1. **(Sanity Check!)** Prove or give a counterexample: for any random variables X and Y, $\mathrm{Var}[\mathrm{X}+\mathrm{Y}] = \mathrm{Var}[\mathrm{X}] + \mathrm{Var}[\mathrm{Y}]$.

   Answer: Anything where $Cov(X,Y) \neq 0$. Probably the best way to illustrate this is by finding out what $Var[X+Y]$ is equal to.

   $$Var[X+Y] = E(((X+Y)-(E[X+Y]))^2)$$
   $$= E((X-E[X])^2 + (Y-E[Y])^2 + 2E[(X-E[X])(Y-E[Y]))]$$
   $$= Var(X) + Var(Y) + 2Cov(X,Y)$$

   We can further examine
   $$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$
   $$= E[XY - XE[Y] - E[X]Y + E[X]E[Y]]$$
   $$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]$$
   $$= E[XY] - E[X]E[Y]$$

   Thus, if for two events, $E[XY] - E[X]E[Y] = 0$, then we call these two events uncorrelated. Note that if X and Y are independent then this is always true.

   Here is an example: Let X be any random variable such that $\mathrm{Var}[\mathrm{X}] \neq 0$. Then, if we set $Y = X$, $\mathrm{Var}[\mathrm{X}+\mathrm{Y}] = \mathrm{Var}[2\mathrm{X}] = \mathbf{E}[(2\mathrm{X})^2] - (\mathbf{E}[2\mathrm{X}])^2 = 4\mathbf{E}[\mathrm{X}^2] - 4(\mathbf{E}[\mathrm{X}])^2 = 4\mathrm{Var}[\mathrm{X}] \neq \mathrm{Var}[\mathrm{X}] + \mathrm{Var}[\mathrm{X}]$.

2. **(Bernoulli and Binomial Distribution)** A random variable X is called a Bernoulli random variable with parameter p if $X = 1$ with probability p and $X = 0$ with probability $1-p$.

   (a) Calculate $\mathrm{E}[X]$ and $\mathrm{Var}[X]$.

   Answer:

   $$\mathrm{E}[X] = p \cdot 1 + (1-p) \cdot 0 = p$$
   $$\mathrm{Var}[X] = \mathbf{E}[(X-\mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = p \cdot 1^2 + (1-p) \cdot 0 - p^2 = p(1-p)$$

   (b) A Binomial random variable with parameters n and p is defined to be the sum of n independent, identically distributed Bernoulli random variables with parameter p. If Z is a Binomial random variable with parameters n and p, what are $\mathrm{E}[Z]$ and $\mathrm{Var}[Z]$?

   Answer: Let X_i be i.i.d. Bernoulli random variables with parameter p. By linearity of expectation,

   $$\mathrm{E}[Z] = \sum_{i=1}^{n} \mathrm{E}[X_i] = np$$

Moreover, since X_i's are independent,

$$\mathrm{Var}[Z] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = np(1-p).$$

3. **(Telebears)** Lydia has just started her Telebears appointment. She needs to register for a marine science class and CS70. There are no waitlists, and she can attempt to enroll once per day in either class or both. The Telebears system is strange and picky, so the probability of enrolling in the marine science class is p_1 and the probability of enrolling in CS70 is p_2. The probabilities are independent. Let M be the number of attempts it takes to enroll in the marine science class, and C be the number of attempts it takes to enroll in CS70.

   (a) What distribution do M and C follow? Are M and C independent?

   Answer: $M \sim \mathrm{Geom}(p)$, $C \sim \mathrm{Geom}(p)$ Yes they are independent.

   (b) For an integer $k \geq 1$, what is $\Pr[C \geq k]$?

   Answer: Question is asking for the probability that it takes at least k tries to enroll in CS70. $(1-p_2)^{k-1}$.

   (c) What is the expected number of classes she will be enrolled in if she must enroll with 14 days (inclusive)?

   Answer: $\Pr[M \leq 14] + \Pr[C \leq 14] = 1-(1-p_1)^{14} + 1-(1-p_2)^{14}$.

   (d) For an integer $k \geq 1$, what is the probability that she is enrolled in both classes before attempt k?

   Answer: Use independence. Let X be the number of attempts before she is enrolled in both. $\Pr[X < k] = \Pr[M < k]\Pr[C < k] = (1-(1-p_1)^{k-1})(1-(1-p_2)^{k-1})$.

4. **(Toujours les poissons)** Use the Poisson distribution to answer these questions.

   (a) Suppose that on average, 20 people ride your roller coaster per day. What is the probability that exactly 7 people ride it tomorrow?

   Answer: $X \sim \mathrm{Poiss}(20)$. $\Pr[X = 7] = \frac{20^7}{7!}e^{-20} \approx 5.23 \cdot 10^{-4}$.

   (b) Suppose that on average, you go to Six Flags twice a year. What is the probability that you will go at most once in 2015?

   Answer: $X \sim \mathrm{Poiss}(2)$. $\Pr[X \leq 1] = \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} \approx 0.41$.

   (c) Suppose that on average, there are 5.7 accidents per day on California roller coasters. (I hope this is not true.) What is the probability there will be *at least* 3 accidents throughout the *next two days* on California roller coasters?

   Answer: Let Y be the number of accidents that occur in the next two days. We can approximate Y as a Poisson distribution $Y \sim \mathit{Poiss}(\lambda = 11.4)$, where λ is the average number of accidents over two days. Now, we compute

$$\begin{aligned}
\Pr[Y \geq 3] &= 1 - \Pr[Y < 3] \\
&= 1 - \Pr[Y = 0 \cup Y = 1 \cup Y = 2] \\
&= 1 - (\Pr[Y=0] + \Pr[Y=1] + \Pr[Y=2]) \\
&= 1 - \left(\frac{11.4^0}{0!}e^{-11.4} + \frac{11.4^1}{1!}e^{-11.4} + \frac{11.4^2}{2!}e^{-11.4}\right) \\
&\approx 0.999.
\end{aligned}$$

We can show what we did above formally with the following claim: the sum of two independent Poisson random variables is Poisson. We won't prove this, but from the above, you should intuitively know why this is true. Now, we can model accidents on day i as a Poisson distribution $X_i \sim Poiss(\lambda = 5.7)$. Now, Let X_1 be the number of accidents that happen on the next day, and X_2 be the number of accidents that happen on the day after next. We are interested in $Y = X_1 + X_2$. Thus, we know $Y \sim Poiss(\lambda = 5.7 + 5.7 = 11.4)$.

5. **(Will I Get My Package?)** A sneaky delivery guy of some company is out delivering n packages to n customers. Not only does he hand a random package to each customer, he tends to open a package before delivering with probability $\frac{1}{2}$ (independently of the choice of the package). Let X be the number of customers who receive their own packages unopened.

   (a) Compute the expectation $\mathrm{E}(X)$.

   Answer: Define $X_i = \begin{cases} 1 & \text{if the } i\text{-th customer gets his/her own package unopened} \\ 0 & \text{otherwise.} \end{cases}$

   Then, $\mathrm{E}(X_i) = \Pr[X_i = 1] = \frac{1}{2n}$, since the i-th customer will get his/her own package with probability $\frac{1}{n}$ and it will be unopened with probability $\frac{1}{2}$.

   By linearity of expectation, $\mathrm{E}[X] = \mathrm{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathrm{E}[X_i] = n \cdot \frac{1}{2n} = \frac{1}{2}$.

   (b) What is the probability that customers i and j both receive their own packages unopened?

   Answer: The probability that both customers receive their own packages is $\frac{(n-2)!}{n!}$, and the probability that both customers receive unopened packages is $\frac{1}{4}$. Since the two events are independent, the probability that both customers receive their own packages unopened is $\frac{1}{4}\frac{(n-2)!}{n!} = \frac{1}{4n(n-1)}$.

   (c) Compute the variance Var[X].

   Answer: Since $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$, we need to first compute $\mathrm{E}(X^2)$.

   Note that $\mathrm{E}(X^2) = \mathrm{E}\big((X_1 + X_2 + \ldots + X_n)^2\big) = \mathrm{E}(\sum_{i,j} X_i X_j) = \sum_{i,j} \mathrm{E}(X_i X_j)$, where the last equality follows from linearity of expectation.

   If $i = j$, then $\mathrm{E}(X_i X_j) = \mathrm{E}(X_i^2) = \frac{1}{2n}$.

   If $i \neq j$, then $\mathrm{E}(X_i X_j) = \Pr(X_i X_j = 1) \cdot 1 + \Pr(X_i X_j = 0) \cdot 0 = \frac{1}{4n(n-1)}$.

   Hence, $\mathrm{E}(X^2) = \sum_{i,j} \mathrm{E}(X_i X_j) = \sum_i \mathrm{E}(X_i^2) + \sum_{i \neq j} \mathrm{E}(X_i X_j) = n \cdot \frac{1}{2n} + n(n-1) \cdot \frac{1}{4n(n-1)} = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$.

   It follows that $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$.