1. **Problem**
Political Party XYZ decides to do a promotional campaign at a mall where they hand out free XYZ branded t-shirts to anyone who is willing to put it on and post a selfie on social media showing them wearing it immediately. 72% of people who are XYZ supporters take up the offer, while only 25% of people who are not XYZ supporters take up the offer. Before this campaign polls showed that 39% of people who come into this mall are XYZ supporters. If you see someone walk out of the mall wearing the new XYZ t-shirt, what is the probability that they are an XYZ supporter? [Give the answer as a proportion accurate to 3 decimals, so 0.###]

Solution
Let A be the event that they are an XYZ supporter, and B be the event that they take up the offer, then

$$\begin{aligned} P[A|B] &= \frac{P[B|A]P[A]}{P[B]} \\ &= \frac{0.72 * 0.39}{0.72 * 0.39 + 0.25 * 0.61} \\ &\approx 0.648 \end{aligned}$$

2. **Problem**
Consider the following sample from an exponential distribution with rate λ (mean $1/\lambda$): 3.2, 0.1. What is the derivative of the log likelihood in the point $\lambda = 3$? [Accurate to 3 decimals]

Solution
$f(x_i) = \lambda \exp(-\lambda x_i)$, so $\ell = \sum_{i=1}^{n}(\log\lambda - \lambda x_i) + c_1 = n\log\lambda - \lambda\sum_{i=1}^{n} x_i + c_1$ and $\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - n\bar{x}$. Thus, if $\lambda = 3$, $n = 2$, and $\bar{x} = 1.65$, then $\frac{d\ell}{d\lambda} \approx -2.6333333$.

3. **Problem**
Given the density function $f(x) = 2.6\exp(2.6x)$, which of the following is **NOT** a valid way to describe this density?

   (a) $x \sim \chi^2(2.6/2)$
   (b) $x \sim Weibull(2.6, 1)$
   (c) $x \sim exp(2.6)$
   (d) $x \sim gamma(1, 2.6)$
   (e) $x \sim GPD(0, 0, 2.6)$

Solution
In a χ^2 distribution the slope/scale parameter is fixed, it is the shape that changes, so we cannot relate it to the exponential density in general.

4. **Problem**
If a random variable X takes on values 5, 6, 7, 8 with probabilities 0.27, 0.36, 0.19, 0.18 respectively, what is $E[X]$?

Solution
$E[X] = \sum_{i=1}^{4} x_i p_i = 6.28$

5. **Problem**
Let $x_1, \ldots, x_{n_x} = 7, 6, 2, 8, 7, 2, 9, 6$ and $y_1, \ldots, y_{n_y} = 3, 1, 4, 2, 9, 3, 8, 4$, then give a p-value flowing from the Welch t-test for the hypothesis $H_0: \mu_X = \mu_Y$ accurate to 3 decimals.

Solution
x <- c(7, 6, 2, 8, 7, 2, 9, 6)

```
y <- c(3, 1, 4, 2, 9, 3, 8, 4)
t.test(x, y)$p.value
```

≈ 0.2494963

6. **Problem**
   Let X be a random variable on the domain 0 to 1, with density function $f(x) = 1.5(x^2+2/3x)$. What is the $P[X > 0.4]$? [Accurate to 3 decimals]

   Solution
   $F(x) = 1.5 \int_0^x u^2 + 2/3u = 1.5 \left[\frac{1}{3}u^3 + \frac{2/3}{2}u^2\right]_0^x = 1.5\frac{2x^3+3*2/3*x^2}{6}$

   $P[X > 0.4] = F(1) - F(0.4) \approx 3.552$

7. **Problem**
   If the lengths of blades of grass on my lawn follow a normal distribution with mean 6.4cm and standard deviation 1.7cm, then what proportion of blades on my lawn are longer than 8.5cm? [Accurate to 3 decimals]

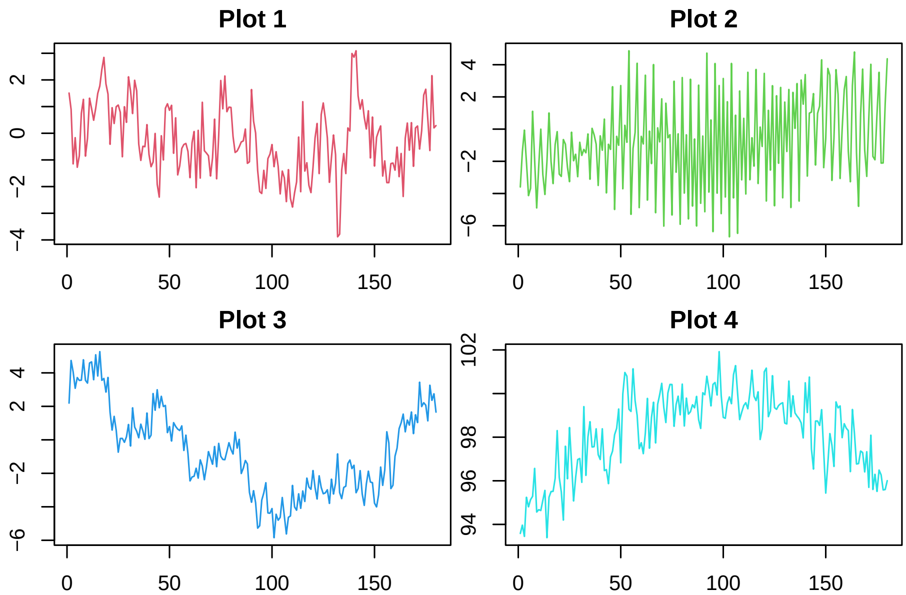
   Solution
   Let $X \sim N(6.4, 1.7^2)$ then $P[X > 8.5] = 1 - \Phi((8.5 - 6.4)/1.7) \approx 0.108$

8. **Problem**
   You do a standard regression involving two factors, A and B, each having two levels, L1 and L2. The interaction of the factors is also modelled. The estimated regression coefficients are reported as below. What is the model prediction for the case where both A and B are on level L2?

Predictor	Coefficient
Intercept	0.18
A_L2	8.14
B_L2	4.89
A_L2*B_L2	-0.75

   Solution
   By default, each coefficient is reported in terms of its contribution relative to the base of that factor, so the coefficient of A=L2 is relative to the intercept (which in this case is A=L1 and B=L1). Similarly, the coefficient of B=L2 is relative to the intercept, and the coefficient of A=L2 and B=L2 is simultaneously relative to A=L2 and B=L2. Thus, the prediction in this case is the sum of all 4 coefficients, which is 12.46.

9. **Problem**
   Which one of the following four time series is most likely to be stationary?



(a) Plot 1

(b) Plot 2

(c) Plot 3

(d) Plot 4

Solution
Plot 1 is a stationary AR(2) series. Plot 3 is a non-standard random walk, it would be a stationary MA(1) series if it were differenced though. Plot 2 shows a strong season pattern at lag 4, rendering it non-stationary. Plot 4 goes up and then down in a long run trend or cycle.

10. **Problem**
Using the random seed 3139, generate 1000 random values from a χ^2 distribution with 30 degrees of freedom, then give the average of these simulations accurate to 2 decimal places. Hint: the answer should be close to the degrees of freedom.

Solution

set.seed(3139)

rchisq(1000, 30) |> mean()

30.2505361