternatively, we could choose the two teams in the grand final first ($\binom{18}{2} = 153$ choices) and then choose the six from the remaining sixteen that play in the finals ($\binom{16}{6} = 8008$) and multiply the two ($153 \times 8008 = 1225224$).

---

8. Prove that

$$\binom{2n}{n} = \sum_{j=0}^{n} \binom{n}{j}^2.$$

(Hint: Use the formula $\displaystyle\binom{n+m}{k} = \sum_{j=0}^{k} \binom{n}{j}\binom{m}{k-j}$).

Solution: Write $\binom{2n}{n}$ as $\binom{n+n}{n}$, then we can apply the formula with $m = n$ and $k = n$ to obtain

$$\binom{2n}{n} = \binom{n+n}{n} = \sum_{j=0}^{n} \binom{n}{j} \binom{n}{n-j},$$

and since $\binom{n}{n-j} = \binom{n}{j}$ it follows that

$$\binom{2n}{n} = \binom{n+n}{n} = \sum_{j=0}^{n} \binom{n}{j} \binom{n}{n-j} = \sum_{j=0}^{n} \binom{n}{j}^2 .$$

---

9. Prove each of the following (hint: use the binomial theorem)

   (a)
   $$\sum_{k=0}^{n} \binom{n}{k} = 2^n ,$$

   (b)
   $$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0 .$$

   (c)
   $$\sum_{k=0}^{n} (-1)^k 2^k \binom{n}{k} = 3^n .$$

   Solution:

   (a) Recall the binomial theorem
   $$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} .$$
   Setting $x = 1$ and $y = 1$ we have $(x+y)^n = (1+1)^n = 2^n$ and thus
   $$2^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = \sum_{k=0}^{n} \binom{n}{k} 1^k 1^{n-k} = \sum_{k=0}^{n} \binom{n}{k},$$
   as required.

   (b) This time set $x = -1$ and $y = 1$, we now have $(x+y)^n = (-1+1)^n = 0^n = 0$ and thus
   $$0 = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (-1)^k 1^{n-k} = \sum_{k=0}^{n} (-1)^k \binom{n}{k},$$
   as required.

(c) This time set $x = 2$ and $y = 1$, we now have $(x+y)^n = (2+1)^n = 3^n$ and thus

$$3^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = \sum_{k=0}^{n} \binom{n}{k} 2^k 1^{n-k} = \sum_{k=0}^{n} 2^k \binom{n}{k},$$

as required.

---

10. We will consider how many ways a committe of any size can be formed from n people with one of the committee members selected to be chair person.

(a) Describe why there are $k\dbinom{n}{k}$ ways to pick a committee of size k from n people with one of the k members selected as chair.

(b) Use the definition of $\dbinom{n}{k}$ to show that $k\dbinom{n}{k} = n\dbinom{n-1}{k-1}$.

(c) Use the above two steps to show that the number of ways to form a committe of any size from n people with one of the committee members selected as chair is $n \times 2^{n-1}$.
(Hint: you'll also need the result $\displaystyle\sum_{k=0}^{n} \binom{n}{k} = 2^n$ from the previous question.)

Solution:

(a) There are $\binom{n}{k}$ ways to choose a committee of k from n people (since a committee is an unordered set without repetition). For a given committee of k committee people there are then k possible choices for the chair person.

(b) Observe that

$$\begin{aligned} k\binom{n}{k} = k\frac{n!}{(n-k)!\,k!} &= k\frac{n \times (n-1)!}{((n-1)-(k-1))!\,k \times (k-1)!} \\ &= n\frac{(n-1)!}{((n-1)-(k-1))!\,(k-1)!} \\ &= n\binom{n-1}{k-1}. \end{aligned}$$

As an aside, note that $n\binom{n-1}{k-1}$ can be interpreted as the number of ways of first choosing a chair person and then choosing the remaining $k-1$ committee members.

(c) To obtain the total number of ways to form a committee size we can consider the number of committees of each size $k = 1, \ldots, n$ and sum the results, that is

$$\sum_{k=1}^{n} k \binom{n}{k} .$$

Now using the result of the second step we have

$$\sum_{k=1}^{n} k \binom{n}{k} = \sum_{k=1}^{n} n \binom{n-1}{k-1} = n \sum_{k=0}^{n-1} \binom{n-1}{k} .$$

Lastly, using the result $\sum_{k=0}^{n} \binom{n}{k} = 2^n$ (but with $n \leftarrow n-1$) we have

$$n \sum_{k=0}^{n-1} \binom{n-1}{k} = n \times 2^{n-1} .$$

---

11. 10 items are to be distributed amongst 3 bins. How many combinations are there if

(a) both the items and bins are distinguishable;

(b) the items are identical and only the bins are distinguishable;

Solution:

(a) We can think of this as selecting one of the three bins for each of the ten items. That is, if we label the bins A, B and C, then how many lists with length 10 can we form with these letters (allowing repetition). This is an ordered permutation with repetition and so there are $3^{10} = 59049$ possible outcomes.

(b) In this case we only care how many of the (identical) items are in each bin. That is, the ordering of the list of bins for each of the 10 items is no longer important. This is then an unordered combination of the bins (with repetition) and thus there are $\binom{10+(3-1)}{(3-1)} = \binom{12}{2} = 66$ possible outcomes.

---

12. We will use induction (on n) to prove the identity

$$\binom{n}{k} = \sum_{i=k}^{n} \binom{i-1}{k-1} .$$

(a) First show the base case $n = k$ is true.

(b) Now assuming the identity holds for n, show that it also holds for $n+1$.

Solution:

(a) With $n = k$ we have $\binom{n}{k} = \binom{k}{k} = 1$ and also

$$\sum_{i=k}^{n} \binom{i-1}{k-1} = \binom{k-1}{k-1} = 1\,,$$

so the base case holds.

(b) For the case $n+1$ we have

$$\binom{n+1}{k} = \sum_{i=k}^{n+1} \binom{i-1}{k-1} = \binom{n}{k-1} + \sum_{i=k}^{n} \binom{i-1}{k-1}\,,$$

and since we may assume the identity holds for the case n then we therefore have

$$\binom{n}{k-1} + \sum_{i=k}^{n} \binom{i-1}{k-1} = \binom{n}{k-1} + \binom{n}{k}\,.$$

The RHS is then equal to $\binom{n+1}{k}$ via the identity from the lecture slides ($\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$).

---

13. Expand $(2x + y^3)^5$.

Solution: Using the binomial theorem we have

$$\begin{aligned}
(2x + y^3)^5 &= \sum_{k=0}^{5} \binom{5}{k} (2x)^k (y^3)^{5-k} \\
&= \binom{5}{0}(2x)^0(y^3)^5 + \binom{5}{1}(2x)^1(y^3)^4 + \binom{5}{2}(2x)^2(y^3)^3 \\
&\quad + \binom{5}{3}(2x)^3(y^3)^2 + \binom{5}{4}(2x)^4(y^3)^1 + \binom{5}{5}(2x)^5(y^3)^0 \\
&= y^{15} + 5 \cdot 2xy^{12} + 10 \cdot 4x^2y^9 + 10 \cdot 8x^3y^6 + 5 \cdot 16x^4y^3 + 32x^5 \\
&= y^{15} + 10xy^{12} + 40x^2y^9 + 80x^3y^6 + 80x^4y^3 + 32x^5\,.
\end{aligned}$$

(You could expand it one factor at a time, but it would be less efficient...)

14. A tiler has n white tiles and m black tiles and wants to lay them in a line to form a path. There are more black tiles than white tiles (i.e. $m > n$). The tiler wants to ensure no two white tiles are next to each other. Come up with a combinatorical formula for how many different patterns the tiler could potentially make?

    Solution: Consider laying down the black tiles first with gaps in between and at the ends, e.g.

    $$\wedge \blacksquare \wedge \blacksquare \wedge \blacksquare \wedge \cdots \wedge \blacksquare \wedge$$

    where $\wedge$ shows the possible locations of white tiles. The problem is then to choose n of the $m+1$ possible locations to place the white tiles (and then move everything together to eliminate the gaps). Thus, since this is an unordered combination (without repetition), there are $\binom{m+1}{n}$ possible patterns the tiler could create.

---

15. Let n be a positive integer. How many solutions are there to

    $$x_1 + x_2 + \cdots + x_r = n$$

    with the condition that

    (a) each x_i is a (strictly) positive integer?
    (Hint: consider laying out the n objects and choosing how to divide them up into r sets having at least one object in each.)

    (b) each x_i is a non-negative integer?

    (c) x_1 is (strictly) positive, but the remaining x_i are non-negative integers?

    Solution:

    (a) Consider n object laid out with spaces in between (excluding the ends). Choosing the x_i is equivalent to choosing $r-1$ of the $n-1$ spaces between the objects in which to place dividers (with no two dividers in the same gap) which separates the n objects into r groups. Thus there are $\binom{n-1}{r-1} = \binom{n-1}{n-r}$ possible solutions (provided $n \geq r > 0$, otherwise there are no solutions).

    (b) In this case the x_i are allowed to be zero. In this case we can again lay out the n objects with spaces in between and at the ends and place our $r-1$ dividers within the $n+1$ spaces but also allowing multiple dividers to occupy the same space. That is, this

is a combination of $n+1$ spaces taken $r-1$ at a time (unordered) with repetition, thus there are $\binom{(r-1)+(n+1)-1}{r-1} = \binom{r+n-1}{r-1}$ possible outcomes (or equivalently $\binom{n+r-1}{n}$).
(Alternatively, we can introduce $y_i = x_i + 1$ so that we have $y_1 + \cdots + y_r = n + r$ with each y_i (strictly) positive such that the problem is now the same as in the first part and there must then be $\binom{n+r-1}{r-1}$ possible outcomes.)

(c) In this case we can lay out the n object with gaps in between and at the end (but not at the start). That is we have n gaps for the $r-1$ dividers which can again occupy the same spaces. It follows that there are $\binom{n+r-2}{r-1} = \binom{n+r-2}{n-1}$ possible outcomes.
