Mathematical Statistics

MAS 713

Tutorial about Chapter 3

Exercise 1

Question:

Let X be a r.v. with cumulative distribution function $F(x)$ and density $f(x) = F'(x)$. Find the probability density function of

a) the maximum of n independent random variables all with cumulative distribution function $F(x)$.
b) the minimum of n independent random variables all with cumulative distribution function $F(x)$.

Solution:

a) We have $X_1, X_2, \ldots, X_n \sim F$, independent of one another.

- Denote $X_{(n)}$ the maximum value. Then, the cdf of $X_{(n)}$ is

$$\begin{aligned} F_{X_{(n)}}(x) &= \mathbb{P}(X_{(n)} \leq x) \\ &= \mathbb{P}\left(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x\right) \\ &\overset{\text{ind.}}{=} \mathbb{P}(X_1 \leq x) \times \mathbb{P}(X_2 \leq x) \times \ldots \times \mathbb{P}(X_n \leq x) \\ &= (F(x))^n \end{aligned}$$

Hence,

$$f_{X_{(n)}}(x) = \frac{d}{dx} F_{X_{(n)}}(x) = n\left(F(x)\right)^{n-1} f(x)$$

b) We have $X_1, X_2, \ldots, X_n \sim F$, independent of one another.

- Denote $X_{(1)}$ the minimum value. Then, the cdf of $X_{(1)}$ is

$$
\begin{aligned}
F_{X_{(1)}}(x) &= \mathbb{P}(X_{(1)} \leq x) \\
&= 1 - \mathbb{P}(X_{(1)} > x) \\
&= 1 - \mathbb{P}\left(X_1 > x, X_2 > x, \ldots, X_n > x\right) \\
&\overset{\text{ind.}}{=} 1 - \mathbb{P}(X_1 > x) \times \mathbb{P}(X_2 > x) \times \ldots \times \mathbb{P}(X_n > x) \\
&= 1 - (1 - F(x))^n
\end{aligned}
$$

Hence,

$$
f_{X_{(1)}}(x) = \frac{d}{dx} F_{X_{(1)}}(x) = -n\,(1 - F(x))^{n-1}\,(-f(x)) = n\,(1 - F(x))^{n-1}\,f(x)
$$

Exercise 2

Question:

An article in the review *Knee Surgery, Sports Traumatology and Arthroscopy* in 2005 cites the following results:

1. a success rate of more than 90% for meniscal tears with a rim width of less than 3mm,
2. but only a 67% success rate for tears of 3–6mm.

If you are unlucky enough to suffer from:
- a meniscal tear of less than 3mm on your **left** knee and
- one of width 3–6mm on your **right** knee,

- what is the probability mass function of the number of successful surgeries? (Assume the surgeries are independent)
- Find the mean and variance of the number of successful surgeries that you would undergo.

Solution:

- Define the random variables

$$X_1 = \begin{cases} 1 & \text{if the surgery on your } \textbf{left} \text{ knee is successful} \\ 0 & \text{if not} \end{cases}$$

and

$$X_2 = \begin{cases} 1 & \text{if the surgery on your } \textbf{right} \text{ knee is successful} \\ 0 & \text{if not} \end{cases}$$

- The total number of successful surgeries that you will undergo is

$$X = X_1 + X_2.$$

- As both X_1 and X_2 can only take the values 0 and 1,
X can only take the values 0 or 1 or 2, so $S_X = \{0, 1, 2\}$.

- $X = 0 \iff X_1 = 0$ and $X_2 = 0$.
- **Because X_1 and X_2 are independent,**

$$\mathbb{P}(X = 0) = \mathbb{P}(X_1 = 0, X_2 = 0) = \mathbb{P}(X_1 = 0) \times \mathbb{P}(X_2 = 0)$$
$$= (1 - 0.9)(1 - 0.67) = 0.033$$

- $X = 2 \iff X_1 = 1$ and $X_2 = 1$.
- **Because X_1 and X_2 are independent,**

$$\mathbb{P}(X = 2) = \mathbb{P}(X_1 = 1, X_2 = 1) = \mathbb{P}(X_1 = 1) \times \mathbb{P}(X_2 = 1)$$
$$= 0.9 \times 0.67 = 0.603$$

- $X = 1$ in any other situations (as $S_X = \{0, 1, 2\}$), so

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 2) = 1 - 0.033 - 0.603 = 0.364$$

$\big($alternatively, use $(X = 1) = (X_1 = 1 \cap X_2 = 0) \cup (X_1 = 0 \cap X_2 = 1)\big)$

- So the **pmf** of X is

$$p(0) = 0.033, \qquad p(1) = 0.364, \qquad p(2) = 0.603$$

(understood that $p(x) = 0$ for any other value of x).

- $\mathbb{E}(X)$ and $\mathbb{V}\mathrm{ar}(X)$ can now be derived directly from the **pmf** of X

$$\mathbb{E}(X) = p(0) \times 0 + p(1) \times 1 + p(2) \times 2 = 1.57$$

$$\mathbb{E}(X^2) = p(0) \times 0^2 + p(1) \times 1^2 + p(2) \times 2^2 = 2.776$$

$$Var(X) = E(X^2) - E(X)^2 = 2.776 - 1.57^2 = 0.3111$$

Exercise 3

Question:

The article "*Error Distribution in Navigation*" (J. Institute of Navigation, 1971) suggests that the distribution of the lateral position error, say X (in nautical miles), which can be either positive or negative, is well approximated by a density like

$$f(x) = c\, e^{-0.2\,|x|} \qquad \text{for } -\infty < x < \infty,$$

for a constant c.

a) Find the value of c which makes f a legitimate density function, and sketch the corresponding density curve.
b) In the long-run, what proportion of errors is negative? At most 2? Between -1 and 2?

Solution:

a) The two conditions for f to be a legitimate density function are
- $f(x) \geq 0$ for all x and
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

- As the exponential is always positive,

$$f(x) \geq 0 \quad \forall x \iff c \geq 0.$$

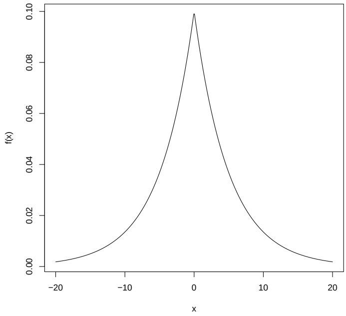
- The second condition gives

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} c\,e^{-0.2\,|x|}\,dx &= \int_{-\infty}^{0} c\,e^{0.2\,x}\,dx + \int_{0}^{\infty} c\,e^{-0.2\,x}\,dx \\
&= c\left(\left[\frac{e^{0.2\,x}}{0.2}\right]_{-\infty}^{0} + \left[\frac{e^{-0.2\,x}}{(-0.2)}\right]_{0}^{\infty}\right) \\
&= c\left(\frac{1}{0.2} + \frac{1}{0.2}\right) = \frac{c}{0.1}
\end{aligned}$$

- As this integral must equal 1, we find $c = 0.1$.

The resulting density is thus

$$f(x) = 0.1\, e^{-0.2\,|x|} \qquad \text{for } -\infty < x < \infty,$$



b)

Question: In the long-run, what proportion of errors is negative?

Solution: We computed c so as the total area under the curve, that is:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

So, by **symmetry** (see picture!), the area under the curve for $x \in (-\infty, 0]$, equal to the probability of X in $(-\infty, 0]$, is equal to 1/2
$\rightsquigarrow$ we expect half of the observed errors to be negative.

b)

Question: In the long-run, what proportion of errors is negative?

Solution: We computed c so as the total area under the curve, that is:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

So, by **symmetry** (see picture!), the area under the curve for $x \in (-\infty, 0]$, equal to the probability of X in $(-\infty, 0]$, is equal to 1/2 $\rightsquigarrow$ we expect half of the observed errors to be negative.

Question: In the long-run, what proportion of errors is at most 2??

Solution: We have that

$$\mathbb{P}(X \leq 2) = \int_{-\infty}^{2} f(x)\,dx$$
$$= 1/2 + 0.1 \int_{0}^{2} e^{-0.2\,x}\,dx$$
$$= 1/2 + 0.1 \left[\frac{e^{-0.2\,x}}{(-0.2)}\right]_0^2 = 0.6648$$

$\rightsquigarrow$ we expect 66.48% of the observed errors to be smaller than 2.

- Note that it could have been easier to write

$$\mathbb{P}(X \leq 2) = 1 - \mathbb{P}(X > 2) = 1 - \int_{2}^{\infty} f(x)\,dx,$$

and we would have found the same answer.

Question: In the long-run, what proportion of errors is at most 2??

Solution: We have that

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \int_{-\infty}^{2} f(x)\, dx \\ &= 1/2 + 0.1 \int_{0}^{2} e^{-0.2\,x}\, dx \\ &= 1/2 + 0.1 \left[\frac{e^{-0.2\,x}}{(-0.2)}\right]_0^2 = 0.6648\end{aligned}$$

$\rightsquigarrow$ we expect 66.48% of the observed errors to be smaller than 2.

- Note that it could have been easier to write

$$\mathbb{P}(X \leq 2) = 1 - \mathbb{P}(X > 2) = 1 - \int_{2}^{\infty} f(x)\, dx,$$

and we would have found the same answer.

Question: In the long-run, what proportion of errors falls between -1 and 2?

Solution:

$$\mathbb{P}(-1 \le X \le 2) = \int_{-1}^{2} f(x)\,dx$$
$$= 0.1\left(\int_{-1}^{0} e^{0.2\,x}\,dx + \int_{0}^{2} e^{-0.2\,x}\,dx\right)$$
$$= \ldots = 0.2555$$

$\rightsquigarrow$ we expect 25.55% of the observed errors to be between -1 and 2.

Question: In the long-run, what proportion of errors falls between -1 and 2?

Solution:

$$\begin{aligned}\mathbb{P}(-1 \le X \le 2) &= \int_{-1}^{2} f(x)\,dx \\ &= 0.1\left(\int_{-1}^{0} e^{0.2\,x}\,dx + \int_{0}^{2} e^{-0.2\,x}\,dx\right) \\ &= \ldots = 0.2555\end{aligned}$$

$\rightsquigarrow$ we expect 25.55% of the observed errors to be between -1 and 2.

Exercise 4

Question:

The probability density function of the weight X (in kg) of packages delivered by a post office is

$$f(x) = \frac{70}{69x^2} \qquad \text{for } 1 < x < 70$$

and 0 elsewhere.

a) Determine the mean and the variance of the weight X.
b) If the shipping cost is \$2.50 per kg, what is the average shipping cost of a package? What is the variance of the shipping cost?
c) In the long-term, what is the proportion of packages whose weight exceeds 50 kg?

Solution:

a) Let X be the random variable "weight of a package". By definition,

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x)\, dx \\ &= \frac{70}{69}\int_1^{70} x\,\frac{1}{x^2}\, dx = \frac{70}{69}\int_1^{70} \frac{1}{x}\, dx = \frac{70}{69}\,[\log(x)]_1^{70} \\ &= \frac{70}{69}\log(70) = 4.31 \text{ (kg)}\end{aligned}$$

$\rightsquigarrow$ on average, the weight of a package delivered by that post office is 4.31 kg.

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\, dx = \frac{70}{69}\int_1^{70} x^2\,\frac{1}{x^2}\, dx \\ &= \frac{70}{69}\int_1^{70} dx = \frac{70}{69}(70-1) = 70 \text{ (kg}^2\text{)},\end{aligned}$$

so

$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 70 - 4.31^2 = 51.42 \text{ (kg}^2\text{)}$$

b)

- Let X be the random variable "weight of a package",
- Denote C the shipping cost for a package. We have that

$$C = 2.5 \times X.$$

- From the properties of the expectation,

$$\mathbb{E}(C) = 2.5 \times \mathbb{E}(X) = 2.5 \times 4.31 = 10.78 \text{ (\$)}$$

- For the variance, we have

$$\mathbb{V}\text{ar}(C) = 2.5^2 \times \mathbb{V}\text{ar}(X) = 6.25 \times 51.42 = 321.4 \text{ (\$}^2\text{)}$$

c)

- Let X be the random variable "weight of a package".
- The long-term proportion of packages with weight exceeding 50 kg is

$$\begin{aligned}\mathbb{P}(X > 50) &= \int_{50}^{\infty} f(x)\,dx \\ &= \frac{70}{69}\int_{50}^{70} \frac{1}{x^2}\,dx \\ &= \frac{70}{69}\left[-\frac{1}{x}\right]_{50}^{70} \\ &= \frac{70}{69}\left(\frac{1}{50} - \frac{1}{70}\right) = 0.0058\end{aligned}$$

Exercise 5

Question:

Compute $\mathbb{E}(x)$ and $\mathbb{V}\text{ar}(x)$ for each of the following probability distributions:

1. $f(x) = ax^{a-1},\ 0 < x < 1, a > 0$
2. $f(x) = \frac{1}{n},\ x = 1, 2, \ldots, n, n > 0$ an integer

Solution:

1)

$$f(x) = ax^{a-1},\ 0 < x < 1, a > 0$$

$$\mathbb{E}(X) = \int_0^1 xax^{a-1}dx = \frac{a}{a+1}$$

$$\mathbb{E}(X^2) = \int_0^1 x^2ax^{a-1}dx = \frac{a}{a+2}$$

$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{a}{a+2} - (\frac{a}{a+1})^2$$

2)

$$f(x) = \frac{1}{n}, \; x = 1, 2, \ldots, n, n > 0 \text{ an integer}$$

$$\mathbb{E}(X) = \sum_{x=1}^{n} x \frac{1}{n} = \frac{n+1}{2}$$

$$\mathbb{E}(X^2) = \sum_{x=1}^{n} x^2 \frac{1}{n} = \frac{(n+1)(2n+1)}{6}$$

$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{(n+1)(2n+1)}{6} - (\frac{n+1}{2})^2 = \frac{n^2+1}{12}$$

Exercise 6

In each of the following situations state whether it is reasonable to use **Binomial** or **Poisson** distributions for X.
If so, tell which one, and (if it is possible) determine what are the values of the parameters:

a) Toss a fair coin 6 times, X is the number of 'Heads'.
b) Toss a fair coin until the first time a head appears, X is the count of the number of tosses you make.
c) A factory makes carpets. Sometimes there are flaws in the carpet. On average a square metre of carpet has 3 flaws. X is the number of flaws in a random square metre of carpet.
d) Most calls made at random by sample surveys don't succeed in talking to a person. Of calls in New York City, only 1/12 succeed. A survey calls 500 randomly selected numbers in New York City, and X is the number that reach a live person.
e) Calls to a telephone exchange come in at an average of 250 an hour, X is the number of calls in a given hour.
f) A die (6 faces, numbered 1,2,3,4,5,6) is tossed twice and X is the number of 6s obtained.

Question: Toss a fair coin 6 times, X is the number of ‘Heads’.

Solution: $X \sim \mathrm{Bin}(6, 0.5)$.

Question: Toss a fair coin 6 times, X is the number of ‘Heads’.

Solution: $X \sim \text{Bin}(6, 0.5)$.

Question: Toss a fair coin until the first time a head appears, X is the count of the number of tosses you make.

Solution: neither Binomial nor Poisson. This distribution is called the **Geometric** distribution.

Question: Toss a fair coin until the first time a head appears, X is the count of the number of tosses you make.

Solution: neither Binomial nor Poisson. This distribution is called the **Geometric** distribution.

Geometric distribution

- Discrete distribution with $S = \{1, 2, 3, \dots\}$ with parameter $\pi \in (0.1)$
- Notation: $X \sim \text{Geom}(\pi)$
- First occurrence of a success in a series of indep. Bernoulli trials with parameter π.
- Probability mass function satisfies

$$\mathbb{P}(X = n) = p(n) = (1 - \pi)^{n-1}\pi, \qquad n \in \{1, 2, \dots\}$$

- $\mathbb{E}(X) = \frac{1}{\pi} \qquad \mathbb{V}\text{ar}(X) = \frac{1-\pi}{\pi^2}$
- Characteristic Function $\qquad \varphi_{Geom(\pi)}(t) = \frac{\pi e^{\mathrm{it}}}{1-(1-\pi)e^{\mathrm{it}}}$

Question: A factory makes carpets. Sometimes there are flaws in the carpet. On average a square metre of carpet has 3 flaws. X is the number of flaws in a random square metre of carpet.

Solution: $X \sim \mathcal{P}(3)$

Question: A factory makes carpets. Sometimes there are flaws in the carpet. On average a square metre of carpet has 3 flaws. X is the number of flaws in a random square metre of carpet.

Solution: $X \sim \mathcal{P}(3)$

Question: Most calls made at random by sample surveys don't succeed in talking to a person. Of calls in New York City, only 1/12 succeed. A survey calls 500 randomly selected numbers in New York City, and X is the number that reach a live person.

Solution: $X \sim \text{Bin}(500, 1/12)$, note that here because n is 'large' and π is 'small', the distribution of X can also be **approximated** by $\mathcal{P}(41.67)$ $(500 \times 1/12 = 41.67)$

Question: Most calls made at random by sample surveys don't succeed in talking to a person. Of calls in New York City, only 1/12 succeed. A survey calls 500 randomly selected numbers in New York City, and X is the number that reach a live person.

Solution: $X \sim \text{Bin}(500, 1/12)$, note that here because n is 'large' and π is 'small', the distribution of X can also be **approximated** by $\mathcal{P}(41.67)$ $(500 \times 1/12 = 41.67)$

Question: Calls to a telephone exchange come in at an average of 250 an hour, X is the number of calls in a given hour.

Solution: $X \sim \mathcal{P}(250)$

Question: Calls to a telephone exchange come in at an average of 250 an hour, X is the number of calls in a given hour.

Solution: $X \sim \mathcal{P}(250)$

Question: A die (6 faces, numbered 1,2,3,4,5,6) is tossed twice and X is the number of 6s obtained.

Solution: $X \sim \mathrm{Bin}(2, 1/6)$

Question: A die (6 faces, numbered 1,2,3,4,5,6) is tossed twice and X is the number of 6s obtained.

Solution: $X \sim \text{Bin}(2, 1/6)$

Exercise 7

Question:

An individual claims to have extrasensory perception (ESP). As a test, a fair coin is tossed ten times, and he is asked to predict in advance the outcome. Our individual gets seven out of ten correct.

- What is the probability he would have done at least this well if he had no ESP?
- Would you believe in his powers?

Solution:

- Denote X the number of correct predictions that the individual gets.
- It is clear that

$$X \sim \text{Bin}(10, \pi),$$

where π is the probability that he predicts a right outcome for a toss.
- Without ESP, π would be equal to 1/2 (random guess between two alternatives), while if he has ESP, then $\pi > 1/2$.
- If $\pi = 1/2$ (no ESP), then probability that he gets seven out of ten correct

$$\mathbb{P}(X \geq 7) = \sum_{x=7}^{10} \mathbb{P}(X = x) = \sum_{x=7}^{10} \binom{10}{x} (1/2)^x \simeq 0.17$$

$\rightsquigarrow$ not that unlikely to have guessed 7 out of 10 outcomes without ESP
$\rightsquigarrow$ there is no clear evidence that he has ESP
$\rightsquigarrow$ we do not really believe in his powers
Note that this does not mean that we are sure that the individual has no ESP!
It is important to understand this way of thinking as it will be the basis of the **hypothesis tests** introduced in a subsequent chapter.

Solution:

- Denote X the number of correct predictions that the individual gets.
- It is clear that

$$X \sim \text{Bin}(10, \pi),$$

where π is the probability that he predicts a right outcome for a toss.

- Without ESP, π would be equal to 1/2 (random guess between two alternatives), while if he has ESP, then $\pi > 1/2$.
- If $\pi = 1/2$ (no ESP), then probability that he gets seven out of ten correct

$$\mathbb{P}(X \geq 7) = \sum_{x=7}^{10} \mathbb{P}(X = x) = \sum_{x=7}^{10} \binom{10}{x} (1/2)^x \simeq 0.17$$

$\rightsquigarrow$ not that unlikely to have guessed 7 out of 10 outcomes without ESP
$\rightsquigarrow$ there is no clear evidence that he has ESP
$\rightsquigarrow$ we do not really believe in his powers

Note that this does not mean that we are sure that the individual has no ESP!

It is important to understand this way of thinking as it will be the basis of the **hypothesis tests** introduced in a subsequent chapter.

Solution:

- Denote X the number of correct predictions that the individual gets.
- It is clear that

$$X \sim \text{Bin}(10, \pi),$$

where π is the probability that he predicts a right outcome for a toss.
- Without ESP, π would be equal to 1/2 (random guess between two alternatives), while if he has ESP, then $\pi > 1/2$.
- If $\pi = 1/2$ (no ESP), then probability that he gets seven out of ten correct

$$\mathbb{P}(X \geq 7) = \sum_{x=7}^{10} \mathbb{P}(X = x) = \sum_{x=7}^{10} \binom{10}{x} (1/2)^x \simeq 0.17$$

$\rightsquigarrow$ not that unlikely to have guessed 7 out of 10 outcomes without ESP
$\rightsquigarrow$ there is no clear evidence that he has ESP
$\rightsquigarrow$ we do not really believe in his powers
Note that this does not mean that we are sure that the individual has no ESP!
It is important to understand this way of thinking as it will be the basis of the **hypothesis tests** introduced in a subsequent chapter.

Solution:

- Denote X the number of correct predictions that the individual gets.
- It is clear that

$$X \sim \text{Bin}(10, \pi),$$

where π is the probability that he predicts a right outcome for a toss.

- Without ESP, π would be equal to 1/2 (random guess between two alternatives), while if he has ESP, then $\pi > 1/2$.
- If $\pi = 1/2$ (no ESP), then probability that he gets seven out of ten correct

$$\mathbb{P}(X \geq 7) = \sum_{x=7}^{10} \mathbb{P}(X = x) = \sum_{x=7}^{10} \binom{10}{x} (1/2)^x \simeq 0.17$$

$\rightsquigarrow$ not that unlikely to have guessed 7 out of 10 outcomes without ESP
$\rightsquigarrow$ there is no clear evidence that he has ESP
$\rightsquigarrow$ we do not really believe in his powers

Note that this does not mean that we are sure that the individual has no ESP!

It is important to understand this way of thinking as it will be the basis of the **hypothesis tests** introduced in a subsequent chapter.

Solution:

- Denote X the number of correct predictions that the individual gets.
- It is clear that

$$X \sim \text{Bin}(10, \pi),$$

where π is the probability that he predicts a right outcome for a toss.
- Without ESP, π would be equal to 1/2 (random guess between two alternatives), while if he has ESP, then $\pi > 1/2$.
- If $\pi = 1/2$ (no ESP), then probability that he gets seven out of ten correct

$$\mathbb{P}(X \geq 7) = \sum_{x=7}^{10} \mathbb{P}(X = x) = \sum_{x=7}^{10} \binom{10}{x} (1/2)^x \simeq 0.17$$

$\rightsquigarrow$ not that unlikely to have guessed 7 out of 10 outcomes without ESP
$\rightsquigarrow$ there is no clear evidence that he has ESP
$\rightsquigarrow$ we do not really believe in his powers

Note that this does not mean that we are sure that the individual has no ESP!
It is important to understand this way of thinking as it will be the basis of the **hypothesis tests** introduced in a subsequent chapter.

Solution:

- Denote X the number of correct predictions that the individual gets.
- It is clear that

$$X \sim \text{Bin}(10, \pi),$$

where π is the probability that he predicts a right outcome for a toss.

- Without ESP, π would be equal to 1/2 (random guess between two alternatives), while if he has ESP, then $\pi > 1/2$.
- If $\pi = 1/2$ (no ESP), then probability that he gets seven out of ten correct

$$\mathbb{P}(X \geq 7) = \sum_{x=7}^{10} \mathbb{P}(X = x) = \sum_{x=7}^{10} \binom{10}{x} (1/2)^x \simeq 0.17$$

$\rightsquigarrow$ not that unlikely to have guessed 7 out of 10 outcomes without ESP
$\rightsquigarrow$ there is no clear evidence that he has ESP
$\rightsquigarrow$ we do not really believe in his powers
Note that this does not mean that we are sure that the individual has no ESP!
It is important to understand this way of thinking as it will be the basis of the **hypothesis tests** introduced in a subsequent chapter.

Exercise 8

Question:

The analysis of results from a leaf transmutation experiment (turning a leaf into a petal) is summarised by type of transformation completed :

		Textural transformation	
		Yes	No
Colour transformation	Yes	243	26
	No	13	18

A naturalist randomly selects three different leaves from this set. Determine the following probabilities :

a) Exactly one has undergone both types of transformations.
b) At least one has undergone both types of transformations.
c) Exactly one has undergone one but not both transformations.
d) At least one has undergone at least one transformation.

Solution:

a) Denote X the number of leaves (in the three selected) which have undergone both types of transformations.

- The experiment concerns 300 leaves, among which 243 have undergone both types of transformations.
- As the leaves are selected **without replacement** (they are different), X follows the Hypergeometric distribution

$$X \sim \text{Hyp}(300, 243, 3).$$

- It follows

$$\mathbb{P}(X = 1) = \frac{\binom{243}{1}\binom{57}{2}}{\binom{300}{3}} = 0.087$$

b)

$$X \sim \text{Hyp}(300, 243, 3).$$

We have

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - \frac{\binom{243}{0}\binom{57}{3}}{\binom{300}{3}} = 1 - 0.006 = 0.994$$

c) Denote Y the number of leaves (in the three selected) which have undergone one but not both types of transformations.

- Out of the 300 leaves, $26 + 13 = 39$ have undergone one but not both types of transformations, so that

$$Y \sim \text{Hyp}(300, 39, 3).$$

$$\text{Hence} \qquad \mathbb{P}(Y = 1) = \frac{\binom{39}{1}\binom{261}{2}}{\binom{300}{3}} = 0.297$$

d) Denote W the number of leaves (in the three selected) which have undergone at least one type of transformation.

- Out of the 300 leaves, $243 + 26 + 13 = 282$ have undergone at least one type of transformation, so that

$$W \sim \text{Hyp}(300, 282, 3).$$

$$\text{So} \quad \mathbb{P}(W \geq 1) = 1 - \mathbb{P}(W = 0) = 1 - \frac{\binom{282}{0}\binom{18}{3}}{\binom{300}{3}} = 1 - 0.0002 = 0.9998$$

Exercise 9

Question:

Suppose a value z is repeatedly randomly chosen from a standard normal distribution :

a) In the long run, what is the proportion of times that z will be at most 2.15? Less than 2.15?
b) What is the long run proportion of times that z will be between -1.23 and 2.85?
c) What is the long run proportion of times that z will exceed 5? Will exceed -5?
d) What is the long run proportion of times that z will satisfy $|z| < 2.50$?

Solution:

- Consider $Z \sim \mathcal{N}(0, 1)$.

a) $\mathbb{P}(Z \leq 2.15) \overset{\text{table}}{=} 0.9842 = \mathbb{P}(Z < 2.15)$

b) $$\begin{aligned}\mathbb{P}(-1.23 < Z < 2.85) &= \mathbb{P}(Z \leq 2.85) - \mathbb{P}(Z < -1.23)\\ &\overset{\text{table}}{=} 0.9978 - 0.1093 = 0.8885\end{aligned}$$

c) $\mathbb{P}(Z > 5) \simeq 0$, $\mathbb{P}(Z > -5) \simeq 1$

d) $$\begin{aligned}\mathbb{P}(|Z| < 2.50) &= \mathbb{P}(-2.50 < Z < 2.50)\\ &= \mathbb{P}(Z < 2.50) - \mathbb{P}(Z < -2.50)\\ &\overset{\text{table}}{=} 0.9938 - 0.0062 = 0.9876\end{aligned}$$

Note: by symmetry of the normal distribution,

$$\mathbb{P}(Z < -2.50) = \mathbb{P}(Z > 2.50) = 1 - \mathbb{P}(Z < 2.50)$$

Exercise 10

Given the four scatter plots for two random variables X, Y in the Figure

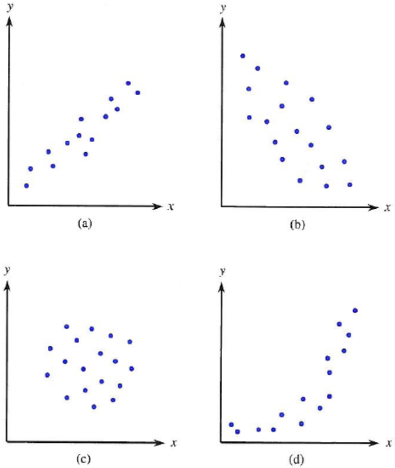


Figure:

Question:

1. which of the plots demonstrates a positive relationship ?
2. which of the plots demonstrates a positive linear relationship ?
3. which of the plots demonstrates a negative relationship ?
4. which of the plots demonstrates no relationship ?
5. for which of the plots would you expect positive correlation ? Negative correlation ? No or little correlation ?

Solution:

1. a) and d)
2. a)
3. b)
4. c)
5. a) large positive correlation, b) moderate negative correlation, c) no correlation, d) moderate positive correlation

Exercise 11

Question:

For each of the following pairs of variables, indicate whether you would expect a positive correlation, a negative correlation, or little or no correlation. Explain your choices.

1. Maximum daily temperature and cooling cost
2. Interest rate and number of loan applications
3. Distance a student doing MAS713 lives from NTU campus and their marks at the Matlab online quizzes

Question: Maximum daily temperature and cooling cost

Solution: One would expect a positive correlation since the hotter the maximum daily temperature the greater the amount of cooling utilised.

Question: Maximum daily temperature and cooling cost

Solution: One would expect a positive correlation since the hotter the maximum daily temperature the greater the amount of cooling utilised.

Question: Interest rate and number of loan applications

Solution: Interest rates directly impact on monthly loan repayment amounts, therefore one would expect a negative correlation between increasing interest rates and decreasing loan applications.

Question: Interest rate and number of loan applications

Solution: Interest rates directly impact on monthly loan repayment amounts, therefore one would expect a negative correlation between increasing interest rates and decreasing loan applications.

Question: Distance a student doing MAS713 lives from NTU campus and their marks at the Matlab online quizzes

Solution: There would likely be little relationship between the marks of a student and their distance lived from campus.

Question: Distance a student doing MAS713 lives from NTU campus and their marks at the Matlab online quizzes

Solution: There would likely be little relationship between the marks of a student and their distance lived from campus.