1 Expected Value

1.1 Summarizing Data

We want to extract information from large amounts of data. Let $x_1, x_2, \ldots, x_n$ be n realizations of a random variable X (such a set is called a *sample*).

1.1.1 Visualizing Data

1. **Frequency Table:** A table containing the number of times each value of X occurred.
2. **Frequency Histogram:** A histogram of the number of times each value of X occurred.

1.1.2 Summary Statistics

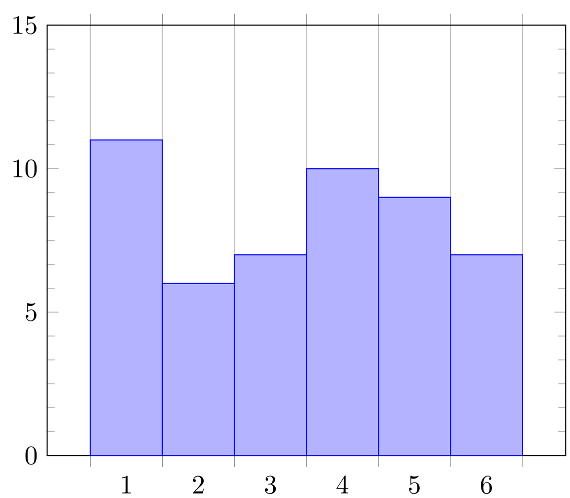
1. **Sample Mean:** The average of all realizations, defined by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. **Sample Median**: A value such that half of the results are below it and the other half above it, when the sample is arranged in numerical order. By convention, we take the middle of two values if there are an even number of observations.
3. **Sample Mode**: The most frequently-occuring values in a sample. There may be more than one.

Example 1. Consider 50 rolls of a dice. The frequency table and corresponding histogram are below

Outcome	1	2	3	4	5	6
Frequency	11	6	7	10	9	7



We also have that the summary statistics are

1. Sample Mean: We see that $\frac{1\times11+2\times6+3\times7+4\times10+5\times9+6\cdot7}{50} = 3.42$.
2. Sample Median: Looking at the 25th and 26th ordered result, we see that the median is 4.
3. Sample Mode: The outcome with the largest frequency is 1.

1.2 Summarizing Random Variables - Expected Value

Instead of computing the summary statistics of large amounts of samples, we can compute its theoretical values if we know the distribution of the random variable. The first key theoretical value is the "balancing point" of the random variable, also known as the *expected value*, *first moment*, or *mean.*

Definition 1. Suppose X is a discrete random variable with probability function $f_X(x)$. The *expected value* of X, denoted by $\mathbb{E}[X]$, is the number

$$\mathbb{E}[X] = \sum_{x \in X(S)} x f_X(x) = \sum_{x \in X(S)} x\,\mathbb{P}(X = x),$$

provided the sum converges absolutely (that is, if $\sum_{x \in X(S)} |x|\ f_X(x) < \infty$).

Remark 1. The sample mean of many realizations of X converges to $\mathbb{E}[X]$. This is called the *law of large numbers.*

1.2.1 Properties

1. Law of the Unconscious Statistician: If $g : \mathbb{R} \to \mathbb{R}$, and X is a random variable with PMF f_X, then $g(X)$ is a random variable taking values $g(X(S))$ and

$$\mathbb{E}[g(X)] = \sum_{x \in X(S)} g(x) f_X(x)$$

2. Linearity: Suppose the random variable X has $\mathbb{E}[X] = \mu$. Then for any constants $a, b \in \mathbb{R}$,

$$\mathbb{E}[aX + b] = a\mu + b = a\,\mathbb{E}[X] + b$$

3. Linearity II:: Suppose that X and Y are random variables. Then

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Remark 2. In general $\mathbb{E}[g(X)] \neq g[\mathbb{E}(X)]$. However, if g is convex then

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]).$$

This is called *Jensen's inequality.*

Remark 3. The fact that a is a constant is very important in the first linearity property. If X and Y are random variables, then in general $\mathbb{E}[XY] \neq \mathbb{E}[X]\,\mathbb{E}[Y]$, unless X and Y are independent.

1.2.2 Expected Value of Common Distributions

1. **Uniform:** If $X \sim \mathrm{U}[a, b]$ then $\mathbb{E}[X] = \frac{a+b}{2}$.
2. **Hypergeometric:** If $X \sim \mathrm{Hyp}(N, r, n)$, then $\mathbb{E}[X] = r\frac{n}{N}$.
3. **Binomial:** If $X \sim \mathrm{Bin}(n, p)$ then $\mathbb{E}[X] = np$.
4. **Negative Binomial:** If $X \sim \mathrm{NegBin}(k, p)$, then $\mathbb{E}[X] = \frac{k(1-p)}{p}$.
5. **Poisson:** If $X \sim \mathrm{Poi}(\mu)$, then $\mathbb{E}[X] = \mu$.

Since $X \sim \mathrm{Ber}(p) \sim \mathrm{Bin}(1, p)$ and $X \sim \mathrm{Geo}(p) \sim \mathrm{NegBin}(1, p)$ we get the following for free,

1. **Bernoulli:** If $X \sim \mathrm{Ber}(p)$ then $\mathbb{E}[X] = p$.
2. **Geometric:** If $X \sim \mathrm{Geo}(p)$ then $\mathbb{E}[X] = \frac{1-p}{p}$.

1.3 Example Problems

1.3.1 Applications

Problem 1.1. Consider a random variable X with PMF

$$f_X(1) = 0.3 \quad f_X(2) = 0.25 \quad f_X(3) = 0.2 \quad f_X(4) = 0.15 \quad f_X(5) = 0.1$$

What is $\mathbb{E}[X]$?

Solution 1.1. By definition,

$$\mathbb{E}[X] = 1 \cdot 0.3 + 2 \cdot 0.25 + 3 \cdot 0.2 + 4 \cdot 0.15 + 5 \cdot 0.1 = 2.5$$

Problem 1.2. A lottery is conducted in which 7 numbers are drawn without replacement between the numbers 1 and 49. A player wins the lottery if the numbers selected on their ticket match all 7 of the drawn numbers. A ticket to play the lottery costs 10 cents, and the jackpot is valued at \$1,000,000. Is the expected return* of this bet positive, i.e., would you play this bet?

Note: The return of a bet is the winnings minus costs.

Solution 1.2. Let R denote the return of the game. The random variable R can take two values, depending on if we win or not:

$$R = \begin{cases} -0.10 & \text{with probability } 1 - \frac{1}{\binom{49}{7}} \\ 999,999.90 & \text{with probability } \frac{1}{\binom{49}{7}} \end{cases}$$

The expected value of R, or the expected return, is then

$$\mathbb{E}[R] = -0.10 \cdot \left(1 - \frac{1}{\binom{49}{7}}\right) + 999,999.90 \cdot \frac{1}{\binom{49}{7}} \approx -0.0884$$

Problem 1.3. Suppose X is a random variable satisfying $a \leq X(\omega) \leq b$ for all $\omega \in S$. Show that $a \leq \mathbb{E}[X] \leq b$.

Solution 1.3. Since $a \leq X(\omega) \leq b$, we have $X(S) \subseteq [a, b]$. There

$$\mathbb{E}[X] = \sum_{x \in X(S)} x f_X(x) \leq \sum_{x \in X(S)} b f_X(x) = b \sum_{x \in X(S)} f_X(x) = b$$

and

$$\mathbb{E}[X] = \sum_{x \in X(S)} x f_X(x) \geq \sum_{x \in X(S)} a f_X(x) = a \sum_{x \in X(S)} f_X(x) = a.$$

Problem 1.4. Let X be the outcome of one die roll with a fair six-sided die.

1. What is the expected value of X?
2. What is the expected value of the square of X?

Solution 1.4.

1. Since $X \sim U[1,6]$ we have
$$\mathbb{E}[X] = \frac{1+6}{2} = 3.5.$$
Alternative Solution: We can compute this directly
$$\mathbb{E}[X] = \frac{1}{6}\sum_{x=1}^{6} x = 3.5.$$

2. The square does not follow any known density, so we compute it directly,
$$\mathbb{E}[X^2] = \frac{1}{6}(1^2 + 2^2 + \cdots + 6^2) = \frac{91}{6}.$$

Problem 1.5. Suppose the discrete random variable X has PMF
$$f_X(-1) = 0.15, \quad f_X(0) = 2c, \quad f_X(1) = 0.5, \quad f_X(2) = 0.05, \quad f_X(3) = c.$$
where c is a constant making f a valid PMF.

1. What is $\mathbb{E}[X]$?
2. What is $\mathbb{E}[e^X]$?

Solution 1.5. Since the probabilities sum to 1, we must have
$$1 = \sum_{x=-1}^{3} f_X(i) = 3c + 0.7 \implies c = 0.1$$

1. By definition,
$$\mathbb{E}[X] = (-1)\cdot 0.15 + 0 \cdot 0.2 + 1 \cdot 0.5 + 2 \cdot 0.05 + 3 \cdot 0.1 = 0.75$$

2. By the law of the unconscious statistician
$$\mathbb{E}[e^X] = e^{-1}\cdot 0.15 + e^0 \cdot 0.2 + e^1 \cdot 0.5 + e^2 \cdot 0.05 + e^3 \cdot 0.1 \approx 3.992$$

Problem 1.6. Suppose two fair six sided die are independently rolled 24 times, and let X denote the number of times the sum of die rolls is 7. What is $\mathbb{E}[X]$?

Solution 1.6. The ways of rolling a 7 out of two die rolls is $\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$ so the probability is $\frac{6}{6^2} = \frac{1}{6}$. This is the probability of a success and there are 24 trials, $X \sim \text{Bin}(24, \frac{1}{6})$ so
$$\mathbb{E}[X] = \frac{24}{6} = 4.$$

Problem 1.7. You are invited to a quiz show! There are two categories of questions, History and Geography. The host lets you pick which category to pick first. If you get your first question right, you are then given the opportunity to answer the second question, otherwise the game is over. Because history is so much harder, the history question is worth \$200 while the Geography question is only worth \$100. Checking back on your high school transcript you estimate that you get a geography question right with probability 70% while you get a history question right with probability 55%. You can assume that knowing the answers to the two questions is independent.

1. Which category should you pick first in order to maximize your expected winnings?
2. Suppose you are indecisive and flip a fair coin before picking the category you answer first. If the coin shows heads, you pick the history question first, otherwise the geography question. Compute the the expected value of your winnings.

Solution 1.7.

1. Let W_h and W_g be the winnings if we pick history first and geography first respectively. If we first pick history, the expected winnings are

$$\mathbb{E}[W_h] = (200 + 100) \cdot 0.55 \cdot 0.7 + (200 + 0) \cdot 0.55 \cdot 0.3 = 148.5$$

If we first pick geography, the expected winnings are

$$\mathbb{E}[W_g] = (100 + 200) \cdot 0.7 \cdot 0.55 + (100 + 0) \cdot 0.7 \cdot 0.45 = 147$$

Since $E[W_h] > E[W_g]$, we should pick history first to maximize the expected winnings.

2. Let W be the winnings if a coin decides our decision. Consider the random variables

$$\mathbb{1}_H = \begin{cases} 1 & \text{we flip a } H \\ 0 & \text{we flip a } T \end{cases} \quad \text{and} \quad \mathbb{1}_T = \begin{cases} 1 & \text{we flip a } T \\ 0 & \text{we flip a } H \end{cases}.$$

We have $W = \mathbb{1}_H W_h + \mathbb{1}_T W_g$, so the linearity of expectation and independence implies that implies that

$$\mathbb{E}[W] = \mathbb{E}[\mathbb{1}_H W_h + \mathbb{1}_T W_g] = \mathbb{E}[\mathbb{1}_H W_h] + \mathbb{E}[\mathbb{1}_T W_g] = \mathbb{E}[\mathbb{1}_H]\,\mathbb{E}[W_h] + \mathbb{E}[\mathbb{1}_T]\,\mathbb{E}[W_g].$$

It is easy to see that $\mathbb{E}[\mathbb{1}(H)] = \mathbb{E}[\mathbb{1}_T] = \frac{1}{2}$ so the results in part (1) implies

$$\mathbb{E}[W] = \mathbb{E}[\mathbb{1}_H W_h] + \mathbb{E}[\mathbb{1}_T W_g] = \frac{1}{2}(\mathbb{E}[W_h] + \mathbb{E}[W_g]) = 147.75.$$

Remark 4. We used the fact that if X and Y are independent that $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$, but we will go more into that in later chapters. In this example, it is not hard to compute the expectation explicitly without using this fact. Since $\mathbb{1}_H W_h = 0$ if we flip a tails, independence implies that

$$\begin{aligned}\mathbb{E}[\mathbb{1}_H W_h] &= \sum_{x \in W_h(S)} 0\,\mathbb{P}(W_h = x, T) + \sum_{x \in W_h(S)} x\,\mathbb{P}(W_h = x, H) \\ &= \sum_{x \in W_h(S)} x\,\mathbb{P}(W_h = x)\,\mathbb{P}(H) = \frac{1}{2}\sum_{x \in W_h(S)} x\,\mathbb{P}(W_h = x) = \frac{1}{2}\,\mathbb{E}[W_h].\end{aligned}$$

The exact same argument implies that $\mathbb{E}[\mathbb{1}_T W_h] = \frac{1}{2}\,\mathbb{E}[W_g]$.

Alternate Solution: We can list out all the cases. Let W denote the winnings.

1. *Heads, first wrong.* We first pick history and get it wrong. We win $W = 0$ with probability $0.5 \cdot 0.45 = 0.225$
2. *Heads, second wrong.* We first pick history, get it right, then geography wrong so we win $W = 200$ with probability $0.5 \cdot 0.55 \cdot 0.3 = 0.0825$.
3. *Heads, no questions wrong.* We win $W = 300$ with probability $0.5 \cdot 0.55 \cdot 0.7 = 0.1925$.

4. *Tails, first wrong.* We first pick geography and get it wrong. We win $W = 0$ with probability $0.5 \cdot 0.3 = 0.15$.

5. *Tails, second wrong.* We first pick geography, get it right, then history wrong so we win $W = 100$ with probability $0.5 \cdot 0.7 \cdot 0.45 = 0.1575$.

6. *Tails, no questions wrong.* We win $W = 300$ with probability $0.5 \cdot 0.7 \cdot 0.55 = 0.1925$.

The PMF of the winnings W is given by

$$f_W(0) = 0.225 + 0.15, \ f_W(100) = 0.1575 \ , f_W(200) = 0.0825, \ f_W(300) = 0.1925 + 0.1925.$$

and we find

$$\mathbb{E}[W] = 100 \cdot 0.1575 + 200 \cdot 0.0825 + 300 \cdot (0.1925 + 0.1925) = 147.75.$$

Problem 1.8. Suppose that calls to the Canadian Tire Financial call center follow a Poisson process with rate 30 calls per minute. Let X denote the number of calls to the center after 1 hour. What is $\mathbb{E}[X/2 - 1]$?

Solution 1.8. The rate of 30 calls per minute is equivalent to a rate of $30 \cdot 60$ calls per hour, so $X \sim \text{Poi}(30 \cdot 60)$. Therefore,

$$\mathbb{E}\left[\frac{X}{2} - 1\right] = \frac{1}{2}\mathbb{E}[X] - 1 = \frac{1800}{2} - 1 = 899.$$

Problem 1.9. Consider a random variable X with PMF $f_X(x) = \frac{1}{x}$ for $x = 2, 4, 8, 16, \ldots$ and 0 otherwise.

1. Show that $\sum_{\text{all x}} f_X(x) = 1$.

2. What is $\mathbb{E}[X]$?

Solution 1.9.

1. The range of X is 2^n where $n \geq 1$. We have by the geometric series that

$$\sum_{\text{all x}} f_X(x) = \sum_{n \geq 1} \frac{1}{2^n} = \sum_{n \geq 0} \frac{1}{2^n} - 1 = \frac{1}{1 - \frac{1}{2}} - 1 = 1.$$

2. For the expected value, we have

$$\mathbb{E}[X] = \sum_{n \geq 1} 2^n f_X(2^n) = \sum_{n=1}^{\infty} 2^n \left(\frac{1}{2^n}\right) = \sum_{n=1}^{\infty} 1 = \infty,$$

so the expected value is infinite, and does not exist in the traditional sense.

Problem 1.10. $(*)$ Suppose that n male-female couples are at a party and that the males and females are randomly paired for a dance. A match occurs if a couple happens to be paired together. What is the expected number of matches?

Solution 1.10. This famous problem is called the matching problem. The expected value is easily computed using the linearity of expectation. We label the males $1, \dots, n$ and let A_i be the event that the male is matched with his wife. We define

$$\mathbb{1}_{A_i} = \begin{cases} 1 & \text{if } A_i \text{ happens} \\ 0 & \text{if } A_i \text{ does not happen} \end{cases} = \begin{cases} 1 & \text{male } i \text{ is matched with his wife} \\ 0 & \text{male } i \text{ is not matched with his wife.} \end{cases}$$

If X denotes the number of matches, then $X = \mathbb{1}(A_1) + \cdots + \mathbb{1}(A_n)$. Therefore,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n}] = \mathbb{E}[\mathbb{1}_{A_1}] + \cdots + \mathbb{E}[\mathbb{1}_{A_n}].$$

By symmetry, we see that $\mathbb{E}[\mathbb{1}_{A_1}] = \mathbb{E}[\mathbb{1}_{A_2}] = \cdots = \mathbb{E}[\mathbb{1}_{A_n}]$ and

$$\mathbb{E}[\mathbb{1}_{A_1}] = 1 \cdot \mathbb{P}(A_1) + 0 \cdot \mathbb{P}(A_n^c) = \mathbb{P}(A_1) = \frac{1}{n}$$

since the probability that the first male is matched with any of the n women is equally likely, and only 1 is his wife. Therefore,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n}] = \mathbb{E}[\mathbb{1}_{A_1}] + \cdots + \mathbb{E}[\mathbb{1}_{A_n}] = n\,\mathbb{P}(A_1) = 1.$$

Remark 5. Notice that the events $A_1, \dots A_n$ are clearly not independent. For example, if $A_1, \dots, A_{n-1}$ were to happen then A_n must be true too since the only men and women left are husband and wife. The linearity of expectation allowed us to decompose the random variable into a sum of possibly dependent events. However, by symmetry we only needed to compute the probability of a single event A_1 in isolation without worrying about the other events $A_2, \dots, A_n$.

1.3.2 Derivations and Proofs

Problem 1.11. Prove the law of the unconscious statistician

Solution 1.11. This follows from a change of variables and a rearrangement of a sum. Given a function g, let

$$D_y = g^{-1}(\{y\}) = \{x : g(x) = y\}.$$

If we let $Y = g(X)$ then

$$f_Y(y) = \mathbb{P}(g(X) = y) = \sum_{x \in D_y} f_X(x).$$

and $Y(S) = g(X(S))$ since

$$Y(S) = \{Y(\omega) : \omega \in S\} = \{g(X(\omega)) : \omega \in S\}.$$

Therefore,

$$\mathbb{E}[g(X)] = \mathbb{E}[Y] = \sum_{y \in Y(S)} y f_Y(y) = \sum_{y \in Y(S)} y \sum_{x \in D_y} f_X(x) = \sum_{y \in Y(S)} \sum_{x \in D_y} g(x) f_X(x) = \sum_{x \in X(S)} g(x) f_X(x).$$

Since $(D_y)_{y \in Y(S)}$ forms a partition of $X(S)$.

Problem 1.12. Prove the linearity property of expectation

Solution 1.12. This follows form the linearity of summation. By the law of the unconscious statistician with $g(x) = ax + b$, we have

$$\mathbb{E}[aX+b] = \sum_{x \in X(S)} (ax+b) f_X(x) = a \sum_{x \in X(S)} x f_X(x) + b \sum_{x \in X(S)} f_X(x) = a\,\mathbb{E}[X] + b$$

since $\sum_{x \in X(S)} f_X(x) = 1$.

Problem 1.13. If $X \sim \mathrm{U}[a,b]$ then $\mathbb{E}[X] = \frac{a+b}{2}$.

Solution 1.13. If $X \sim \mathrm{U}[a,b]$ then the sum of positive integers implies

$$\mathbb{E}[X] = \sum_{x=a}^{b} \frac{x}{b-a+1} = \frac{1}{b-a+1}\left(\sum_{x=1}^{b} x - \sum_{x=1}^{a-1} x\right) = \frac{1}{b-a+1}\left(\frac{b(b+1)}{2} - \frac{a(a-1)}{2}\right)$$
$$= \frac{1}{b-a+1}\left(\frac{b^2 - a^2 + (b+a)}{2}\right) = \frac{a+b}{2}.$$

Alternative Solution: If $X \sim \mathrm{U}[0,n]$ then the sum of positive integers implies

$$\mathbb{E}[X] = \sum_{x=0}^{n} \frac{x}{n+1} = \frac{n(n+1)}{2(n+1)} = \frac{n}{2}.$$

Notice that $X \sim \mathrm{U}[a,b] \sim a + \mathrm{U}[0,b-a]$. Therefore, if we let $Y \sim \mathrm{U}[0,b-a]$, then $X = a + Y$ so linearity implies that

$$\mathbb{E}[X] = \mathbb{E}[a+Y] = a + \mathbb{E}[Y] = a + \frac{b-a}{2} = \frac{a+b}{2}.$$

Problem 1.14. $(*)$ If $X \sim \mathrm{Hyp}(N,r,n)$, then $\mathbb{E}[X] = r\frac{n}{N}$.

Solution 1.14. This proof uses a trick called the linearity of expectation. To simplify notation, suppose that we have r blue balls and $N-r$ red balls, then $X \sim \mathrm{Hyp}(N,r,n)$ denotes the number of blue balls we drew from a sample of n balls without replacement.

We label the blue balls $1,\dots,r$ and let A_i denote the event that the blue ball labeled i was drawn. Consider the random variable

$$\mathbb{1}_{A_i} = \begin{cases} 1 & \text{if we drew the blue ball labeled } i \\ 0 & \text{if we did not draw the blue ball labeled } i. \end{cases}$$

If $X \sim \mathrm{Hyp}(N,r,n)$ then $X = \sum_{i=1}^{r} \mathbb{1}(A_i)$ which is the total number of blue balls that we drew. We have by the linearity of expectation that

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{r} \mathbb{1}_{A_i}\right] = \sum_{i=1}^{r} \mathbb{E}[\mathbb{1}_{A_i}].$$

Next, for any i, we have

$$\mathbb{E}[\mathbb{1}_{A_i}] = 1 \cdot \mathbb{P}(A_1) + 0 \cdot (1 - \mathbb{P}(A_1)) = \mathbb{P}(\text{we drew the blue ball labeled } i).$$

By symmetry (we are not more likely to draw a particular ball over another one), for all $i \le r$

$$\mathbb{P}(A_i) = \mathbb{P}(A_1) = \mathbb{P}(\text{we drew the blue ball labeled } 1) = \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$$

so

$$\mathbb{E}[X] = \sum_{i=1}^{r} \mathbb{E}[\mathbb{1}_{A_i}] = r\,\mathbb{P}(A_1) = \frac{rn}{N}.$$

Problem 1.15. If $X \sim \text{Bin}(n,p)$ then $\mathbb{E}[X] = np$.

Solution 1.15. If $X \sim \text{Bin}(n,p)$ then $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ so

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{n} x \cdot \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \cdot \frac{n!}{(n-x)!x!} \cdot p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n \cdot (n-1)!}{((n-1)-(x-1))!(x-1)!} p \cdot p^{x-1} (1-p)^{n-1-(x-1)} && \text{add and subtract 1} \\
&= np \sum_{y=0}^{n-1} \frac{(n-1)!}{((n-1)-y)!y!} \cdot p^y (1-p)^{(n-1)-y} && \text{re-index sum} \\
&= np \underbrace{\sum_{y=0}^{n-1} \binom{n-1}{y} \cdot p^y (1-p)^{(n-1)-y}}_{=1 \text{sum of PMF of Bin}(n-1,p)} \\
&= np.
\end{aligned}$$

Alternative Solution: If $X \sim \text{Bern}(p)$ then

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1-p) = p.$$

Now suppose that $X \sim \text{Bin}(n,p)$. Since $X = X_1 + \cdots + X_n$ where X_i are independent and $X \sim \text{Bern}(p)$ (the number of successes in n trials is equal to the sum of n successful trials), linearity implies that

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \dots X_n] = n\,\mathbb{E}[X_1] = np.$$

Problem 1.16. $(*)$ If $X \sim \text{NegBin}(k,p)$, show that $\mathbb{E}[X] = \frac{k(1-p)}{p}$.

Solution 1.16. We first consider the geometric random variable. We will use many times throughout this derivation the identity

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$

for $|q| < 1$. If $X \sim \text{Geo}(p) \sim \text{NegBin}(1, p)$ then $f_X(x) = p(1-p)^x$ so

$$
\begin{aligned}
\mathbb{E}[X] = \sum_{x=0}^{\infty} xp(1-p)^x &= p\sum_{x=1}^{\infty}\sum_{k=1}^{x}(1-p)^x && 0 \cdot e^p(1-p)^0 = 0, \sum_{k=1}^{x} = x \\
&= p\sum_{k=1}^{\infty}\sum_{x=k}^{\infty}(1-p)^x && 1 \le k \le x < \infty \\
&= p\sum_{k=1}^{\infty}(1-p)^k\sum_{x=0}^{\infty}(1-p)^x && \sum_{x=k}^{\infty}(1-p)^x = (1-p)^k\sum_{x=0}^{\infty}(1-p)^x \\
&= p\sum_{k=1}^{\infty}\frac{(1-p)^k}{1-(1-p)} && \sum_{x=0}^{\infty}(1-p)^x = \frac{1}{1-(1-p)} \\
&= (1-p)\sum_{k=0}^{\infty}(1-p)^k && \sum_{x=1}^{\infty}(1-p)^x = (1-p)\sum_{x=0}^{\infty}(1-p)^x \\
&= \frac{1-p}{p}. && \sum_{x=0}^{\infty}(1-p)^x = \frac{1}{1-(1-p)}
\end{aligned}
$$

Now suppose that $X \sim \text{NegBin}(k, p)$. For $1 \le i \le k$, let X_i denote the number of fails between the $(i-1)$st success and the ith success. Since X_i counts the number of fails until the next success, we have $X_i \sim \text{Geo}(p)$ for all i. By definition, $X = X_1 + \dots + X_k$ since the total fails until k successes is equal to the sum of the number fails between successes, linearity implies that

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \dots X_k] = k\,\mathbb{E}[X_1] = \frac{k(1-p)}{p}.$$

Problem 1.17. If $X \sim \text{Poi}(\mu)$, show that $\mathbb{E}[X] = \mu$.

Solution 1.17. If $X \sim \text{Pois}(\mu)$ then $f_X(x) = e^{-\mu}\frac{\mu^x}{x!}$ so

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{\infty} x \cdot e^{-\mu}\frac{\mu^x}{x!} \\
&= \sum_{x=1}^{\infty} x \cdot e^{-\mu}\frac{\mu^x}{x!} && 0 \cdot e^{-\mu}\frac{\mu^0}{0!} = 0 \\
&= \mu\sum_{x=1}^{\infty} e^{-\mu}\frac{\mu^{x-1}}{(x-1)!} \\
&= \mu \underbrace{\sum_{y=0}^{\infty} e^{-\mu}\frac{\mu^y}{y!}}_{=1 \text{ sum of PMF of Poi}(\mu)} && \text{re-index sum} \\
&= \mu.
\end{aligned}
$$