1. (20 points) In a survey conducted by the City of San Diego, 120 of 400 interviewed San Diegans confirmed they had at least one encounter with coyotes in the past six months.

Using the above data, construct a 90% confidence interval for the unknown fraction of the city population who encountered coyotes in the past six months. Clearly state (in words) your conclusion.

[You may leave your answer in terms of \( \Phi \) or use the table provided on the last page of the exam.]

**Solution.** Denote by \( p \) the fraction of San Diegans who encountered at least one coyote in the past six months. In order to construct the confidence interval for \( p \), take the estimate \( \hat{p} = \frac{120}{400} = 0.3 \) and find the smallest \( \varepsilon > 0 \) for which

\[
P(|p - \hat{p}| < \varepsilon) \geq 0.9.
\]

Using the normal approximation of the binomial distribution we estimate the above probability as

\[
P(|p - \hat{p}| < \varepsilon) \geq 2\Phi(2\varepsilon \sqrt{n}) - 1,
\]

where \( n \) is the number of interviewed San Diegans. The 90% confidence interval is obtained by taking \( \varepsilon \) such that

\[
2\Phi(2\varepsilon \sqrt{n}) - 1 = 0.9.
\]

Since the CDF of the standard normal distribution \( \Phi \) is strictly increasing, we can solve the above equation with respect to \( \varepsilon \)

\[
\varepsilon = \frac{\Phi^{-1}(0.95)}{2 \cdot 20}.
\]

Conclusion. The 90% confidence interval for the fraction of the San Diego population that encountered coyotes is

\[
\left(0.3 - \frac{\Phi^{-1}(0.95)}{40}, 0.3 + \frac{\Phi^{-1}(0.95)}{40}\right).
\]
2. (20 points) Let $X$ be a random variable describing the lifetime of a certain component. We know that $X$ is a continuous random variable with PDF (in years)

$$f_X(x) = \begin{cases} \frac{2}{x^3}, & x \in [1, +\infty), \\ 0, & \text{otherwise}. \end{cases} \quad (6)$$

The component is replaced either when it fails, or after 4 years of service even if it is still operational.

Denote by $Y$ the time the component is in operation (i.e., from being installed to being replaced). Compute $E(Y)$.

**Solution.** The time in operation $Y$ is equal to $X$ if $X$ is less than 4 years, or equal to 4 if $X$ is greater than 4 years. Therefore, we can express $Y$ as a function of $X$

$$Y = g(X), \quad (7)$$

where

$$g(x) = \begin{cases} x, & x \leq 4, \\ 4, & x > 4. \end{cases} \quad (8)$$

We can thus compute $E(Y)$ as $E(g(X))$

$$E(Y) = \int_{\mathbb{R}} g(x) f_X(x) \, dx = \int_1^4 x \cdot \frac{2}{x^3} \, dx + \int_4^{+\infty} 4 \cdot \frac{2}{x^3} \, dx \quad (9)$$

The first integral gives

$$\int_1^4 \frac{2}{x^2} \, dx = -\frac{2}{x} \bigg|_1^4 = 2 - 1 = \frac{3}{2}. \quad (10)$$

the second integral gives

$$\int_4^{+\infty} \frac{8}{x^3} \, dx = -\frac{8}{2x^2} \bigg|_4^\infty = \frac{1}{4}. \quad (11)$$

Therefore,

$$E(Y) = \frac{3}{2} + \frac{1}{4} = \frac{7}{4}. \quad (12)$$
3. (20 points) According to the US Department of Treasury, one in every 10,000 US dollar notes is counterfeit.

A cash-in-transit van operating in San Diego area transports 20,000 US dollar notes from a supermarket to a bank.

Estimate the probability that there are at least 3 counterfeit notes in this van.

[Explain your choice of approximation. You may leave the answers in terms of $\Phi(x)$ or $e^x$. Do not use the continuity correction.]

**Solution.** We model the number of the counterfeit notes in the van using a random variable $X$ having binomial distribution with parameters $n = 20000$ and $p = 10^{-4}$, $X \sim \text{Bin}(20000, 10^{-4})$. We are asked to compute the probability that $X \geq 3$. For this, we use the Poisson approximation with $\lambda = np = 20000 \cdot 10^{-4} = 2$, so that for any $k \in \{0, 1, 2, \ldots\}$

$$P(X = k) \approx \frac{2^k}{k!}e^{-2}. \quad (13)$$

We choose the Poisson approximation since $np^2 = 2 \cdot 10^{-4}$ is much smaller than 1, and

$$np(1 - p) = 20000 \cdot 10^{-4}(1 - 10^{-4}) \approx 2, \quad (14)$$

which is not enough to guarantee a good normal approximation.

In order to estimate $P(X \geq 3)$ we use the complement formula together with the Poisson approximation

$$P(X \geq 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \approx 1 - \left[e^{-2} + \frac{2}{1!}e^{-2} + \frac{4}{2!}e^{-2}\right] = 1 - 5e^{-2}. \quad (15)$$
4. (20 + extra 5 points) Let $X$ be a random variable taking values in the set \{1, 2, 3, \ldots\}. Let $p = P(X = 1)$ satisfy $0 < p < 1$.

Suppose that for random variable $X$

$$P(X = k + n \mid X > n) = P(X = k)$$  \hfill (16)

for any $n, k \geq 1$.

(a) Consider the identity (16) with $k = 1$ and $n = 1$

$$P(X = 2 \mid X > 1) = P(X = 1).$$

Use it to express $P(X = 2)$ in terms of $p$. [Hint: Notice that \{X = 2\} $\subset$ \{X > 1\}.]

**Solution.**

Using (16) for $k = 1$ and $n = 1$ and the definition of the conditional probability we have

$$p = P(X = 2 \mid X > 1) = \frac{P(X = 2, X > 1)}{P(X > 1)} = \frac{P(X = 2)}{1 - P(X = 1)} = \frac{P(X = 2)}{1 - p},$$ \hfill (17)

from which we find that

$$P(X = 2) = (1 - p)p.$$ \hfill (18)

(b) Consider the identity (16) with $k = 2$ and $n = 1$

$$P(X = 3 \mid X > 1) = P(X = 2).$$

Use it together with the result of (a) to express $P(X = 3)$ in terms of $p$.

**Solution.** Using (16) for $k = 2$ and $n = 1$, equation (18) and the definition of the conditional probability we have

$$(1 - p)p \overset{(18)}{=} P(X = 2) \overset{(16)}{=} P(X = 3 \mid X > 1) = \frac{P(X = 3, X > 1)}{P(X > 1)} = \frac{P(X = 3)}{1 - p},$$ \hfill (19)

from which we find that

$$P(X = 3) = (1 - p)^2 p.$$ \hfill (20)
(c) Use the identity (16) with general \( k \geq 1 \) and \( n = 1 \) to show that
\[
P(X = k + 1) = P(X = k)P(X > 1),
\]
and determine the distribution of \( X \).

**Solution.** Repeating the same argument as in parts (a) and (b) for general \( k \geq 1 \) and \( n = 1 \) we get
\[
P(X = k) \overset{(16)}{=} P(X = k + 1 | X > 1) = \frac{P(X = k + 1, X > 1)}{P(X > 1)} = \frac{P(X = k + 1)}{1 - p}, \tag{21}
\]
from which it follows that
\[
P(X = k + 1) = (1 - p)P(X = k). \tag{22}
\]
Fix \( l \geq 2 \). By applying (22) \( l - 1 \) times we find
\[
P(X = l) = (1 - p)P(X = l - 1) \tag{23}
\]
\[
= (1 - p)^2P(X = l - 2) \tag{24}
\]
\[
= \cdots \tag{25}
\]
\[
= (1 - p)^{l-1}P(X = 1) \tag{26}
\]
\[
= (1 - p)^{l-1}p. \tag{27}
\]
We conclude that \( X \) has geometric distribution with parameter \( p = P(X = 1) \), \( X \sim \text{Geom}(p) \).

**Remark.** Property (16) is the discrete version of the memoryless property. We have seen in the lectures that if \( X \sim \text{Geom}(p) \), then \( X \) satisfies the memoryless property (16). In this problem we have proven the converse: if \( X \) is a discrete random variable satisfying (16), then \( X \sim \text{Geom}(p) \). Therefore, geometric distribution is the only discrete distribution on the set \( \{1, 2, \ldots \} \) that satisfies the memoryless property (16).