1 Variance

1.1 Summarizing Random Variables - Variance

The expected value is insufficient to capture all the interesting behavior of a random variable. The second key theoretical value is the "deviations" of the random variable from its expected value. There are several ways to measure this deviation. Let $\mathbb{E}[X] = \mu$.

1. **Deviation**
$$\mathbb{E}[(X-\mu)] = \mathbb{E}[X] - \mu = 0$$
2. **Absolute deviation**
$$\mathbb{E}\left[|X-\mu|\right]$$
3. **Squared deviation**
$$\mathbb{E}\left[(X-\mu)^2\right]$$

Definition 1. 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.

1.1.1 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* $\mathbb{P}(X = \mathbb{E}[X]) = 1$.

1.1.2 Variance of Common Distributions

1. **Uniform:** If $X \sim \mathrm{U}[a,b]$ then $\mathrm{Var}(X) = \frac{(b-a+1)^2-1}{12}$.
2. **Hypergeometric:** If $X \sim \mathrm{Hyp}(N,r,n)$, then $\mathrm{Var}(X) = n\frac{r}{N}\left(1-\frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$.
3. **Binomial:** If $X \sim \mathrm{Bin}(n,p)$ then $\mathrm{Var}(X) = np(1-p)$.
4. **Negative Binomial:** If $X \sim \mathrm{NegBin}(k,p)$, then $\mathrm{Var}(X) = \frac{k(1-p)}{p^2}$.
5. **Poisson:** If $X \sim \mathrm{Poi}(\mu)$, then $\mathrm{Var}(X) = \mu$.

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 $\mathrm{Var}(X) = p(1-p)$.
2. **Geometric:** If $X \sim \mathrm{Geo}(p)$ then $\mathrm{Var}(X) = \frac{1-p}{p^2}$.

1.2 Higher Order Moments

The expectation and the variance give a simple summary of the distribution giving the center and variability of the distribution. The generalizations of these concepts give more summaries of the behavior of distributions

- **Moments:** The kth moment of the distribution of X, is $E[X^k]$.
- **Central Moments:** The kth central moment of the distribution of X is $\mathbb{E}[(X-\mathbb{E}(X))^k]$.
- There are also other statistics such as
  - *Skewness* (measures asymmetry)
  $$\mathbb{E}\left[\left(\frac{(X-\mathbb{E}(X))}{\sqrt{\mathrm{Var}(X)}}\right)^3\right]$$
  - *Kurtosis* (measures heavy tailedness)
  $$\mathbb{E}\left[\left(\frac{(X-\mathbb{E}(X))}{\sqrt{\mathrm{Var}(X)}}\right)^4\right]$$

1.3 Example Problems

1.3.1 Applications

Problem 1.1. 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 1.1. Using the formulas for mean and variance, we have

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

while

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

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 $\{1, 2, \ldots, 6\}$, so Y should have larger variance even though both random variables have the same mean.

Problem 1.2. 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 1.2. We have $X \sim \mathrm{Bin}(1000, 0.5)$, so

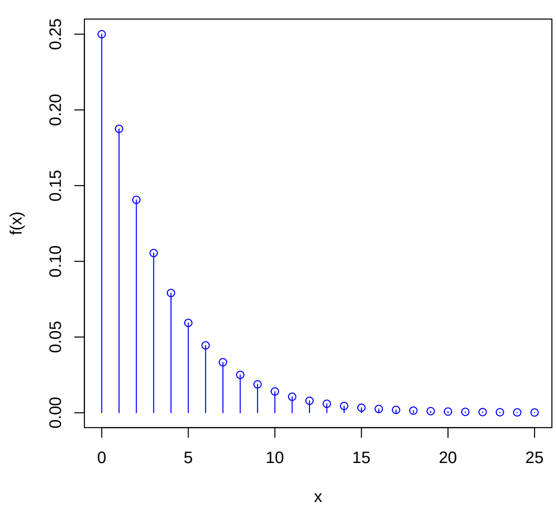
$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)} = \sqrt{1000(0.5)(1 - 0.5)} \approx 15.81.$$

Problem 1.3. Suppose that X has variance $\mathrm{Var}(X) = 2$. Compute the variance of Y, where $Y = -2X + 3$.

Solution 1.3. By the variance of linear maps,

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

Problem 1.4. Which PMF is the figure most likely showing?



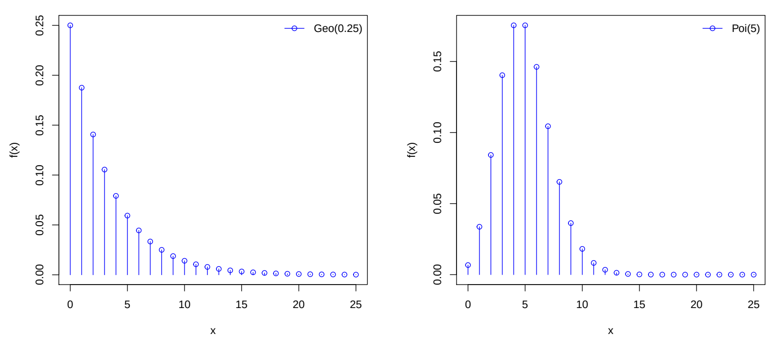
1. Poi(1)
2. Geo(0.25)

3. $\text{NegBin}(5, 0.75)$
4. $\text{Bin}(25, 0.25)$

Solution 1.4. The PDF is strictly decreasing so it cannot be negative binomial (for $k \geq 2$) or binomial since the binomial coefficient means the PMFs have a bump. The $\text{Poi}(\mu)$ distribution is strictly decreasing for $\mu < 1$ and has a bump for $\mu > 1$. When $\mu = 0$ then it turns out that $\mathbb{P}(X = 0) = \mathbb{P}(X = 1)$. This means that the $\text{Geo}(0.25)$ is the only possibility.

Alternative Solution: We can also look at $f_X(0)$ which is 0.25 to determine if it is Poisson or geometric random variable. If $X \sim \text{Geo}(0.25)$ then $f_X(0) = 0.25$. While if $X \sim \text{Poi}(1)$ then $f_X(0) = e^{-1} \approx 0.36$, so the PMF is the one of the Geometric random variable.

Problem 1.5. Consider the following PMF



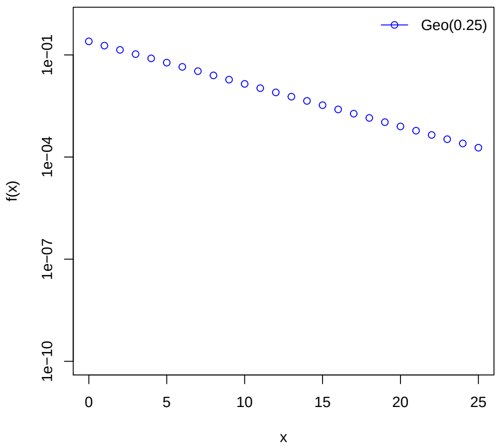
Which is TRUE?

1. $\mathbb{P}(X = 25)$ is much larger for the distribution on the left than for the distribution on the right.
2. $\mathbb{P}(X = 25)$ is much larger for the distribution on the right than for the distribution on the left.
3. $\mathbb{P}(X = 25)$ is about the same for the distributions on the left and on the right.

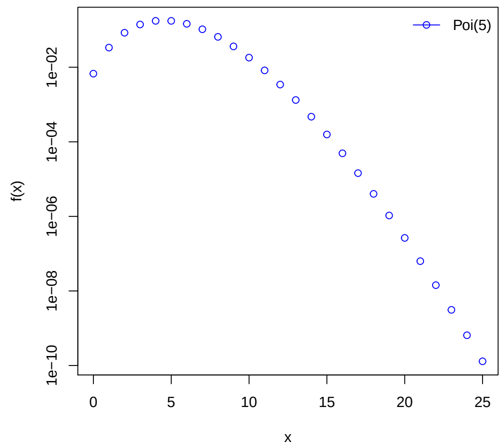
Solution 1.5. It is hard to see from the picture, but we can explicitly compute the probabilities. Let $X \sim \text{Geo}(0.25)$ and $Y \sim \text{Poi}(5)$.

$$f_X(0.25) = 0.25(1 - 0.25)^{25} \approx 0.0002 \qquad f_Y(25) = e^{-5}\frac{5^{25}}{25!} \approx 1.3 \times 10^{-10}.$$

The PMFs on the log scale is shown below



Probability function of Geo(0.25) with log−scale on y axis



Probability function of Poi(5) with log−scale on y axis

1.3.2 Derivations and Proofs

Problem 1.6. Show that

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

Solution 1.6. To simplify notation, we define $\mathbb{E}[X] = \mu$. Then by the linearity of expectation,

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

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

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

Problem 1.7. Show that for any constants a and b,

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

Solution 1.7. We can use the definition of the variance. Let $Y = aX + b$,

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

Remark 1. 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 1.8. Prove Theorem 1.

Solution 1.8. Let $\mathbb{E}[X] = \mu$.

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

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

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

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

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 1.9. If $X \sim \mathrm{U}[a,b]$ then $\mathbb{E}[X] = \frac{(b-a+1)^2-1}{12}$.

Solution 1.9. If $X \sim \mathrm{U}[0,n]$ 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

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

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

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

Problem 1.10. $(*)$ If $X \sim \mathrm{Hyp}(N, r, n)$, then $\mathrm{Var}(X) = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$.

Solution 1.10. 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 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 $1, \ldots, r$ 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

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

Remark 2. We can set $p = \frac{r}{N}$ which denotes the probability of a successful draw. If $X \sim \text{Hyp}(N, r, n)$ and $Y \sim \text{Bin}(n, p)$ then

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

and

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

The hypergeometric and binomial random variable have the same mean, and they have the same variance except for a factor $\frac{N-n}{N-1} \le 1$. 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 1.11. If $X \sim \text{Bin}(n, p)$ then $\text{Var}(X) = np(1-p)$.

Solution 1.11. If $X \sim \text{Bin}(n, p)$ then $f_X(x) = \binom{n}{x}p^x(1-p)^{n-x}$. We use the formula

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

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

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

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

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

Problem 1.12. ($*$) If $X \sim \mathrm{NegBin}(k,p)$, show that $\mathrm{Var}(X) = \frac{k(1-p)}{p^2}$.

Solution 1.12. 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 $X \sim \mathrm{Geo}(p) \sim \mathrm{NegBin}(1,p)$ then $f_X(x) = p(1-p)^x$ 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,

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

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, $X = X_1 + \cdots + X_k$ since the total fails until k successes is equal to the sum of the number fails between successes, linearity implies that

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

Problem 1.13. If $X \sim \text{Poi}(\mu)$, show that $\operatorname{Var}(X) = \mu$.

Solution 1.13. If $X \sim \text{Pois}(\mu)$ then $f_X(x) = e^{-\mu}\frac{\mu^x}{x!}$. We use the formula

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

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

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

Then

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

2 Continuous Random Variables

Recall the two "informal" classifications of random variables we consider in this course.

- We say that a random variable is *discrete* if its range is a discrete subset of $\mathbb{R}$ (i.e., a finite or a countably infinite set).
- A random variable is *continuous* if its range is an interval that is a subset of $\mathbb{R}$ (e.g. $[0,1], (0,\infty), \mathbb{R}$).

The definition of the CDF of discrete and continuous random variables are identical.

Definition 2. The *cumulative distribution function* (CDF) of a random variable X is

$$F_X(x) = \mathbb{P}(X \leq x) := \mathbb{P}(\{\omega \in S : X(\omega) \leq x\}), \quad x \in \mathbb{R}.$$

From the point of CDFs, we have the following natural classifications of random variables.

Definition 3. If the CDF of X is

1. a *piecewise constant* function, then X is a *discrete* random variable.
2. a *continuous* function, then X is a *continuous* random variable.

Remark 3. It is possible that a CDF does not fall under either of these categories, such as mixed random variables with have CDFs with both jump discontinuities and strictly increasing parts.

Since the intervals $(a,b]$ and $(-\infty,a]$ are disjoint,

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F_X(b) - F_X(a).$$

If X is a continuous random variable, then for any $x \in \mathbb{R}$,

$$\mathbb{P}(X = x) = \lim_{\epsilon \to 0} \mathbb{P}(x - \epsilon < X \leq x) = F_X(x) - F_X(x^-) = 0$$

by continuity of F_X. This implies that the inequalities don't matter for continuous random variables

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a).$$

Remark 4. If X is discrete, then the inequalities matter, so

$$\mathbb{P}(a < X \leq b),\ \ \mathbb{P}(a < X < b),\ \ \mathbb{P}(a \leq X < b),\ \ \mathbb{P}(a \leq X \leq b)$$

can be different since $\mathbb{P}(X = a)$ or $\mathbb{P}(X = b)$ may be non-zero.

2.1 Probability Density Function (PDF)

We can try to define the PMF for a continuous random variable, but the fact that $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$ is problematic. We get around this by defining a new way to encode how likely a certain value x is without defining it as a probability.

Definition 4. We say that a continuous random variable X with distribution function F_X is *absolutely continuous* if it is the antiderivative of some function f_X,

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

Remark 5. It might be the case that $F_X(x)$ is not differentiable at isolated points. However, if $f_X(x)$ is continuous at x, then

$$F'_X(x) = \frac{d}{dx}F_X(x) = f_X(x).$$

Likewise, jump discontinuities of f_X correspond to non-differentiable points of F_X.

Definition 5. The function f_X is called the *probability density function* (PDF) $f_X(x)$ and it satisfies the following properties

1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$;
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$;

The *support* of f_X (or X) is the closure of the set of non-zero values of the PDF,

$$\operatorname{supp}(f_X) = \operatorname{cl}(\{x \in \mathbb{R} : f_X(x) \neq 0\}).$$

Remark 6. By convention we take the closure of the set, which means that we always include the endpoints of intervals in the support. This is not a big issue since $f_X(x)$ can be arbitrarily defined at the endpoints since the area under the PDF does not change if we redefine the endpoints.

The PDF $f_X(x)$ is proportional to the probability that X lies in a small interval around x in the sense that for ϵ small

$$\mathbb{P}\left(x - \frac{\epsilon}{2} \leq X \leq x + \frac{\epsilon}{2}\right) = \int_{x-\frac{\epsilon}{2}}^{x+\frac{\epsilon}{2}} f_X(x)\,dt \approx \epsilon f_X(x).$$

So $f_X(x)$ isn't the probability that $X = x$, but it encodes the likelihood of x compared to other values.

Remark 7. In this course, unless otherwise stated, all continuous random variables will be absolutely continuous so we often drop the prefix "absolutely".

2.2 Expected Value and Variance

The expected value and variance of a random variable can be defined analogously to the discrete random variables where the sum is now replaced by an integral.

Definition 6. If X is a continuous random variable with PDF $f_X(x)$, and $g : \mathbb{R} \to \mathbb{R}$ is a function, then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)\,dx,$$

provided the expression exists.

It follows that for continuous random variables,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf_X(x)dx$$

and

$$\operatorname{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f_X(x)dx.$$

This is analogous to how the expected value and variance of discrete random variables were defined, but the sum over the PMF is replaced with an integral over the PDF.

2.3 Example Problems

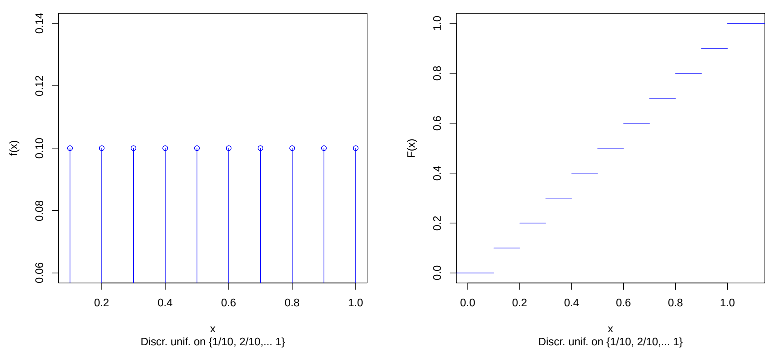
2.3.1 Applications

Problem 2.1. Suppose we are cutting a stick of length 1 randomly and denote by X the cutting point. The random variable X is continuous with range $[0, 1]$.

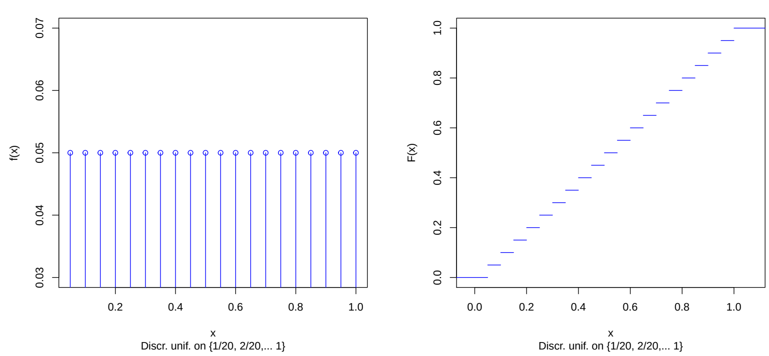
1. Approximate X with a discrete uniform distribution on $\{\frac{1}{10}, \frac{2}{10} \dots, 1\}$. Draw the PMF and CDF.
2. Approximate X with a discrete uniform distribution on $\{\frac{1}{20}, \frac{2}{20} \dots, 1\}$. Draw the PMF and CDF.
3. Approximate X with a discrete uniform distribution on $\{\frac{1}{100}, \frac{2}{100} \dots, 1\}$. Draw the PMF and CDF.
4. Approximate X with a discrete uniform distribution on $\{\frac{1}{10000}, \frac{2}{10000} \dots, 1\}$. Draw the PMF and CDF.

Solution 2.1.

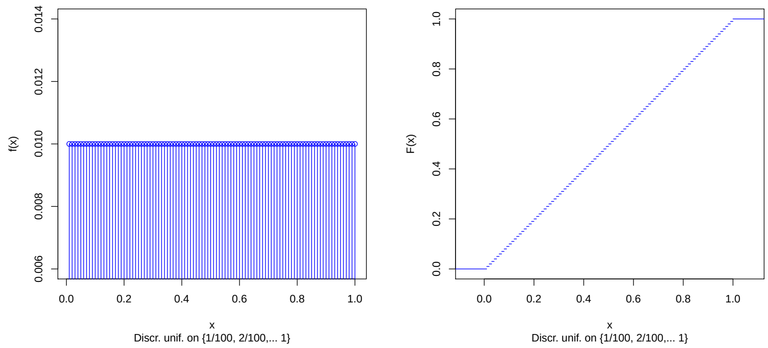
1. We have $f_X(x) = \frac{1}{10}$ for all $x \in \{\frac{1}{10}, \frac{2}{10} \dots, 1\}$.



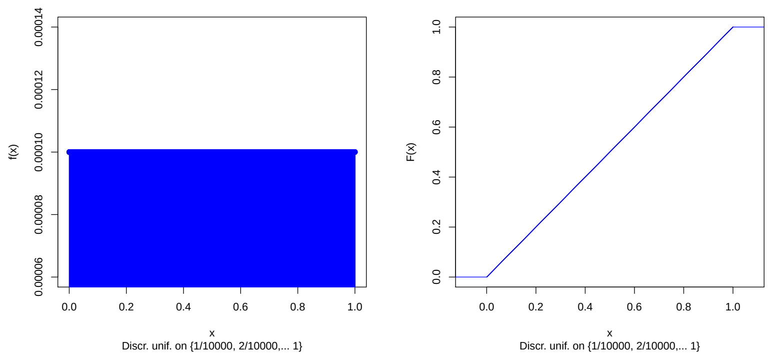
2. We have $f_X(x) = \frac{1}{20}$ for all $\{\frac{1}{20}, \frac{2}{20} \dots, 1\}$.



3. We have $f_X(x) = \frac{1}{100}$ for all $x \in \{\frac{1}{100}, \frac{2}{100} \dots, 1\}$



4. We have $f_X(x) = \frac{1}{10000}$ for all $x \in \{\frac{1}{10000}, \frac{2}{10000} \ldots, 1\}$.



Remark 8. Notice that the PMF gradually tends to 0 and the CDF gets closer to CDF of the continuous uniform distribution on $[0, 1]$.

Problem 2.2. Suppose that X is a continuous random variable with PDF

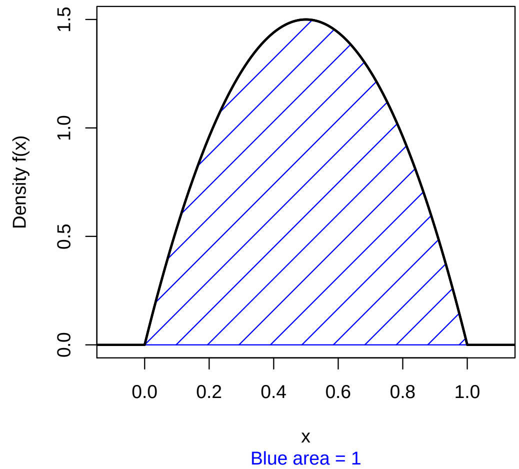
$$f_X(x) = \begin{cases} cx(1-x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

1. Compute c so that this is a valid pdf.
2. Compute the cdf $F_X(x)$.
3. Compute $\mathbb{P}(1/4 \le X \le 3/4)$

Solution 2.2.

1. We need $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$, so

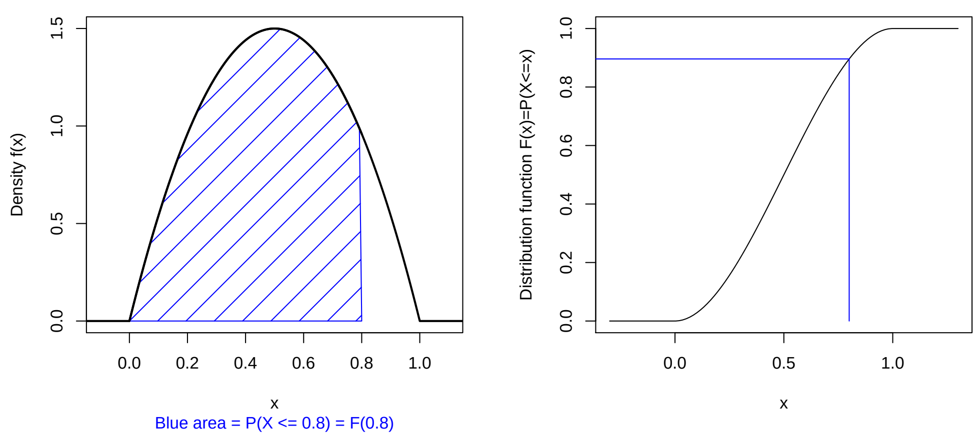
$$\int_0^1 cx(1-x)\,dx = \frac{c}{6} = 1 \implies c = 6.$$



2. The CDF $F_X(x)$ is

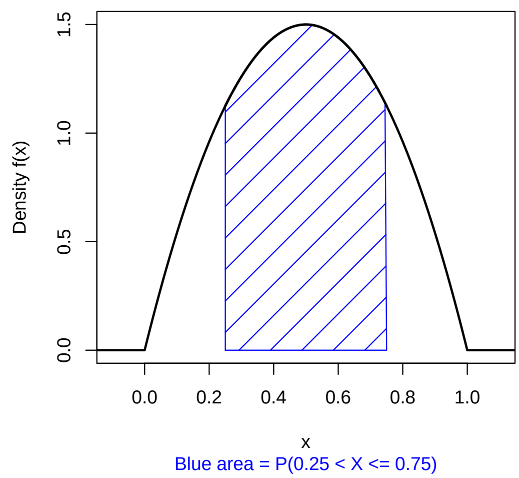
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x 6t(1-t)\,dt = 3x^2 - 2x^3, \quad x \in [0,1]$$

and $F_X(x) = 0$ for $x \le 0$ and $F_X(x) = 1$ for $x \ge 1$.



3. We compute the probability using the CDF:

$$\mathbb{P}\left(\frac{1}{4} \le X \le \frac{3}{4}\right) \overset{\text{cont.}}{=} \mathbb{P}\left(\frac{1}{4} < X \le \frac{3}{4}\right) = \mathbb{P}(X \le 3/4) - P(X \le 1/4)$$
$$= F_X(3/4) - F_X(1/4) = 11/16.$$



Alternative Solution: By integrating the PDF

$$\mathbb{P}(1/4 \leq X \leq 3/4) = \int_{1/4}^{3/4} f_X(t)\, dt = 11/16.$$

Problem 2.3. Suppose the random variable X has PDF

$$f_X(x) = \begin{cases} 6x(1-x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

Compute $\mathbb{E}(X)$ and $\mathrm{Var}(X)$.

Solution 2.3. By definition,

$$\mathbb{E}(X) = \int_0^1 x \cdot 6x(1-x)\, dx = \frac{1}{2}$$

and

$$\mathbb{E}(X^2) = \int_0^1 x^2 \cdot 6x(1-x)\, dx = \frac{3}{10}.$$

Therefore,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{3}{10} - \left(\frac{1}{2}\right)^2 = \frac{1}{20} = 0.05$$

Problem 2.4. Suppose X has PDF $f_X(x)$, and f_X is an even function about the origin on $\mathbb{R}$ (i.e. $f_X(x) = f_X(-x)$). If $\mathbb{E}[X]$ is well defined, show that $\mathbb{E}[X] = 0$.

Solution 2.4. Since $f_X(x) = f_X(-x)$, the "positive and negative areas" cancel since $x \cdot f_X(x)$ is an odd function. To see this directly, notice that

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx \\
&= \int_{-\infty}^{0} x \cdot f_X(x)\, dx + \int_{0}^{\infty} x \cdot f_X(x)\, dx \\
&= \int_{0}^{\infty} (-x) \underbrace{f_X(-x)}_{=f_X(x)}\, dx + \int_{0}^{\infty} x \cdot f_X(x)\, dx \\
&= -\int_{0}^{\infty} x \cdot f_X(x)\, dx + \int_{0}^{\infty} x \cdot f_X(x)\, dx \\
&= 0
\end{aligned}$$