MAT 2377
Probability and Statistics for Engineers

Chapter 2
Discrete Distributions

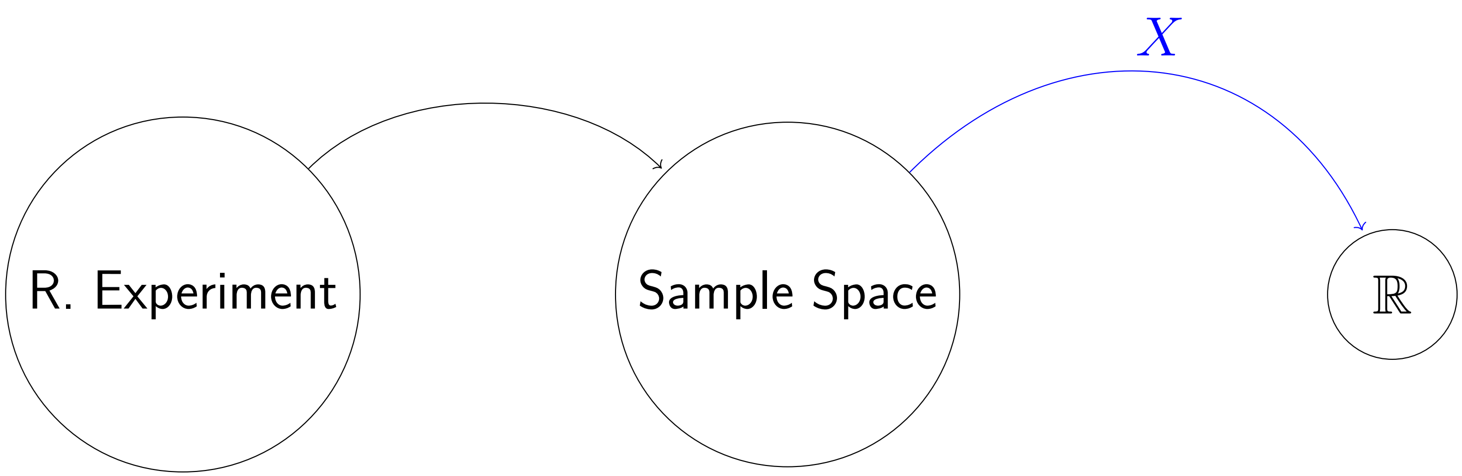
Iraj Yadegari (uOttawa)

Fall 2021

Contents

Random Variables

Recall that, for any random "experiment," the set of all possible outcomes is denoted by $\mathcal{S}$.



A **random variable** (r.v.) is a function $X : \mathcal{S} \to \mathbb{R}$, i.e. it is a rule that associates a (real) number to every outcomes of the experiment.

$\mathcal{S}$ is the **domain** of the r.v. X; $X(\mathcal{S}) \subseteq \mathbb{R}$ is its **range**.

Notation for R.V.

We use the follwoing notation throughout:

- capital roman letters (X, Y, etc.) to denote r.v.
- corresponding lower case roman letters (x, y, etc.) to denote *generic values taken* by the r.v.

Probability Mass Function

A discrete r.v. can be used to define events: if X takes values $X(\mathcal{S}) = \{x_i\}$, then we can define events $A_i = \{s \in \mathcal{S} : X(s) = x_i\}$.

Since outcomes of a random experiment randomly, the random variable takes its values by chance.

- The Probability Mass Function (p.m.f) of X is

$$\boxed{f(x) = P\Big(\{s \in \mathcal{S} : X(s) = x\}\Big) := P(X = x).} \tag{1}$$

When $\mathcal{S}$ is discrete, we say that X is a **discrete r.v.** and the p.d.f. is called a **probability mass function** (p.m.f.).

Example: Flip a fair coin. The outcome space is $\mathcal{S} = \{\text{Head}, \text{Tail}\}$. Let $X : S \to \mathbb{R}$ be defined by $X(\text{Head}) = 1$ and $X(\text{Tail}) = 0$. Then X is a discrete random variable (as a convenience, we write $X = 1$ and $X = 0$).

$$X = \begin{cases} 0 & ;\text{for Tail} \\ 1 & ;\text{for Head} \end{cases}$$

If the coin is fair, the p.m.f. of X is $f : \mathbb{R} \to \mathbb{R}$, where

$$f(x) = P(X = x) = \begin{cases} \frac{1}{2} & ;\text{for } x = 0, 1, \\ \\ 0 & ;\text{Otherwise} \end{cases}$$

Cumulative Distribution Function

The **Cumulative Distribution Function (C.D.F)** of X is

$$\boxed{F(x) = P(X \leq x)} \tag{2}$$

$F : R \rightarrow [0, 1]$ is a non-decreasing function. It is a stepwise function for discrete r.v.

Properties of the P.M.F. and the C.D.F.

If X is a discrete random variable with p.m.f. $f(x)$ and c.d.f. $F(x)$, then

- $\boxed{0 < f(x) \leq 1, \text{ for all } x \in X(\mathcal{S})}$;
- $\boxed{\sum_{x \in X(\mathcal{S})} f(x) = 1}$;

NOTE: Any p.m.f must satisfy these tow conditions.

- for any event $A \subseteq \mathcal{S}$, $P(X \in A) = \Sigma_{x \in A} f(x)$; So

$$\boxed{F(a) = P(X \le a) = \sum_{x \le a} f(x)}$$

- for any $a, b \in \mathbb{R}$,

$$\boxed{P(X > a) = 1 - F(a)}$$

Since $\{X > a\}^c = \{X \le a\}$, therefore, we have

$$P(X > a) = 1 - P(X \le a) = 1 - F(a).$$

- for any $a, b \in \mathbb{R}$,

$$\boxed{P(X < b) = F(b) - f(b)}$$

From $\{X \leq b\} = \{X < b\} \cup \{X = b\}$, we have

$$P(X < b) = P(X \leq b) - P(X = b) = F(b) - f(b).$$

- for any $a, b \in \mathbb{R}$,

$$\boxed{P(X \geq a) = 1 - F(a) + f(a)}$$

because

$$\begin{aligned} P(X \geq a) &= 1 - P(X < a) = 1 - \Big(P(X \leq a) - P(X = a)\Big) \\ &= 1 - F(a) + f(a) \end{aligned}$$

We can use these results to compute the probability of a **discrete** r.v. X falling in various intervals. The basic property is

$$\boxed{P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)}$$

Using this property, we can obtain the following equations

$$\boxed{P(a \leq X \leq b) = P(a < X \leq b) + P(X = a) = F(b) - F(a) + f(a)}$$

$$\boxed{P(a < X < b) = P(a < X \leq b) - P(X = b) = F(b) - F(a) - f(b)}$$

$$\boxed{P(a \leq X < b) = P(a \leq X \leq b) - P(X = b) = F(b) - F(a) + f(a) - f(b)}$$

Examples:

1. Roll a fair die. The outcome space is

$$\mathcal{S} = \{1, \ldots, 6\}.$$

Let $X : \mathcal{S} \to \mathbb{R}$ be defined by $X(i) = i$ for $i = 1, \ldots, 6$. Then X is a discrete r.v.

If the die is fair, the p.m.f. of X is $f : \mathbb{R} \to \mathbb{R}$, where

$$f(i) = P(X = i) = \begin{cases} 1/6 & ; \text{for } \ \text{i} = 1, \ldots, 6, \\ \\ 0 & ; \text{Otherwise} \end{cases}$$

2. For the random variable X from the previous example, the c.d.f. is $F : \mathbb{R} \to \mathbb{R}$, where

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 1 \\ i/6 & \text{if } i \leq x < i+1, \text{ for } i = 1, \ldots, 6 \\ 1 & \text{if } x \geq 6 \end{cases}$$

3. For the same random variable, we can compute $P(3 \leq X \leq 5)$ directly:

$$P(3 \leq X \leq 5) = P(X = 3) + P(X = 4) + P(X = 5) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2},$$

or we can use the c.d.f.

$$P(3 \leq X \leq 5) = F(5) - F(3) + f(3) = \frac{5}{6} - \frac{3}{6} + \frac{1}{6}$$

4. The number of calls received over a specific time period, X, is a discrete random variable, with potential values $0, 1, 2, \ldots$.

   We will discuss p.m.f. for this type of variables in this course

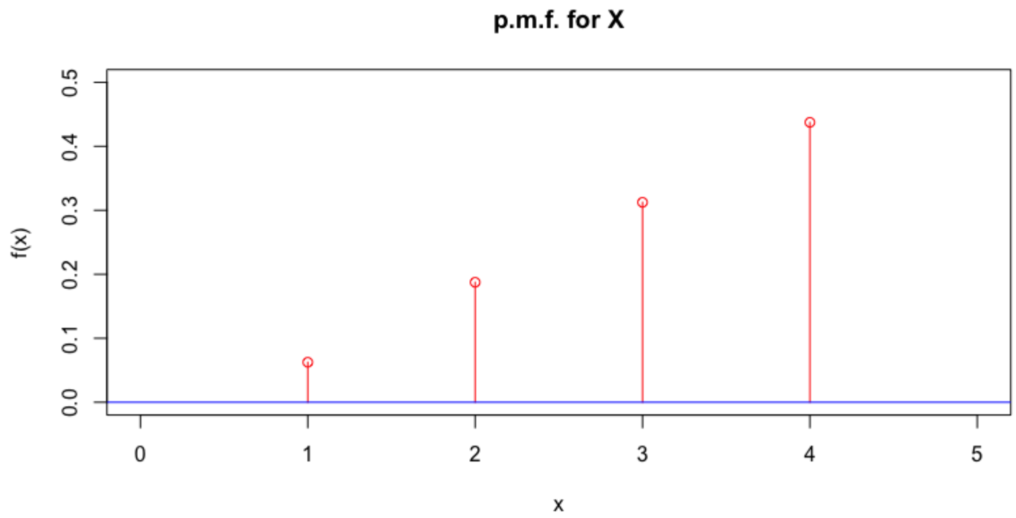
5. Find the c.d.f. $F(x)$ of a discrete random variable X with p.m.f. $f(x)$ defined by $f(x) = 0.1x$ if $x = 1, 2, 3, 4$ and $f(x) = 0$ otherwise.

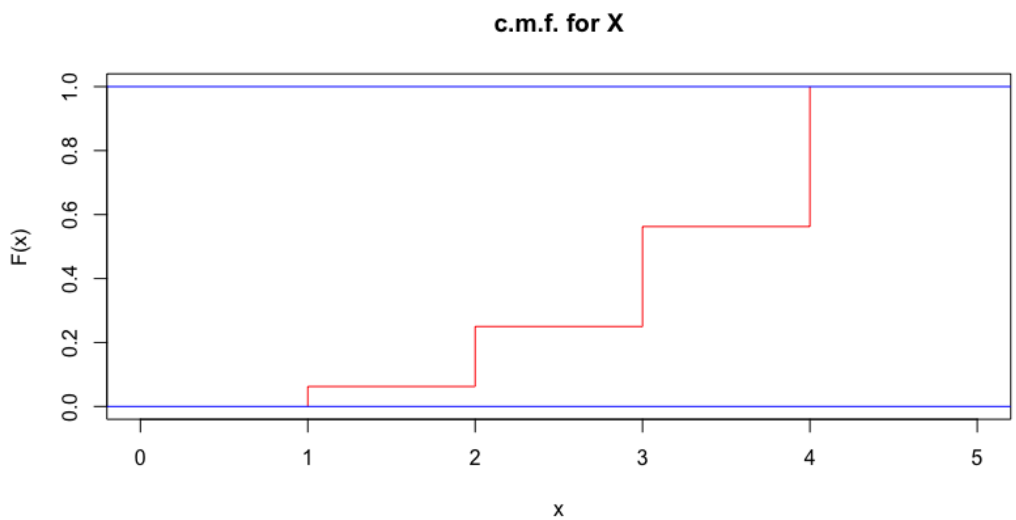
   Solution: $f(x)$ is indeed a p.m.f. because $0 < f(x) \leq 1$ for all x and

$$\sum_{x=1}^{4} 0.1x = 0.1(1+2+3+4) = 0.1\frac{4(5)}{2} = 1.$$

   We have

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 0.1 & \text{if } 1 \leq x < 2 \\ 0.3 & \text{if } 2 \leq x < 3 \\ 0.6 & \text{if } 3 \leq x < 4 \\ 1 & \text{if } x \geq 4 \end{cases}$$


p.m.f. for X
f(x)
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4
5
x


c.m.f. for X
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
x

Expectation of a Discrete R.V.

The Expectation of a discrete random variable X is defined as

$$\boxed{\mathrm{E}[X] = \sum_x x \cdot P(X = x) = \sum_x x f(x)\,,}$$

where the sum extends over all values of x taken by X.

The definition can be extended to a general function of X:

$$\boxed{\mathrm{E}[u(X)] = \sum_x u(x) P(X = x) = \sum_x u(x) f(x).}$$

As an important example, $\mathrm{E}[X^2] = \sum_x x^2 P(X = x) = \sum_x x^2 f(x)$.

Examples:

1. What is the expectation on the roll Z of fair $6-$sided die?

   Solution: $\mathrm{E}[Z] = \sum_{z=1}^{6} z \cdot P(Z = z) = \frac{1}{6} \sum_{z=1}^{6} z = \frac{1}{6} \cdot \frac{6(6+1)}{2} = 3.5.$

2. For each $1\$$ bet in a gambling game, a player can win $3\$$ with probability $\frac{1}{3}$ and lose $1\$$ with probability $\frac{2}{3}$. Let X be the net gain/loss from the game. Find the expected value of the game.

   Solution: X can take on the value $2\$$ (for a win) and $-2\$$ for a loss (outcome $-$ bet). The expected value of X is thus

$$\mathrm{E}[X] = 2 \cdot \frac{1}{3} + (-2) \cdot \frac{2}{3} = -\frac{2}{3}.$$

3. If Z is the number showing on a roll of a fair $6-$sided die, find $\mathrm{E}[Z^2]$ and $\mathrm{E}[(Z-3.5)^2]$.

Solution:

$$\mathrm{E}[Z^2] = \sum_z z^2 P(Z=z) = \frac{1}{6}\sum_{z=1}^{6} z^2 = \frac{1}{6}(1^2+\cdots+6^2) = \frac{91}{6}$$

$$\begin{aligned}\mathrm{E}[(Z-3.5)^2] &= \sum_{z=1}^{6}(z-3.5)^2 P(Z=z) = \frac{1}{6}\sum_{z=1}^{6}(z-3.5)^2 \\ &= \frac{(1-3.5)^2+\cdots+(6-3.5)^2}{6} = \frac{35}{12}\end{aligned}$$

Interpretation of Expectation

We can interpret the expectation as the average or the **mean** of X,

$$\boxed{\textbf{Expectation} = \textbf{Mean} = \textbf{Average},}$$

which we often denote by $\mu = \mu_X$.

For instance, in the example of the fair die, $\mu_Z = \mathrm{E}[Z] = 3.5$.

Note that in the final example, we could have written

$$\mathrm{E}[(Z - 3.5)^2] = \mathrm{E}[(Z - \mathrm{E}[Z])^2].$$

This is an important quantity associated to a random variable X.
It can be interpreted as the **average distance from the mean**.

Variance of a Discrete R.V.

The variance of a discrete random variable X is the **average of squared difference from the mean**:

$$\boxed{\mathrm{Var}(X) = \mathrm{E}[(X-\mu_X)^2] = \sum_x (x-\mu_X)^2 P(X=x).}$$

We can simplify this to

$$\begin{aligned}
\sum_x (x-\mu_X)^2 P(X=x) &= \sum_x \left(x^2 - 2x\mu_X + \mu_X^2\right) f(x) \\
&= \sum_x x^2 f(x) - 2\mu_X \sum_x x f(x) + \mu_X^2 \sum_x f(x) \\
&= \mathrm{E}[X^2] - 2\mu_X\mu_X + \mu_X^2 \cdot 1 = \mathrm{E}[X^2] - \mu_X^2.
\end{aligned}$$

Therefore, Variance is can be written as $\boxed{\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}^2[X]}$.

Standard Deviation of a Discrete R.V.

The **standard deviation** of a discrete random variable X is defined directly from the variance:

$$\boxed{\mathrm{SD}[X] = \sqrt{\mathrm{Var}[X]}\,.}$$

The mean gives some idea as to where the **bulk** of a distribution is located $\Rightarrow$ measure of **centrality** (more on this later).

The variance and standard deviation provide information about the **spread**; distributions with higher variance/SD are **more spread out about the average**.

NOTE: SD has the same UNIT as the data.

Examples: let X and Y be random variables with the following p.d.f.:

x	$P(X=x)$
-2	$1/5$
-1	$1/5$
0	$1/5$
1	$1/5$
2	$1/5$

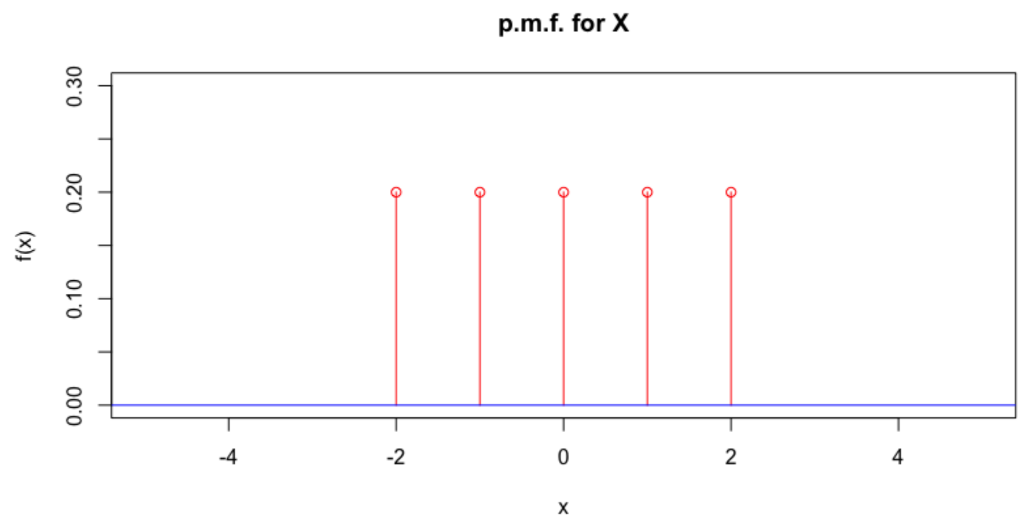
y	$P(Y=y)$
-4	$1/5$
-2	$1/5$
0	$1/5$
2	$1/5$
4	$1/5$

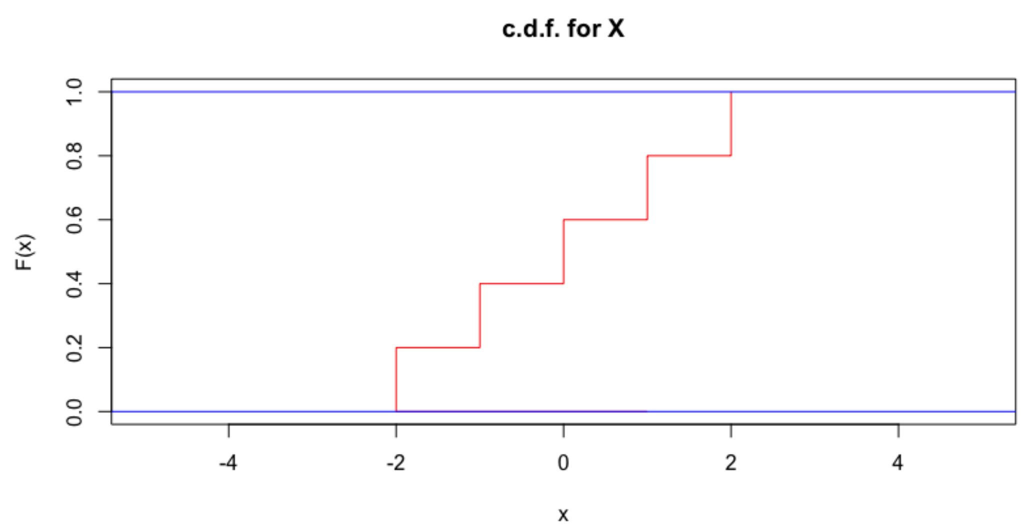
Compute the expected values and compare the variances.

Solution: We have $\mathrm{E}[X]=\mathrm{E}[Y]=0$ and $2=\mathrm{Var}[X]<\mathrm{Var}[Y]=8$, meaning that we would expect both distributions to be centered at 0, but Y should be more spread-out than X.

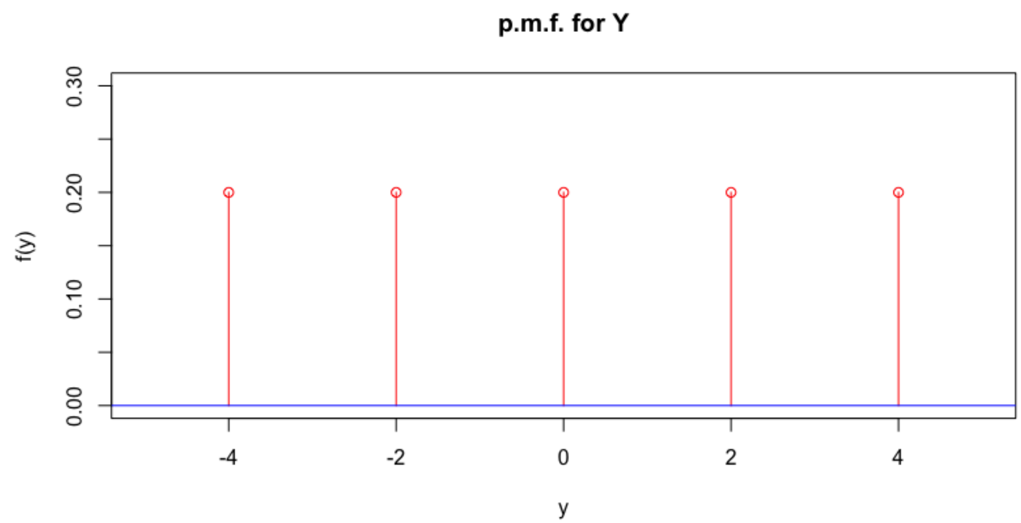
Let check the P.M.F and C.D.F of X and Y to understand variation in a r.v.

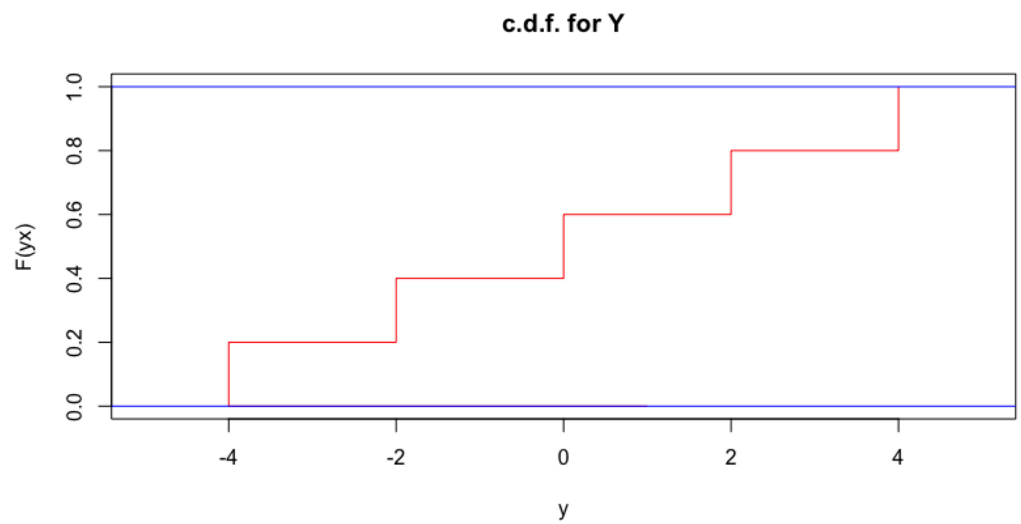
P.M.F and C.D.F of X:





P.M.F and C.D.F of Y:





Properties of Expectation and Variance

For all constant values $a, b \in \mathbb{R}$:

- $\mathrm{E}[a + bX] = a + b\mathrm{E}[X]$;

  $$\Big(\mathrm{E}[a + X] = a + \mathrm{E}[X]\ ;\quad \mathrm{E}[bX] = b\mathrm{E}[X]\ ;\quad \mathrm{E}[a] = a\Big)$$

- $\mathrm{Var}[a + bX] = b^2\mathrm{Var}[X]$;

  $$\Big(\mathrm{Var}[a + X] = \mathrm{Var}[X]\ ;\quad \mathrm{Var}[bX] = b^2\mathrm{Var}[X]\ ;\quad \mathrm{Var}[a] = 0\Big)$$

  Always $\quad \mathrm{Var}[X] \geq 0$

Hypergeometric Distribution

Consider an Urn that contains N balls, K balls are blue and $N-K$ are red. We take a sample of size n, **without replacement**. Let X be the number of blue balls in the sample. Therefore, its probability mass function is

$$\boxed{f(x) = P(X=x) = \frac{\binom{K}{x}\cdot\binom{N-K}{n-x}}{\binom{N}{n}},}$$

for $\max(0, n+K-N) < x < \min(n, K)$.

For this random variable, we have

$$\boxed{\mathrm{E}[X] = np \quad \text{and} \quad \mathrm{Var}[X] = np\,(1-p)\left(\frac{N-n}{N-1}\right), \quad \text{where} \quad p = \frac{K}{N}}$$

Example: Consider a $5-$card poker hand consisting of cards selected at random from a $52-$card deck. Find the probability distribution of X, where X indicates the number of red cards ($\diamondsuit$ and $\heartsuit$) in the hand.

Solution: in all there are $\binom{52}{5}$ ways to select a 5-card poker hand from a 52-card deck (Hearts $\heartsuit$, Diamonds $\diamondsuit$, spades $\spadesuit$, and Clubs $\clubsuit$).

By construction, X can take on values $x = 0, 1, 2, 3, 4, 5.$

If $X = 0$, then none of the 5 cards in the hands are $\diamondsuit$ or $\heartsuit$, and all of the 5 cards in the hands are $\spadesuit$ or $\clubsuit$. Thus $\binom{26}{0} \cdot \binom{26}{5}$ is the number of cases that 5-card hands only contain black cards, and

$$P(X = 0) = \frac{\binom{26}{0} \cdot \binom{26}{5}}{\binom{52}{5}}.$$

In general, if $X = x$, $x = 0, 1, 2, 3, 4, 5$, there are $\binom{26}{x}$ ways of having x red, $\diamondsuit$ or $\heartsuit$, in the hand AND $\binom{26}{5-x}$ ways of having $5 - x$ black, $\spadesuit$ or $\clubsuit$, in the hand, so that

$$f(x) = P(X = x) = \begin{cases} \frac{\binom{26}{x} \cdot \binom{26}{5-x}}{\binom{52}{5}} & \text{for } \text{x=0,1,2,3,4,5} \\ \\ 0 & \text{Otherwise} \end{cases}$$

Bernoulli Experiment

A **Bernoulli random trial (experiment)** is a random experiment with two possible outcomes, "success" and "failure". Let p denote the probability of a success.

$$\boxed{\text{(1): F/S}}$$

Examples:

- fail/pass a course;
- Getting/Not-getting COVID-19;
- satisfactory/defective items on a production line;

Binomial Experiment

A **binomial experiment** consists of n repeated independent Bernoulli trials, each with the same probability of success, p.

(1): F/S	(2): F/S	$\cdots$	(n-1): F/S	(n): F/S

Examples:

- Number of courses you can pass among 5 course;
- Number of guests that get the COVID-19 in a party, among 100 people;
- Number of satisfactory/defective items on a production line, among 200 items;

Binomial Distribution

In a binomial experiment of n independent events, each with probability of success p, the number of successes X is a discrete random variable that follows a **binomial distribution** with parameters (n, p):

$$\boxed{f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, \ldots, n.}$$

This is often abbreviated to "$X \sim \mathcal{B}(n, p)$" or "$X \sim Bin(n, p)$".

If $X \sim Bin(1, p)$ then $P(X = 0) = 1 - p$, $P(X = 1) = p$, and so

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

Explanation

X : Number of successes in n independent trials $x = 0, 1, 2, \cdots, n$

Sequence of trials: | (1): **F/S** | (2): **F/S** | $\cdots$ | (n-1): **F/S** | (n): **F/S** |

$$P(X = x) = \underbrace{\binom{n}{x}}_{\#\text{ of permutation of } x \text{ success in n trials}} \times \underbrace{p^x (1-p)^{n-x}}_{\text{prob of 1 permutation of } x \text{ success in } n \text{ trials}}$$

$$= \binom{n}{x} p^x (1-p)^{n-x}$$

Recall the number of unordered samples of size r from a set of size n is:

$$\boxed{{}_nC_r = \binom{n}{r} = \frac{n!}{(n-r)!r!}.}$$

This is called Binomial Coefficient.

- $2! \times 4! = (1 \times 2) \times (1 \times 2 \times 3 \times 4) = 48$, but $(2 \times 4)! = 8! = 40320$.
- $\binom{5}{1} = \frac{5!}{1! \times 4!} = \frac{1\times2\times3\times4\times5}{1\times(1\times2\times3\times4)} = \frac{5}{1} = 5$; In general: $\binom{n}{1} = n$. Also $\binom{n}{0} = 1$
- $\binom{6}{2} = \frac{6!}{2!\times4!} = \frac{4!\times5\times6}{2!\times4!} = \frac{5\times6}{2} = 15$; 0!=1; 1!=1;
- $\binom{27}{22} = \frac{27!}{22!\times5!} = \frac{22!\times23\times24\times25\times26\times27}{5!\times22!} = \frac{23\times24\times25\times26\times27}{120}$

Expectation and Variance for Binomial

If $X \sim Bin(n,p)$, then $P(X=x)$ is as on the previous slide and it can be shown that

$$\boxed{\mathrm{E}[X] = \sum_{x=0}^{n} xP(X=x) = np,}$$

and

$$\boxed{\mathrm{Var}[X] = \mathrm{E}\left[(X-np)^2\right] = \sum_{x=0}^{n}(x-np)^2 P(X=x) = np(1-p)\,.}$$

Later we will see an easier way to derive these by interpreting X as a sum of other discrete random variables.

Sampling with or without replacement?

In the problem with an Urn that contains N balls, K balls are blue and $N-K$ are blue. Probability of selecting a red ball at random is $P = \frac{K}{N}$.

We take a sample of size n. Let X be the number of red balls in the sample.

- If sampling is **WITH replacement**, then X has Binomial distribution, $X \sim Bin(n, P)$.
- If sampling is **WITHOUT replacement**, then X has hypergeometric distribution.

Examples:

1. Suppose that each water sample taken in some well-defined region has a 10% probability of being polluted.

   If 12 samples are selected independently, then it is reasonable to model the number X of polluted samples as $Bin(12, 0.1)$. Find

   a) $\mathrm{E}[X]$ and $\mathrm{Var}[X]$;
   b) $P(X = 3)$;
   c) $P(X \leq 3)$.

   Solution:

   a) If $X \sim Bin(n, p)$ then $\mathrm{E}[X] = np$ and $\mathrm{Var}[X] = np(1 - p)$, so

$$\mathrm{E}[X] = 12 \times 0.1 = 1.2; \quad \mathrm{Var}[X] = 12 \times 0.1 \times 0.9 = 1.08\,.$$

b) By definition, $P(X=3) = \binom{12}{3}(0.1)^3(0.9)^9 \approx 0.0852.$

c) By definition,

$$P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = \cdots$$

However, for $X \sim Bin(12, 0.1)$, $P(X \leq 3)$ is tabulated on p.430 of text (Table A.1), and is ≈ 0.9744.

The table can also be used to compute

$$P(X=3) = P(X \leq 3) - P(X \leq 2) = 0.9744 - 0.8891 \approx 0.0853.$$

Note the rounding error.

Table A.1 (continued) Binomial Probability Sums $\sum_{x=0}^{r} b(x; n, p)$

		p									
n	*r*	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90
12	0	0.2824	0.0687	0.0317	0.0138	0.0022	0.0002	0.0000			
	1	0.6590	0.2749	0.1584	0.0850	0.0196	0.0032	0.0003	0.0000		
	2	0.8891	0.5583	0.3907	0.2528	0.0834	0.0193	0.0028	0.0002	0.0000	
	3	0.9744	0.7946	0.6488	0.4925	0.2253	0.0730	0.0153	0.0017	0.0001	
	4	0.9957	0.9274	0.8424	0.7237	0.4382	0.1938	0.0573	0.0095	0.0006	0.0000
	5	0.9995	0.9806	0.9456	0.8822	0.6652	0.3872	0.1582	0.0386	0.0039	0.0001
	6	0.9999	0.9961	0.9857	0.9614	0.8418	0.6128	0.3348	0.1178	0.0194	0.0005
	7	1.0000	0.9994	0.9972	0.9905	0.9427	0.8062	0.5618	0.2763	0.0726	0.0043
	8		0.9999	0.9996	0.9983	0.9847	0.9270	0.7747	0.5075	0.2054	0.0256
	9		1.0000	1.0000	0.9998	0.9972	0.9807	0.9166	0.7472	0.4417	0.1109
	10				1.0000	0.9997	0.9968	0.9804	0.9150	0.7251	0.3410
	11					1.0000	0.9998	0.9978	0.9862	0.9313	0.7176
	12						1.0000	1.0000	1.0000	1.0000	1.0000

Figure 1: Table of c.d.f. $F(r) = P(X \leq r)$ for $X \sim Bin(12, p)$, $p = 0.1, \ldots, 0.9$.

2. An airline sells 101 tickets for a flight with 100 seats. Each passenger with a ticket is known to have a $p = 0.97$ probability of showing up for their flight. What is the probability of 101 passengers showing up (and the airline being caught overbooking)? Make appropriate assumptions. What if the airline sells 125 tickets?

   Solution: let X be the number of passengers that show up. We want to compute $P(X > 100)$.

   If all passengers show up independently of one another (no families or late bus?), we can model $X \sim Bin(101, 0.97)$ and

$$P(X > 100) = P(X = 101) = \binom{101}{101}(0.97)^{101}(0.03)^0 \approx 0.046$$

If the airline sells $n = 125$ tickets, we can model $X \sim Bin(125, 0.97)$ and

$$P(X > 100) = 1 - P(X \leq 100) = 1 - \sum_{x=0}^{100} \binom{125}{x} (0.97)^x (0.03)^{125-x}.$$

This is harder to compute directly, but is very nearly 1 (try it in `R`).

Geometric Distribution

Consider a sequence of independent Bernoulli trials, with probability p of success at each step. Let X be the number of steps until the first success occurs. Then X is a **geometric** random variable and its probability mass function is

$$\boxed{f(x) = P(X = x) = p\,(1-p)^{x-1}, \quad x = 1, \ldots}$$

This is often abbreviated to "X$\sim$ Geo(p)".

For this random variable, we have

$$\boxed{\mathrm{E}[X] = \frac{1}{p} \quad \text{and} \quad \mathrm{Var}[X] = \frac{1-p}{p^2}.}$$

Explanation

X : Number of independent trials until first success

$$x = 1, 2, 3, \cdots$$

Sequence of trials:

(1): **F**	(2): **F**	$\cdots$	(x-1): **F**	(x): **S**

$$f(x) = P(X = x) = \underbrace{(1-p)(1-p)\cdots(1-p)}_{x-1 \text{ failures}} \times \underbrace{\quad p \quad}_{1 \text{ success}} = (1-p)^{x-1}\,p$$

Examples:

- A fair $6-$sided die is thrown until it shows a 6. What is the probability that 5 throws are required?

  Solution: If 5 throws are required, we have to compute $P(X = 5)$, where X is geometric Geo$(1/6)$:

$$P(X = 5) = (1 - p)^{5-1}p = (5/6)^4(1/6) \approx 0.0804.$$

- In the example above, how many throws would you expect to need?

  Solution: $\mathrm{E}[X] = \frac{1}{1/6} = 6.$

Negative Binomial Distribution

Consider a sequence of independent Bernoulli trials, with probability p of success at each step. Let X be the number of steps until the k-th success occurs. Then X is a **negative binomial** random variable and its probability mass function is

$$\boxed{f(x) = P(X = x) = \binom{x-1}{k-1}(1-p)^{x-k}p^k, \quad x = k, \ldots}$$

This is often abbreviated to "X$\sim$ NegBin$(p, k)''or``X \sim$ NBin$(p, k)''$.

For this random variable, we have

$$\boxed{\mathrm{E}[X] = \frac{k}{p} \quad \text{and} \quad \mathrm{Var}[X] = \frac{k(1-p)}{p^2}.}$$

Explanation

X : Number of independent trials until k-th success

$$x = k, k+1, k+2, \cdots$$

Sequence of trials:

(1): **F/S**	(2): **F/S**	$\cdots$	(x-1): **F/S**	(x): **S**

$$f(x) = P(X = x) = \underbrace{\binom{x-1}{k-1}(1-p)^{x-k}p^{k-1}}_{k-1 \text{ success in } x-1 \text{ trials}} \times \underbrace{p}_{1 \text{ success in last trial}}$$

$$= \binom{x-1}{k-1}(1-p)^{x-k}p^{k}$$

Examples:

- A fair $6-$sided die is thrown until it three 6's are rolled. What is the probability that 5 throws are required?

  Solution: If 5 throws are required, we have to compute $P(X=5)$, where X is geometric NegBin$(1/6,3)$:

$$P(X=5)=\binom{5-1}{3-1}(1-p)^{5-3}p^3=\binom{4}{2}(5/6)^2(1/6)^3\approx 0.0193.$$

- In the example above, how many throws would you expect to need?

  Solution: $\mathrm{E}[X]=\frac{3}{1/6}=18.$

Poisson Process

We count the number of "changes" that occur in a continuous interval of time or space (such as $\#$ of defects on a production line over a 1 hr period, $\#$ of customers that arrive at a teller over a 15 min interval, etc.).

Let λ be the **average number of changes in the Unit interval** of time or space.

We have a **Poisson Process** with rate λ, denoted by $\mathcal{P}(\lambda)$, if:

a) the number of changes occurring in non-overlapping intervals are independent;

b) the probability of exactly one change in a short interval of length h is approximately λh, and

c) The probability of $2+$ changes in a sufficiently short interval is essentially 0.

Consider a random experiment that satisfies the above conditions

Let X be the number of changes in a **unit interval** (this could be 1 day, or 15 minutes, or 10 years, etc.).

What is $P(X = x)$, for $x = 0, 1, \ldots$?

Partition the unit interval into n disjoint sub-intervals of length $1/n$.

1. By condition b), the probability of one change occurring in one of the sub-intervals is approximately λ/n.
2. By condition c), the probability of $2+$ changes is ≈ 0.
3. By condition a), we have a sequence of n independent Bernoulli trials with probability $p = \lambda/n$.

Thus we approximate the distribution of X by the binomial distribution $Bin(n, \lambda/n)$.

Therefore,

$$
\begin{aligned}
f(x) = P(X = x) &\approx \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{\lambda^x}{x!}\cdot\underbrace{\frac{n!}{(n-x)!}\cdot\frac{1}{n^x}}_{\text{term }1}\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\text{term }2}\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{\text{term }3}.
\end{aligned}
$$

Letting $n \to \infty$, we get

$$
\begin{aligned}
P(X = x) &= \lim_{n\to\infty}\frac{\lambda^x}{x!}\cdot\underbrace{\frac{n!}{(n-x)!}\cdot\frac{1}{n^x}}_{\text{term }1}\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\text{term }2}\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{\text{term }3} \\
&= \frac{\lambda^x}{x!}\cdot 1\cdot\exp(-\lambda)\cdot 1 \\
&= \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, \ldots
\end{aligned}
$$

Poisson Distribution

Let X be the number of changes in a **unit interval** of time or space and λ be the **its average in the Unit interval**.

We say that X has Poisson distribution with parameter λ if its probability mass function is

$$\boxed{P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, \ldots}$$

This is often abbreviated to "X$\sim \mathcal{P}(\lambda)$".

It can be shown that

$$\boxed{\mathrm{E}[X] = \lambda \quad \text{and} \quad \mathrm{Var}[X] = \lambda,}$$

that is, the mean and the variance of a Poisson random variable are identical!

Examples:

1. A traffic flow is typically modeled by a Poisson distribution. It is known that the traffic flowing through an intersection is 6 cars/minute, on average.

   What is the probability of no cars entering the intersection in a 30 second period?

   Solution: Our unit interval is 30 second not a minute.
   6 cars/min $= 3$ cars/30 sec. Thus $\lambda = 3$ for our unit interval, and we need to compute

$$P(X = 0) = \frac{3^0 e^{-3}}{3!} = \frac{e^{-3}}{6} \approx 0.0498.$$

2. A hospital needs to schedule night shifts in the maternity ward.

   It is known that there are 3000 deliveries per year; if these happened randomly round the clock (is this a reasonable assumption?), we would expect 1000 deliveries between the hours of midnight and 8.00 a.m., a time when much of the staff is off-duty.

   It is thus important to ensure that the night shift is sufficiently staffed to allow the maternity ward to cope with the workload on any particular night, or at least, on a high proportion of nights.

   The average number of deliveries per night is $\lambda = 1000/365.25 \approx 2.74$.

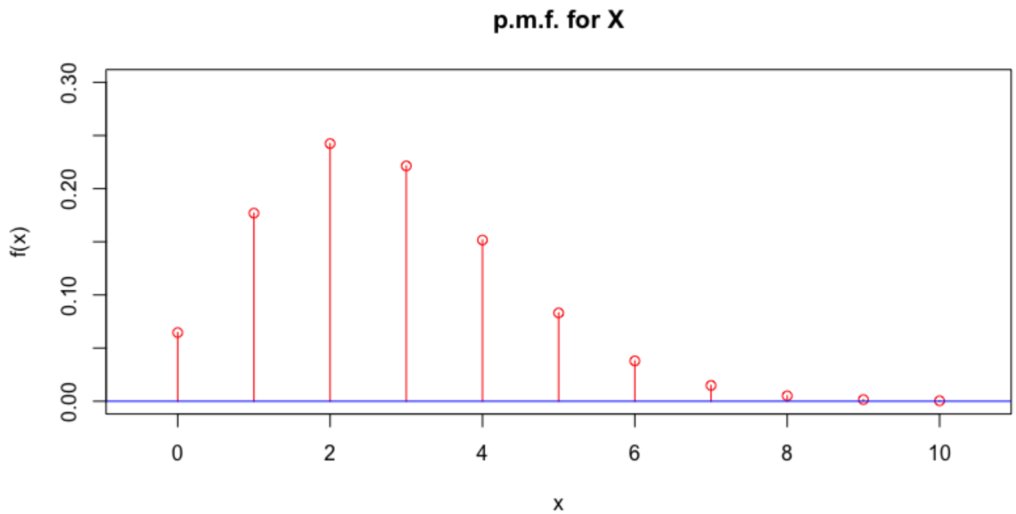
   If the daily number X of night deliveries follows a Poisson process $\mathcal{P}(\lambda)$, we can compute the probability of delivering $x = 0, 1, 2, \ldots$ babies on each night.

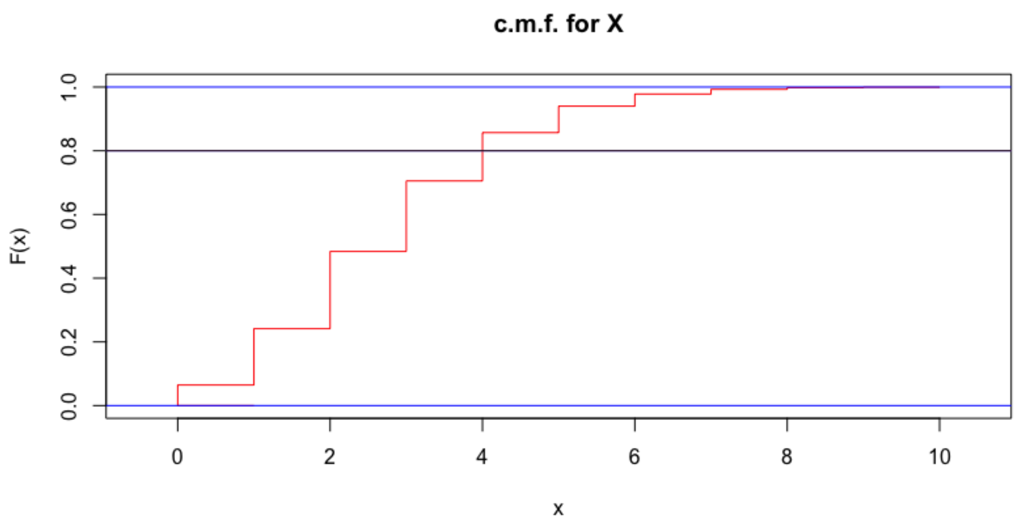
Some of the probabilities are:

$P(X = x)$	$\lambda^x \cdot \exp(-\lambda)/x!$
$P(X = 0)$	$2.74^0 \cdot \exp(-2.74)/0! = 0.065$
$P(X = 1)$	$2.74^1 \cdot \exp(-2.74)/1! = 0.177$
$P(X = 2)$	$2.74^2 \cdot \exp(-2.74)/2! = 0.242$
$\ldots$	$\ldots$

3. If the maternity ward wants to prepare for the greatest possible traffic on 80% of the nights, how many deliveries should be expected?

   Solution: we search an x for which $P(X \leq x-1) \leq 0.80 \leq P(X \leq x)$: since `ppois(3,2.74)=.705` and `ppois(4,2.74)=.857`, if they prepare for 4 deliveries a night, they will be ready for the worst on at least 80% of the nights (closer to 85.7%, actually). Note that this is different than asking how many deliveries are expected nightly (namely, $\mathrm{E}[X] = 2.74$).


p.m.f. for X
f(x)
0.00
0.10
0.20
0.30
0
2
4
6
8
10
x


c.m.f. for X
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
x

4. On how many nights in the year would 5 or more deliveries be expected?

   Solution: we need to evaluate

$$\begin{aligned} 365.25 \cdot P(X \geq 5) &= 365.25(1 - P(X \leq 4)) \\ &= \mathtt{365.25} * (\mathtt{1} - \mathtt{ppois}(\mathtt{4}, \mathtt{2.74})) \approx 52.27. \end{aligned}$$

5. Over the course of one year, what is the greatest number of deliveries expected in any night?

   Solution: we want largest value of x for which $365.25 \cdot P(X = x) > 1$.

```
> nights=c() # initializing vector
> for(j in 0:10){nights[j+1]=365.25*dpois(j,2.74)}; # c.m.f.
> max(which(nights>1))-1 # identify largest index
```

$\Rightarrow x = 8$.

Joint Distributions

Let X, Y be two discrete random variables. The **joint probability mass function** (joint p.m.f.) of X, Y is a function $f(x, y)$ satisfying

1. $f(x, y) \geq 0$, for all x, y;
2. $\Sigma_{-\infty}^{\infty}\Sigma_{-\infty}^{\infty} f(x, y)\, dxdy = 1$.

For any $A \subseteq \mathbb{R}^2$, $P(A) = \Sigma\Sigma_A f(x, y)\, dxdy$.

Example: Roll a pair of unbiased dice. For each of the 36 possible outcomes, let X denote the smaller roll, and Y the larger roll.

a) How many outcomes correspond to the event $A = \{(X = 2, Y = 3)\}$?

Solution: the rolls $(3, 2)$ and $(2, 3)$ both give rise to event A.

b) What is $P(A)$?

Solution: there are 36 possible outcomes, so $P(A) = \frac{2}{36} \approx 0.0556$.

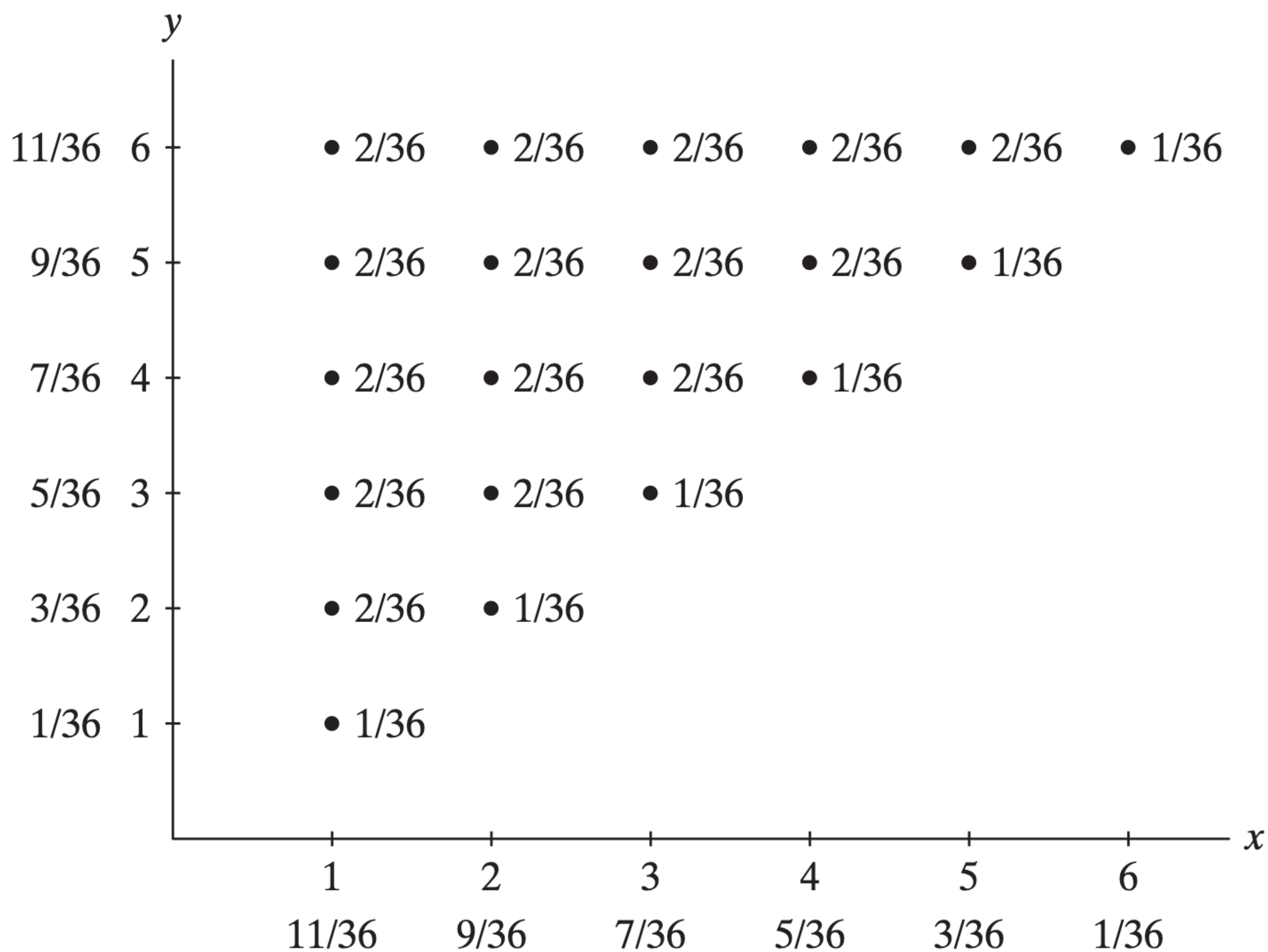
c) What is the joint p.m.f. of X, Y?

Solution: there is only one outcome $(X = a, Y = a)$ that gives rise to $\{X = Y = a\}$. For every other event $\{X \neq Y\}$, two outcomes do the trick: (X, Y) and (Y, X). The joint p.m.f. is thus

$$f(x,y) = \begin{cases} 1/36 & 1 \leq x = y \leq 6 \\ 2/36 & 1 \leq x < y \leq 6 \end{cases}$$

The first property is automatically satisfied, as is the third (by construction). There are only 6 outcomes for which $X = Y$, all the remaining outcomes (of which there are 15) have $X < Y$. Thus,

$$\sum_{x=1}^{6} \sum_{y=x}^{6} f(x,y) = 6 \cdot \frac{1}{36} + 15 \cdot \frac{2}{36} = 1.$$


y
x
11/36 6
9/36 5
7/36 4
5/36 3
3/36 2
1/36 1
2/36 2/36 2/36 2/36 2/36 1/36
2/36 2/36 2/36 2/36 1/36
2/36 2/36 2/36 1/36
2/36 2/36 1/36
2/36 1/36
1/36
1 2 3 4 5 6
11/36 9/36 7/36 5/36 3/36 1/36

d) Compute $P(X = a)$ and $P(Y = b)$, for $a, b = 1, \ldots, 6$.

Solution: for every $a = 1, \ldots, 6$, the event $\{X = a\}$ corresponds to the following union of events:

$$\{X = a, Y = a\} \cup \{X = a, Y = a + 1\} \cup \cdots \cup \{X = a, Y = 6\}.$$

These events are mutually exclusive, so that

$$P(X = a) = \sum_{y=a}^{6} P(\{X = a, Y = y\}) = \frac{1}{36} + \sum_{y=a+1}^{6} \frac{2}{36}$$
$$= \frac{1}{36} + \frac{2(6-a)}{36}, \quad a = 1, \ldots, 6.$$

Similarly, we get $P(Y = b) = \frac{1}{36} + \frac{2(b-6)}{36}$, $b = 1, \ldots, 6$. These **marginal probabilities** can be found in the margins of the p.m.f.

e) Compute $P(X = 3|Y > 3)$ and $P(Y \leq 3|X \geq 4)$.

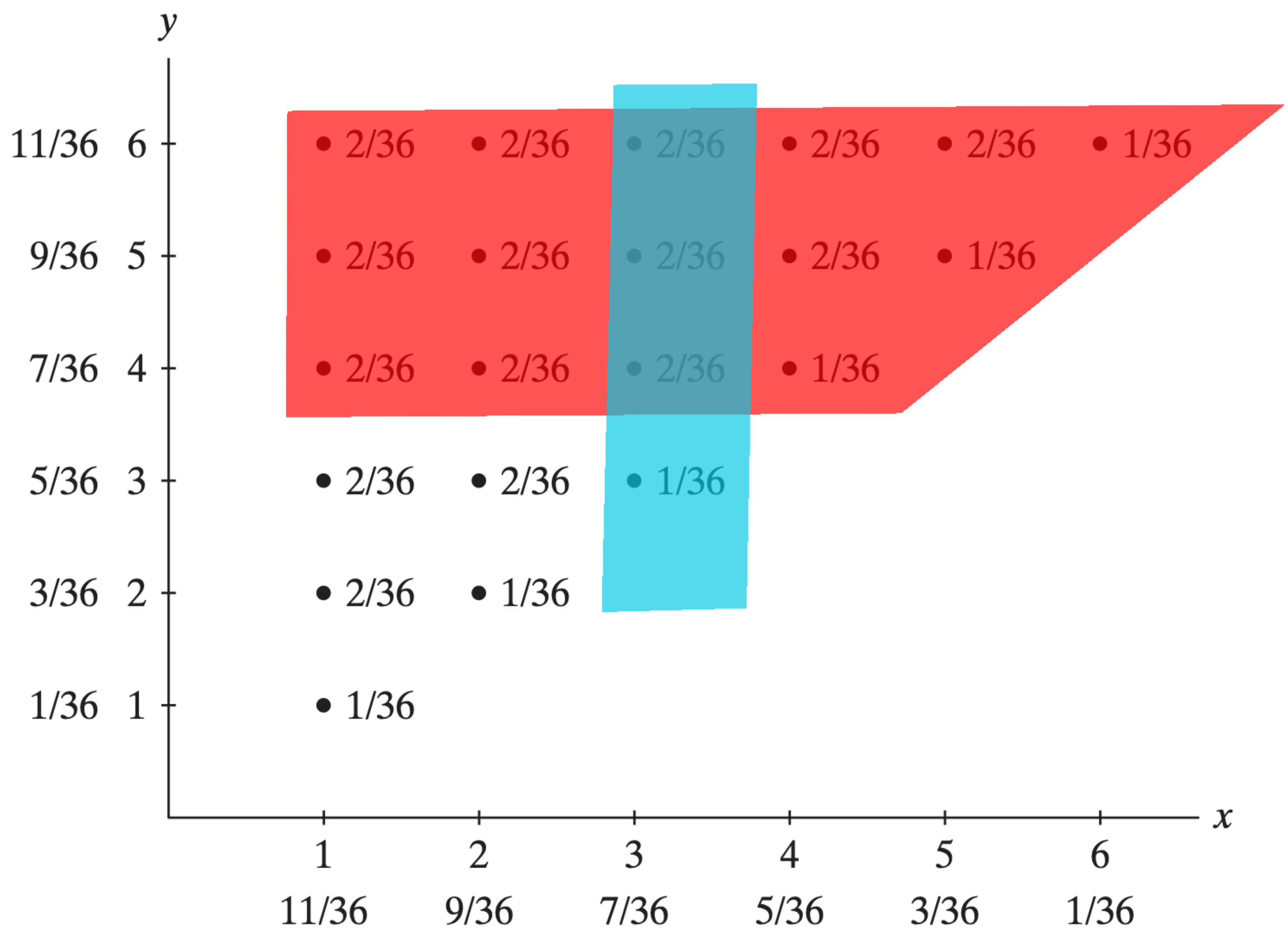
Solution: the notation suggests how to compute these **conditional probabilities**:

$$P(X = 3|Y > 3) = \frac{P(X = 3 \cap Y > 3)}{P(Y > 3)}$$

The region corresponding to $P(Y > 3) = \frac{27}{36}$ is shaded in red (see next slide); the region corresponding to $P(X = 3) = \frac{7}{36}$ is shaded in blue.

The region corresponding to $P(X = 3 \cap Y > 3) = \frac{6}{36}$ is the intersection of the blue and the red regions, so

$$P(X = 3|Y > 3) = \frac{6/36}{27/36} = \frac{6}{27} \approx 0.2222.$$


y
11/36 6
9/36 5
7/36 4
5/36 3
3/36 2
1/36 1
2/36 2/36 2/36 2/36 2/36 1/36
2/36 2/36 2/36 2/36 1/36
2/36 2/36 2/36 1/36
2/36 2/36 1/36
2/36 1/36
1/36
x
1 2 3 4 5 6
11/36 9/36 7/36 5/36 3/36 1/36

Since $P(Y \leq 3 \cap X \geq 4) = 0$, $P(Y \leq 3 | X \geq 4) = 0$.

f) Are X and Y independent?

Solution: why don't we simply use the multiplicative rule to compute $P(X = 3 \cap Y > 3) = P(X = 3)P(Y > 3)$?

Well, we don't yet know if X and Y are **independent**, that is, we don't know if

$$P(X = x, Y = y) = P(X = x)P(Y = y) \quad \text{for all allowable } x, y.$$

As it is, $P(X = 1, Y = 1) = \frac{1}{36}$, but $P(X = 1)P(Y = 1) = \frac{11}{36} \cdot \frac{1}{36}$, so X and Y are **dependent** (this is often the case when the domain of the joint p.d.f./p.m.f. is not rectangular).

Summary

X	**Description**	$P(X=x)$	**Domain**	$\mathrm{E}[X]$	$\mathrm{Var}[X]$
Uniform (Discrete)	Equally likely outcomes	$\frac{1}{b-a+1}$	$a,\dots,b$	$\frac{a+b}{2}$	$\frac{(b-a+2)(b-a)}{12}$
Binomial	Number of successes in n independent trials	$\binom{n}{x}p^x(1-p)^{n-x}$	$0,\dots,n$	np	$np(1-p)$
Poisson	Number of arrivals in a fixed period of time	$\frac{\lambda^x\exp(-\lambda)}{x!}$	$0,1,\dots$	λ	λ

Summary

X	**Description**	$P(X=x)$	**Domain**	$\mathrm{E}[X]$	$\mathrm{Var}[X]$
Geometric	Number of trials until 1^{st} success	$(1-p)^{x-1}p$	$1,2,\ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	Number of trials until k^{th} successes	$\binom{x-1}{k-1}(1-p)^{x-k}p^k$	$k,k+1,...$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$