Random Variables

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib import cm
from scipy.special import comb
np.random.seed(999)
```

The Concept of Random Variables

In any experiment, there are numerous characteristics that can be observed or measured, but in most cases the experimenter will focus on some specific aspect or aspects of a sample. For example, in a study of commuting patterns in a metropolitan area, each individual in a sample might be asked about commuting distance and the number of people commuting in the same vehicle, but not about IQ, income, family size, and other such characteristics.

In general, **each outcome of an experiment can be associated with a number by specifying a rule of association** (e.g., the number among the sample of ten components that fail to last 1000 hours or the total weight of baggage for a sample of 25 airline passengers). Such a rule of association is called a random variable—a variable because different numerical values are possible and random because the observed value depends on which of the possible experimental outcomes results.

> **Definition of Random Variables:** For a given sample space S of some experiment, a **random variable (rv)** is any rule that associates a number with each outcome in S. Mathematically speaking, a random variable is a **function** whose domain is the sample space and whose range is the set of real numbers called the range space of the rv.
>
> $$X: S \to \mathrm{R}.$$

Customarily, we denoted a random variable with uppercase letters, such as X and Y. In contrast to most uses of a lowercase letter, such as x, to denote a variable, we will now use lowercase letters to represent some particular value of the corresponding random variable. The notation $X(s) = x$ means that x is the value associated with the outcome s by the rv X.

Examples of Random Variables

Example 1:

> Consider tossing a coin two times, where the sample space is $S = \{HH, HT, TH, TT\}$. Define X as a random variable representing the number of heads that occur. The value assigned to X for each outcome is as follows:
>
> $$X(HH) = 2,\ X(HT) = X(TH) = 1,\ X(TT) = 0.$$

Example 2:

> Consider an experiment in which batteries are tested until one with a voltage outside the acceptable range is found. Let A denote a battery with a voltage in the acceptable range and F denote a battery with a voltage outside the acceptable range. The sample space for this experiment is a *countably infinite* set:
>
> $$S = \{F, AF, AAF, AAAF, \ldots\}.$$
>
> Define the random variable X as:
>
> $$X(s) = \text{the number of batteries tested before the experiment terminates.}$$
>
> The random variable X assigns values as follows:
>
> $$X(F) = 1,\ X(AF) = 2,\ X(AAF) = 3, \ldots.$$
>
> Thus, every positive integer is a possible value of X, and the range space of X is N (the set of natural numbers).

Example 3:

> In an experiment of tossing two dice, the sample space is:
>
> $$S = \{(1, 1), (1, 2), \ldots, (6, 6)\},$$
>
> which contains 36 outcomes. Many random variables can be defined for this experiment. For example:
>
> - Let X represent the **sum of the face-up values** of the two dice:
>
> $$X(s) = x \text{ where } x \in \{2, 3, 4, \ldots, 12\}.$$
>
> - Let Y represent the **number of dice showing an even number**:
>
> $$Y(s) = y \text{ where } y \in \{0, 1, 2\}.$$
>
> - For a more complex random variable, define Z as:

> $$Z = \begin{cases} 0 & \text{if the sum of the two dice equals 7,} \\ 1 & \text{otherwise.} \end{cases}$$
>
> For example, $Z((1,2)) = 1,\ Z((5,3)) = 0,\ Z((6,1)) = 0.$

Example 4:

> Consider an experiment where the temperature of a chemical solution is measured. The possible outcomes of this experiment are real numbers in an interval, say $S = [20, 30]$, where the temperature is measured in degrees Celsius. Define the random variable X as:
>
> $$X = \text{the temperature of the solution (in °C)}.$$
>
> Since the temperature can take any value in the interval $[20, 30]$, X is a *continuous random variable*. Its range is:
>
> $$\text{Range of } X = [20, 30].$$
>
> X is continuous because the range is *uncountably infinite* since every values in the interval $[20, 30]$ is a possible value of X. For instance, X might be $21, 24.5, 26.134,$ or $28 + \frac{1}{3}$. We will discuss about the continuity of random variables in the next section.

Two Types of Random Variables

From the previous section, we have seen examples of different type of rv's. Notice that some rv's have *discrete* set of possible values and some have *continuous* range of possible values. We can formally classify them in such way as follows:

> A **discrete random variable** is an rv whose range space is a **countable** set. If an rv has a **uncountable** set as the range space, then it is a **continuous random variable**.

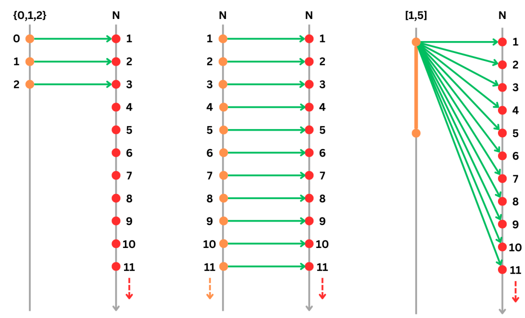
Some may not recognize the term **countable** and **uncountable**. A countable infinite set is a set whose elements can either be put into **one-to-one** correspondence with natural numbers (in this case, it's called countably infinite) or whose element can be count in a finite amount (in this case, it's called finite). Otherwise it is uncountable.

For examples:

- $\text{\set}8, 9, 10$ is a finite set.
- $\text{\set}1, 2, 3, \ldots$ is a countably infinite set.
- $\text{\set}\frac{1}{2}, \frac{2}{2}, \frac{3}{2}, \ldots$ is a countably infinite set.

- \set$\ldots, -2, -1, 0, 1, 2, \ldots$ is a countably infinite set.
- The range $[20, 30]$ and the range $[1, 5) \cup [8, 12]$ are infinite sets.

From the examples in previous section, the rv's in example 1 to 3 are all discrete and the rv in example 4 is continuous. The examples of mapping diagrams of each range space to the natural numbers is shown below. We can see that such a one-to-one correspondence with natural numbers is impossible for the range space of the rv in example 4 since a continuous range is infinitely dense.



The Concept of Probability Distributions

Once we define a random variable (rv), the next important step is to describe how probabilities are assigned to its possible values. This is done using a **probability distribution**.

The probability distribution provides a mathematical description of the likelihood of each possible outcome of the random variable. For example, consider the experiment where we toss two dice and define the rv X to be the sum of the values of the two dice. The range space of X is \set$2, 3, 4, \ldots, 12$. The probability distribution describes how the total probability of the sample space (which is 1 or 100%) is subdivided and assigned to each outcome in the range space. It specifies how much probability is associated with 2, how much is associated with 3, and so on.

We use the following notation for the probability assigned to each value:

$$\begin{aligned}
P(X = 2) &= \text{ the probability that the value of } X \text{ is equal to } 2 = p(2) \\
P(X = 3) &= \text{ the probability that the value of } X \text{ is equal to } 3 = p(3) \\
&\vdots \\
P(X = 12) &= \text{ the probability that the value of } X \text{ is equal to } 12 = p(12)
\end{aligned}$$

In general, $P(X = x)$ or $p(x)$ denotes the probability assigned to the value x. In this case, the value of $p(x)$ assigned to each value x is shown by the following table:

x	2	3	4	5	6	7	8	9	10	11	12	other values
$P(X = x)$	0.028	0.056	0.083	0.111	0.139	0.167	0.139	0.111	0.083	0.056	0.028	0

Note that these values are rounded so they might not be very accurate (in case you wonder why the probability add up to 1.001).

Discrete Probability Distributions

Probability Mass Function

A **discrete probability distribution** is the probability distribution of a **discrete random variable** which can be defined using the *probability mass function*.

> The function $f(x)$ is called the **probability distribution** or **probability mass function (pmf)** of a *discrete* rv X if it satisfies the following conditions:
>
> 1. When evaluated at x it gives the probability that X is equal to x:
>
> $$f(x) = P(X = x).$$
>
> 2. The function satisfies the non-negativity axiom of probability:
>
> $$f(x) \geq 0 \qquad \text{for any value } x.$$
>
> 3. The function satisfies the normalization axiom of probability, ensuring the total probability over all possible values equals 1:
>
> $$\sum_{\forall x} f(x) = 1.$$

Example Suppose a fair coin is tossed three times and we are interested in the number of heads. So we define an rv X to be the number of heads. To find the probability distribution of X, first consider the sample space of this experiment:

$$S = \backslash\mathrm{set} HHH, HHT, HTH, THH, HTT, THT, TTH, TTT.$$

We can clearly see that the range space of X is $\backslash\mathrm{set}0, 1, 2, 3$. To find the probability distribution, we need to find the probability of each value in the range space. This can be done as follows:

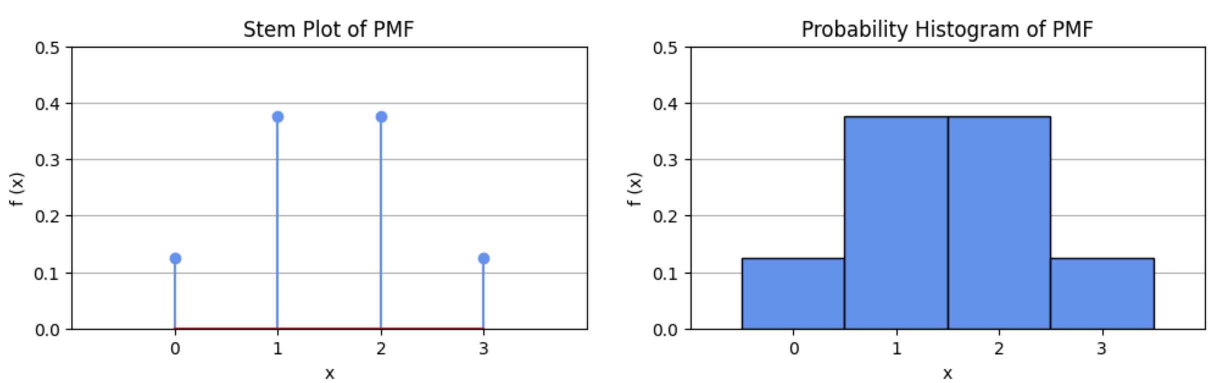
$$f(0) = P(X = 0) = P(\backslash\mathrm{set}TTT) = 1/8,$$
$$f(1) = P(X = 1) = P(\backslash\mathrm{set}HTT, THT, TTH) = 3/8,$$
$$f(2) = P(X = 2) = P(\backslash\mathrm{set}HHT, HTH, THH) = 3/8,$$
$$f(3) = P(X = 3) = P(\backslash\mathrm{set}HHH) = 1/8.$$

Hence, the probability distribution or the pmf of X is

$$f(x) = \begin{cases} \frac{1}{8} & \text{if } x = 0, 3 \\ \frac{3}{8} & \text{if } x = 1, 2 \\ 0 & \text{elsewhere} \end{cases}$$

The figures shown below illustrate the *line plot* (or *stem plot*) and *probability histogram* of the pmf of X.

In [2]:
```
fig, axs = plt.subplots(1, 2, figsize=(12,3))

x = [0,1,2,3]
fx = [1/8, 3/8, 3/8, 1/8]

axs[0].stem(x, fx, linefmt='cornflowerblue')
axs[1].bar(x, fx, width=1, edgecolor='black', color='cornflowerblue')

for ax in axs:
    ax.set(
        xlabel='x', ylabel='f (x)',
        xlim=[-1,4], ylim=[0,0.5],
        xticks=[0,1,2,3], yticks=[0,0.1,0.2,0.3,0.4,0.5],
        axisbelow=True
    )
    ax.grid(True, axis='y')

axs[0].set_title("Stem Plot of PMF")
axs[1].set_title("Probability Histogram of PMF");
```



Note that the probability histogram is drawn slightly differently from a standard histogram: instead of dividing the values into bins, each rectangle directly corresponds to the probability of a specific value x and is drawn directly above the point x.

Cumulative Distribution Function for Discrete RV's

For some fixed value x, we often wish to compute the probability that the observed value of X will be at most x ($X \leq x$).

For example, consider the pmf from the previous example:

$$f(x) = \begin{cases} 1/8 & \text{if } x = 0, 3 \\ 3/8 & \text{if } x = 1, 2 \\ 0 & \text{elsewhere} \end{cases}$$

The probability that X is at most 1 is then

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{1}{8} + \frac{3}{8} = \frac{1}{2}.$$

In this context, we can also ask for the probability that X is at most 1.5 which is the same as the probability that X is at most 1:

$$P(X \leq 1.5) = P(X \leq 1) = \frac{1}{2}.$$

And in fact for any $1 \leq X < 2$, the probability that X is at most x is the same value $P(X \leq 1)$.

In general, We can defined a function that specifies the value $P(X \leq x)$ for every values x. Such function is called the *cumulative distribution function*.

> The **cumulative distribution function (cdf)** $F(x)$ of a discrete rv X with pmf $f(x)$ is a function that when evaluated at x gives the probability that X is less than or equal to x and is defined for every real number x by
>
> $$F(x) = P(X \leq x) = \sum_{t \leq x} f(t).$$

Example Consider tossing a fair coin three times. We already determine the probability distribution of the random variable X representing number of heads as

$$f(x) = \begin{cases} 1/8 & \text{if } x = 0, 3 \\ 3/8 & \text{if } x = 1, 2 \\ 0 & \text{elsewhere} \end{cases}$$

Let's determine the value of cdf $F(x)$ for each value x in the range space of X:

$$F(0) = P(X \le 0) = P(X = 0) = 1/8,$$
$$F(1) = P(X \le 1) = P(X = 0) + P(X = 0) = 1/2,$$
$$F(2) = P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2) = 7/8,$$
$$F(3) = P(X \le 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 1.$$

As we discussed earlier, for any other values of x, $F(x)$ will be equal to the value of F at the closet value to the left of x that is in the range space of X. For instances,

$$F(1.999) = P(X \le 1.999) = P(X \le 1) = F(1) = 1/8,$$
$$F(10000) = P(X \le 10000) = P(X \le 3) = F(3) = 1.$$

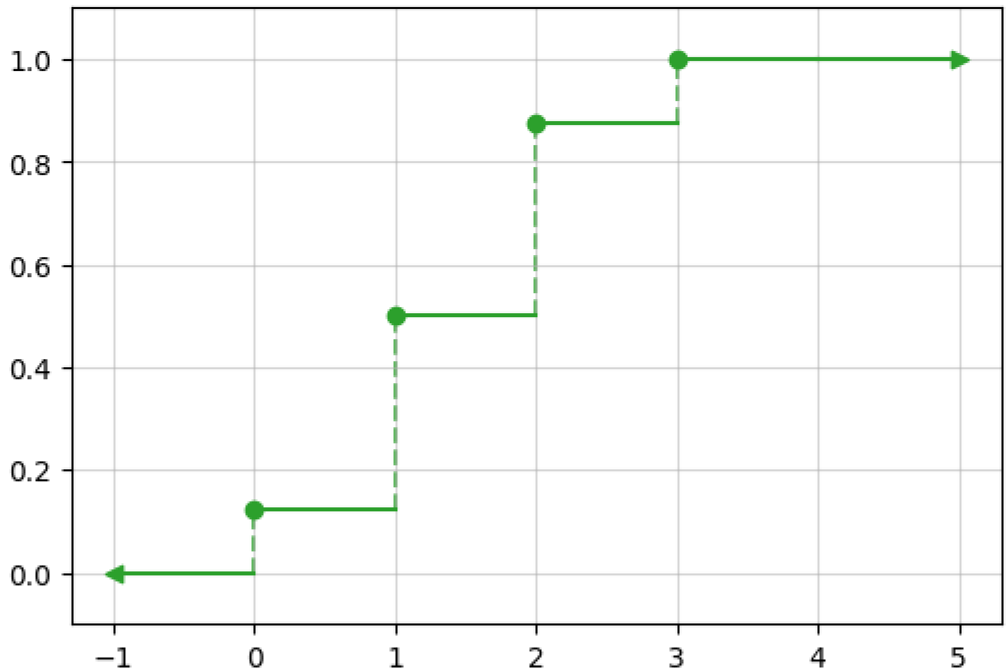
Therefore, the cdf is

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/8 & \text{if } 0 \le x < 1 \\ 1/2 & \text{if } 1 \le x < 2 \\ 7/8 & \text{if } 2 \le x < 3 \\ 1 & \text{if } x \ge 3 \end{cases}$$

And the graph of this cdf is as shown below:

In [3]:
```
fig, ax = plt.subplots(figsize=(6,4))

rx = np.array([-1, 0  , 1   ,2   , 3, 4])
Fx = np.array([0 , 1/8, 1/2 ,7/8 , 1, 1])

for x,F in zip(rx,Fx):
    xs = np.linspace(x, x+1, 10)
    Fs = np.array([F for _ in range(10)])
    ax.plot(xs, Fs, color='tab:green')

for x,F_,F in zip(rx[1:-1], Fx[:-2], Fx[1:-1]):
    ax.plot([x,x], [F_,F], color='tab:green', linestyle='--', alpha=0.75, zorder=-1

ax.scatter(rx[1:-1], Fx[1:-1], c='tab:green', marker='o')
ax.scatter([-1], [0], c='tab:green', marker='<')
ax.scatter([ 5], [1], c='tab:green', marker='>')

ax.set(
    ylim=[-0.1,1.1],
    axisbelow=True
)
ax.grid(True, alpha=0.5)
```

Processing math: 100%



Properties of Cumulative Distribution Functions for Discrete RV's

The cumulative distribution function (cdf) $F(x)$ of a *discrete* rv $X(s) = x$ with probability mass function (pmf) $f(x)$ satisfies the following properties:

1. $F(x)$ is a **step function**.

2. $F(x)$ is a monotone **non-decreasing function**.

3. The limit of $F(x)$ to the left is $1 = 0$ and to right is 1, i.e.

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1.$$

   This further implies $0 \leq F(x) \leq 1$.

4. For any two numbers a and b where $a \leq b$

$$\begin{aligned} P(a \leq X \leq b) &= P(X \leq b) - P(X < a) \\ &= F(b) - F(a^-) \end{aligned}$$

   where $F(a^-) = \lim_{x \to a^-} F(x)$. By taking $a = b$ we get $P(X = a) = F(a) - F(a^-)$.

Example Consider testing batteries until one with a voltage outside the acceptable range is found. Let A denote a battery with a voltage in the acceptable range and F denote a battery with a voltage outside the acceptable range. The sample space for this experiment is:

$$S = \backslash\text{set}F, AF, AAF, \ldots.$$

Let X be an rv representing the number of batteries tested before a batteries with unacceptable voltage is found. If 1 in 10 batteries have a voltage outside the acceptable range. then,

$$X(\overbrace{AA\ldots A}^{n-1}F) = n, \quad \text{for } n = 1, 2, 3, \ldots$$

Suppose that the voltage of batteries is independent of each other. We can calculate the pmf of X using multiplication rule:

$$f(n) = P(X = n) = P(\text{\setAA}\overbrace{\ldots A}^{n-1}F) = (0.9)^{n-1} \cdot 0.1, \quad \text{for } n = 1, 2, 3, \ldots$$

And the cdf can be calculate as follows:

$$\begin{aligned} F(n) = P(X \leq n) &= \sum_{i=1}^{n} (0.9)^{i-1} \cdot 0.1 \\ &= \frac{0.1 \cdot (1 - 0.9^n)}{1 - 0.9} \\ &= 1 - 0.9^n, \quad \text{for } n = 1, 2, 3, \ldots \end{aligned}$$

For other values of x, the cdf is defined as $F(x) = F(\lfloor x\ rfloor)$, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x.

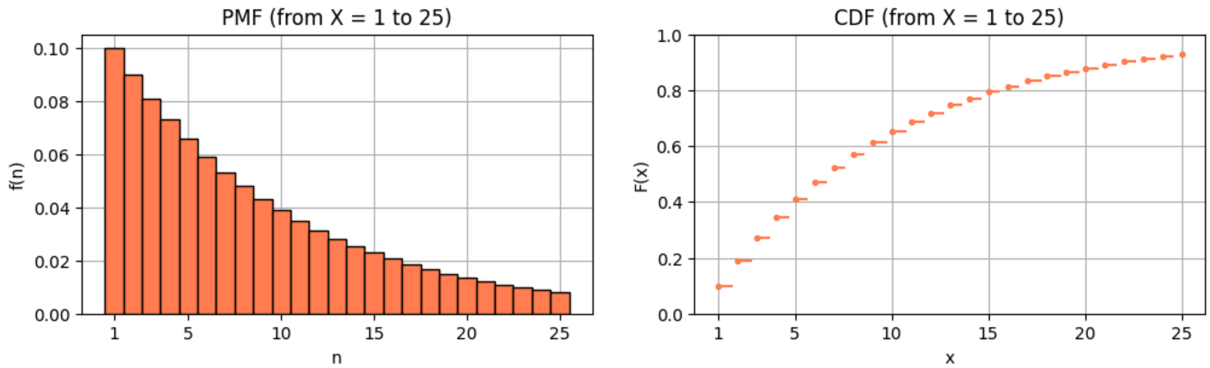
The graphs of both the pmf and the cdf of X are shown below:

In [4]:
```
def AF_pmf(n):
    return 0.1 * (0.9 ** (n * (n == np.floor(n)) - 1))

def AF_cdf(n, dis=False):
    y = 1 - (0.9 ** np.floor(n))
    if dis:
        y[:-1][np.diff(y) >= 1e-6] = np.nan
    return y
```

In [5]:
```
fig, axs = plt.subplots(1, 2, figsize=(12,3))

xs = np.arange(1,26)
xss = np.linspace(1,25,100)

axs[0].bar(
    xs, AF_pmf(xs),
    width=1, color='coral' ,edgecolor='black'
)
axs[0].set(axisbelow=True, xticks=[1] + list(range(5,26,5)), xlabel='n', ylabel='f(
axs[0].grid(True)

axs[1].plot(
    xss, AF_cdf(xss, True),
    color='coral'
)
axs[1].scatter(
```

Processing math: 100%

```
    xs, AF_cdf(xs),
    c='coral', s=8
)
axs[1].set(axisbelow=True, xticks=[1] + list(range(5,26,5)), yticks=np.arange(0,1.0
axs[1].grid(True)
```



Now if we want to know the probability that we will have to test 5 to 10 batteries before an unacceptable battery is found, we calculate:

$$
\begin{aligned}
P(5 \le X \le 10) &= F(10) - F(5^-) \\
&= F(10) - F(4.999\ldots) \\
&= F(10) - F(4) \\
&\approx 0.6561 - 0.3487 = 0.3074.
\end{aligned}
$$

Continuous Probability Distributions

Probability Density Function

For a continuous random variable (rv), we cannot use a probability mass function (pmf) to describe the probability assigned to each possible value of the rv. The key reason lies in the nature of continuous random variables: their range space is an infinite set of numbers. If we attempt to subdivide the total probability of 1 and assign it to each individual outcome in the range space, *we would need to divide 1 into infinitely many parts*. ***This results in the probability of any single specific value being 0***.

This creates a seeming *paradox*: we know that these values are possible outcomes of the random variable, but their probabilities are 0 if we focus on exactly any one of them. To resolve this, we describe the **probability of a continuous random variable over an interval of values** rather than a specific value, using a **probability density function (pdf)** instead of a pmf. The pdf allows us to calculate probabilities by integrating over intervals, providing a coherent way to describe distributions of continuous random variables.

> The function $f(x)$ is a **probability distribution** or **probability density function (pdf)** for the *continuous* random variable X, defined over the set of real numbers, if
>
> 1. The area under the curve of f in the interval $[a, b]$ is equal to the probability that X falls within that interval:
>
> $$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$
>
> 2. The function satisfies the non-negativity axiom of probability:
>
> $$f(x) \geq 0 \qquad \text{for all } x \in \mathrm{R}.$$
>
> 3. The function satisfies the normalization axiom of probability, that is the total area under the curve of f over the entire real line must equal 1, ensuring that the probability over all possible outcomes sums to 1:
>
> $$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Since we define pdf using integral, in doesn't matter whether the endpoints of the interested interval are included or not, that is

$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = \int_a^b f(x)\, dx.$$

Example Suppose that the error in the reaction temperature, in $^\circ C$, for a controlled laboratory experiment is a continuous random variable X having the probability density function

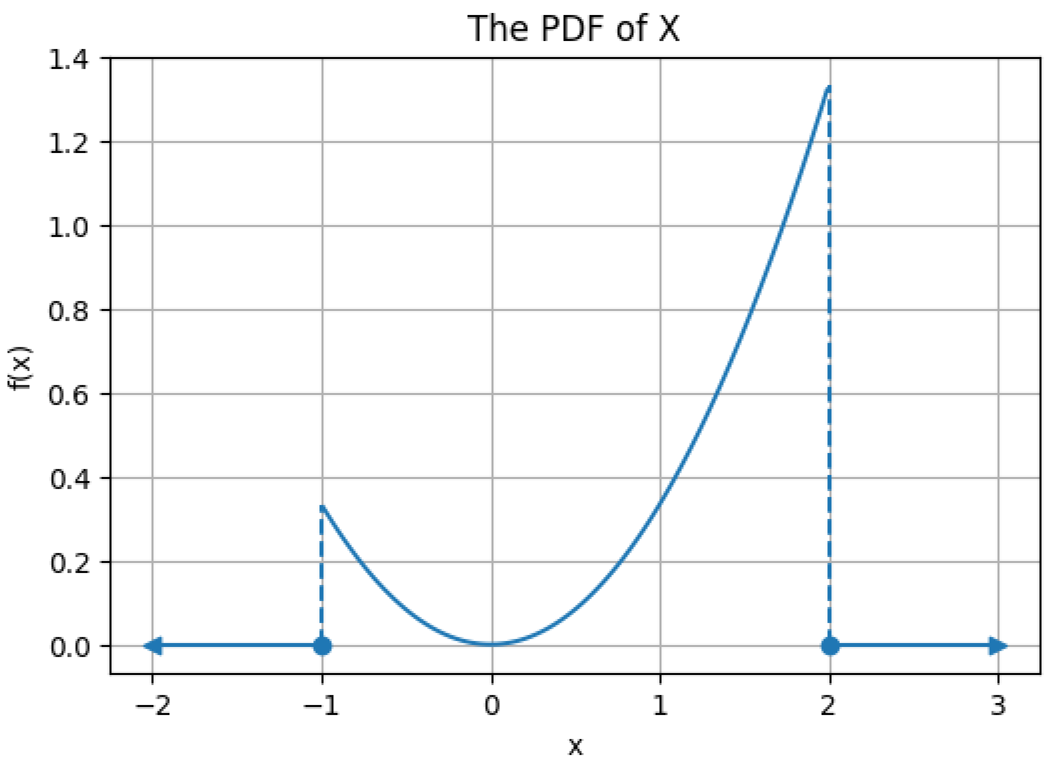
$$f(x) = \begin{cases} \frac{x^2}{3} & \text{if} -1 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

The graph of this pdf is shown below.

In [6]:
```
def temp_pdf(x):
    y = (((x > -1) & (x < 2)) * x) ** 2 / 3
    return y
```

In [7]:
```
fig, ax = plt.subplots(figsize=(6,4))

xs1 = np.concatenate((np.linspace(-2,-1,10), np.array([np.nan]), np.linspace(2,3,10
xs2 = np.linspace(-0.99,1.99,100)

ax.plot(xs1, temp_pdf(xs1), color='tab:blue')
ax.plot(xs2, temp_pdf(xs2), color='tab:blue')
ax.plot([-1,-1,np.nan,2,2], [0,1/3,np.nan,0,4/3], color='tab:blue', linestyle='--')
```

```python
ax.scatter([-1,2], [0,0])
ax.scatter([-2], [0], marker='<', color='tab:blue')
ax.scatter([3 ], [0], marker='>', color='tab:blue')

ax.set(
    axisbelow=True,
    title='The PDF of X', xlabel='x', ylabel='f(x)'
)
ax.grid(True)
```



To verify that f is a valid pdf, consider:

1. Clearly, $f(x) \geq 0$ for all $x \in \mathbb{R}$.
2. Consider the total probability

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-1}^{2} \frac{x^2}{3}\, dx$$

$$= \left[\frac{x^3}{9}\right]_{-1}^{2}$$

$$= \frac{8}{9} - (-\frac{1}{9}) = 1.$$

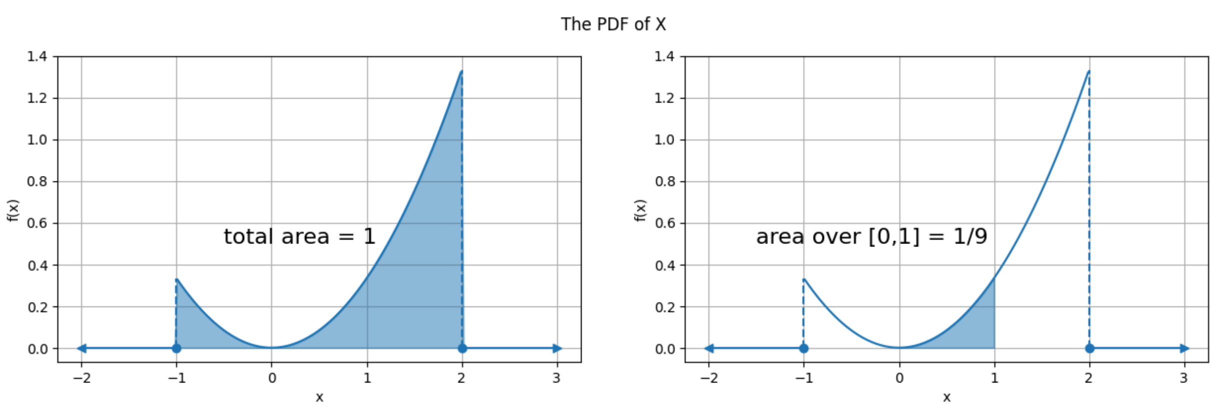
Thus, f is a valid pdf for X.

Suppose we want to know the probability that the temperature falls between 0 and 1, we calculate

Processing math: 100%

$$P(0 \le X \le 1) = \int_0^1 f(x)\, dx = \int_0^1 \frac{x^2}{3}\, dx = \frac{1}{9}.$$

In [8]:
```
fig, axs = plt.subplots(1, 2, figsize=(15,4))

xs1 = np.concatenate((np.linspace(-2,-1,10), np.array([np.nan]), np.linspace(2,3,10
xs2 = np.linspace(-0.99,1.99,100)

for ax in axs:
    ax.plot(xs1, temp_pdf(xs1), color='tab:blue')
    ax.plot(xs2, temp_pdf(xs2), color='tab:blue')
    ax.plot([-1,-1,np.nan,2,2], [0,1/3,np.nan,0,4/3], color='tab:blue', linestyle='
    ax.scatter([-1,2], [0,0])
    ax.scatter([-2], [0], marker='<', color='tab:blue')
    ax.scatter([3 ], [0], marker='>', color='tab:blue')

    ax.set(axisbelow=True, xlabel='x', ylabel='f(x)')
    ax.grid(True)

xs = np.linspace(-2,3,200)
axs[0].fill_between(xs, xs*0, temp_pdf(xs), color='tab:blue', alpha=0.5)
axs[0].text(-0.5, 0.5, 'total area = 1', fontsize=16)

xs = np.linspace(0,1,200)
axs[1].fill_between(xs, xs*0, temp_pdf(xs), color='tab:blue', alpha=0.5)
axs[1].text(-1.5, 0.5, 'area over [0,1] = 1/9', fontsize=16)
fig.suptitle('The PDF of X');
```



Cumulative Density Function for Continuous RV's

Like with discrete rv's, we might interest in the probability that a continuous rv will be at most some number x (P(X\le x)).

For example, consider a continuous rv X with pdf given by

Processing math: 100%

$$f(x) = \begin{cases} \frac{1}{5} & \text{if } 0 \le x \le 5 \\ 0 & \text{elsewhere} \end{cases}$$

The probability that X will be at most 2.5 is

$$P(X \le 2.5) = P(-\infty \le X \le 2.5) = \int_{-\infty}^{2.5} f(x)\, dx$$

Thus,

$$P(X \le 2.5) = \int_0^{2.5} \frac{1}{5}\, dx = 2.5 \cdot \frac{1}{5} = \frac{1}{2}.$$

In general, we can defined a function that specifies the value $P(X \le x)$ for every values x as a *cumulative distribution function* similar to cdf for discrete rv's but defined differently.

> The **cumulative distribution function** of a *continuous* random variable X with density function $f(x)$ is a function $F(x)$ define for every real numbers x as
>
> $$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt.$$

Example Consider an rv X discussed in the previous example whose pdf is given by

$$f(x) = \begin{cases} \frac{x^2}{3} & \text{if } -1 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

So the cdf of X for $-1 < x < 2$ is

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-1}^{x} \frac{t^2}{3}\, dt$$

$$= \left[\frac{t^3}{9}\right]_{-1}^{x} = \frac{x^3 + 1}{9}.$$

Therefore,

$$F(x) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{x^3+1}{9} & \text{if } -1 \le x < 2 \\ 1 & \text{if } x \ge 2 \end{cases}$$

Suppose we want to know the probability that X is at most 1, then we calculate

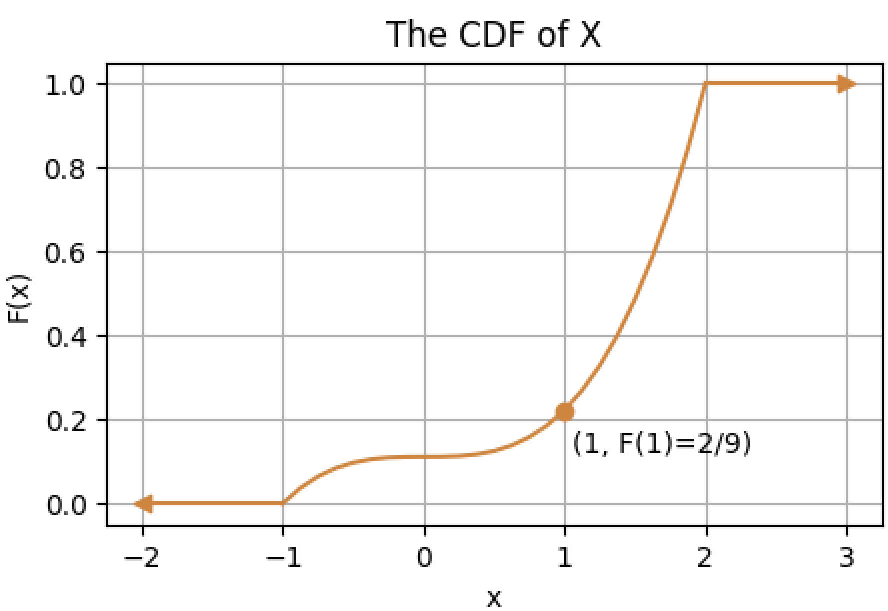
$$P(X \leq 1) = F(1) = \frac{1^3 + 1}{9} = \frac{2}{9}.$$

For another example, if we want to know $P(X > 1)$, then we can calculate

$$P(X > 1) = 1 - F(1) = 1 - \frac{2}{9} = \frac{7}{9}.$$

In [9]:
```
def temp_cdf(x):
    return ((x ** 3) + 1) / 9
```

In [10]:
```
fig, ax = plt.subplots(figsize=(5,3))

xs = np.linspace(-1,2,25)

ax.plot(xs, temp_cdf(xs), color='peru')
ax.plot([-2,-1], [0,0], color='peru')
ax.plot([2,3], [1,1], color='peru')
ax.scatter([-2], [0], color='peru', marker='<')
ax.scatter([3], [1], color='peru', marker='>')
ax.scatter([1], [temp_cdf(1)], color='peru')
ax.text(1.05, temp_cdf(1) - 0.1, "(1, F(1)=2/9)")
ax.set(axisbelow=True, title="The CDF of X", xlabel="x", ylabel="F(x)")
ax.grid(True)
```



Properties of Cumulative Distribution Functions for Discrete RV's

The cumulative distribution function (cdf) $F(x)$ of a *continuous* rv $X(s) = x$ with probability density function (pdf) $f(x)$ satisfies the following properties:

1. $F(x)$ is a **continuous function** over the real numbers.

2. $F(x)$ is a monotone **non-decreasing function**.

Processing math: 100%

3. The limit of $F(x)$ to the left is $1 = 0$ and to right is 1, i.e.

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1.$$

This further implies $0 \le F(x) \le 1$.

4. The derivative of $F(x)$ is $f(x)$ (if the derivative exist). This property comes directly from the fundamental theorem of calculus.

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}.$$

5. Since the total probability is always equal to 1 then $P(X \le x) + P(X > x) = 1$, that is

$$P(X > x) = 1 - F(x).$$

6. For any two numbers a and b where $a \le b$

$$\begin{aligned} P(a \le X \le b) &= P(a < X \le b) = P(a \le X < b) = P(a < X < b). \\ &= F(b) - F(a). \end{aligned}$$

Parameters of a Probability Distribution

The "parameters of a probability distribution" are specific **numerical values** that **define the shape and characteristics** of a given distribution, such as its central tendency (mean) and spread (variance).

> Suppose a pdf/cdf $f(x)$ depends on a quantity that can be assigned any one of a number of possible values (excluding x itself), with each different value determining a different probability distribution. Such a quantity is called a **parameter** of the distribution. The collection of all probability distributions for different values of the parameter is called a **family of probability distributions**.

Suppose $f(x)$ depends on $a_1, a_2, \ldots,$ and a_n, we write $f(x; a_1, a_2, \ldots, a_n)$ rather than just $f(x)$.

Discrete Example Consider tossing a coin which can either be *fair* or *unfair*. Suppose that the probability of tossing a head is p and the probability of tossing a tail is $1 - p$. Let X be a discrete rv representing the number of heads from 3 tosses, so the range space of X is \set$0, 1, 2, 3$ and the probability of each value is shown in the table below:

x	$P(X = x)$
0	$(1 - p)^3$
1	$p(1 - p)^2$

x	$P(X = x)$
2	$p^2(1-p)$
3	p^3

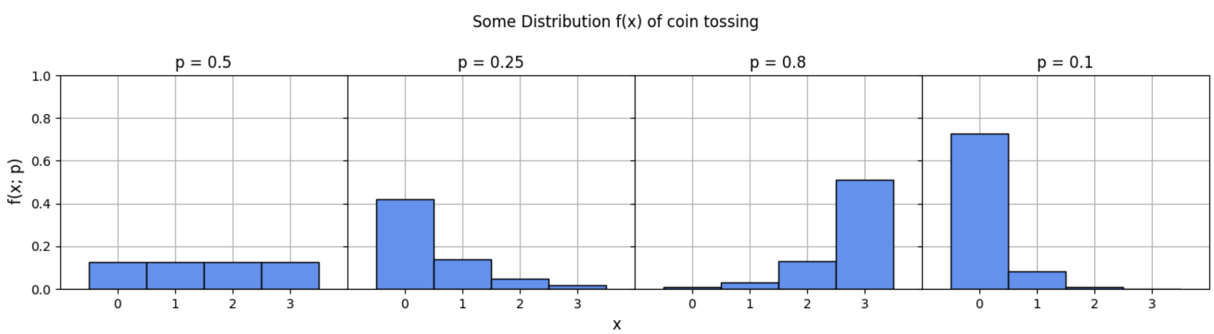
So the pmf of X is given by

$$f(x;p) = \begin{cases} (1-p)^3 & \text{if } x = 0 \\ p(1-p)^2 & \text{if } x = 1 \\ p^2(1-p) & \text{if } x = 2 \\ p^3 & \text{if } x = 3 \\ 0 & \text{elsewhere} \end{cases} = \begin{cases} p^x(1-p)^{3-x} & \text{if } x = 0, 1, 2, 3 \\ 0 & \text{elsewhere} \end{cases}$$

We can see that this distribution is determined by one parameter p.

In [11]:
```
def coin_pmf(x, p=0.5):
    return (p ** x) * ((1 - p) ** (3 - x))
```

In [12]:
```
fig, axs = plt.subplots(1, 4, figsize=(16,3), sharey=True)
fig.subplots_adjust(wspace=0)
fig.suptitle("Some Distribution f(x) of coin tossing", x=0.5, y=1.1)
fig.supxlabel("x", y=-0.05)
fig.supylabel("f(x; p)", x=0.09)

xs = np.array([0,1,2,3])

for i,p in enumerate([0.5,0.25,0.8,0.1]):
    axs[i].bar(xs, coin_pmf(xs, p), width=1, color='cornflowerblue', edgecolor='bla
    axs[i].set(title=f"p = {p}", ylim=[0,1], xlim=[-1,4], xticks=[0,1,2,3], axisbel
    axs[i].grid()
```



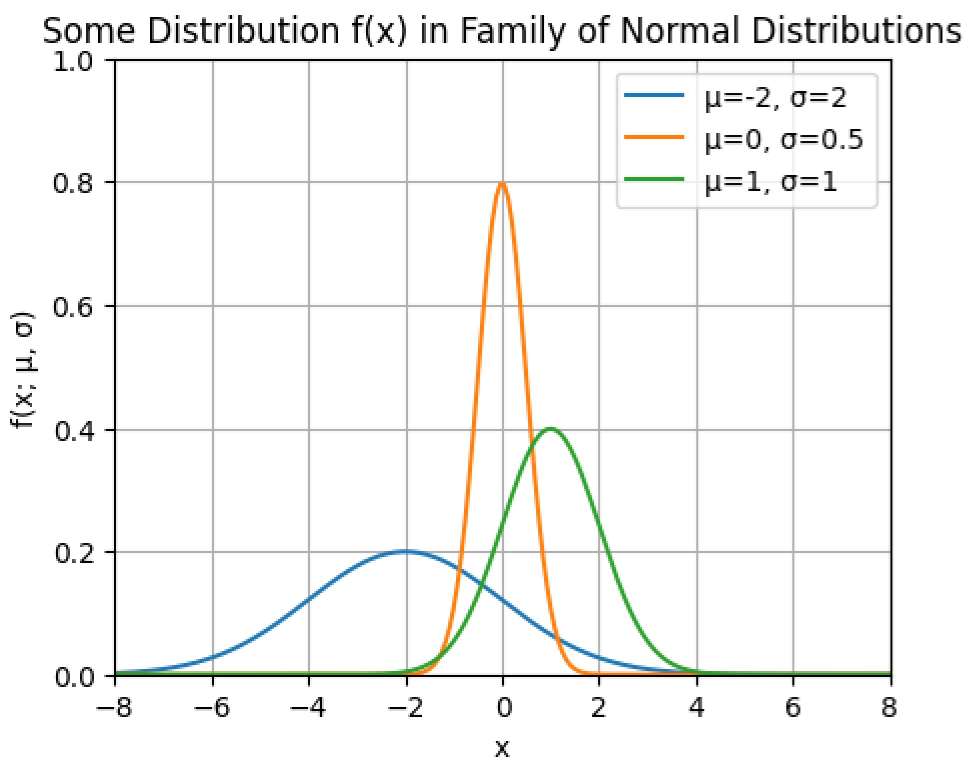
Continuous Example *Normal Distribution* is one of the most important probability distribution in statistics (which we will go in detail in the next chapter). The pdf of a normal distributed random variable is given by

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

This distribution is determined by two parameters: the mean μ and the standard deviation σ of the random variable.

In [13]:
```
def norm_pdf(x, mu=0, sigma=1):
    return np.exp((-((x - mu) / sigma)**2) / 2) / (np.sqrt(2 * np.pi * sigma**2))
```

In [14]:
```
fig, ax = plt.subplots(figsize=(5,4))

xs = np.linspace(-8,8,200)
ms = [-2,   0, 1]
ss = [ 2, 0.5, 1]

for mu,sigma in zip(ms, ss):
    ax.plot(xs, norm_pdf(xs, mu, sigma), label=f"μ={mu}, σ={sigma}")

ax.set(
    title="Some Distribution f(x) in Family of Normal Distributions",
    xlabel='x', ylabel='f(x; μ, σ)',
    xlim=[-8,8], ylim=[0,1],
    axisbelow=True
)
ax.legend()
ax.grid()
```



Joint Probability Distributions

Processing math: 100%

There are many experimental situations in which more than **one random variable will be of interest** to an investigator. We first consider *joint probability distributions* for two discrete rv's, then for two continuous variables, and finally for more than two variables.

Two Discrete Random Variables

If X and Y are two discrete random variables, the probability distribution for their simultaneous occurrence can be represented by a function $f(x, y)$ for any pair of values (x, y) within the range of the random variables X and Y. It is customary to refer to this function as the **joint probability distribution** of X and Y.

> The function $f(x, y)$ is a **joint probability mass function** of discrete random variables X and Y if
>
> 1. $f(x, y)$ gives the probability that outcomes x and y occur at the same time.
>
> $$f(x, y) = P(X = x \text{ and } Y = y).$$
>
> For a shorter notation, we can write $P(X = x \text{ and } Y = y) = P(X = x, Y = y)$.
>
> 2. $f(x, y)$ is non-negative.
>
> $$f(x, y) \geq 0 \quad \text{for all } (x, y).$$
>
> 3. $f(x, y)$ is normalized, ensuring that total probability over all possible values of x and y equals 1.
>
> $$\sum_{\forall x} \sum_{\forall y} f(x, y) = 1.$$

Let A be any set (also called ***region***) consisting of pairs of values (x, y) (e.g., $A = \set(x, y) | x + y = 5$ or $A = \set(x, y) | \max(x, y) \leq 3$). Then we can write the probability of A as $P[(X, Y) \in A]$ obtained by summing the joint pmf over all pairs in A:

$$P[(X, Y) \in A] = \sum_{(x,y) \in A} \sum f(x, y).$$

Example Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected.

The possible values of pair (x, y) are $(0, 0), (1, 0), (0, 1), (1, 1), (2, 0),$ and $(0, 2)$. The probability of each pair can be calculated as follows:

Processing math: 100%

$$n(S) = \binom{8}{2} = 28$$

$$P(X=0, Y=0) = \frac{\binom{3}{2}}{28} = \frac{3}{28}$$

$$P(X=1, Y=0) = \frac{\binom{3}{1}\binom{3}{1}}{28} = \frac{9}{28}$$

$$P(X=0, Y=1) = \frac{\binom{2}{1}\binom{3}{1}}{28} = \frac{6}{28}$$

$$P(X=1, Y=1) = \frac{\binom{3}{1}\binom{2}{1}}{28} = \frac{6}{28}$$

$$P(X=2, Y=0) = \frac{\binom{3}{2}}{28} = \frac{3}{28}$$

$$P(X=0, Y=2) = \frac{\binom{2}{2}}{28} = \frac{1}{28}$$

So we might define joint pmf as

$$f(x,y) = \begin{cases} 1/28 & \text{if } (x,y) = (0,2) \\ 3/28 & \text{if } (x,y) = (0,0), (2,0) \\ 6/28 & \text{if } (x,y) = (0,1), (1,1) \\ 9/28 & \text{if } (x,y) = (1,0) \\ 0 & \text{elsewhere} \end{cases}$$

Or in general, we can write:

$$f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{28}$$

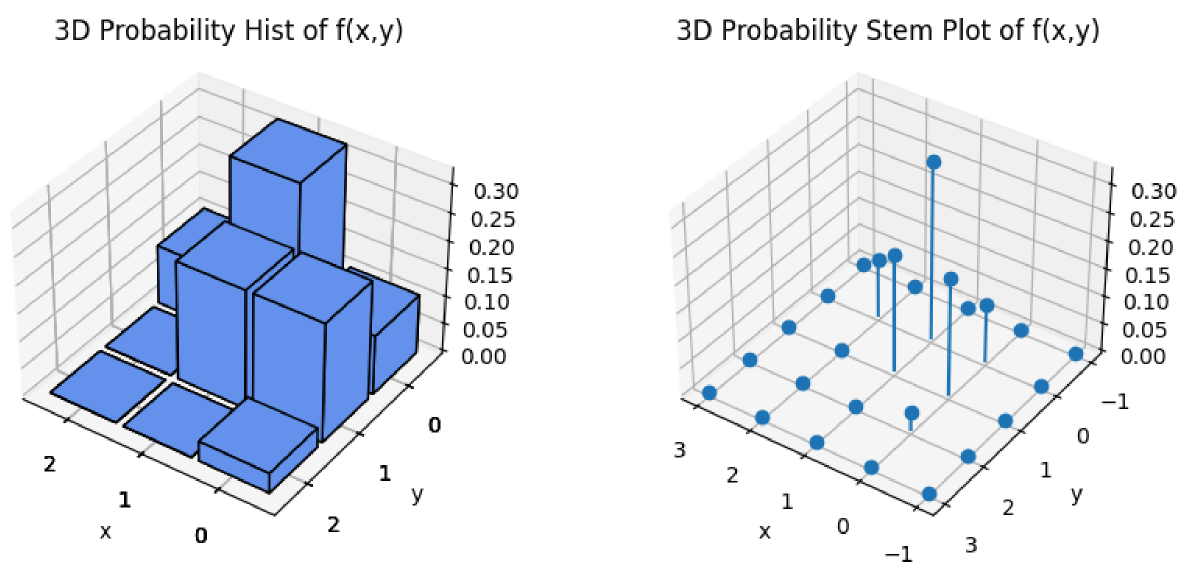
for $x = 0, 1, 2$, $y = 0, 1, 2$ and $x + y \leq 2$ or else $f(x,y) = 0$.

In [15]:
```
def pen_pmf(x, y):
    binom_3_x = comb(3, x)
    binom_2_y = comb(2, y)
    binom_3_2_xy = comb(3, 2 - x - y)
    return (x + y <= 2) * (binom_3_x * binom_2_y * binom_3_2_xy) / 28
```

In [16]:
```
fig = plt.figure(figsize=(10,4))

ax = fig.add_subplot(121, projection='3d')
x = [0,1,2]
```

Processing math: 100%

```
y = [0,1,2]
x, y = np.meshgrid(x, y)
x, y = x.ravel(), y.ravel()
z = np.zeros_like(x)
height = pen_pmf(x, y)

ax.bar3d(x, y, z, 0.9, 0.9, height, shade=False, color='cornflowerblue', edgecolor=
ax.view_init(elev=36.5 ,azim=125)
ax.set(xticks=x+0.5, yticks=y+0.5, xticklabels=x, yticklabels=y, xlabel='x', ylabel

ax2 = fig.add_subplot(122, projection='3d')
x = np.arange(-1,4)
y = np.arange(-1,4)
x, y = np.meshgrid(x, y)
x, y = x.ravel(), y.ravel()
z = pen_pmf(x, y)

ax2.stem(x, y, z, basefmt=" ")
ax2.view_init(elev=36.5 ,azim=125)
ax2.set(xlabel='x', ylabel='y', title='3D Probability Stem Plot of f(x,y)');
```



Two Continuous Random Variables

When X and Y are continuous random variables, the joint density function f(x, y) is a surface lying above the xy-plane, and $P[(X, Y) \in A]$, where A is any region in the xy-plane, is equal to the volume of under the $f(x, y)$ surface bounded by the region A.

> The function $f(x, y)$ is a **joint probability density function** of continuous random variables X and Y if
>
> 1. The volume under $f(x, y)$ bounded by the region A gives the probability that outcomes (x, y) falls within A.

$$P[(X, Y) \in A] = \iint_A f(x, y)\, dxdy.$$

2. $f(x, y)$ is non-negative.

$$f(x, y) \geq 0 \quad \text{for all } (x, y).$$

3. $f(x, y)$ is normalized, ensuring that area under $f(x, y)$ over R^2 is 1.

$$\iint_{\mathrm{R}^2} f(x, y)\, dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dxdy = 1.$$

Example Define a joint probability density function of continuous rv's X and Y as

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This function is clearly non-negative. So, to verify that this is legitimate pdf, consider

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dxdy &= \int_0^1\int_0^1 \frac{6}{5}(x + y^2)\, dxdy \\
&= \int_0^1\int_0^1 \frac{6}{5}x + \frac{6}{5}y^2\, dxdy \\
&= \int_0^1\int_0^1 \frac{6}{5}x\, dxdy + \int_0^1\int_0^1 \frac{6}{5}y^2\, dxdy \\
&= \int_0^1 \left[\frac{6}{5}xy\right]_{y=0}^1 dx + \int_0^1 \left[\frac{6}{5}y^2x\right]_{x=0}^1 dy \\
&= \int_0^1 \frac{6}{5}x\, dx + \int_0^1 \frac{6}{5}y^2\, dy \\
&= \frac{6}{5}\left[\frac{x^2}{2}\right]_0^1 + \frac{6}{5}\left[\frac{y^3}{3}\right]_0^1 \\
&= \frac{6}{10} + \frac{6}{15} \; = \; 1.
\end{aligned}$$

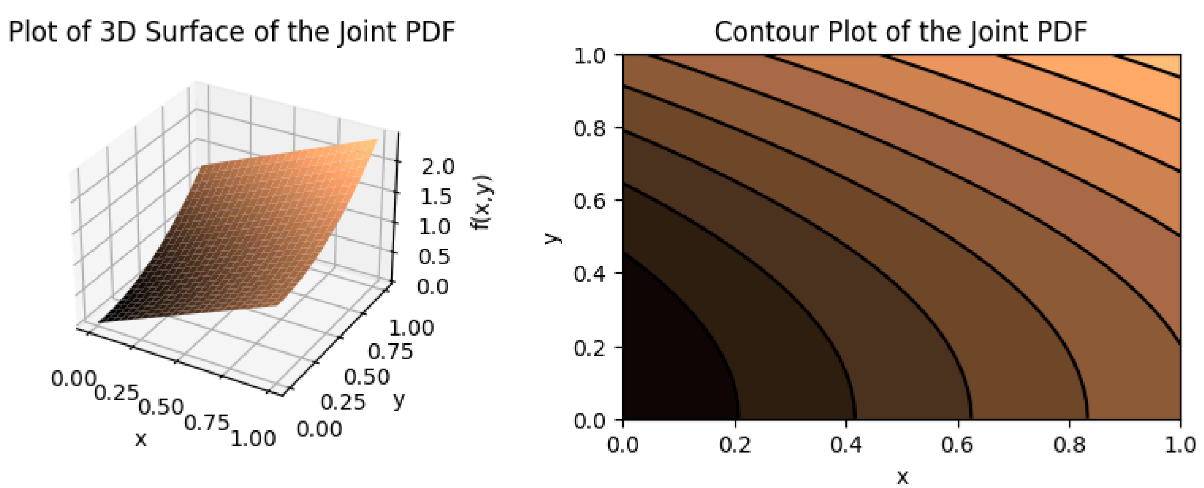
So $f(x, y)$ is normalized. Hence, it is a valid pdf.

Suppose we want to know $P(0 \leq X \leq \frac{1}{4}, 0 \leq Y \leq \frac{1}{4})$, then we can calculate:

$$
\begin{aligned}
P(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}) &= \int_0^{\frac{1}{4}}\int_0^{\frac{1}{4}} \frac{6}{5}(x+y^2)\, dxdy \\
&= \frac{6}{5}\int_0^{\frac{1}{4}}\int_0^{\frac{1}{4}} x\, dxdy + \frac{6}{5}\int_0^{\frac{1}{4}}\int_0^{\frac{1}{4}} y^2\, dxdy \\
&= \frac{6}{20}\int_0^{\frac{1}{4}} x\, dx + \frac{6}{20}\int_0^{\frac{1}{4}} y^2\, dy \\
&= \frac{6}{20}\left[\frac{x^2}{2}\right]_0^{\frac{1}{4}} + \frac{6}{20}\left[\frac{y^3}{3}\right]_0^{\frac{1}{4}} \\
&= \frac{3}{320} + \frac{1}{640} = \frac{7}{640} \approx \ = 0.0109
\end{aligned}
$$

In [17]:
```
def joint_pdf(x, y):
    dom = (x >= 0) & (x <= 1) & (y >= 0) & (y <= 1)
    return (6 / 5) * (x + y * y) * dom
```

In [18]:
```
fig = plt.figure(figsize=(10,3))

ax1 = fig.add_subplot(121, projection='3d')
ax2 = fig.add_subplot(122)

X = np.linspace(0,1,25)
Y = np.linspace(0,1,25)
X, Y = np.meshgrid(X, Y)
Z = joint_pdf(X, Y)

ax1.plot_surface(X, Y, Z, cmap=cm.copper)
ax1.set(title='Plot of 3D Surface of the Joint PDF', xlabel='x', ylabel='y', zlabel

ax2.contour(X, Y, Z, colors='k', levels=10)
ax2.contourf(X, Y, Z, cmap=cm.copper, levels=10)
ax2.set(title='Contour Plot of the Joint PDF', xlabel='x', ylabel='y');
```

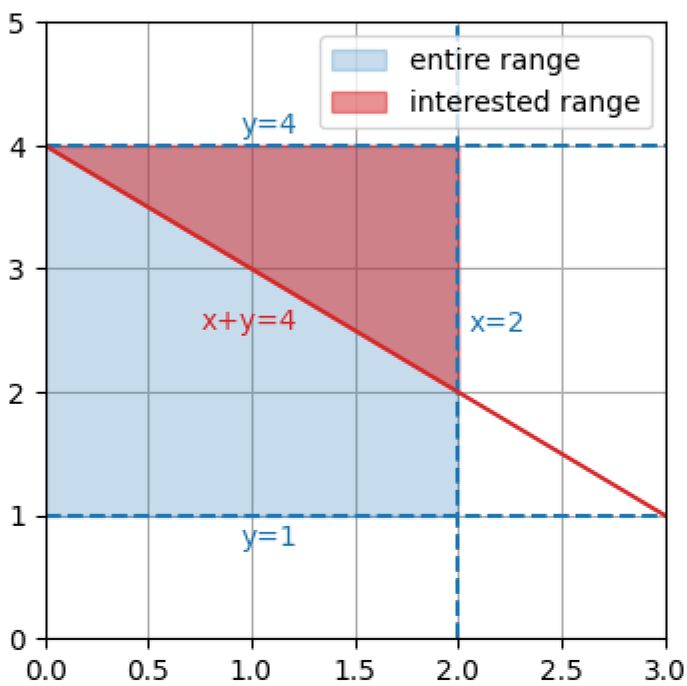


Example If the joint probability density function of X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{50}(x^2+y^2) & \text{if } x \in (0,2), y \in (1,4) \\ 0 & \text{elsewhere} \end{cases}$$

and we ask for the probability $P(X+Y>4)$. First consider the area in the range space where the condition $x+y>4$ holds:

In [19]:
```
fig, ax = plt.subplots(figsize=(4,4))

ax.fill_between([0,2], [1,1], [4,4], alpha=0.25, color='tab:blue', label='entire ra
ax.fill_between([0,2], [4,2], [4,4], alpha=0.5, color='tab:red', label='interested

ax.plot([2,2], [0,5], linestyle='--', color='tab:blue')
ax.text(2.05, 2.5, 'x=2', color='tab:blue')
ax.plot([0,3], [1,1], linestyle='--', color='tab:blue')
ax.text(0.95, 0.75, 'y=1', color='tab:blue')
ax.plot([0,3], [4,4], linestyle='--', color='tab:blue')
ax.text(0.95, 4.1, 'y=4', color='tab:blue')

ax.plot([0,3], [4,1], color='tab:red')
ax.text(0.75, 2.5, 'x+y=4', color='tab:red')

ax.set(xlim=[0,3], ylim=[0,5], axisbelow=True)
ax.legend()
ax.grid()
```



So, we can set the integral to calculate the probability as follows:

Processing math: 100%

$$
\begin{aligned}
P(X+Y>4) &= \int_0^2\int_{4-x}^4 f(x,y)\,dydx = \int_0^2\int_{4-x}^4 \frac{1}{50}(x^2+y^2)\,dydx \\
&= \int_0^2 \frac{1}{50}\left[x^2y+\frac{y^3}{3}\right]_{y=4-x}^4 dx \\
&= \int_0^2 \frac{1}{50}\left(x^3-\frac{(4-x)^3}{3}+\frac{64}{3}\right)dx \\
&= \frac{1}{200}\left[x^4\right]_0^2+\frac{1}{600}\left[(4-x)^4\right]_0^2+\frac{64}{150}[x]_0^2 \\
&= \frac{2^4}{200}+\frac{(4-2)^4-(4-0)^4}{600}+\frac{64(2)}{150} \\
&= \frac{8}{15}
\end{aligned}
$$

Therefore, the probability that $X+Y$ is greater than 4 is $8/15 \approx 0.533$.

Marginal Probability Distributions

Given a joint probability $f(x,y)$ of X and Y. The probability distribution $g(x)$ of X alone can be obtained by summing $f(x,y)$ over all values of Y or integrating over all values of Y for discrete and continuous random variables respectively. Similarly, the probability distribution $h(y)$ of Y alone can be obtained by summing or integrating $f(x,y)$ over all possible values of X. We called this kind of distribution a **marginal distribution**.

> The **marginal probability mass functions** of X alone and Y alone are
>
> $$g(x)=\sum_{\forall y} f(x,y) \quad \text{and} \quad h(y)=\sum_{\forall x} f(x,y)$$
>
> for X and Y being discrete rv's and the **marginal probability density functions** of X alone and Y alone are
>
> $$g(x)=\int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad h(y)=\int_{-\infty}^{\infty} f(x,y)\,dx$$
>
> for X and Y being continuous rv's.

The properties of marginal probability distribution are the same as the ordinary probability distribution function (representing probability, non-negative, normalized, etc.)

Example Consider the joint probability mass function of X and Y discussed earlier

$$f(x,y) = \begin{cases} 1/28 & \text{if } (x,y) = (0,2) \\ 3/28 & \text{if } (x,y) = (0,0), (2,0) \\ 6/28 & \text{if } (x,y) = (0,1), (1,1) \\ 9/28 & \text{if } (x,y) = (1,0) \\ 0 & \text{elsewhere} \end{cases}$$

The marginal distribution function of X is

$$g(x) = \sum_{y \in \backslash\text{set}0,1,2} f(x,y) = f(x,0) + f(x,1) + f(x,2).$$

We can see that

$$g(0) = \sum_{y \in \backslash\text{set}0,1,2} f(0,y) = f(0,0) + f(0,1) + f(0,2) = \frac{3}{28} + \frac{6}{28} + \frac{1}{28} = \frac{10}{28}.$$

$$g(1) = \sum_{y \in \backslash\text{set}0,1,2} f(1,y) = f(1,0) + f(1,1) + f(1,2) = \frac{9}{28} + \frac{6}{28} + 0 = \frac{15}{28}.$$

$$g(2) = \sum_{y \in \backslash\text{set}0,1,2} f(2,y) = f(2,0) + f(2,1) + f(2,2) = \frac{3}{28} + 0 + 0 = \frac{3}{28}.$$

So we can write the marginal distribution $g(x)$ of X as

$$g(x) = \begin{cases} 10/28 & \text{if } x = 0 \\ 15/28 & \text{if } x = 1 \\ 3/28 & \text{if } x = 2 \\ 0 & \text{elsewhere} \end{cases}$$

In similar manner, we can calculate the marginal probability $h(y)$ of Y:

$$h(0) = \sum_{x \in \backslash\text{set}0,1,2} f(x,0) = f(0,0) + f(1,0) + f(2,0) = \frac{3}{28} + \frac{9}{28} + \frac{3}{28} = \frac{15}{28}.$$

$$h(1) = \sum_{x \in \backslash\text{set}0,1,2} f(x,1) = f(0,1) + f(1,1) + f(2,1) = \frac{6}{28} + \frac{6}{28} + 0 = \frac{12}{28}.$$

$$h(2) = \sum_{x \in \backslash\text{set}0,1,2} f(x,2) = f(0,2) + f(1,2) + f(2,2) = \frac{1}{28} + 0 + 0 = \frac{1}{28}.$$

Therefore,

$$h(y) = \begin{cases} 15/28 & \text{if } y = 0 \\ 12/28 & \text{if } y = 1 \\ 1/28 & \text{if } y = 2 \\ 0 & \text{elsewhere} \end{cases}$$

Example Consider another probability distribution from the past example

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

So we can calculate the marginal probability density function of X as follows:

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 \frac{6}{5}(x + y^2)\, dy$$

$$= \frac{6}{5}x + \frac{6}{5}\left[\frac{y^3}{3}\right]_0^1$$

$$\text{so} \quad g(x) = \begin{cases} \frac{6}{5}x + \frac{2}{5} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Similarly for the marginal distribution $h(y)$ of Y:

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^1 \frac{6}{5}(x + y^2)\, dx$$

$$= \frac{6}{5}\left[\frac{x^2}{2}\right]_0^1 + \frac{6}{5}y^2$$

$$\text{so} \quad h(y) = \begin{cases} \frac{3}{5} + \frac{6}{5}y^2 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Conditional Probability Distribution & Independent RV's

Recall from the previous chapter that the *conditional probability* of an event A given B is defined as

$$P(A \mid B) = \frac{P(A \cup B)}{P(B)}, \quad P(B) \neq 0.$$

If we define the events A and B to be $Y = y$ and $X = x$ respectively, then

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x, y)}{g(x)}, \quad g(x) \neq 0$$

> Let X and Y be two random variables, *discrete or continuous*. The **conditional distribution** of the random variable Y given that $X = x$ is

> $$f(y \mid x) = \frac{f(x, y)}{g(x)}, \quad g(x) \neq 0$$
>
> where $f(x, y)$ is the joint probability distribution of X and y and $g(x)$ is the marginal probability distribution of X.
>
> Similarly, the conditional distribution of X given that $Y = y$ is
>
> $$f(x \mid y) = \frac{f(x, y)}{h(y)}, \quad h(y) \neq 0$$
>
> where $h(y)$ is the marginal probability distribution of Y.

Clearly, conditional probability distributions have the same properties as the ordinary probability distributions (such as probability representation, non-negativity, and normalization).

If we wish to find the probability that the discrete random variable X falls between a and b when it is known that the discrete variable $Y = y$, we evaluate

$$P(a < X < b \mid Y = y) = \sum_{a<x<b} f(x \mid y),$$

where the summation extends over all values of X between a and b. When X and Y are continuous, we evaluate

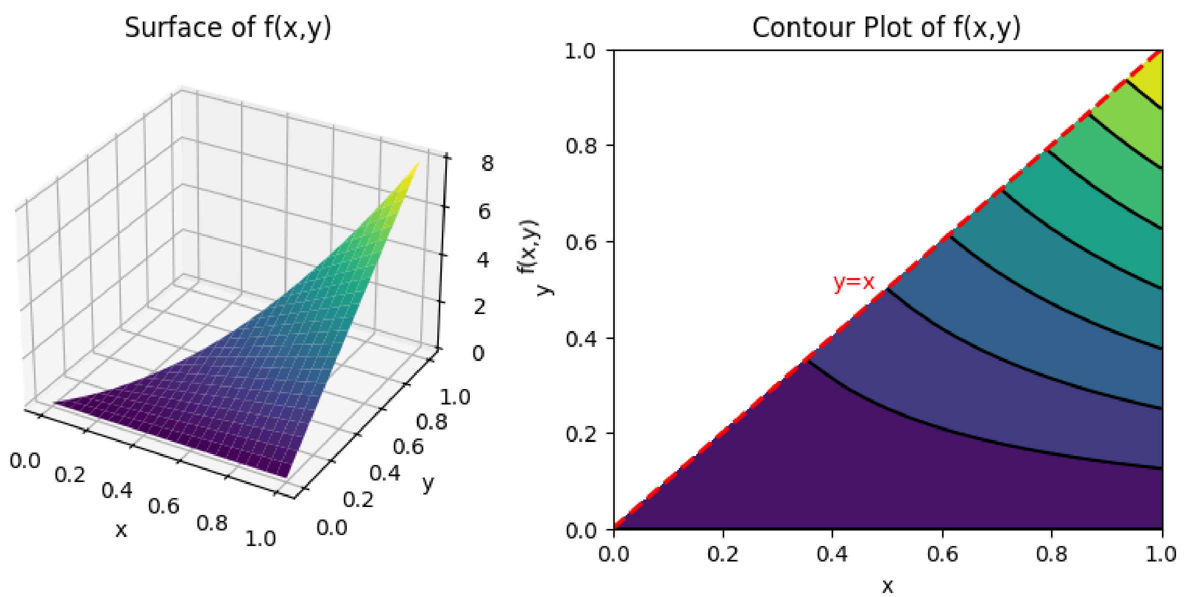
$$P(a < X < b \mid Y = y) = \int_a^b f(x \mid y)\, dx$$

In [20]:
```
def example_pdf(X, Y, na=np.nan):
    Z = np.where(Y <= X, 1, na)
    return Z * 8 * X * Y
```

In [21]:
```
fig = plt.figure(figsize=(10,4))

ax1 = fig.add_subplot(121, projection='3d')
ax2 = fig.add_subplot(122)

X = np.linspace(0,1,25)
Y = np.linspace(0,1,25)
X, Y = np.meshgrid(X, Y)
Z = example_pdf(X, Y)

ax1.plot_surface(X, Y, Z, cmap=cm.viridis)

ax2.contour(X, Y, Z, colors='k')
ax2.contourf(X, Y, Z, cmap=cm.viridis);
ax2.plot([0,1], [0,1], color='red', linestyle='--', linewidth=2)
ax2.text(0.4, 0.5, "y=x", color='r')
```

Processing math: 100%

```
ax1.set(title="Surface of f(x,y)", xlabel='x', ylabel='y', zlabel='f(x,y)')
ax2.set(title="Contour Plot of f(x,y)", xlabel='x', ylabel='y');
```



Example The figure shown above is the plots of the joint probability distribution of continuous rv's X and Y defined as

$$f(x, y) = \begin{cases} 8xy & \text{if } 0 < x < 1, 0 < y < x \\ 0 & \text{elsewhere} \end{cases}$$

To determine the conditional probability distribution $f(y|x)$, we first calculate the marginal distribution $g(x)$ of X. Since the range space of X and Y is the region $R = \set(x, y) | 0 < x < 1, 0 < y < x$, then

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^x 8xy\, dy = 4x^3, \quad 0 < x < 1$$

So the conditional distribution $f(y|x)$ is

$$f(y|x) = \frac{f(x, y)}{g(x)} = \frac{8xy}{4x^3} = \frac{2y}{x^2}$$

Suppose we want to know $P(Y < \frac{1}{8} | X = \frac{1}{2})$, we calculate:

$$\begin{aligned} P(Y < \frac{1}{8} | X = \frac{1}{2}) &= \int_0^{\frac{1}{8}} f(y | x = \frac{1}{2})\, dy \\ &= \int_0^{\frac{1}{8}} \frac{2y}{1/4}\, dy \\ &= \left[4y^2\right]_0^{1/8} = \frac{1}{16}. \end{aligned}$$

If we want to know $P(Y < \frac{1}{8} | X < \frac{1}{2})$, we can calculate:

$$P(Y < \frac{1}{8} | X < \frac{1}{2}) = \frac{P(X < \frac{1}{2}, Y < \frac{1}{8})}{P(X < \frac{1}{2})}$$

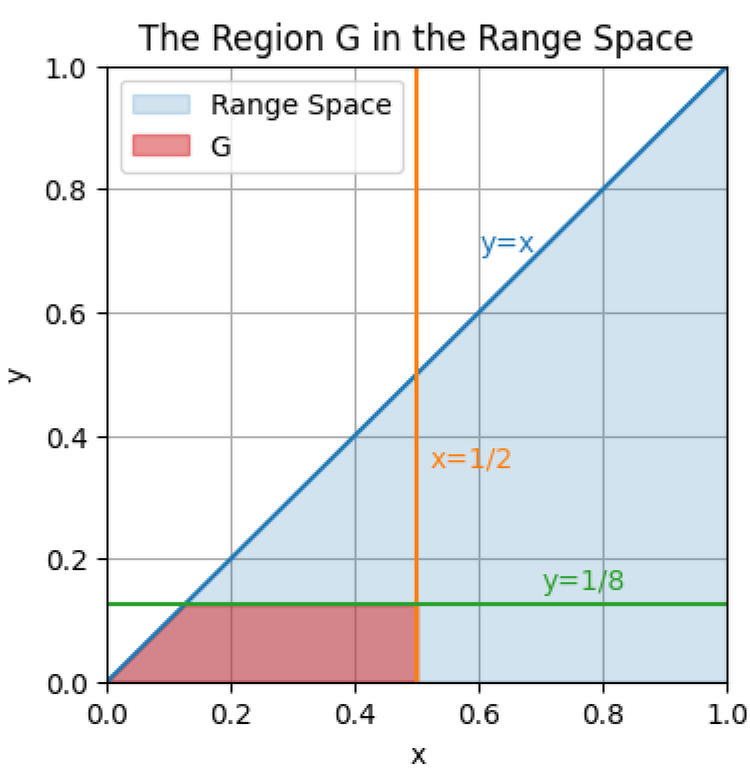
Let's break this down into two parts, first consider

$$P(X < \frac{1}{2}, Y < \frac{1}{8}) = \iint_G f(x, y)\, dxdy$$

where G is the region in the range space where $x < \frac{1}{2}$ and $y < \frac{1}{8}$, that is

$G = \text{\set}(x, y) | 0 < x < \frac{1}{2}, 0 < y < \frac{1}{8}, y < x$

In [22]:
```
fig, ax = plt.subplots(figsize=(4,4))

ax.plot([0,1], [0,1], color='tab:blue')
ax.plot([1/2, 1/2], [0,1], color='tab:orange')
ax.plot([0,1], [1/8, 1/8], color='tab:green')
ax.fill_between([0, 1], [0, 0], [0, 1], label='Range Space', alpha=0.2, color='tab:
ax.fill_between([0, 1/8, 1/2], [0, 0, 0], [0, 1/8, 1/8], label='G', alpha=0.5, colo

ax.text(0.6, 0.7, 'y=x', color='tab:blue')
ax.text(0.52, 0.35, 'x=1/2', color='tab:orange')
ax.text(0.7, 0.15, 'y=1/8', color='tab:green')

ax.set(xlim=[0,1], ylim=[0,1], axisbelow=True, xlabel='x', ylabel='y', title='The R
ax.legend()
ax.grid()
```



Thus,

$$P(X < \frac{1}{2}, Y < \frac{1}{8}) = \int_0^{\frac{1}{8}} \int_y^{\frac{1}{2}} 8xy \, dxdy$$

$$= \int_0^{\frac{1}{8}} y - 4y^3 \, dy$$

$$= \frac{1}{128} - \frac{1}{4096} = \frac{31}{4096}$$

Now consider another part

$$P(X < \frac{1}{2}) = \int_0^{\frac{1}{2}} g(x) \, dx$$

$$= \int_0^{\frac{1}{2}} 4x^3 \, dx = \frac{1}{16}$$

Therefore,

$$P(Y < \frac{1}{8} | X < \frac{1}{2}) = \frac{P(X < \frac{1}{2}, Y < \frac{1}{8})}{P(X < \frac{1}{2})} = \frac{31/4096}{1/16} = \frac{31}{256} \approx 0.12$$

Independent Random Variables

> Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x, y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if
>
> $$f(x, y) = g(x)h(y)$$
>
> for all (x, y) within their range space.

Example From the previous example where

$$f(x, y) = \begin{cases} 8xy & \text{if } 0 < x < 1, 0 < y < x \\ 0 & \text{elsewhere} \end{cases}$$

is the joint probability density function of X and Y. The marginal probability distributions is given by

$$g(x) = \int_0^x 8xy \, dy = 4x^3, \quad 0 < x < 1$$

$$h(y) = \int_y^1 8xy \, dx = 4y(1 - y^2), \quad 0 < y < 1$$

We can see that X and Y are not statistically independent since

$$f(x,y) = 8xy \neq 16x^3y(1-y^2) = g(x)h(y).$$

Example The discrete joint probability distribution

$$f(x,y) = \begin{cases} 1/28 & \text{if } (x,y) = (0,2) \\ 3/28 & \text{if } (x,y) = (0,0), (2,0) \\ 6/28 & \text{if } (x,y) = (0,1), (1,1) \\ 9/28 & \text{if } (x,y) = (1,0) \\ 0 & \text{elsewhere} \end{cases}$$

with marginal probability distribution of X and Y defined as

$$g(x) = \begin{cases} 10/28 & \text{if } x = 0 \\ 15/28 & \text{if } x = 1 \\ 3/28 & \text{if } x = 2 \\ 0 & \text{elsewhere} \end{cases} \quad \text{and} \quad h(y) = \begin{cases} 15/28 & \text{if } y = 0 \\ 12/28 & \text{if } y = 1 \\ 1/28 & \text{if } y = 2 \\ 0 & \text{elsewhere} \end{cases}$$

We can see that X and y are not statistically independent since

$$f(0,0) = \frac{3}{28} \neq \frac{10}{28} \times \frac{15}{28} = \frac{75}{392} = g(0)h(0).$$

Multiple RV's

To model the joint behavior of more than two random variables, we extend the concept of a joint distribution of two variables.

> If $X_1, X_2, \ldots, X_n$ are all discrete random variables, the **joint pmf** of the variables is the function
>
> $$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f(x_1, x_2, \ldots, x_n).$$
>
> If the variables are continuous, the **joint pdf** of $X_1, \ldots, X_n$ is the function $f(x_1, x_2, \ldots, x_n)$ such that for any n intervals $[a_1, b_1], \ldots, [a_n, b_n]$,
>
> $$P(a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n) = \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1.$$

Likewise, we can extend the concept of marginal probability to marginal probability distribution of more than one rv's called *joint marginal distribution*. The same extension can be done for conditional probability distribution and independent rv's as well.

Let $f(x_1, x_2, \ldots, x_n)$ be the joint probability distribution function of $X_1, X_2, \ldots, X_n$. The **joint marginal probability distribution** of $X_1, X_2, \ldots, X_k$ $(k < n)$ is the function $\phi(x_1, x_2, \ldots, x_k)$ for which

$$\phi(x_1, x_2, \ldots, x_k) = \sum_{\forall x_{k+1}} \cdots \sum_{x_n} f(f(x_1, x_2, \ldots, x_n))$$

if the rv's are discrete and

$$\phi(x_1, x_2, \ldots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(f(x_1, x_2, \ldots, x_n))\, dx_{k+1} \ldots dx_n$$

if the rv's are continuous.

The **joint probability distribution** of $X_1, X_2, \ldots, X_k$ given that $X_{k+1} = x_{k+1}, X_{k+2} = x_{k+2} \ldots, X_n = x_n$ is defined as

$$f(x_1, x_2, \ldots, x_k | x_{k+1}, x_{k+2}, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n)}{\phi(x_{k+1}, x_{k+2}, \ldots, x_n)}$$

Let $f(x_1), f(x_2), \ldots, f(x_n)$ be the marginal distributions of $X_1, X_2, \ldots, X_n$ respectively. Then $X_1, X_2, \ldots, X_n$ are said to be **statistically independent** if

$$f(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2) \cdots f(x_n)$$

for all $(x_1, x_2, \ldots, x_n)$ within their range.