1 Expected Value and Variance

In this section, we will introduce two summary statistics which will encode the typical value and the spread of the random variable.

1.1 Expected Value

The first key theoretical value is the "typical value" of the random variable, also known as the *expected value*, *first moment*, or *mean*. It is given by the *weighted average* of the random variable against its density.

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

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

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

The expected value can be interpreted as the integral of the random variable with respect to its probability distribution.

Remark 1. The average of many realizations of X converges to $\mathbb{E}[X]$. This is called the *law of large numbers*. The expected value is also the number that minimizes the squared error (see Problem 1.13).

1.1.1 Properties

1. Law of the Unconscious Statistician: If $g : \mathbb{R} \to \mathbb{R}$, and X is a random variable with PMF p_X, then $g(X)$ is a random variable taking values $g(X(\Omega))$ and
$$\mathbb{E}[g(X)] = \sum_{x \in X(\Omega)} g(x) p_X(x)$$
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
$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$
4. Jensen's Inequality: If g is convex then
$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]).$$
5. Equivalent Formula: If $X \geq 0$, then
$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq t)\, dt.$$

Remark 2. In general $\mathbb{E}[g(X)] \neq g[\mathbb{E}(X)]$. However, if g is a linear function or X is a constant random variable then equality holds.

Remark 3. In general $\mathbb{E}[XY] \neq \mathbb{E}[X]\,\mathbb{E}[Y]$. However, if X and Y are independent then equality holds.

1.2 Variance

The second key theoretical value is the "deviations" of the random variable from its expected value.

Definition 2 (Variance)**.** The *variance* of X, denoted by $\mathrm{Var}[X]$, is the non-negative number

$$\mathrm{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

The variance is in the squared units, so the standard deviation defined by

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}[X]}$$

measures the deviation in the original units.

1.2.1 Properties

1. Equivalent formula I:
$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$
2. Equivalent formula II:
$$\mathrm{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2$$
3. Variance of Linear Functions: For any constants $a, b \in \mathbb{R}$,
$$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$$
4. Variance of Linear Functions II: If X and Y are independent then
$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$
5. Zero Variance: Suppose a random variable X has $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = 0$. This means, X does not "vary" from its mean at all, and is constant with probability 1

   Theorem 1

   $\mathrm{Var}(X) = 0$ *if and only if* $\mathbb{P}(X = \mathbb{E}[X]) = 1$.

Remark 4. The simple formulas to compute the variance is one of the main reasons why we use the squared deviations in the definition of the variance over other measures of deviation. We will see later that squared deviations match our natural ways of measuring distances, so it will allow us to build some geometric intuition of results.

1.3 Expected Value and Variance of Common Distributions

Distribution	**PMF / PDF**	**Mean**	**Variance**
$\mathrm{DUnif}(a,b)$	$\frac{1}{b-a+1}, \quad x = \{a, a+1, \dots, b\}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
$\mathrm{Bern}(p)$	$p^x(1-p)^{1-x}, \quad x \in \{0,1\}$	p	$p(1-p)$
$\mathrm{Bin}(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}, \quad x = 0, 1, \dots, n$	np	$np(1-p)$
$\mathrm{Geo}(p)$	$(1-p)^{x-1}p, \quad x = 1, 2, \dots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
$\mathrm{NegBin}(k,p)$	$\binom{x+k-1}{x}p^k(1-p)^x, \quad x = 0, 1, 2, \dots$	$\frac{k(1-p)}{p}$	$\frac{k(1-p)}{p^2}$
$\mathrm{Poi}(\lambda)$	$\dfrac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots$	λ	λ
$\mathrm{Hyp}(N,r,n)$	$\dfrac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \ \max\{0, n-(N-r)\} \le x \le \min\{r, n\}$	$r\frac{n}{N}$	$n\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$

1.4 Example Problems

Problem 1.1. Consider a random variable X with PMF

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

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 $\omega \in \Omega$. Show that $a \leq \mathbb{E}[X] \leq b$.

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

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

and

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

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
$$p_X(-1) = 0.15, \quad p_X(0) = 2c, \quad p_X(1) = 0.5, \quad p_X(2) = 0.05, \quad p_X(3) = c.$$
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
$$1 = \sum_{x=-1}^{3} p_X(i) = 3c + 0.7 \implies c = 0.1$$

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, $X \sim \mathrm{Bin}(24, \frac{1}{6})$ 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

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

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 5. We used the fact that if X and Y are independent then $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$, 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

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

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

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

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 $X \sim \mathrm{Poi}(30 \cdot 60)$. 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 $p_X(x) = \frac{1}{x}$ for $x = 2, 4, 8, 16, \dots$ and 0 otherwise.

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

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

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

2. For the expected value, we have

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

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

Problem 1.10. Consider the random variables

- X is a r.v. representing the outcome of one fair 6-sided die roll
- Y is a r.v. representing the number of phone calls over 1 minute at Lenovo call centre, with the rate of 3.5 calls per minute

Compute the mean and variance of X and Y.

Solution 1.10. We have that $X \sim \mathrm{DUnif}[1,6]$ and $Y \sim \mathrm{Poi}(3.5)$. Using the formulas for mean and variance, we have

$$\mathbb{E}[X] = 3.5, \ \operatorname{Var}(X) = \frac{6^2 - 1}{12} \approx 2.9$$

while

$$\mathbb{E}[Y] = 3.5, \ \operatorname{Var}(Y) = 3.5.$$

This makes intuitive sense because both X and Y have the same mean, but the fact that Y can take values on $\mathbb{N}$ while X can only take values on $\{1, 2, \dots, 6\}$, so Y should have larger variance even though both random variables have the same mean.

Problem 1.11. Suppose a fair coin is flipped 1,000 times, and let X denote the number of heads observed. What is the standard deviation of X?

Solution 1.11. We have $X \sim \mathrm{Bin}(1000, 0.5)$, so

$$\mathrm{SD}(X) = \sqrt{\operatorname{Var}(X)} = \sqrt{1000(0.5)(1-0.5)} \approx 15.81.$$

Problem 1.12. Suppose that X has variance $\operatorname{Var}(X) = 2$. Compute the variance of Y, where $Y = -2X + 3$.

Solution 1.12. By the variance of linear maps,

$$\operatorname{Var}(Y) = \operatorname{Var}(-2X + 3) = (-2)^2 \operatorname{Var}(X) = 8.$$

1.5 Proofs of Key Results

Problem 1.13. For any constant c, show that

$$\mathbb{E}[(X-c)^2] \geq \mathbb{E}[(X - \mathbb{E}[X])^2].$$

In particular, the expected value is the constant that minimizes the mean squared error. Furthermore, the value of the mean squared error is given by the variance.

Solution 1.13. This proof follows directly from the properties of the expected value. We have

$$\mathbb{E}[(X-c)^2] = \mathbb{E}[(X - \mathbb{E}[X] + \mathbb{E}[X] - c)^2] = \mathbb{E}[X^2] - 2c\,\mathbb{E}[X] + c^2.$$

Therefore,

$$\frac{d}{dc}\,\mathbb{E}[(X-c)^2] = -2\,\mathbb{E}[X] + 2c.$$

The critical point of this function is attained when $c = \mathbb{E}[X]$, and this point is a global minimum since $c \mapsto \mathbb{E}[(X-c)^2]$ is an upward facing parabola.

Alternative Proof: We can also do this proof without calculus. This proof follows directly from the properties of the expected value. By adding and subtracting $\mathbb{E}[X]$, we see that

$$\begin{aligned}\mathbb{E}[(X-c)^2] &= \mathbb{E}[(X - \mathbb{E}[X] + \mathbb{E}[X] - c)^2] \\ &= \mathbb{E}[(X - \mathbb{E}[X])^2] + \mathbb{E}[(\mathbb{E}[X] - c)^2] + 2\,\mathbb{E}[(X - \mathbb{E}[X])(\mathbb{E}[X] - c)]\end{aligned}$$

Since $\mathbb{E}[X] - c$ is not random, we see that the cross terms vanish

$$\mathbb{E}[(X - \mathbb{E}[X])(\mathbb{E}[X] - c)] = (\mathbb{E}[X] - c)\,\mathbb{E}[(X - \mathbb{E}[X])] = (\mathbb{E}[X] - c)(\mathbb{E}[X] - \mathbb{E}[X]) = 0.$$

Since $\mathbb{E}[(\mathbb{E}[X] - c)^2] \geq 0$, we conclude that

$$\mathbb{E}[(X - c)^2] = \mathbb{E}[(X - \mathbb{E}[X])^2] + \mathbb{E}[(\mathbb{E}[X] - c)^2] \geq \mathbb{E}[(X - \mathbb{E}[X])^2]$$

as required.

Problem 1.14. Prove the law of the unconscious statistician

Solution 1.14. 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

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

and $Y(\Omega) = g(X(\Omega))$ since

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

Therefore,

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

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

Problem 1.15. Prove the linearity property of expectation

Solution 1.15. This follows form the linearity of summation. By the law of the unconscious statistician with $g(x) = ax + b$, we have

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

since $\sum_{x \in X(\Omega)} p_X(x) = 1$.

Problem 1.16. Prove the equivalent formula for the expected value when X is non-negative.

Solution 1.16. If $x \geq 0$, we can write it as

$$x = \int_0^x dt = \int_0^\infty \mathbb{1}(t \leq x)\, dt.$$

Using this fact in the definition of the expected value gives us

$$\begin{aligned}
\mathbb{E}[X] = \sum_{x \in X(\Omega)} x\,\mathbb{P}(X = x) &= \sum_{x \in X(\Omega)} \int_0^\infty \mathbb{1}(t \leq x)\,\mathbb{P}(X = x)\,dt \\
&= \int_0^\infty \sum_{x \in X(\Omega)} \mathbb{1}(t \leq x)\,\mathbb{P}(X = x)\,dt \\
&= \int_0^\infty \sum_{x \in X(\Omega): x \geq t} \mathbb{P}(X = x)\,dt \\
&= \int_0^\infty \mathbb{P}(X \geq t)\,dt.
\end{aligned}$$

The interchange of the sum and integral can always be done if $X(\Omega)$ is finite. If $|X(\Omega)|$ is infinite, then it is also true and it can be justified by the monotone convergence theorem.

Problem 1.17. Show that

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2.$$

Solution 1.17. To simplify notation, we define $\mathbb{E}[X] = \mu$. Then by the linearity of expectation,

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] = \mathbb{E}[X^2] - 2\mu\,\mathbb{E}[X] + \mu^2 = \mathbb{E}[X^2] - \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

To conclude the second equality, notice that

$$\mathbb{E}[X(X-1)] = \mathbb{E}[X^2 - X] = \mathbb{E}[X^2] - \mathbb{E}[X] \implies \mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X]$$

so substituting this into the formula above implies

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2.$$

Problem 1.18. Show that for any constants a and b,

$$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X).$$

Solution 1.18. We can use the definition of the variance. Let $Y = aX + b$,

$$\begin{aligned}
\mathrm{Var}(aX+b) = \mathrm{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] &= \mathbb{E}[(aX + b - \mathbb{E}[aX+b])^2] \\
\mathbb{E}[aX+b] = a\,\mathbb{E}[X] + b \qquad &= \mathbb{E}[(aX - a\,\mathbb{E}[X])^2] \\
&= a^2\,\mathbb{E}[(X - \mathbb{E}[X])^2] = a^2\,\mathrm{Var}(X).
\end{aligned}$$

Remark 6. This makes intuitive sense because shifting a random variable by b does not change the spread of the random variables. However, scaling the random variable by a will change the spread by a factor of a^2 since we are measuring the squared deviations, so the scaling factor is squared.

Problem 1.19. Prove Theorem 1.

Solution 1.19. Let $\mathbb{E}[X] = \mu$.

($\implies$) Suppose $\mathbb{P}(X = \mu) = 1$. In other words, $p_X(\mu) = 1$ and there are no other non-zero values of the PMF, so by definition of the variance,

$$\mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mu^2\,\mathbb{P}(X=\mu) - (\mu\,\mathbb{P}(X=\mu))^2 = 0.$$

($\impliedby$) Conversely, suppose $Var(X) = 0$. Then, again by definition of the variance,

$$0 = \mathrm{Var}(X) = \mathbb{E}((X-\mu)^2) = \sum_{\text{all } x} \underbrace{(x-\mu)^2}_{\geq 0}\underbrace{\mathbb{P}(X=x)}_{=p_X(x)\geq 0}.$$

Suppose for the sake of contradiction that there exists a $\nu \neq \mu$ such that $\mathbb{P}(X = \nu) > 0$. In this case, we have

$$\mathrm{Var}(X) \geq (\nu-\mu)^2\,\mathbb{P}(X=\nu) > 0$$

which contradicts the fact that $\mathrm{Var}(X) = 0$. Therefore, we must have $\mathbb{P}(X = \mu) = 1$.

Problem 1.20. If $X \sim \mathrm{DUnif}[a, b]$ then $\mathbb{E}[X] = \frac{a+b}{2}$.

Solution 1.20. If $X \sim \mathrm{DUnif}[a, b]$ then the sum of positive integers implies

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

Alternative Solution: If $X \sim \mathrm{DUnif}[0, n]$ 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 $X \sim \mathrm{DUnif}[a, b] \sim a + \mathrm{DUnif}[0, b-a]$. Therefore, if we let $Y \sim \mathrm{DUnif}[0, b-a]$, 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 1.21. If $X \sim \mathrm{Hyp}(N, r, n)$, then $\mathbb{E}[X] = r\frac{n}{N}$.

Solution 1.21. 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 $X \sim \text{Hyp}(N, r, n)$ 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 1.22. If $X \sim \text{Bin}(n, p)$ then $\mathbb{E}[X] = np$.

Solution 1.22. If $X \sim \text{Bin}(n, p)$ then $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ 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 1.23. If $X \sim \text{NegBin}(k, p)$, show that $\mathbb{E}[X] = \frac{k(1-p)}{p}$.

Solution 1.23. 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 $p_X(x) = p(1-p)^x$ so

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

Now suppose that $X \sim \text{NegBin}(k,p)$. For $1 \leq i \leq k$, 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, $X = X_1 + \cdots + X_k$ since the total fails until k successes is equal to the sum of the number fails between successes, linearity implies that

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

Problem 1.24. If $X \sim \text{Poi}(\lambda)$, show that $\mathbb{E}[X] = \lambda$.

Solution 1.24. If $X \sim \text{Poi}(\lambda)$ then $p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!}$ so

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

Problem 1.25. If $X \sim \text{DUnif}[a,b]$ then $\text{Var}(X) = \frac{(b-a+1)^2-1}{12}$.

Solution 1.25. If $X \sim \text{DUnif}[0,n]$ then the sum of positive integers implies

$$\mathbb{E}[X^2] = \sum_{x=0}^{n} \frac{x}{n+1} = \frac{n(n+1)}{2(n+1)} = \frac{n}{2}, \quad \mathbb{E}[X^2] = \sum_{x=0}^{n} \frac{x^2}{n+1} = \frac{n(n+1)(2n+1)}{6(n+1)} = \frac{n(2n+1)}{6}$$

so

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{n(2n+1)}{6} - \frac{n^2}{4} = \frac{(n+1)^2-1}{12}.$$

Notice that $X \sim \text{DUnif}[a,b] \sim a + \text{DUnif}[0, b-a]$. Therefore, if we let $Y \sim \text{DUnif}[0, b-a]$, then $X = a + Y$ so the variance of a linear function implies that

$$\text{Var}(X) = \text{Var}(a+Y) = \text{Var}(Y) = \frac{(b-a+1)^2-1}{12}.$$

Problem 1.26. If $X \sim \text{Hyp}(N,r,n)$, then $\text{Var}(X) = n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$.

Solution 1.26. 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 $X \sim \text{Hyp}(N,r,n)$ denotes the number of blue balls we drew from a sample of n balls without replacement. We first compute $\mathbb{E}[X(X-1)]$. It suffices to compute

$$\mathbb{E}\left[\binom{X}{2}\right] = \mathbb{E}\left[\frac{X(X-1)}{2}\right]$$

which denotes the expected value of the number of pairs of blue balls we drew.

We label the successful balls $1,\dots,r$ and let A_{ij} denote the event that the pair of balls labeled i and j was drawn where. Consider the random variable

$$\mathbb{1}_{A_{ij}} = \begin{cases} 1 & \text{if we drew the pair of blue balls } i \text{ and } j \\ 0 & \text{if we did not draw the pair of blue balls } i \text{ and } j. \end{cases}$$

If $X \sim \text{Hyp}(N,r,n)$ then $\binom{X}{2} = \sum_{1\le i<j\le r} \mathbb{1}_{A_{ij}}$ which is the total number of pairs of blue balls that we drew. We have by the linearity of expectation that

$$\mathbb{E}\left[\binom{X}{2}\right] = \mathbb{E}\left[\sum_{1\le i<j\le r} \mathbb{1}_{A_{ij}}\right] = \sum_{1\le i<j\le r} \mathbb{E}[\mathbb{1}_{A_{ij}}].$$

Next, for any i,j, we have

$$\mathbb{E}[\mathbb{1}_{A_{ij}}] = 1\,\mathbb{P}(A_{ij}) + 0\cdot(1-\mathbb{P}(A_{ij})) = \mathbb{P}(A_{ij}).$$

By symmetry (we are not more likely to draw a particular ball over another one), for all $1 \le i < j \le r$

$$\mathbb{P}(A_{ij}) = \mathbb{P}(A_{12}) = \mathbb{P}(\text{ we drew the pair of blue balls } i \text{ and } j\ ) = \frac{\binom{2}{2}\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

so

$$\mathbb{E}\left[\binom{X}{2}\right] = \sum_{1\le i<j\le r} \mathbb{E}[\mathbb{1}(A_{ij})] = \frac{r(r-1)}{2}\,\mathbb{P}(A_{12}) = \frac{r(r-1)}{2}\,\frac{n(n-1)}{N(N-1)}.$$

Multiplying by two implies that

$$\mathbb{E}[X(X-1)] = 2\,\mathbb{E}\left[\binom{X}{2}\right] = 2\,\mathbb{E}\left[\frac{X(X-1)}{2}\right] = \frac{r(r-1)n(n-1)}{N(N-1)}.$$

Since we know that $\mathbb{E}[X] = r\frac{n}{N}$ we have

$$\mathrm{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \frac{r(r-1)n(n-1)}{N(N-1)} + \frac{rn}{N} - \frac{r^2n^2}{N^2} = n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Remark 7. We can set $p = \frac{r}{N}$ which denotes the probability of a successful draw. If $X \sim \mathrm{Hyp}(N, r, n)$ and $Y \sim \mathrm{Bin}(n, p)$ then

$$\mathbb{E}[X] = np \qquad \mathrm{Var}(X) = np(1-p)\left(\frac{N-n}{N-1}\right).$$

and

$$\mathbb{E}[Y] = np \qquad \mathrm{Var}(Y) = np(1-p).$$

The hypergeometric and binomial random variable have the same mean, and they have the same variance except for a factor $\frac{N-n}{N-1} \leq 1$. The variance of the hypergeometric is slightly less because sampling without replacement reduces the "spread" since our sample space shrinks with each draw.

When $n = 1$ then the factor $\frac{N-n}{N-1} = 1$, so the variance of a hypergeometric and binomial random variables are the same, since sampling one object with or without replacement is the same. When $N \to \infty$ then the factor $\frac{N-n}{N-1} = 1$, which is again consistent because sampling with or without replacement from a large population is essentially the same.

Problem 1.27. If $X \sim \mathrm{Bin}(n, p)$ then $\mathrm{Var}(X) = np(1-p)$.

Solution 1.27. If $X \sim \mathrm{Bin}(n, p)$ then $p_X(x) = \binom{n}{x}p^x(1-p)^{n-x}$. We use the formula

$$\mathrm{Var}(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2$$

and note we already know $\mathbb{E}(X) = np$. By definition

$$\begin{aligned}
\mathbb{E}(X(X-1)) &= \sum_{x=0}^{n} x(x-1)\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(n-2-(x-2))!(x-2)!}p^{x-2}(1-p)^{n-2-(x-2)} \\
&= n(n-1)p^2 \underbrace{\sum_{y=0}^{n-2} \frac{(n-2)!}{(n-2-y)!y!}p^y(1-p)^{n-2-y}}_{=1 \text{ sum of PMF of } (n-2,p)} \\
&= n(n-1)p^2.
\end{aligned}$$

Then

$$\mathrm{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

Alternative Solution: If $X \sim \mathrm{Bern}(p)$ then by definition

$$\mathrm{Var}(X) = \mathbb{E}[(X-p)^2] = (1-p)^2p + (-p)^2(1-p) = p(1-p).$$

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

$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + \ldots X_n) = n\,\mathrm{Var}(X_1) = np(1-p).$$

Problem 1.28. If $X \sim \text{NegBin}(k,p)$, show that $\text{Var}(X) = \frac{k(1-p)}{p^2}$.

Solution 1.28. 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 $|q| < 1$. From this identity, we can recover higher order versions of this by differentiating the power series term by term with respect to q, that is

$$\frac{d}{dq}\sum_{k=0}^{\infty} q^k = \frac{d}{dq}\frac{1}{1-q} \implies \sum_{k=1}^{\infty} kq^{k-1} = \frac{1}{(1-q)^2}$$

and

$$\frac{d^2}{dq^2}\sum_{k=0}^{\infty} q^k = \frac{d}{dq^2}\frac{1}{1-q} \implies \sum_{k=2}^{\infty} k(k-1)q^{k-2} = \frac{2}{(1-q)^3}$$

We will use these identities to give a different derivative of the first and second moments of a geometric random variable.

If $X \sim \text{Geo}(p) \sim \text{NegBin}(1,p)$ then $p_X(x) = p(1-p)^x$ so the first derivative identity implies that

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} xp(1-p)^x = \sum_{x=1}^{\infty} xp(1-p)^x = p(1-p)\sum_{x=1}^{\infty} x(1-p)^{x-1} = \frac{p(1-p)}{(1-(1-p))^2} = \frac{(1-p)}{p}.$$

Likewise, the second derivative identity implies that

$$\mathbb{E}[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)p(1-p)^x = p(1-p)^2\sum_{x=2}^{\infty} x(x-1)(1-p)^{x-2} = \frac{2p(1-p)^2}{(1-(1-p))^3} = \frac{2(1-p)^2}{p^2}.$$

Therefore,

$$\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 = \frac{2(1-p)^2}{p^2} + \frac{(1-p)}{p} - \frac{(1-p)^2}{p^2} = \frac{(1-p)}{p^2}.$$

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 and the X_i are independent. 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

$$\text{Var}(X) = \text{Var}(X_1 + \dots X_k) = k\,\text{Var}(X_1) = \frac{k(1-p)}{p^2}.$$

Problem 1.29. If $X \sim \text{Poi}(\lambda)$, show that $\text{Var}(X) = \lambda$.

Solution 1.29. If $X \sim \text{Poi}(\lambda)$ then $p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$. We use the formula

$$\text{Var}(X) = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2$$

and note we already know $\mathbb{E}(X) = \lambda$. By definition

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1) \cdot e^{-\lambda} \frac{\lambda^x}{x!} \\
&= \sum_{x=2}^{\infty} x(x-1) \cdot e^{-\lambda} \frac{\lambda^x}{x!} && 0 \cdot e^{-\lambda} \frac{\lambda^0}{0!} = 0 \\
&= \lambda^2 \sum_{x=2}^{\infty} e^{-\lambda} \frac{\lambda^{x-2}}{(x-2)!} \\
&= \lambda^2 \underbrace{\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!}}_{=1 \text{ sum of PMF of Poi}(\lambda)} && \text{re-index sum} \\
&= \lambda^2.
\end{aligned}$$

Then

$$\mathrm{Var}(X) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - (\mathbb{E}(X))^2 = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$