1. (a) What is the definition of independent random variables?

   (b) Roll a 6-sided die twice. Let

$$X = \text{first outcome}$$
$$Y = \text{first outcome} - \text{second outcome}.$$

   Check that X and Y are not independent by

   i. Using the definition from (a), or

   ii. Computing $\mathrm{Cov}(X,Y)$, or equivalently, $E[X]$, $E[Y]$, and $E[XY]$.

   (For i, try to compute $P(X \le 3)$, $P(Y \le 0)$, and $P((X \le 3) \cap (Y \le 0))$.

   (c) It is known that if two normal distributions are independent, then sum of them is also a normal distribution (I guess you learned this!). Let X and Y be two normal distributions of mean and standard deviation $(\mu, \sigma) = (1,3)$ and $(-3,4)$. What is the mean and variance of $X+Y$? Can you write down the pdf of it?

(a) Two random variables X, Y are independent if

$$P((X \le a) \cap (Y \le b)) = P(X \le a)P(Y \le b)$$

for any a, b.

(b) i. Choose $a = 3$ and $b = 0$ as suggested (you can do with other choices if you want). Then

$$P(X \le 3) = \frac{3}{6} = \frac{1}{2}$$
$$P(Y \le 0) = \frac{21}{36} = \frac{7}{12}$$
$$P((X \le 3) \cap (Y \le 0)) = \frac{15}{36} = \frac{5}{12}$$

and $P((X \le 3) \cap (Y \le 0)) \ne P(X \le 3)P(Y \le 0)$. So they are not independent.

ii. You can compute expectations directly, but here's a shortcut. Let X_1, X_2 be outcomes for the first and second roll, respectively. Then both are uniform random variables on $\{1, \dots, 6\}$ that are independent. Then $X = X_1$ and $Y = X_1 - X_2$, so the covariance is

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X_1, X_1 - X_2) = \mathrm{Cov}(X_1, X_1) - \mathrm{Cov}(X_1, X_2) = \mathrm{Var}(X_1) \ne 0$$

so X, Y are not independent. ($\mathrm{Cov}(X_1, X_2) = 0$ since X_1, X_2 are independent, and $\mathrm{Cov}(X_1, X_1) = \mathrm{Var}(X_1)$.

(c) $E[X+Y] = E[X] + E[Y] = -2$ and $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) = 3^2 + 4^2 = 25 \Rightarrow \sigma_{X+Y} = \sqrt{25} = 5$. Since it is a normal distribution, pdf would be

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot 5} e^{-\frac{(x+2)^2}{2 \cdot 25}}.$$

2. Let X be a geometric random variable with unknown p (hence $P(X = i) = (1-p)^{i-1}p$, $i = 1, 2, 3, \ldots$). You sampled numbers from X 10 times, and get the following outputs:

$$1, 8, 2, 4, 7, 3, 4, 2, 5, 4.$$

   (a) What is the sample mean?

   (b) What is the sample variance?

   (c) Now, let's try to *guess* p from our experiments. What is the probability of getting the above outputs, i.e.

$$P(X = 1)P(X = 8)P(X = 2)P(X = 4)P(X = 7)\cdots P(X = 4)$$

   as a function in p?

   (d) Use derivatives to find p that maximizes the probability in (c).

(a)

$$\frac{1+8+2+4+7+3+4+2+5+4}{10} = 4.$$

(b)

$$\frac{3^2+4^2+2^2+0^2+3^2+1^2+0^2+2^2+1^2+0^2}{9} = \frac{44}{9}$$

(c) $(1-p)^{40-10}p^{10} = (1-p)^{30}p^{10}$.

(d) The derivative is

$$-30(1-p)^{29}p^{10} + (1-p)^{30}10p^9 = (1-p)^{29}p^9(-30p + 10(1-p)) = 0 \Rightarrow p = \frac{10}{40} = \frac{1}{4}.$$

This is the maximum likelihood estimator for p. In general, if the outcomes are $x_1, \ldots, x_n$, then

$$\hat{p} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^{-1}.$$

Note that this is a biased estimator - $E[\hat{p}] \neq 1/p$.[1]

[1] https://math.stackexchange.com/questions/473190/unbiased-estimator-for-geometric-distribution