1. Suppose that 45% of the phone calls you receive are scam calls. Assume that the probability of a scam call is independent from one call to the next. Let X be the number of calls you receive until (and including) the next scam call.

(a) What is $P(X = 3)$?

$$P(X=3) = \underbrace{(0.55)^2}_{\text{not scam calls}}\underbrace{(0.45)}_{\text{scam call}} \approx 0.136$$

(b) If n is any positive integer, what is $P(X = n)$?

$$P(X=n) = \underbrace{(0.55)^{n-1}}_{\text{not scam calls}}\underbrace{(0.45)}_{\text{scam call}}$$

(c) What is $E(X)$?

$$E(X) = \frac{1}{0.45} \approx 2.22$$

2. Let Y be the number of calls until (and including) the fourth scam call.

(a) What is $P(Y = n)$?

$$P(Y=n) = \underbrace{\binom{n-1}{3}}_{\text{3 scam calls in } n-1 \text{ calls}}\underbrace{(0.45)^3}_{\text{prob. 3 scam calls}}\underbrace{(0.55)^{n-4}}_{\text{prob. } n-4 \text{ non-scam calls}}\underbrace{(0.45)}_{\text{last scam call}} = \binom{n-1}{3}(0.45)^4(0.55)^{n-4}$$

(b) What is $E(Y)$?

$$E(Y) = \frac{4}{0.45} \approx 8.89$$

3. An interviewer must find 10 people who agree to be interviewed. Suppose that when asked, a person agrees to be interviewed with probability 0.4.

(a) What is the probability that the interviewer will have to ask exactly 20 people?

Let $X \sim NB(r=10,\ p=0.4)$, so $P(X=20) = \binom{20-1}{10-1}(0.4)^{10}(0.6)^{10} \approx 0.0586$

(b) What are the mean and standard deviation of the number of people that the interviewer will have to ask?

$$E(X) = \frac{r}{p} = \frac{10}{0.4} = 25$$

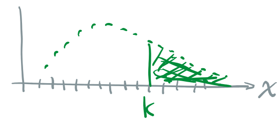
$$Var(X) = \frac{r(1-p)}{p^2} = \frac{10(0.6)}{0.4} = 37.5$$

$$\sigma_X = \sqrt{37.5} \approx 6.12$$

4. If X has a geometric distribution with parameter p, and k is a positive integer, what is $P(X > k)$?

$X > k$ means that the first k trials are all failures.

Thus, $\boxed{P(X > k) = (1-p)^k}$ ← **GEOMETRIC TAIL PROBABILITY**



5. Scam calls, again. Suppose that 45% of the phone calls you receive are scam calls.

(a) What is the probability that *none* of the first 4 calls are scam calls?

$X \sim \text{Geometric}(0.45) \qquad P(X > 4) = (0.55)^4 \approx 0.092$

(b) If none of the first 4 calls are scam calls, what is the probability that none of the first 7 calls are scam calls?

$$P(X > 7 \mid X > 4) = \frac{P(X > 7 \text{ and } X > 4)}{P(X > 4)} = \frac{P(X > 7)}{P(X > 4)} = \frac{(0.55)^7}{(0.55)^4} = \underbrace{(0.55)^3}_{\text{This is } P(X>3)} \approx 0.166$$

(c) If none of the first 4 calls are scam calls, what is the probability that none of the first $4 + k$ calls are scam calls?

$$P(X > 4 + k \mid X > 4) = \frac{P(X > 4 + k)}{P(X > 4)} = \frac{(0.55)^{4+k}}{(0.55)^4} = (0.55)^k = P(X > k)$$

OBSERVATION: For $X \sim \text{Geo}(p)$ and integers $0 < s < t$,

$$\boxed{P(X > t \mid X > s) = P(X > t - s)}$$

← **MEMORYLESS PROPERTY** of a geometric rv

INTERPRETATION: The waiting time until the next success does not depend on how many failures you have already seen.

BONUS: Show that $\sum_{k=1}^{\infty} k(1-p)^{k-1} p = \frac{1}{p}$. This proves that the mean of a geometric random variable is $\frac{1}{p}$.

Recall the geometric series: $a + ar + ar^2 + ar^3 + \cdots = \frac{a}{1-r}$ for $|r| < 1$

Let $a = r = 1-p$: $\sum_{k=1}^{\infty} (1-p)^k = (1-p) + (1-p)^2 + (1-p)^3 + \cdots = \frac{1-p}{p}$

so: $\sum_{k=1}^{\infty} (1-p)^k = \frac{1}{p} - 1$

Differentiate with respect to p:

$$\sum_{k=1}^{\infty} -k(1-p)^{k-1} = -\frac{1}{p^2}$$

Multiply by $-p$: $\sum_{k=1}^{\infty} k(1-p)^{k-1} p = \frac{1}{p}$