Random Variables

Chapter Intended Learning Outcomes:

(i) Understand discrete and continuous random variables

(ii) Know common distributions of random variables

(iii) Able to generate common random variables

(iv) Able to compute expectation of random variables

(v) Able to apply random variables and common probability distributions to solve problems

Random Variables

A random variable (RV) X is a real-valued number assigned to every outcome of an experiment.
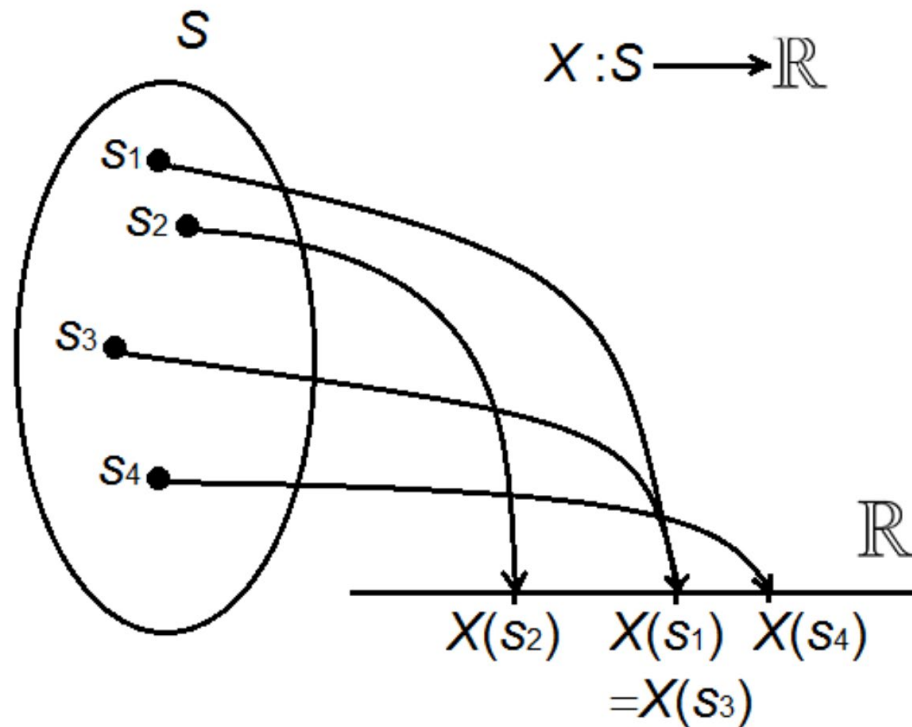
RV is a function from the sample space to a number, allowing numerical designation for events which are set functions.

Recall dice rolling, event with odd number is a set {1, 3, 5}. In this case, we may assign RV $X = 1$ if the outcome is an odd number, i.e., $X(1) = X(3) = X(5) = 1$.

Traditionally, RV is denoted by an upper-case letter (e.g., X) and any specific value of RV is a lower-case letter (e.g., x)

While we may assign RV $X = 0$ for even number. Hence X is random, with value 0 or 1.

For an unbiased dice, we have:

$$P(X = 0) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

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

$$P(0 \leq X \leq 1) = 1 \quad \text{and} \quad P(X < 0) = P(X > 1) = 0$$

Note that we can assign/define RVs in many different ways for a particular sample space, e.g., we may also set $X = 10$ and $X = -10$ for odd and even number events, respectively.

But we can define the RVs based on a specific problem so that it will be solved in an easier manner.

RV can be classified as:
- Discrete: its range is countable, with finite or a countably infinite number of values on the real line.
- Continuous: It has a range with uncountably infinite number of values on the real line.

Example 2.1
Consider the experiment of flipping a coin until a tail occurs. Define a RV for this experiment and discuss its property.

We can define a RV X which is the number of trials until the experiment ends (i.e., tail occurs). Hence the RV is discrete with only positive integer values:

$$x = 1, 2, \cdots \Rightarrow 1 \leq X < \infty$$

Example 2.2

Can we define a discrete RV on a continuous sample space? And can we define a continuous RV on a discrete sample space? For each case, use an example for illustration.

Yes, for both. First, suppose the sample space is any point in the interval (0,1]. We can define a RV that maps (0,0.5] to 0 and (0.5,1] to 1.

Consider the outcome of flipping a coin, which is either head (H) or tail (T). We can define a RV which is selected in (0,0.5] to represent that H occurs, while (0.5,1] for T.

Distribution Functions for Random Variables

For discrete RV X, we use probability mass function (PMF) $p(x) = P(X = x) = P_X(x)$ to assign probability to each RV value.

Suppose X must be one of the values x_1, x_2, x_3, ... then:

$$p(x_i) = P(X = x_i) \geq 0, \quad i = 1, 2, \cdots \qquad (2.1)$$

$$\sum_{x \in R_x} p(x) = 1, \quad R_x = \{x_1, x_2, \cdots\} \qquad (2.2)$$

where R_x is the range. Suppose $A \subset R_x$, the probability of $X \in A$ is thus equal to:

$$P(X \in A) = \sum_{x \in A} p(x)$$

Apart from PMF, the cumulative distribution function (CDF) can also be used to describe the RV probability.

Assuming that $x_1 < x_2 < x_3 < \cdots$, the CDF is defined as:

$$F(x) = P(X \le x) = P_X(X \le x) \qquad (2.3)$$

Consider rolling a dice again, we may directly assign the outcomes "1" to "6" as $x_1 = 1$ to $x_6 = 6$. Hence we have:

$$p(x_i) = P(X = x_i) = \frac{1}{6}, \quad i = 1, 2, \cdots, 6$$

$$\sum_{x \in R_x} p(x) = 1, \quad R_x = \{x_1, x_2, \cdots, x_6\}$$

If A refers to the event with odd number, then $A = \{x_1, x_3, x_5\}$ and

$$P(X \in A) = \sum_{x \in A} p(x) = \frac{1}{2}$$

Also,

$$F(x_3) = P_X(X \le x_3) = p(x_1) + p(x_2) + p(x_3) = \frac{1}{2}$$

is the probability that the outcome is less than or equal to 3.

Apparently, CDF is a non-decreasing function with minimum and maximum values 0 and 1, respectively.

If PMF is known, then CDF can be determined and vice versa.

Example 2.3
Consider the experiment of rolling 2 fair dice and the outcome is the sum of the faces. Assign RV for this experiment and then plot the PMF and CDF. Compute the probability that the sum is between 6 and 8.

To facilitate the probability computation, we assign X as the sum value, i.e., $2 \leq X \leq 12$.

Applying the probability theory, e.g., the sum is 2 only when both dice give "1", hence $p(2) = 1/6 \times 1/6$, we obtain:

X, Sum of the Two Faces	Number of Ways that X Can Occur	Probability Distribution, $P_X(x)$
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

From the computed probabilities, we construct PMF:
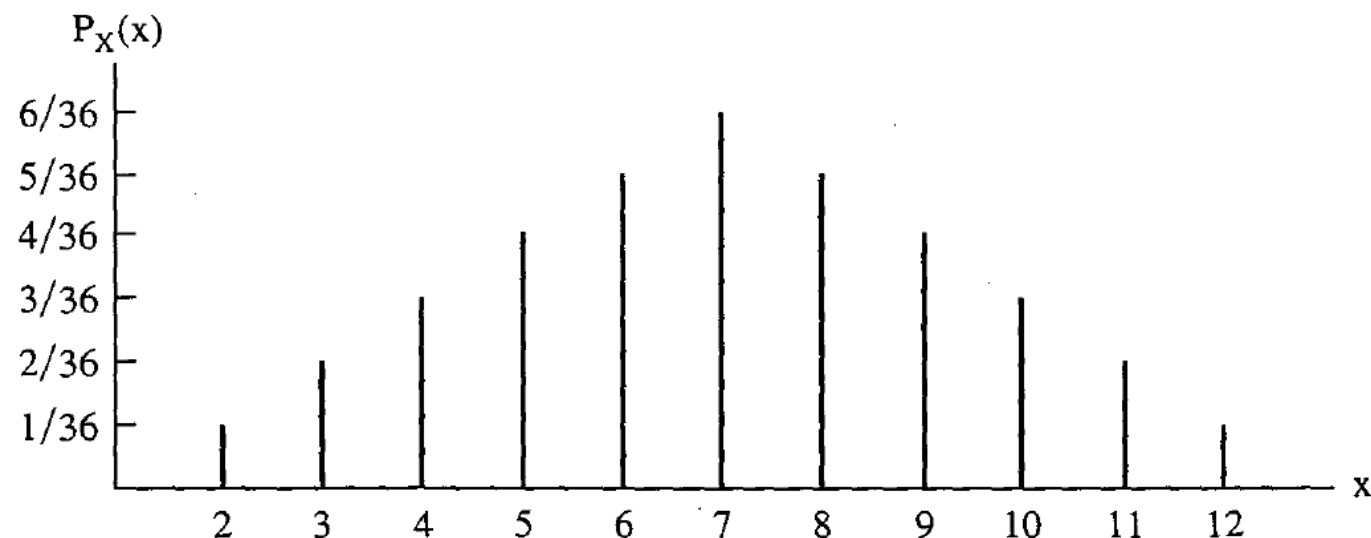
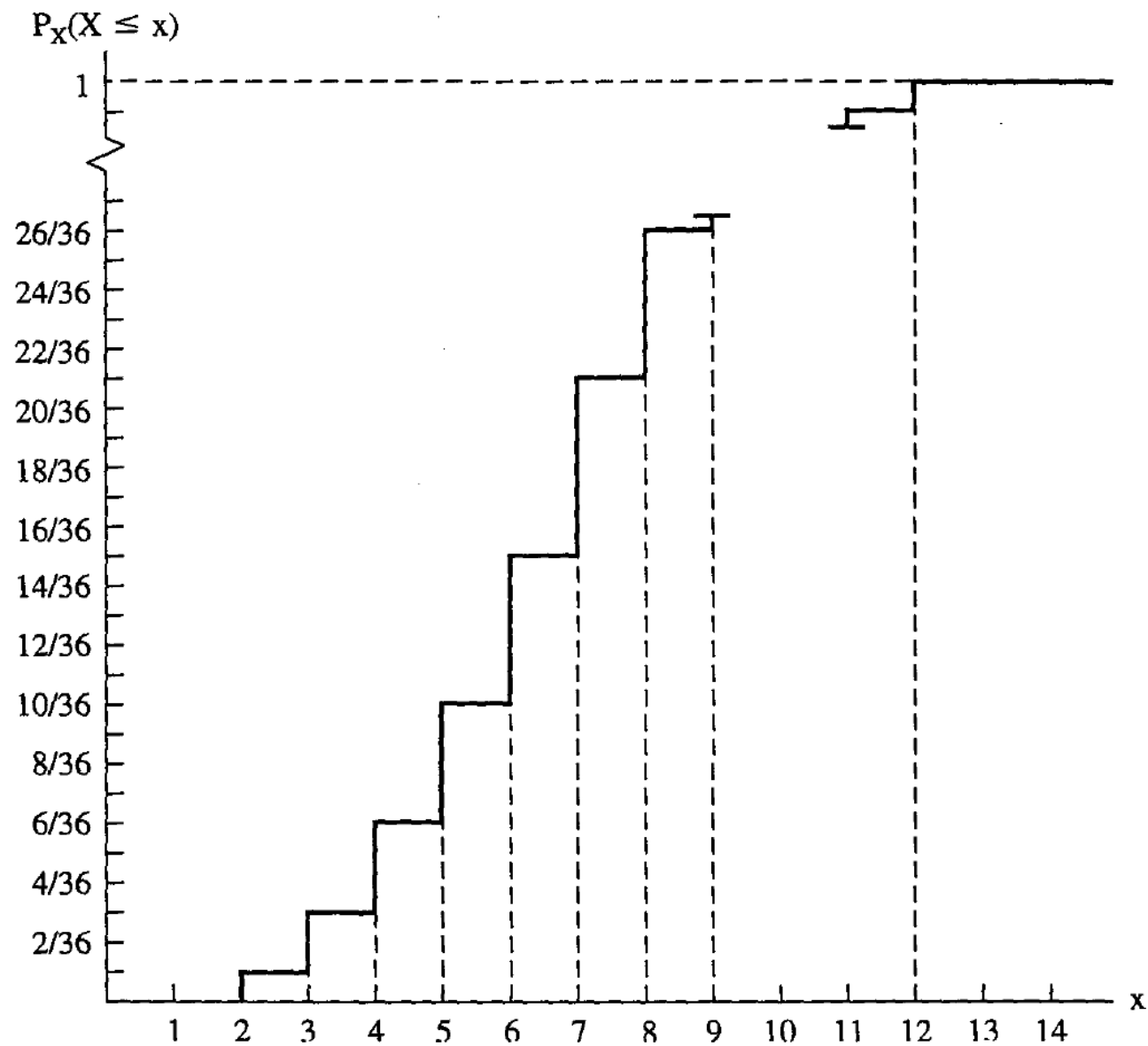The required probability is:

$$p(6) + p(7) + p(8) = \frac{4}{9}$$

Or

$$P_X(X \le 8) - P_X(X \le 5) = \frac{26}{36} - \frac{10}{36} = \frac{4}{9}$$

The CDF is:

Example 2.4

Consider the experiment of flipping a coin until a head occurs. Assume the probability of coming up a head is p, and successive flips are independent. Design a suitable RV and find the probability that a head occurs at the nth flip.

Following Example 2.1, we define X as the RV which is the number of trials until the head (H) appears. Also, it is clear that the probability of getting tail (T) is $1 - p$.

We have:
$$p(1) = P(X = 1) = P(\{H\}) = p$$
$$p(2) = P(X = 2) = P(\{T, H\}) = (1 - p)p$$
$$p(3) = P(X = 3) = P(\{T, T, H\}) = (1 - p)^2 p$$

Hence the required probability is:
$$p(n) = P(X = n) = (1 - p)^{n-1} p$$

What is P(a head occurs within infinite trials)?

Binomial Distribution

Consider an experiment which results in two possible outcomes: success with probability p, and failure with probability $q = 1 - p$.

It is called a Bernoulli trial. If we assign RV $X = 1$ for success and $X = 0$ for failure, then:

$$p(0) = P(X = 0) = 1 - p$$

$$p(1) = P(X = 1) = p$$

(2.4)

where X is called Bernoulli RV.

Binomial RV extends the Bernoulli RV by combining n independent Bernoulli trials, where we are interested in the number of successes. Assigning RV X to represent the success number, we have $0 \leq X \leq n$.

The PMF and CDF of the binomial RV are:

$$p(r) = P(X = r) = C(n, r)p^r(1 - p)^{n-r}, \quad 0 \le r \le n$$

$$\text{(2.5)}$$

$$F(r) = P(X \le r) = \sum_{i=0}^{r} C(n, i)p^i(1 - p)^{n-i}$$

where $C(n, r)$ accounts for the number of combinations. The binomial distribution is characterized by n and p.

Example 2.5
4 fair coins are flipped and assuming that the outcomes are independent. Compute the probability of obtaining 2 heads. How about the probability of obtaining at most 2 heads?

Let RV X be the number of heads. With $n = 4$ and $p = 0.5$, the required probability of the former is:

$$p(2) = C(4, 2)0.5^2(1 - 0.5)^{4-2} = 0.375$$

For the second, we use:

$$F(3) = p(0) + p(1) + p(2) = \sum_{i=0}^{2} C(4, i)(0.5)^i (1 - 0.5)^{4-i}$$

Using MATLAB, $p(0)$ and $p(1)$ are:

```
n=4;
p=0.5;
k=0;
nchoosek(n,k)*p^k*(1-p)^(n-k)
= 0.0625

k=1;
nchoosek(n,k)*p^k*(1-p)^(n-k)
= 0.2500
```
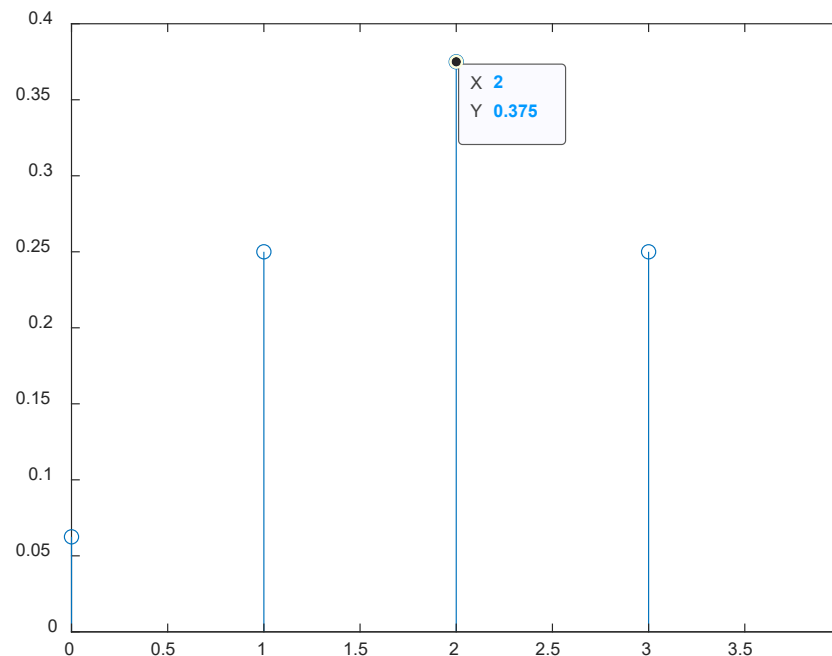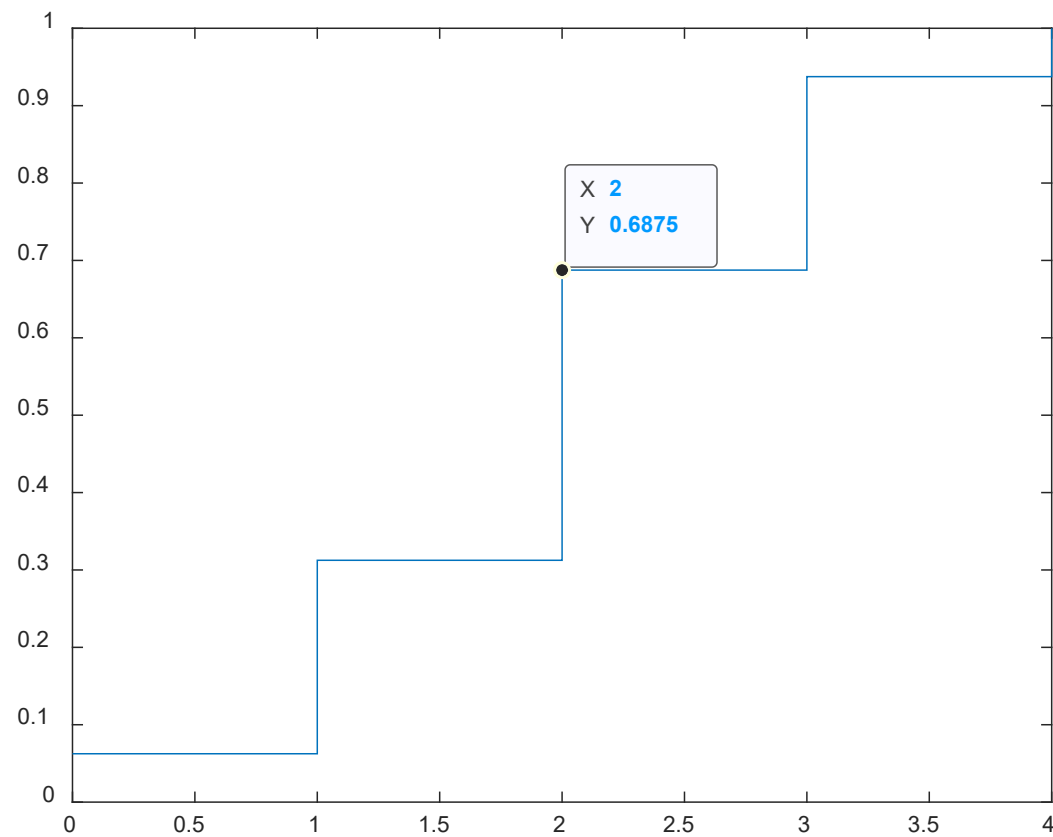
Hence

$$F(3) = 0.6875$$

```
n=4;
p=0.5;
for k=0:n;
P(k+1)=nchoosek(n,k)*p^k*(1-p)^(n-k);
end
stem(0:n,P)
```


Which combination of 4 coins is most probable? Any relation with *n* and *p*?

```
C(1)=P(1);
for k=1:n;
C(k+1)=C(k)+P(k+1);
end
stairs(0:n,C)
```

Note that the probability is the same if we consider flipping a coin 4 times in sequence.

To generate binomial RVs, the MATLAB command is `random`.

Example 2.6

Use `random` to generate binomial RVs with $(n, p) = (1, 0.4)$ and $(10, 0.4)$. Investigate the performance of this generator.

The first corresponds to Bernoulli RVs with $p = 0.4$. Consider generating 10 RVs, we use:

```
random('bino', 1, 0.4, 1,10)
= 0    1    0    1    0    0    0    0    1    1
```

Each of them is either 0 or 1. In this realization, 4 "1" are resulted, which are consistent with $p = 4/10$.

For the second one, the command is:

```
random('bino', 10, 0.4, 1,10)
=  3     5     6     4     2     4     5     3     6  =  4
```

Each entry has a value between 0 and 10.

To investigate the performance of `random`, more experiments are needed and `histogram` can be used to plot the empirical PDF and CDF.
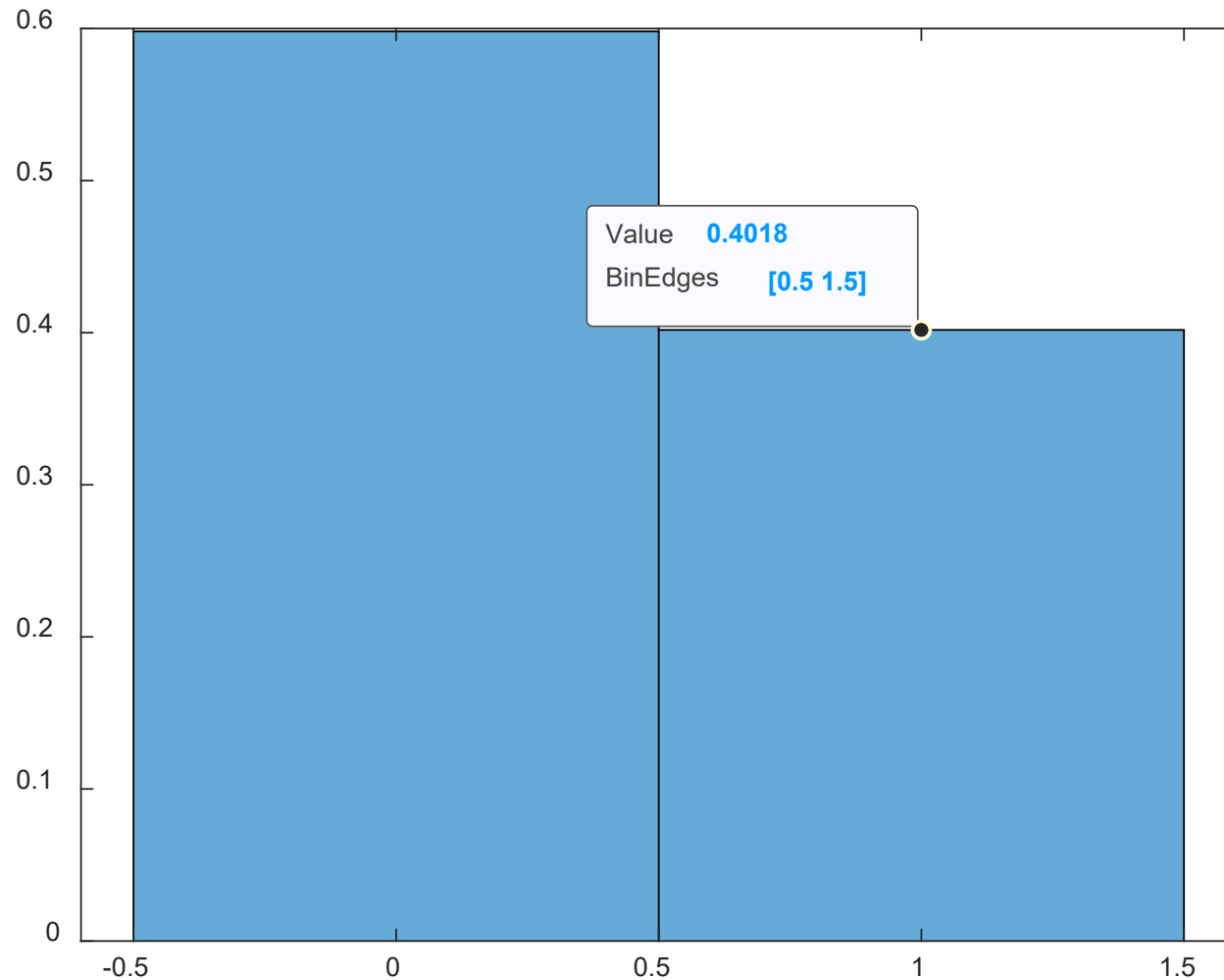
```
r1=random('bino', 1, 0.4, 1,100000);
r2=random('bino', 10, 0.4, 1,100000);
```
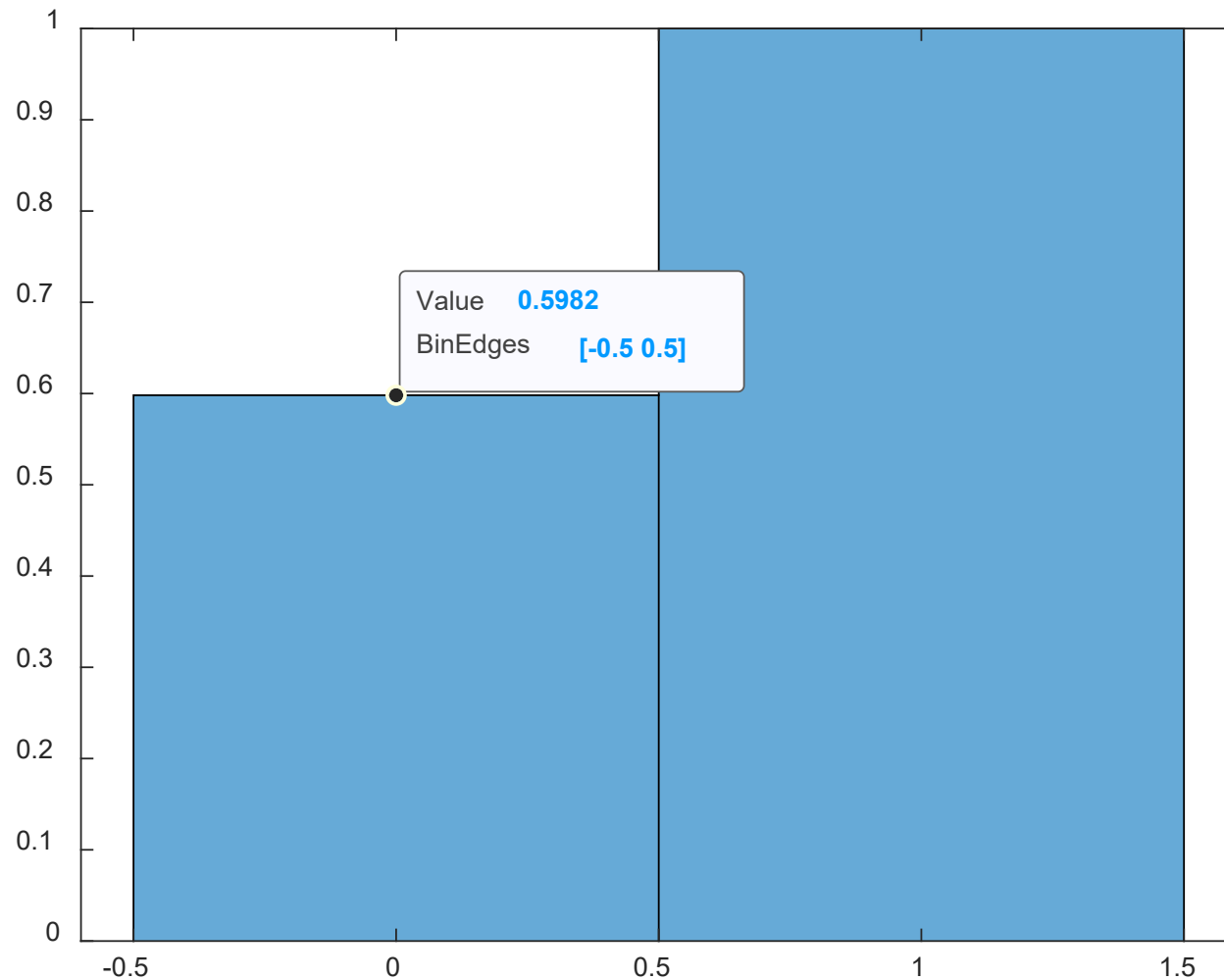
To plot PDF, we use:
```
histogram(r1, 'Normalization', 'pdf')
```
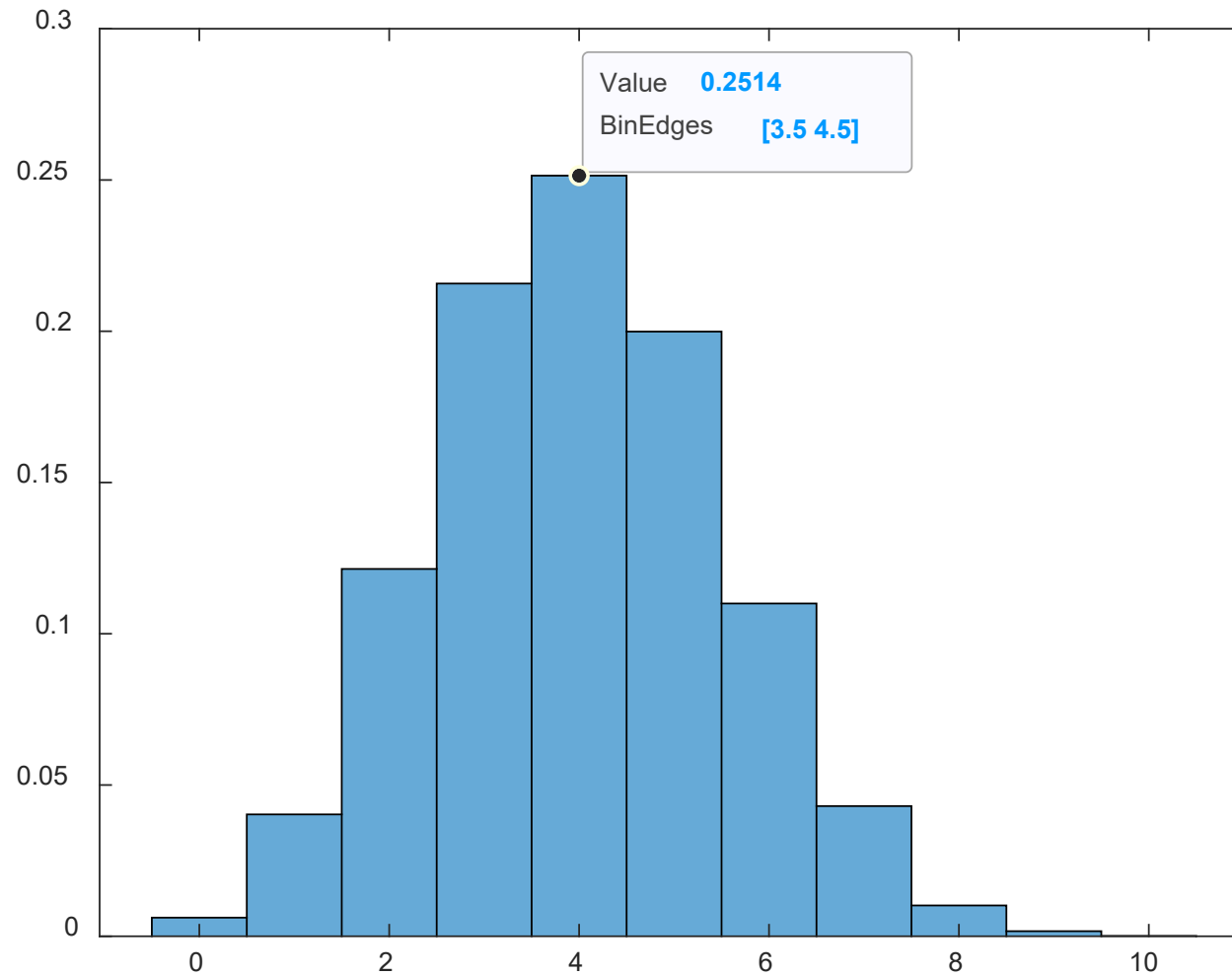
For CDF:
```
histogram(r1, 'Normalization', 'cdf')
```

It can be seen that the empirical PDF contains only 2 impulses at "0" and "1".

Is it close to the desired PDF/CDF? How can you make it closer?

Is it consistent with the desired PDF?

Geometric Distribution

Considering independent trials, each having probability p of being a success (and failure with probability $q = 1 - p$), are performed until a success occurs.

Define RV X to be the number of trial(s) until the experiment ends, which is called geometric RV.

From Example 2.1, we know $1 \leq X < \infty$. As in Example 4, the PMF is:

$$p(r) = P(X = r) = (1 - p)^{r-1}p, \quad 1 \leq r < \infty \qquad (2.6)$$

Why combination as in (2.1) is not considered?
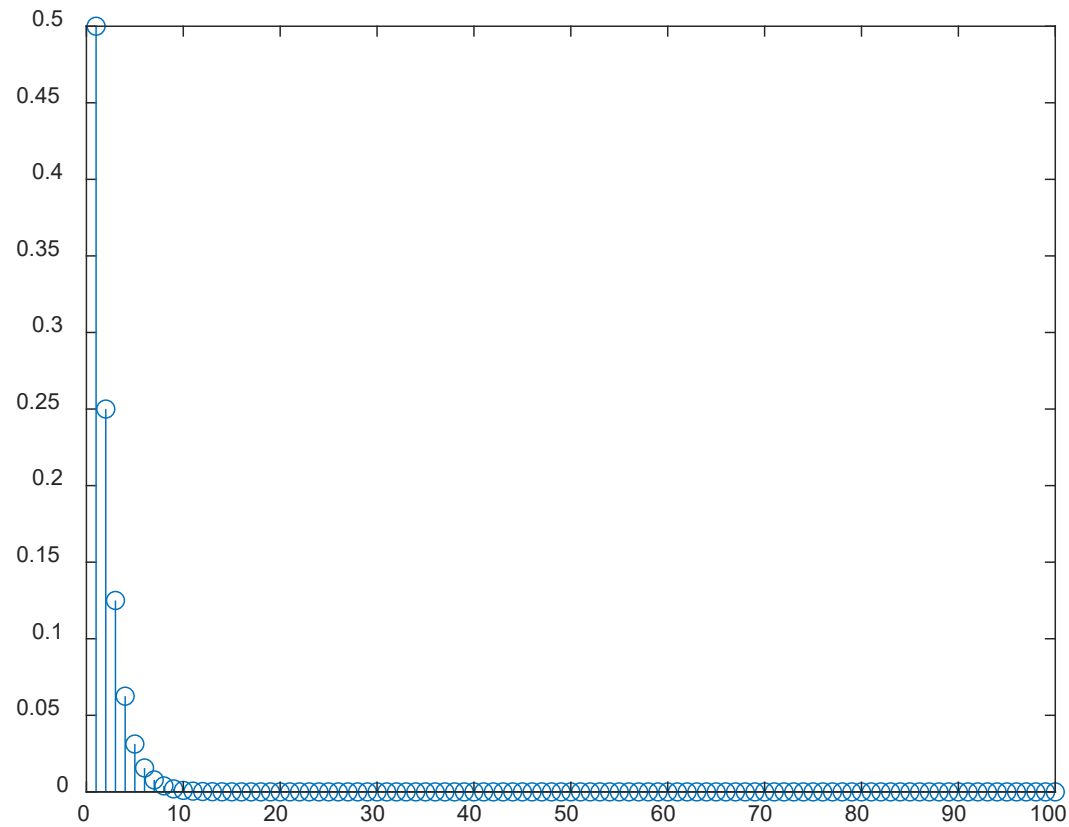**What is/are the parameter(s) of geometric RV?**

Example 2.7

Determine CDF of geometric RV with success probability p.

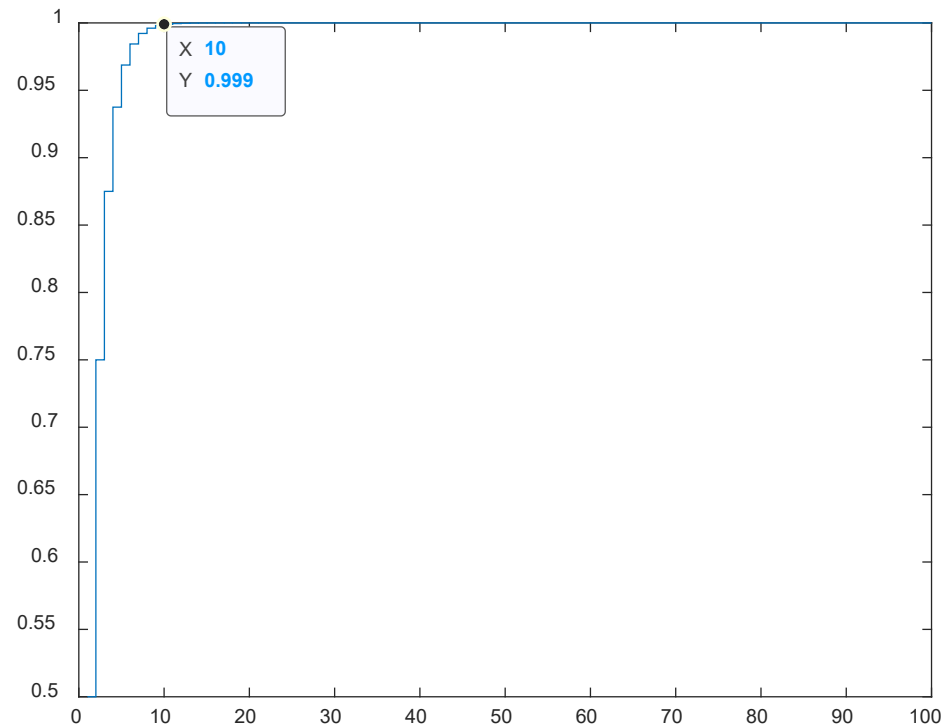$$F(r) = P(X \leq r) = \sum_{i=1}^{r} q^{i-1} p = \frac{p(1-q^r)}{1-q} = 1 - (1-p)^r$$

For example, the probability of tossing a fair coin which appears a tail in 10 trials is:

$$F(10) = 1 - (1-0.5)^{10} = 0.999$$

```
n=100; %only show the first 100 PMFs
p=0.5;
for k=1:n;
P(k)= (1-p)^(k-1)*p;
end
stem(1:n,P)
```

```
C(1)=P(1);
for k=2:n;
C(k)=C(k-1)+P(k);
end
stairs(1:n,C)
```

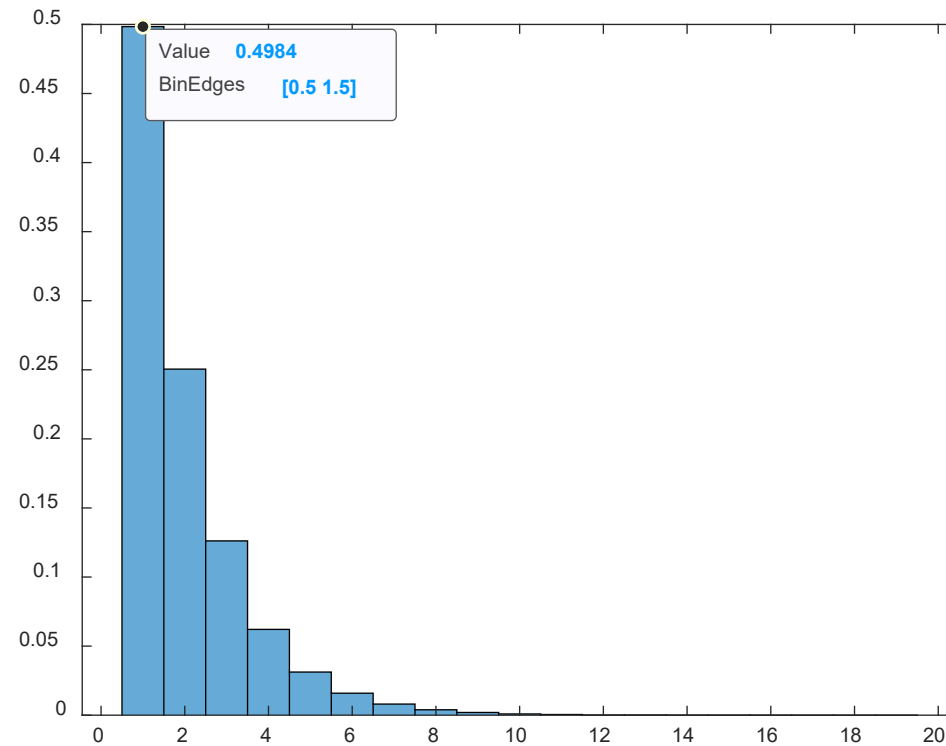To generate 10 geometric RVs with $p = 0.5$, we can use:

```
geornd(0.5,1,10)+1
= 2    1    1    3    1    1    1    3    1    2
```

Note that 1 is added because `geornd` starts at 0.

```
r=geornd(0.5,1, 100000)+1;
histogram(r, 'Normalization', 'pdf')
```

The displayed data tip shows:

Value **0.4984**
BinEdges **[0.5 1.5]**

Poisson Distribution

A RV $X \in \mathbb{Z}^{0+}$ is said to be Poisson RV with parameter $\lambda > 0$ if its PMF has the form of:

$$p(r) = P(X = r) = e^{-\lambda}\frac{\lambda^r}{r!}, \quad 0 \le r < \infty \qquad (2.7)$$

The Poisson function is the limiting case of the binomial function for large n and small p, with $\lambda = np$.

Hence r can be used to represent the number of successes in n trials if n is large enough.

Using (2.5) and considering $n \to \infty$:

$$p(r) = \frac{n!}{(n-r)!\,r!} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r}$$

$$= \frac{n(n-1)\cdots(n-r+1)}{n^r} \frac{\lambda^r}{r!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^r} \approx e^{-\lambda} \frac{\lambda^r}{r!}$$

because

$$\lim_{n\to\infty} \frac{n(n-1)\cdots(n-r+1)}{n^r} = 1$$

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}, \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^r = 1$$

Example 2.8

Investigate approximating the binomial RV using Poisson RV for two settings, $n = 10$ and $p = 0.4$, and $n = 100$ and $p = 0.04$.

```
n=10;
p=0.4;
lambda=n*p;
for k=0:n;
Pb(k+1)=nchoosek(n,k)*p^k*(1-p)^(n-k);
Pp(k+1)=exp(-lambda)*lambda^k/factorial(k);
end
stem(0:n,Pb,'r');
hold
stem(0:n,Pp,'b');
hold
```

The approximation is not good for small value of n:
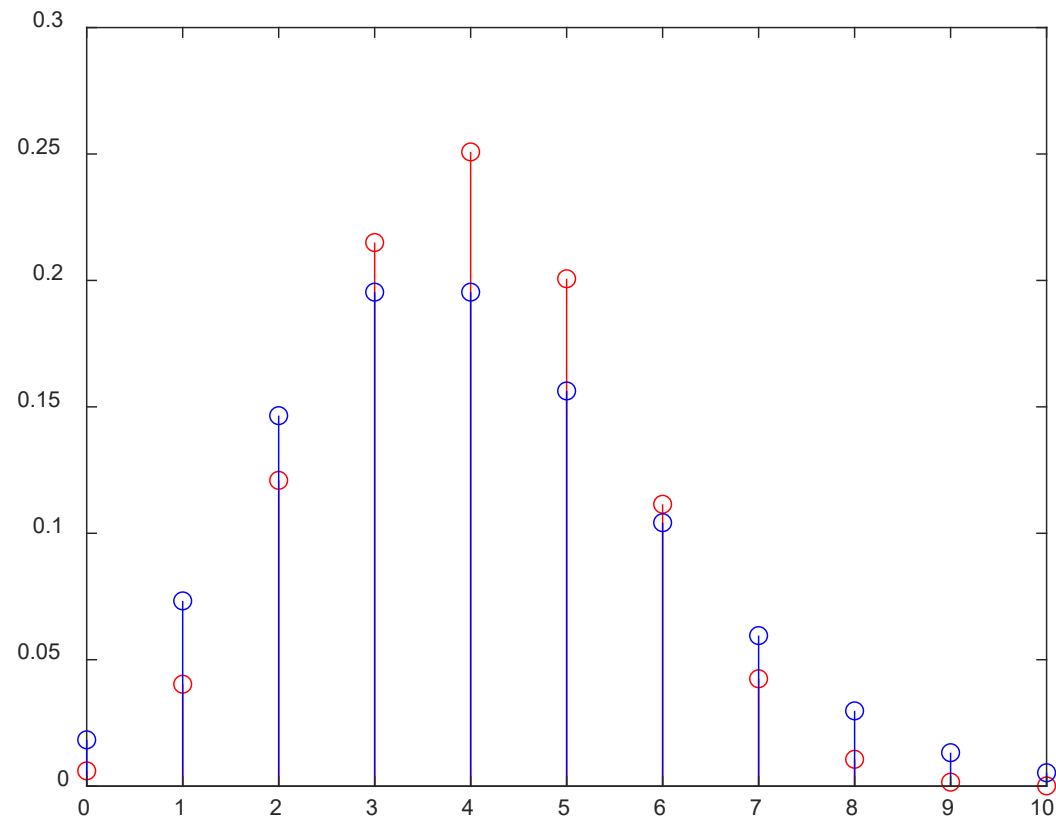
Which binomial RV is most probable to occur? Any relation with n and p?

Repeat with:

```
n=100;
p=0.04;
```
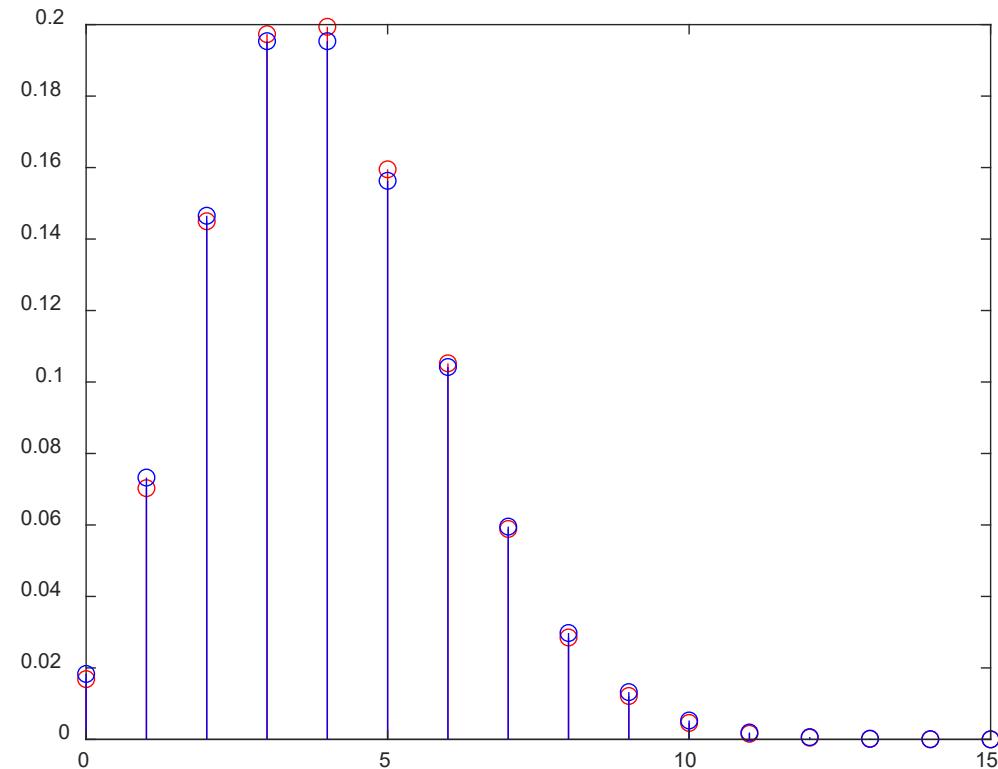
Good approximation is observed for large value of n. Note also that it is not accurate to compute the binomial RV, e.g.,

```
>> nchoosek(100,15)
Warning: Result may not be exact. Coefficient is greater
than 9.007199e+15 and is only accurate to 15 digits
ans = 2.5334e+17
```

To generate RV with $\lambda = np = 4$, we can use:
```
poissrnd(4)
r=poissrnd(4,1, 100000);
histogram(r, 'Normalization', 'pdf')
```

The Poisson RV can be interpreted as the number of events in a given time interval.

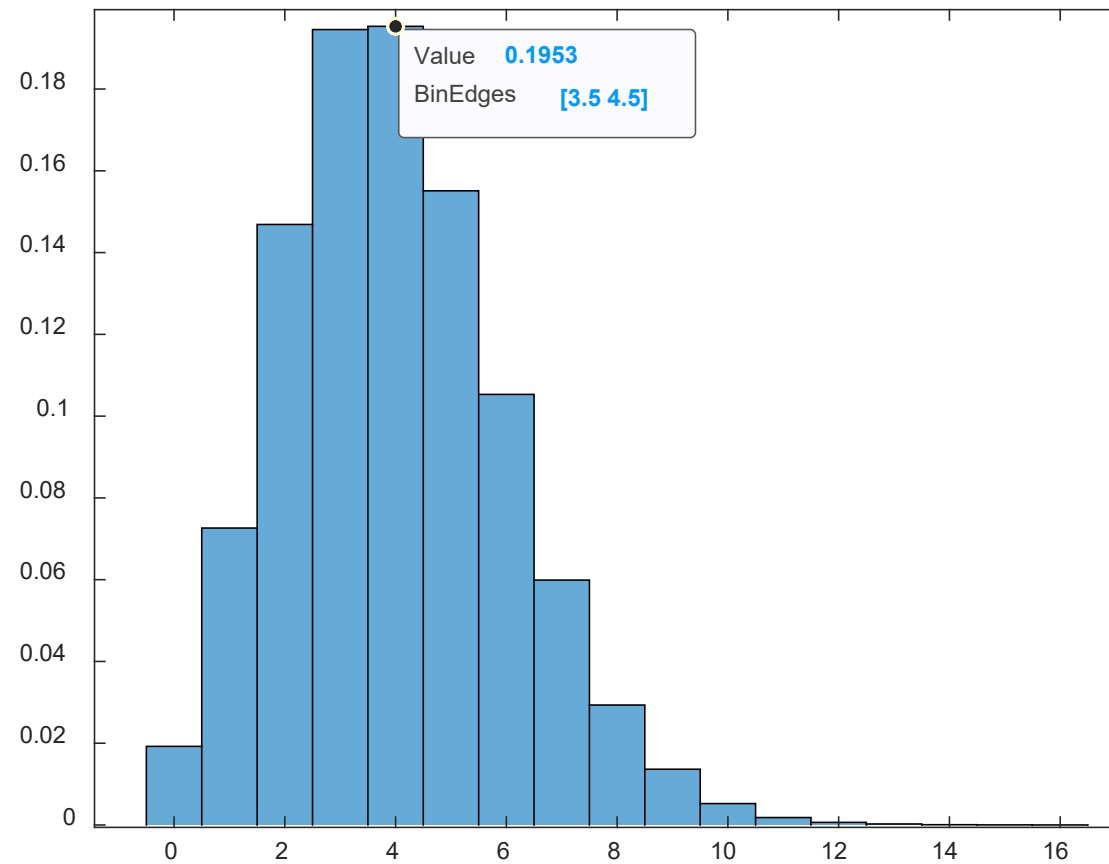Consider a time interval of t and we divide it as n segments where each has a duration of $\varepsilon = t/n$.

length $= \varepsilon$

0

t

number of intervals $= t/\varepsilon \equiv n$
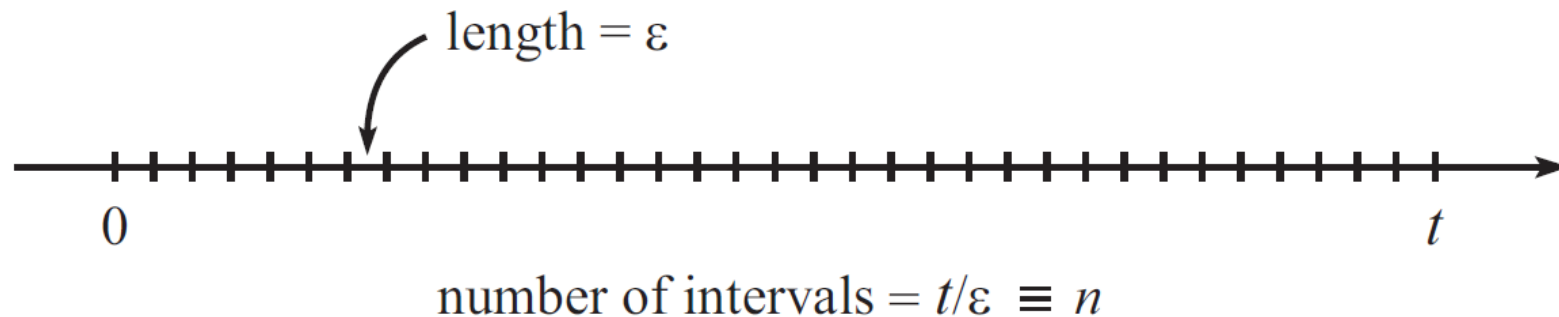
Let $p = \mu\varepsilon$ be the probability of exactly 1 event (success) occurring in the time interval ε. We basically assume that ε is so small such that at most 1 event can occur, and hence the probability of 0 event (failure) is $1 - p$, corresponding to Bernoulli distribution.

Assuming independence among all n intervals, following (2.5), the probability that there are/is exactly r event(s), $0 \le r \le n$ in the time interval of t, is then:

$$p(r) = C(n, r)(\mu\varepsilon)^r(1 - \mu\varepsilon)^{n-r}, \quad 0 \le r \le n \qquad (2.8)$$

Recalling the approximations for $n \to \infty$ while $\varepsilon \to 0$ or $p = \mu\varepsilon \to 0$, (2.8) becomes:

$$p(r) \approx e^{-\mu\varepsilon n}\frac{(\mu\varepsilon n)^r}{r!} = e^{-\mu t}\frac{(\mu t)^r}{r!}, \quad 0 \le r \le n \qquad (2.9)$$

Setting $\lambda = \mu t$ yields the Poisson distribution in (2.7). Considering that $\mu = p/\varepsilon$ is the expected number of events per unit time or average rate of events, λ can be viewed as the average number of events in the time interval t.

Note also that $\mu t = (p/\varepsilon)(n\varepsilon) = np$.

Example 2.9

The number of hits at a website in any time interval is a Poisson RV with an average rate of 2 hits/s. What is the probability that there are no hits in an interval of 0.25s? What is the probability that there are no more than 2 hits in an interval of 1s?

For no hits, it means $r = 0$ in (2.9). Together with $\lambda = \mu t = 2 \cdot 0.25 = 0.5$, the probability is:

$$p(0) = e^{-0.5}\frac{0.5^0}{0!} = e^{-0.5} = 0.6065$$

When $\lambda = \mu t = 2 \cdot 1 = 2$, the probability with no more than 2 hits in 1s is:

$$F(2) = P(X \leq 2) = \sum_{i=0}^{2} p(i) = e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!}\right) = e^{-2}(1+2+2) = 0.6767$$

Example 2.10

Suppose on average, 3.2 α-particles are given off in an interval of 1 second by a radioactive material. What is a good approximation of the probability that no more than 2 α-particles appear in 1 second?

If we consider that the material consists of a large number of atoms n, each of which has probability of $p = 3.2/n$ in giving off an α-particle. As n is large and p is small, we thus can use Poisson distribution to perform the probability approximation with $\lambda = np = 3.2$. The required probability is:

$$P(X \leq 2) = e^{-3.2}\left(1 + 3.2 + \frac{(3.2)^2}{2}\right) \approx 0.38$$

Example 2.11

The number of database queries processed by a computer in any 10s is a Poisson RV X with an average of 5 queries. What is the probability that there will be no queries processed in a 10-second interval? What is the probability that at least 2 queries will be processed in a 2-second interval?

For 10-second interval, $\lambda = 5$. Hence we have:

$$p(0) = e^{-5}\frac{5^0}{0!} = e^{-5} = 0.0067$$

Noting that $\mu = \lambda/t = 5/10 = 0.5$, and for a 2-second interval, $\lambda = 0.5 \cdot 2 = 1$, then the required probability is:

$$1 - (p(0) + p(1)) = 1 - e^{-1}\left(\frac{1^0}{0!} + \frac{1^1}{1!}\right) = 1 - 2e^{-1} = 0.2642$$

Example 2.12

Suppose a customer service counter that processes an order in 1 min., while the average number of customers arrived is 70 per hour. If it is required that no more than one person waiting in line 95% of the time, can the customer service counter meet this requirement?

The requirement means that the probability of 1 person or less is waiting is at least 0.95. Since the processing time is 1 min., at most 2 arrivals in each time slot of 1 min. is allowed: one is just serviced while the other waits a maximum of 1 min. Using Poisson approximation, we have $\lambda = 70/60 \cdot 1 = 7/6$, and compute the probability of 2 or fewer customers arriving at the counter:

$$P(X \leq 2) = e^{-7/6} \left(1 + 7/6 + \frac{(7/6)^2}{2} \right) = 0.8866 < 0.95$$

Hence the requirement cannot be met with 1 service counter. If we add one more counter, this implies that the rate is reduced to 35 per hour, or $\lambda = 7/12$.

The probability of 2 or fewer customers arriving at each counter is now:

$$P(X \leq 2) = e^{-7/12} \left(1 + 7/12 + \frac{(7/12)^2}{2} \right) = 0.9785$$

Considering independence, the probability of 2 or fewer customers arriving at both counters will be $0.9785^2 = 0.9575$ and the requirement is met.

On the other hand, binomial RV can also be applied. Given that the average number of customers arrived is 70 per hour, the probability that a customer arrived in 1s is thus $p = 70/(60 \cdot 60) = 7/360$.

Considering independence in the period of 60s and assuming that in each 1s, there is at most 1 customer is waiting, then according to (2.5) with $p = 7/360$, $n = 60$ and $r = 2$:

$$F(r) = P(X \le r) = \sum_{i=0}^{r} C(n,i)p^i(1-p)^{n-i} = 0.8884$$

```
n=60;
p=7/360;
for k=0:2;
P(k+1)=nchoosek(n,k)*p^k*(1-p)^(n-k);
end
sum(P)
```

It is seen that the calculations based on Poisson and binomial modeling are consistent.

For continuous RV X, where X is any real number in $(-\infty, \infty)$, we use probability distribution function (PDF) $p(x) = P(X = x) = P_X(x)$ and CDF $F(x) = P(X \le x) = P_X(X \le x)$.

The PDF and CDF are related as:

$$p(x) = \frac{d}{dx}F(x) = \frac{d}{dx}P(X \le x) \qquad (2.10)$$

$$F(x) = P(X \le x) = P(X \in (-\infty, x]) = \int_{-\infty}^{x} p(u)\,du \quad (2.11)$$
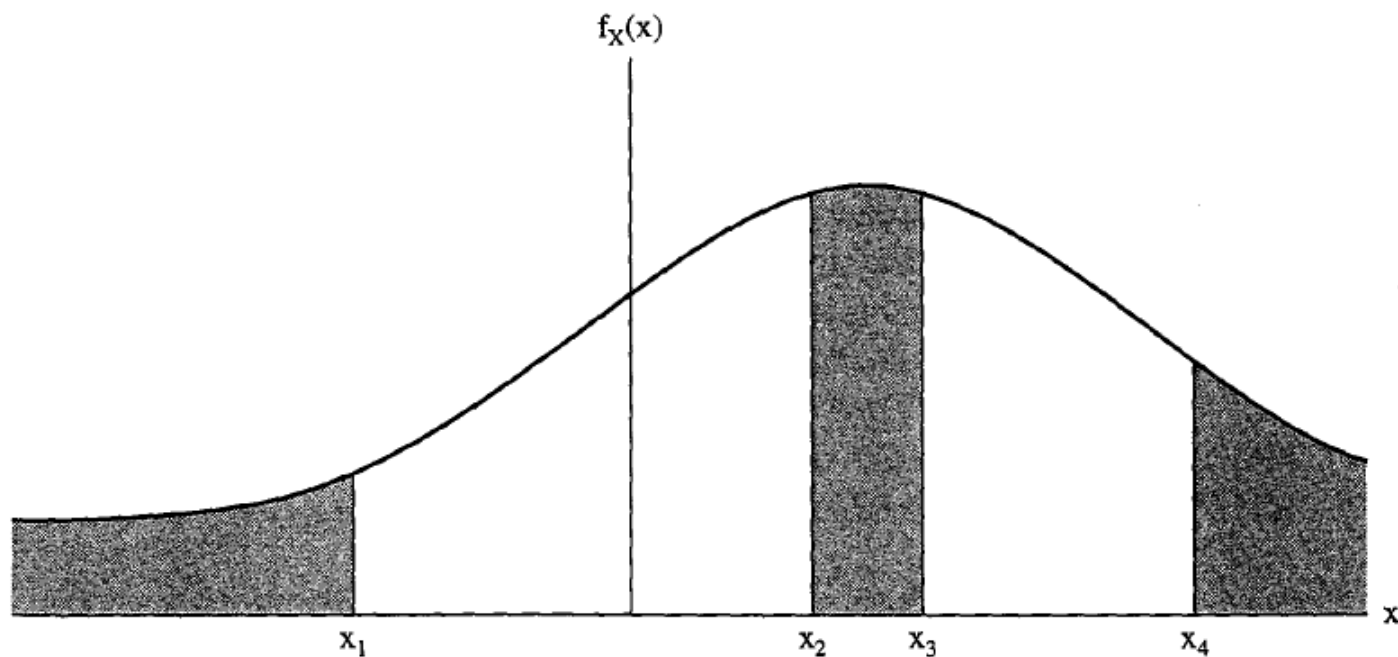
indicating that $F(x)$ is differentiable.

As in the discrete RV case, if PDF is known, then CDF can be determined and vice versa.

From (2.10)-(2.11):

$$P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} p(x)dx = 1$$

$$P(X = a) = \int_{a}^{a} p(x)dx = 0$$

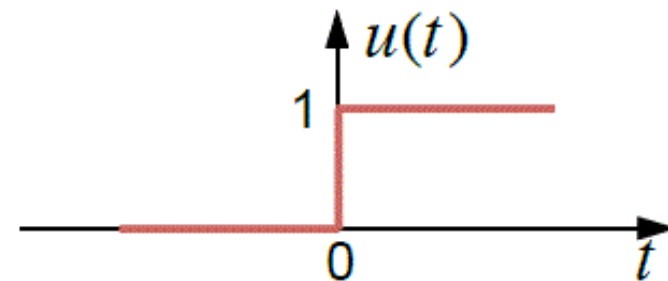$$P(a < X \le b) = \int_{a}^{b} p(x)dx$$

Uniform Distribution

When RV X is uniformly distributed over the interval (a, b), its value can be any value in this range with equal probability of occurrence.

We write $X \sim \mathcal{U}(a, b)$, and the PDF is:

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases} = \frac{1}{b-a}[u(x-a) - u(x-b)] \qquad (2.12)$$

where the unit step function $u(t)$ is defined as:

$$u(t) = \begin{cases} 1, & t > 0 \\ 0, & t < 0 \end{cases}$$

Apparently, a uniform RV can be characterized by a and b.
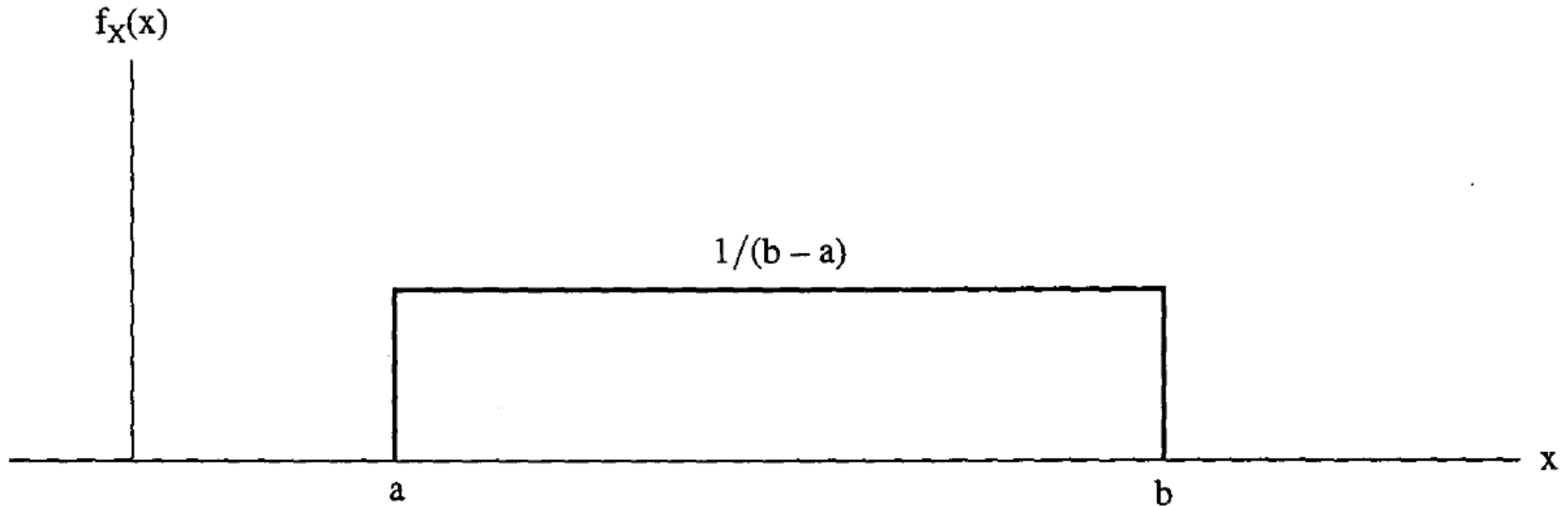
It is commonly used in modeling quantization error, say, rounding a real number to its nearest integer, e.g., $2.6 \to 3$ and $3.2 \to 3$ or $[2.5, 3.5) \to 3$, and we have $a = -0.5$ and $b = 0.5$:

$$p(x) = \begin{cases} 1, & -0.5 < x < 0.5 \\ 0, & \text{otherwise} \end{cases}$$

Why p(x) =1 when x is between (-0.5, 0.5)?

For $X \sim \mathcal{U}(0,1)$, it is called a standard uniform RV.

MATLAB command `rand` can generate standard uniform RV, e.g., each realization of `stem(0:20,rand(1,21))` gives a distinct and random sequence, with values are bounded between 0 and 1.
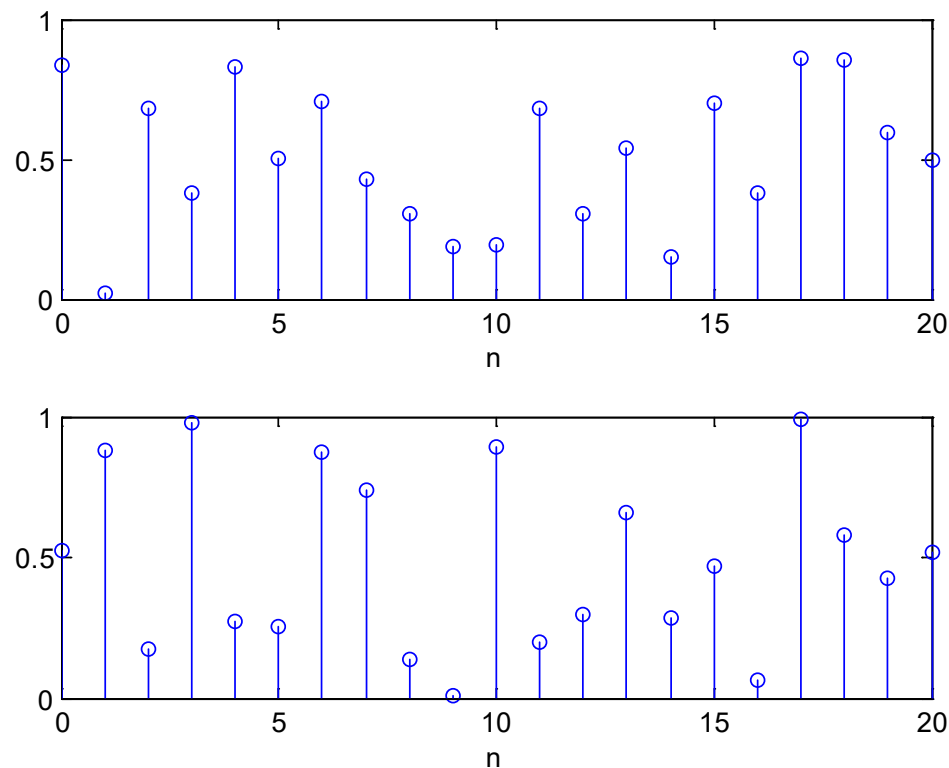
We can easily shift and scale a standard uniform X to produce different uniform RVs.

To obtain $Y \sim \mathcal{U}(a, b)$ from $X \sim \mathcal{U}(0, 1)$, we can use:

$$Y = a + (b - a)X$$

How to get RV with U(-0.5,0.5) from standard one?

How to get RV with U(0,10) from standard one? What is p(x) in the range of (0,10)?

`stem(0:20,rand(1,21)-0.5)`
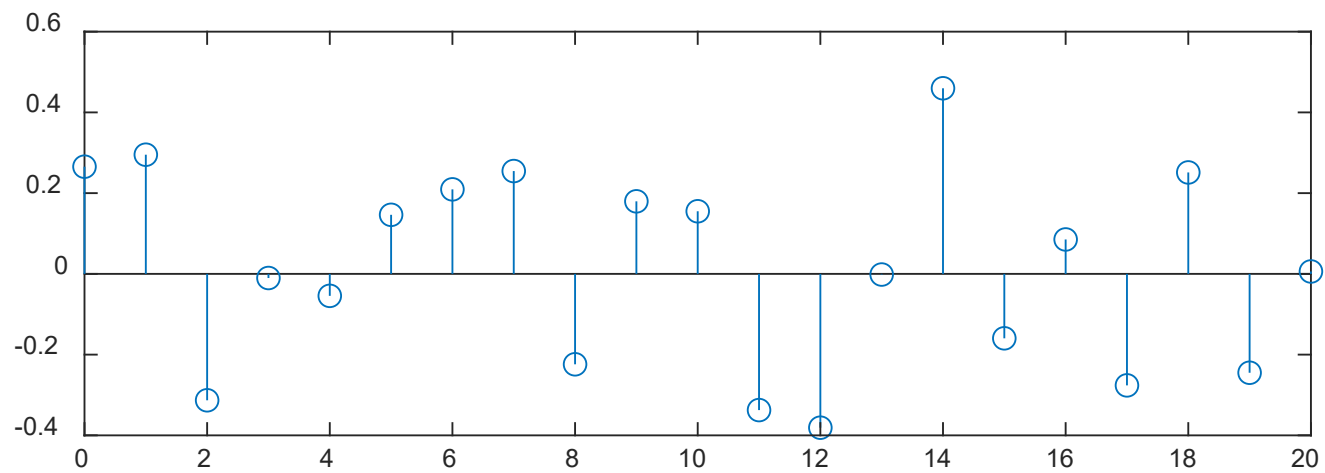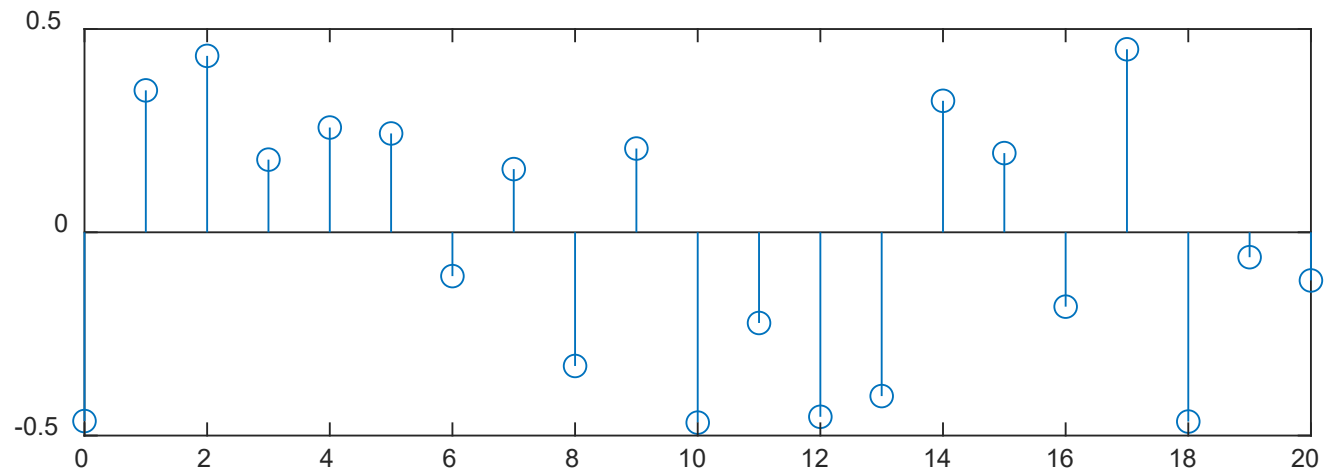
Example 2.13

Find the CDF of a RV uniformly distributed between (a, b). If X is uniformly distributed over (0,10), calculate the probability that $2 \leq X < 3$.

Using (2.11), we get:

$$F(x) = \begin{cases} 0, & a \geq x \\ \dfrac{x-a}{b-a}, & a < x < b \\ 1, & x \geq b \end{cases}$$

Setting $a = 0$ and $b = 10$, the required probability is $F(3) - F(2)$ or $P(2 < X \leq 3)$ is:

$$P(2 < X \leq 3) = F(3) - F(2) = \frac{3-0}{10-0} - \frac{2-0}{10-0} = 0.1$$

Gaussian Distribution

Gaussian (or normal) RV is the most popular continuous RV which is characterized by two parameters, namely, mean μ and standard deviation $\sigma > 0$ (or variance σ^2).

We write $X \sim \mathcal{N}(\mu, \sigma^2)$ and its PDF is:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad (2.13)$$
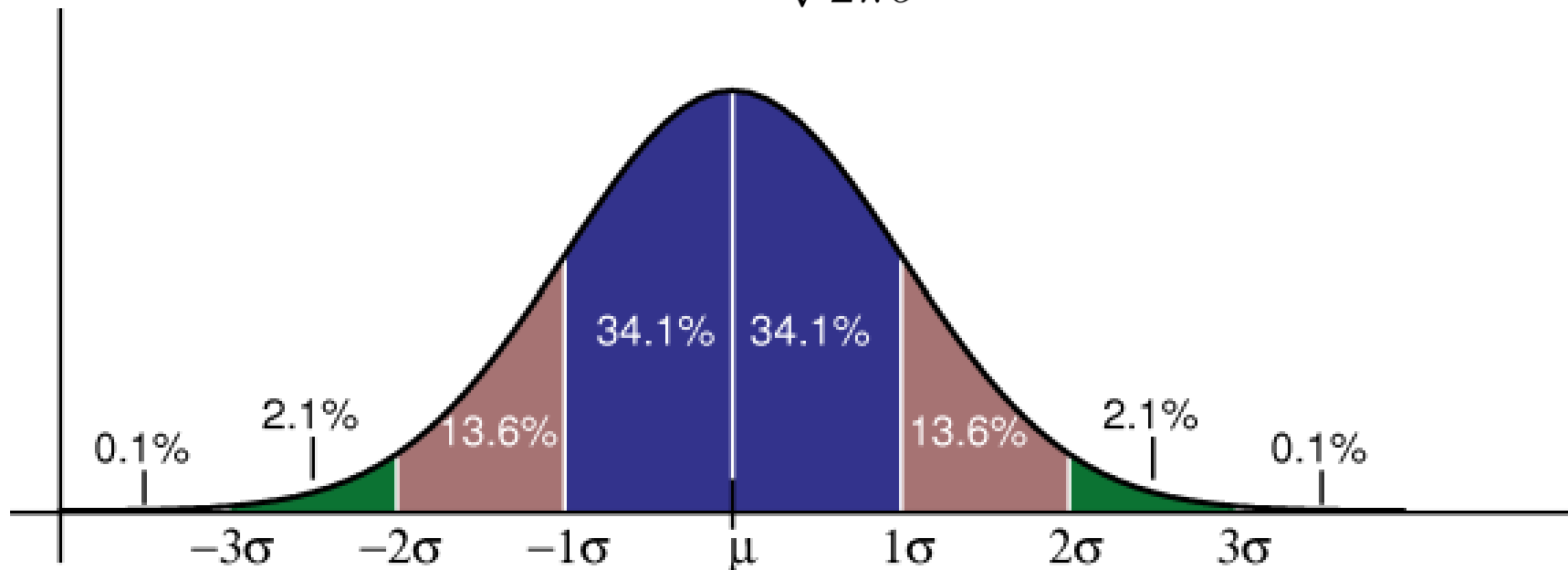
[Gaussian Distribution (princeton.edu)](princeton.edu)

Note that $p(x)$ is symmetric around μ.

Many real-world data particularly when the size is sufficiently large, e.g., population's heights, population's weights, students' examination marks, follow a bell shape aligning with the Gaussian PDF.

For example, the heights of adult men in US are approximately normally distributed with a mean of 70 inches and a standard deviation of 3 inches. Then we may estimate that 99.7% of US adult men have heights between 61 and 79 inches. And we expect the larger the σ, the larger the spread.

For $X \sim \mathcal{N}(0,1)$, it is called a standard Gaussian RV:

$$p(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \Leftrightarrow F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du \quad (2.14)$$

The standard normal complementary PDF is:

$$Q(x) = P(X > x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du = 1 - F(x) \qquad (2.15)$$

Standard normal distribution
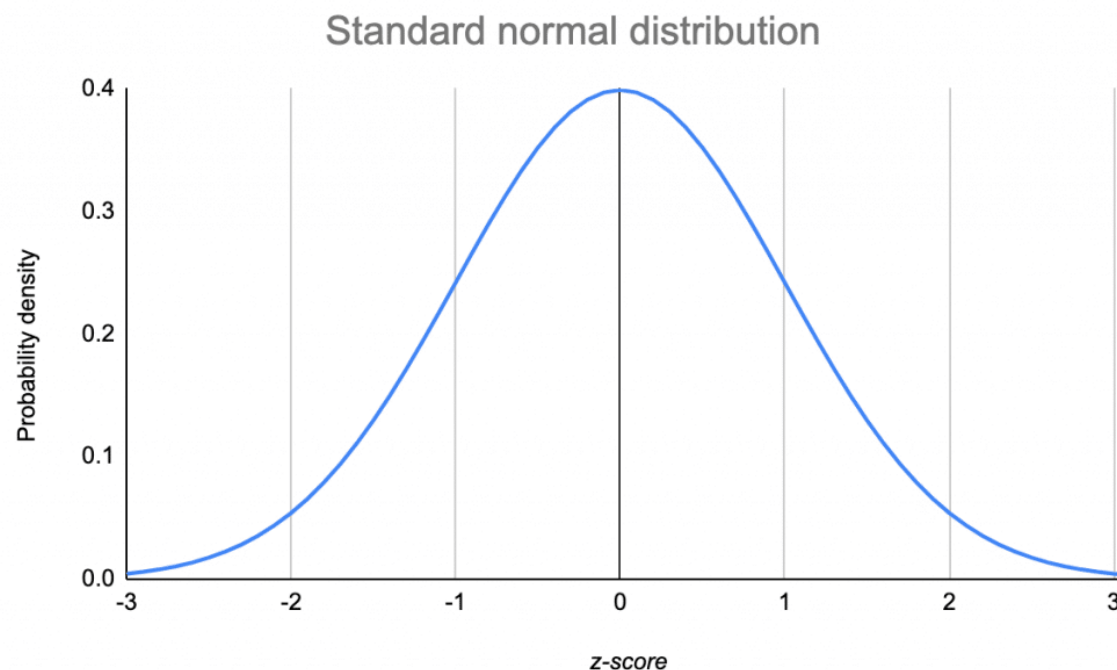
[The Standard Normal Distribution | Examples, Explanations, Uses (scribbr.com)](scribbr.com)

We can see that for $X \sim \mathcal{N}(0,1)$:

$$p(-x) = 1 - p(x) \qquad (2.16)$$

The MATLAB command `randn` can generate standard Gaussian RV where $\mu = 0$ and $\sigma = 1$, e.g., each realization of `stem(0:20,randn(1,21))` gives a distinct and random sequence, whose values are fluctuating around 0.
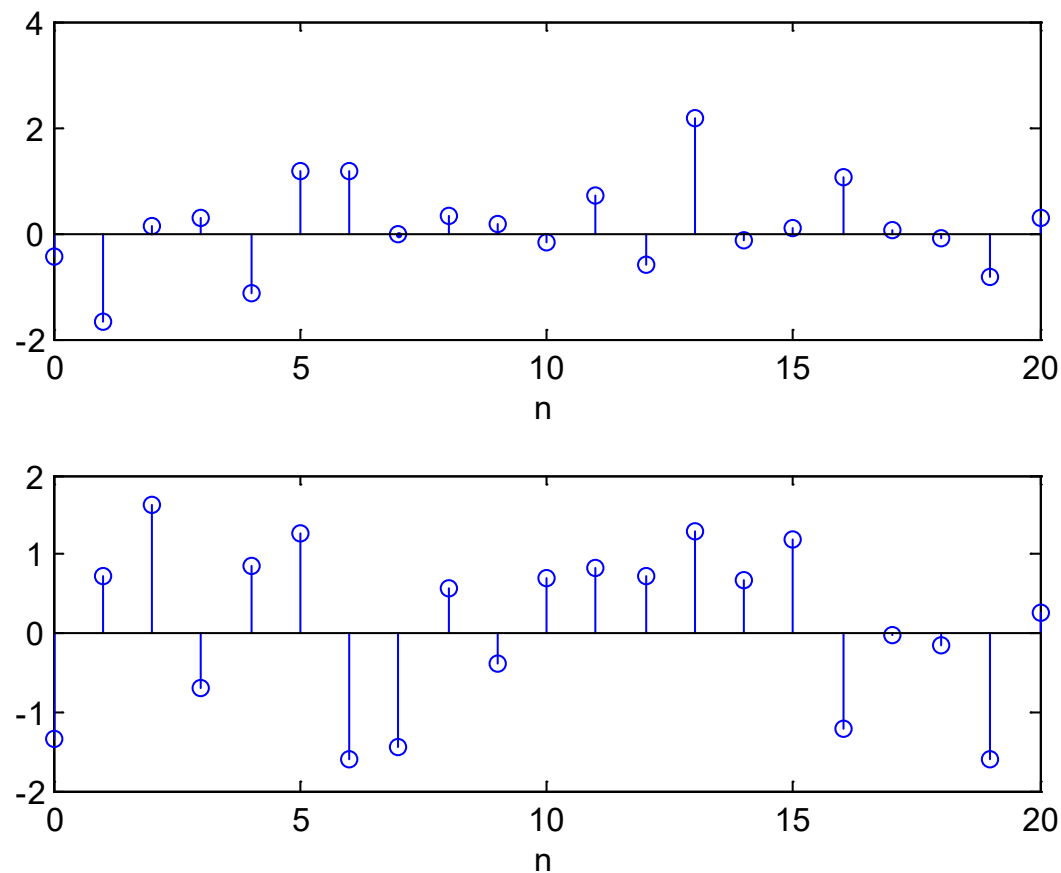
A general Gaussian RV $Y \sim \mathcal{N}(\mu, \sigma^2)$ is related to $X \sim \mathcal{N}(0,1)$:

$$X = \frac{Y - \mu}{\sigma} \Leftrightarrow Y = \mu + \sigma X$$

This implies that we can easily shift and scale a standard Gaussian RV to produce different Gaussian RVs.

Example 2.14
Consider a Gaussian RV $Y \sim \mathcal{N}(61, 100)$. What is $P(Y \leq 46)$? Furthermore, what is $P(51 < Y \leq 71)$?

According to (2.13)-(2.14), the probability is

$$P(Y \leq 46) = \int_{-\infty}^{46} \frac{1}{\sqrt{200\pi}} e^{-\frac{1}{200}(u-61)^2} du = P(X \leq (46 - 61)/10)$$

There are at least 2 ways to compute the numerical value using the MATLAB command `normcdf`.

```
normcdf(46,61,10)
```

```
= 0.0668
```

Or converting to standard Gaussian RV:

```
normcdf((46-61)/10)
```

For the second one, we have:

$$
\begin{aligned}
P(51 < Y \le 71) &= \int_{51}^{71} \frac{1}{\sqrt{200\pi}} e^{-\frac{1}{200}(u-61)^2} du \\
&= \int_{-\infty}^{71} \frac{1}{\sqrt{200\pi}} e^{-\frac{1}{200}(u-61)^2} du - \int_{-\infty}^{51} \frac{1}{\sqrt{200\pi}} e^{-\frac{1}{200}(u-61)^2} du
\end{aligned}
$$

```
normcdf(71,61,10) - normcdf(51,61,10)
```

```
= 0.6827
```

Example 2.15

In an optical fiber transmission system, the probability of a bit error is $Q(\sqrt{\gamma/2})$ where γ is the signal-to-noise ratio. What is the minimum value of γ that produces a bit error rate not exceeding 10^{-6}?

Using (2.15):

$$Q(\sqrt{\gamma/2}) = \int_{\sqrt{\gamma/2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \leq 10^{-6} \Leftrightarrow F(\sqrt{\gamma/2}) > 1 - 10^{-6}$$

The command `norminv` can be used to find $\sqrt{\gamma/2}$:

```
norminv(1-1e-6)

= 4.7534
```

Hence

$$\sqrt{\gamma/2} \geq 4.7534 \Rightarrow \gamma \geq 45.1901 \Rightarrow \gamma_{\min} = 45.1901$$

Exponential Distribution

A RV X is called exponential RV if its PDF has the form:

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad (2.17)$$

where it is characterized by the parameter $\lambda > 0$. The corresponding CDF is:

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad (2.18)$$

Exponential distribution may be interpreted as a continuous version of the geometric distribution. That is, the former models the amount of time, while the latter corresponds to the number of trial(s), until a specific event occurs.

Example 2.16

Suppose the probability that a telephone call lasts no more than t min. is modelled as an exponential RV:

$$p(t) = \begin{cases} \dfrac{1}{3}e^{-t/3}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

Find the probability that a call lasts between 2 and 4 min.

Applying (2.18) with $\lambda = 1/3$, the probability is:

$$P(2 \leq X \leq 4) = P(X \leq 4) - P(X \leq 2) = -e^{-4/3} + e^{-2/3} = 0.2498$$

Expectation of Random Variable

The expectation of a RV X refers to the average values of the RV, which includes mean (1st moment) and power or mean square value (2nd moment), denoted by $\mathbb{E}\{X\}$ and $\mathbb{E}\{X^2\}$, respectively, where $\mathbb{E}$ is the expectation operator.

For $\mathbb{E}\{X\}$, the definitions of discrete and continuous cases are:

$$\mathbb{E}\{X\} = \sum_{x \in R_x} x p(x) \tag{2.19}$$

$$\mathbb{E}\{X\} = \int_{-\infty}^{\infty} x p(x) dx \tag{2.20}$$

These can be generalized to nth moment, e.g., for (2.20):

$$\mathbb{E}\{X^n\} = \int_{-\infty}^{\infty} x^n p(x) dx \tag{2.21}$$

The moment can also be computed about a point x_0, i.e.,

$$\mathbb{E}\{(X - x_0)^n\} = \int_{-\infty}^{\infty} (x - x_0)^n p(x) dx \qquad (2.22)$$

In particular, when $x_0 = \mathbb{E}\{X\}$, (2.22) is the central moment.

The expected values are important because the properties of RV can be obtained, e.g., we may make the best decision under uncertainty (most probable value and risk), and characterize the RV such as normal RV which depends only on mean μ and 2nd central moment or variance σ^2.

Example 2.17
Determine the mean of the Bernoulli RV with success probability p. When $p = 0.1$, what is the expected number of experiment(s) to obtain one success?

Using (2.19):

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

Thus the expected number of successes $\mathbb{E}\{X\}$ is just the probability that the trial will be a success.

Similar to experimental probability, the mean can also be estimated via multiple trials, i.e., average:

$$m_X = \frac{1}{N} \sum_{i=1}^{N} x_i, \quad m_X \to \mathbb{E}\{X\}, \ N \to \infty$$

where x_i denotes the outcome of the ith experiment.

```
r1=random('bino', 1, 0.1, 1,100);
mean(r1)
= 0.0700
```

```
r2=random('bino', 1, 0.1, 1,1000000);
mean(r2)
= 0.1005
```

When $p = 0.1$, the expected number of experiments to obtain one success is $1/p = 10$.

Extending the computation to binomial RV with parameters n and p, we can obtain $\mathbb{E}\{X\} = np$.

When $n = 10$ and $p = 0.1$, we then have $\mathbb{E}\{X\} = np = 1$, corresponding to the consistent result.

```
r1=random('bino', 10, 0.1, 1,100);
mean(r1)
= 0.9100
```

```
r2=random('bino', 10, 0.1, 1, 1000000);
mean(r2)
= 1.00100
```

Example 2.18

Suppose the PMF of RV X is: $p(0) = 0.2$, $p(1) = 0.5$, and $p(2) = 0.3$. Calculate $\mathbb{E}\{X\}$ and $\mathbb{E}\{X^2\}$.

$$\mathbb{E}\{X\} = 0 \cdot 0.2 + 1 \cdot 0.5 + 2 \cdot 0.3 = 1.1$$
$$\mathbb{E}\{X^2\} = 0^2 \cdot 0.2 + 1^2 \cdot 0.5 + 2^2 \cdot 0.3 = 1.7$$

Note that $\mathbb{E}\{X^2\} \neq (E\{X\})^2$.

Example 2.19

Find the expected value of the sum of the faces in Example 2.3.

$$\mathbb{E}\{X\} = \frac{(2 + 12) + 2(3 + 11) + 3(4 + 10) + 4(5 + 9) + 5(6 + 8) + 6 \cdot 7}{36} = 7$$

Hence for betting about the sum value, we should choose "7". This may also be indicated from the maximum value of $p(7)$.
Can we always use the maximum p(x) for E{X}?

Example 2.20

Calculate $\mathbb{E}\{X\}$ for Poisson RV with the use of $\sum_{k=0}^{\infty} \lambda^k/k! = e^\lambda$.

Based on (2.7):

$$\begin{aligned}
\mathbb{E}\{X\} &= \sum_{r=0}^{\infty} r e^{-\lambda} \frac{\lambda^r}{r!} = \sum_{r=1}^{\infty} r e^{-\lambda} \frac{\lambda^r}{r!} = \sum_{r=1}^{\infty} e^{-\lambda} \frac{\lambda^r}{(r-1)!} \\
&= \lambda e^{-\lambda} \sum_{r=1}^{\infty} \frac{\lambda^{r-1}}{(r-1)!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \cdot e^\lambda = \lambda
\end{aligned}$$

Example 2.21

Calculate $\mathbb{E}\{X\}$ and $\mathbb{E}\{X^2\}$ for uniform RV with PDF:

$$p(x) = \frac{1}{b-a}\left[u(x-a) - u(x-b)\right]$$

$$\mathbb{E}\{X\} = \int_{-\infty}^{\infty} x p(x) dx = \int_{a}^{b} \frac{x}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

$$\mathbb{E}\{X^2\} = \int_{-\infty}^{\infty} x^2 p(x) dx = \int_{a}^{b} \frac{x^2}{b-a} dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$

The average value of X aligns with our intuitive understanding. For a standard uniform RV $X \sim \mathcal{U}(0,1)$, the mean and power are 0.5 and 1/3.

Example 2.22
Use MATLAB to generate a sequence of 10000 random numbers uniformly distributed between −0.5 and 0.5 based on the command `rand`. Verify its characteristics.

The mean and variance are computed as:

$$\mathbb{E}\{X\} = \mu = \int_{-\infty}^{\infty} x p(x) dx = \int_{-0.5}^{0.5} x dx = \left.\frac{x^2}{2}\right|_{-0.5}^{0.5} = 0$$

$$\mathbb{E}\{(X-\mu)^2\} = \mathrm{var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx = \int_{-0.5}^{0.5} x^2 dx = \left.\frac{x^3}{3}\right|_{-0.5}^{0.5} = \frac{1}{12}$$

We can use the MATLAB command `rand` which produces a uniform RV between $(0,1)$ to validate these results. To produce X, we need to shift `rand` by 0.5, and employ 10,000 realizations to empirically calculate the mean and variance:

$$\hat{\mu} = \frac{1}{10000} \sum_{i=1}^{10000} X_i \qquad \text{and} \qquad \frac{1}{10000} \sum_{i=1}^{10000} (X_i - \hat{\mu})^2$$

where X_i represents the ith element of `X`.
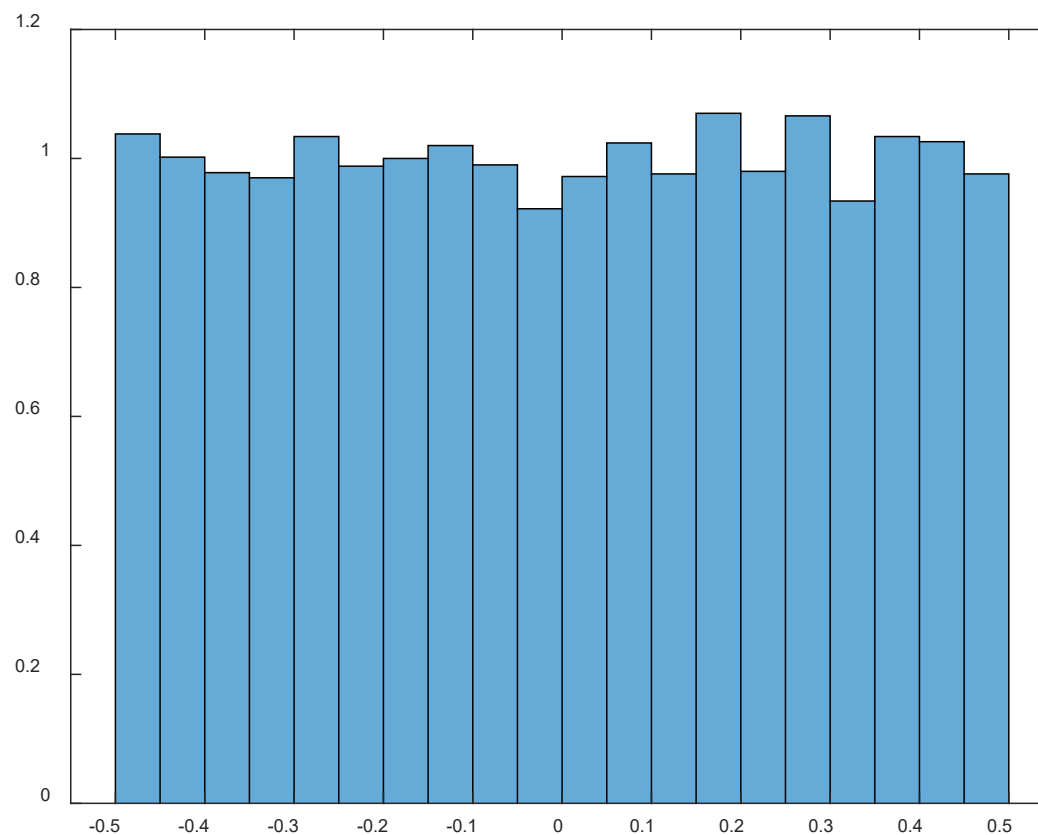
```
X=rand(1,10000)-0.5; %each entry is one realization

mean(X) % mean performs averaging over 10000 values
= 4.5185e-04

mean((X-mean(X)).*(X-mean(X)))
= 0.0838

histogram(X,'Normalization','pdf')
```

We can see that all bins with width 0.05 is a rectangle with height close to 1, e.g.,

$$\int_{-0.5}^{-0.45} p(x)dx \approx \cdots \approx \int_{0.45}^{0.5} p(x)dx \approx 0.05$$

Example 2.23

Calculate $\mathbb{E}\{X\}$ and variance for Gaussian RV.

Using (2.13) and (2.20), we have:

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

$$\begin{aligned}
\mathbb{E}\{X\} &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu) e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx + \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \mu e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} y e^{-\frac{1}{2\sigma^2}y^2} dy + \mu \int_{-\infty}^{\infty} p(x) dx = \mu
\end{aligned}$$

Note that by the symmetric property, the first integral is 0.

The variance is:

$$\text{var}(X) = \mathbb{E}\{(X-\mu)^2\} = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy, \quad y = (x-\mu)/\sigma$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left(-ye^{-y^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2} dy \right), \quad u = y, dv = ye^{-y^2/2} dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = \sigma^2$$

where the Gaussian integral formula $\int_{-\infty}^{\infty} e^{x^2} dy = \sqrt{\pi}$ is used.

Hence a Gaussian RV X is characterized by its mean and variance, and we can write $X \sim \mathcal{N}(\mu, \sigma^2)$.

Example 2.24

Consider a continuous RV X with PDF $p(x)$. Express $\mathrm{var}(X)$ in terms of $\mathbb{E}\{X\}$ and $\mathbb{E}\{X^2\}$.

Let $\mu = \mathbb{E}\{X\}$ for notational simplification. We have:

$$
\begin{aligned}
\mathrm{var}(X) &= \mathbb{E}\{(X-\mu)^2\} = \int_{-\infty}^{\infty}(x-\mu)^2 p(x)dx \\
&= \int_{-\infty}^{\infty} x^2 p(x)dx - 2\mu \int_{-\infty}^{\infty} xp(x)dx + \mu^2 \int_{-\infty}^{\infty} p(x)dx \\
&= \mathbb{E}\{X^2\} - 2\mu \cdot \mu + \mu^2 \\
&= \mathbb{E}\{X^2\} - (\mathbb{E}\{X\})^2
\end{aligned}
$$

Note that the identity holds for both continuous and discrete cases:

$$
\mathrm{var}(X) = \mathbb{E}\{X^2\} - (\mathbb{E}\{X\})^2 \tag{2.23}
$$

Example 2.25

Use the MATLAB command `randn` to generate a standard Gaussian sequence of length 10000. Verify its characteristics.

For a standard Gaussian RV, we have $X \sim \mathcal{N}(0, 1)$. Hence:

$$\mathbb{E}\{X\} = \mu = 0$$

and

$$\text{var}(X) = \mathbb{E}\{(X - \mu)^2\} = \mathbb{E}\{X^2\} = \sigma^2 = 1 \Rightarrow \sigma = 1$$
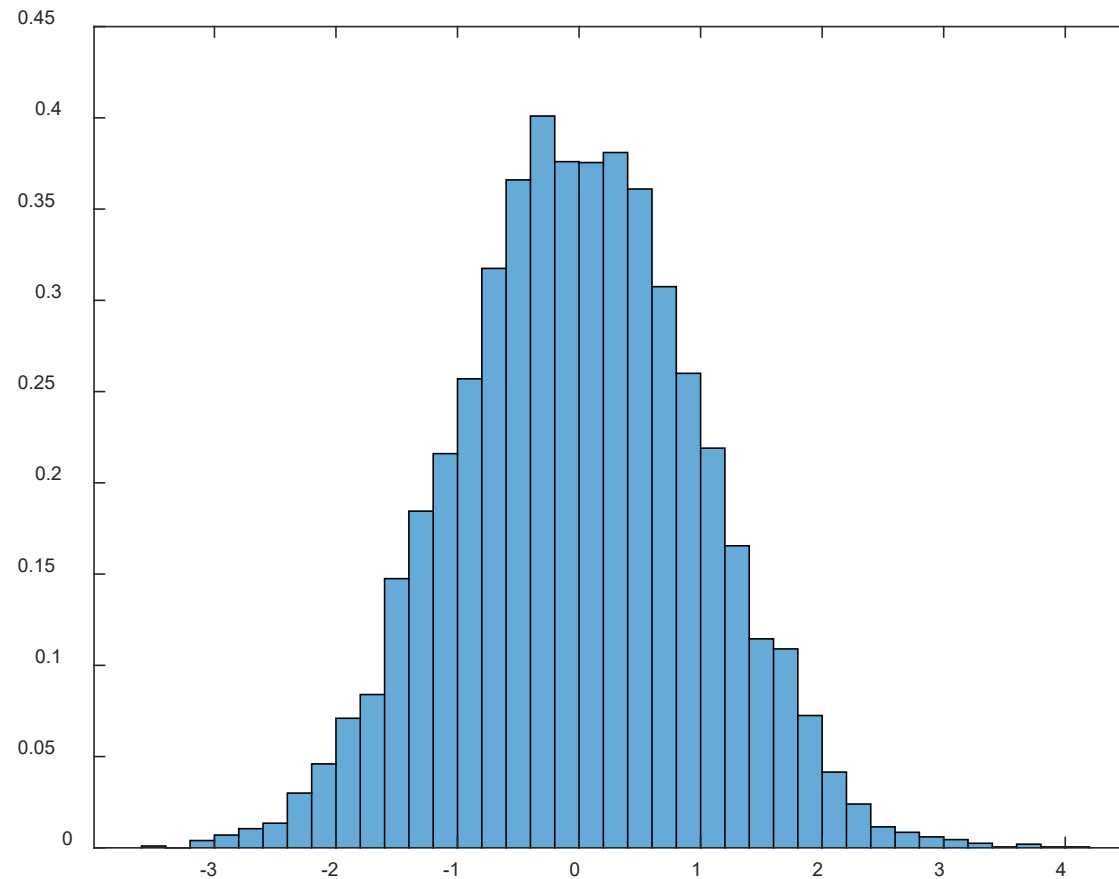
```
X=randn(1,10000); %each entry is one realization

mean(X) % mean performs averaging over 10000 values
= -0.0109

mean((X-mean(X)).*(X-mean(X)))
= 0.9999

histogram(X,'Normalization','pdf')
```

The distribution aligns with Gaussian RV which is indicated by the bell shape.

We also see that most of the values are within $\pm 3\sigma$.

The expectation operator can be generalized to a function of RV X, i.e., $Y = g(X)$ and it is clear that Y is a RV:

$$\mathbb{E}\{Y\} = \mathbb{E}\{g(X)\} = \sum_{x \in R_x} g(x)p(x) \qquad (2.24)$$

$$\mathbb{E}\{Y\} = \mathbb{E}\{g(X)\} = \int_{-\infty}^{\infty} g(x)p(x)dx \qquad (2.25)$$

For example, when $g(X) = X$ and $g(X) = (X - x_0)^n$, (2.25) will become (2.20) and (2.22), respectively.

Example 2.26
Consider a uniform RV $X \sim \mathcal{U}(0, 1)$. Calculate $\mathbb{E}\{X^3\}$.

Applying (2.25), we get

$$\mathbb{E}\{X^3\} = \int_{-\infty}^{\infty} x^3 p(x)dx = \int_0^1 x^3 dx = \left.\frac{x^4}{4}\right|_0^1 = \frac{1}{4}$$

Alternatively, we can start from the CDF. Using the result in Example 2.13 with $a = 0$ and $b = 1$:

$$F(x) = \begin{cases} 0, & 0 \le x \\ x, & 0 < x < 1 \\ 1, & x \ge 1 \end{cases}$$

That is, $F(x) = P(X \le x) = x$ for $0 < x < 1$.

Let $Y = X^3$. As $X \in (0, 1)$, we have $Y \in (0, 1)$.

$$F_Y(y) = P(Y \le y) = P(X^3 \le y) = P(X \le y^{1/3}) = y^{1/3}, \quad 0 < y < 1$$

Applying (2.10)

$$p_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{3}y^{-2/3}, \quad 0 < y < 1$$

Hence according to (2.20):

$$\mathbb{E}\{X^3\} = \mathbb{E}\{Y\} = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y \cdot \frac{1}{3} y^{-2/3} dy = \frac{1}{3} \int_0^1 y^{1/3} dy = \frac{1}{4}$$

Writing $Y = g(X)$, its PDF can also be obtained as:

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

Now $Y = g(X) = X^3 \Rightarrow g^{-1}(y) = x = y^{1/3}$. Again, with $Y \in (0,1)$, we then compute:
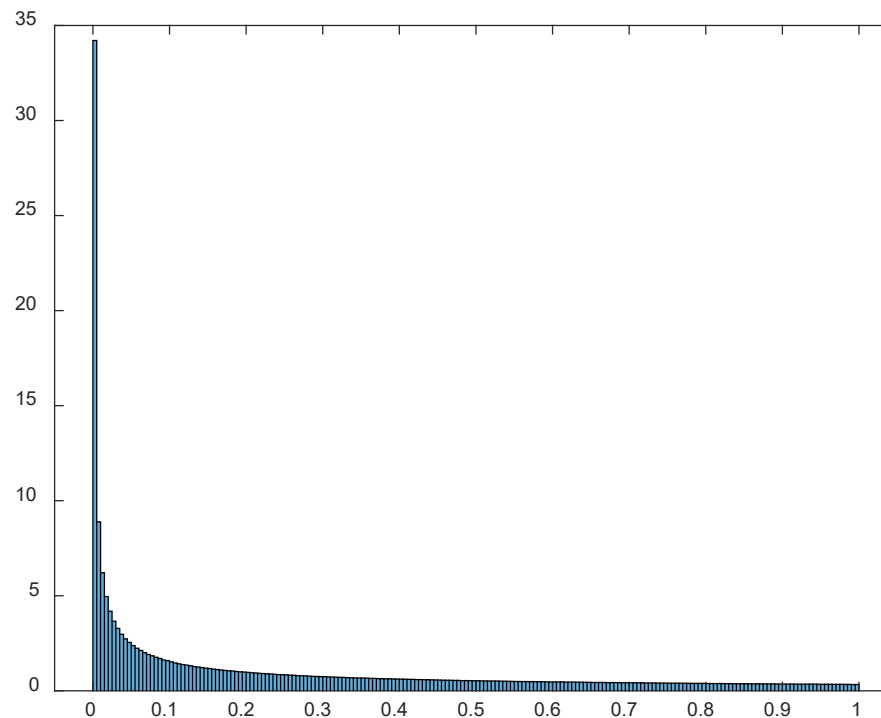
$$p_X(g^{-1}(y)) = p_X(x) = 1, \quad 0 < y < 1$$

$$\frac{dg^{-1}(y)}{dy} = \frac{dy^{1/3}}{dy} = \frac{1}{3} y^{-2/3} = \left| \frac{1}{3} y^{-2/3} \right|, \quad 0 < y < 1$$

Combining the results, we get the same $p_Y(y)$.

The same result is obtained. Comparing the two approaches, the first method is much simpler while the second method can provide the CDF and PDF.

```
x=rand([1,10000000]);
y=x.^3;
histogram(y,'Normalization','pdf')
```

For discrete RV, the PMF of $Y = g(X)$ is:

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x) \qquad (2.26)$$

Example 2.27

Suppose the amplitude V of signal is a discrete RV with PMF:

$$p_V(v) = \begin{cases} \dfrac{1}{7}, & v = -3, -2, \cdots, 3 \\ 0, & \text{otherwise} \end{cases}$$

Let $Y = V^2/2$ be the signal power. Find $p_Y(y)$ and determine $\mathbb{E}\{Y\}$.

Y only has 4 possible values, 4.5 if $v = -3$, 3, 2 if $v = -2$, 2, 0.5 if $v = -1$, 1 and 0 if $v = 0$.

Using (2.26) together with the given $p_V(v)$, we have:

$$p_Y(y) = \begin{cases} \dfrac{1}{7}, & y = 0 \\ \dfrac{2}{7}, & y = 0.5, \ 2, \ 4.5 \\ 0, & \text{otherwise} \end{cases}$$

Note that $Y = V^2/2$ implies $V = \pm\sqrt{2Y}$, i.e., $p_Y(0) = p_V(0) = 1/7$, $p_Y(y) = p_V(-\sqrt{2y}) + p_V(\sqrt{2y}) = 2/7$, $y = 0.5, 2, 4.5$. Applying (2.19), $\mathbb{E}\{Y\}$ is:

$$\mathbb{E}\{Y\} = \frac{1}{7} \cdot 0 + \frac{2}{7}(0.5 + 2 + 4.5) = 2$$

Alternatively, we can use (2.24):

$$\mathbb{E}\{Y\} = \frac{1}{7}(4.5 + 2 + 0.5 + 0 + 0.5 + 2 + 4.5) = 2$$

Example 2.28

Given $Y = aX + b$ where X is a continuous RV while a and b are constants. Determine $\mathbb{E}\{Y\}$.

According to (2.20), we have:

$$
\begin{aligned}
\mathbb{E}\{Y\} = E\{aX + b\} &= \int_{-\infty}^{\infty} (ax + b)p(x)dx \\
&= a \int_{-\infty}^{\infty} xp(x)dx + b \int_{-\infty}^{\infty} p(x)dx \\
&= a\mathbb{E}\{X\} + b
\end{aligned}
$$

That is,
$$
\mathbb{E}\{aX + b\} = a\mathbb{E}\{X\} + b \tag{2.27}
$$

Note that (2.27) also holds for discrete RV.

Moment Generating Function

The moment generating function of a continuous RV X is defined as the expected value of e^{tX}, denoted by $\phi(t)$:

$$\phi(t) = \mathbb{E}\{e^{tX}\} = \int_{-\infty}^{\infty} e^{tx} p(x) dx \qquad (2.28)$$

$$\phi'(t) = \frac{d\phi(t)}{dt} = \frac{dE\{e^{tX}\}}{dt} = \mathbb{E}\{Xe^{tX}\} \Rightarrow \phi'(0) = \mathbb{E}\{X\}$$

In general:

$$\phi^n(t) = \frac{d^n \phi(t)}{dt^n} = \frac{d^n E\{e^{tX}\}}{dt^n} = \mathbb{E}\{X^n e^{tX}\} \Rightarrow \phi^n(0) = \mathbb{E}\{X^n\}$$

Hence we can use $\phi(t)$ for moment computation:

$$\phi^n(0) = \mathbb{E}\{X^n\}, \quad n \geq 1 \qquad (2.29)$$

Note that the characteristic function $\psi(t) = \mathbb{E}\{e^{jtX}\}$, which is similar to $\phi(t)$, has a relationship with $p(x)$, which is analogous to Fourier transform:

$$\psi(t) = \int_{-\infty}^{\infty} e^{jtx} p(x) dx \leftrightarrow p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-jtx} \psi(t) dt \qquad (2.30)$$

For discrete RV X, the moment generating function is:

$$\phi(t) = \mathbb{E}\{e^{tX}\} = \sum_{x \in R_x} e^{tx} p(x) dx \qquad (2.31)$$

where (2.29) is still valid.

Example 2.29
Repeat Example 2.23 to calculate $\mathbb{E}\{X\}$ and variance for a Gaussian RV $X \sim \mathcal{N}(\mu, \sigma^2)$ using (2.28).

We first consider $Y \sim \mathcal{N}(0,1)$. Based on (2.28):

$$\mathbb{E}\{e^{tY}\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty} e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2-2ty)/2} dy$$

$$= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-t)^2/2} dy = e^{t^2/2}$$

Since $Y \sim \mathcal{N}(0,1)$, then $X = \sigma Y + \mu \sim \mathcal{N}(\mu, \sigma^2)$. Hence:

$$\phi(t) = \mathbb{E}\{e^{tX}\} = \mathbb{E}\{e^{t(\sigma Y+\mu)}\} = e^{t\mu}\mathbb{E}\{e^{t\sigma Y}\} = e^{(\mu t + \sigma^2 t^2/2)}$$

Applying chain rule in differentiation yields:

$$\phi'(t) = (\mu + \sigma^2 t)e^{(\mu t + \sigma^2 t^2/2)} \Rightarrow \phi'(0) = \mathbb{E}\{X\} = \mu$$

$$\phi''(t) = (\mu + \sigma^2 t)^2 e^{(\mu t + \sigma^2 t^2/2)} + \sigma^2 e^{(\mu t + \sigma^2 t^2/2)} \Rightarrow \phi''(0) = \mathbb{E}\{X^2\} = \mu^2 + \sigma^2$$

From (2.23):

$$\mathrm{var}(X) = \mathbb{E}\{X^2\} - (\mathbb{E}\{X\})^2 = \sigma^2$$

Example 2.30

Calculate $\mathbb{E}\{X\}$ and variance for binomial RV which is characterized by n and p using (2.31).

Using (2.5) and (2.31), we have:

$$\phi(t) = \mathbb{E}\{e^{tX}\} = \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n}\binom{n}{k}(e^t p)^k(1-p)^{n-k}$$
$$= (e^t p + 1 - p)^n$$

$$\phi'(t) = n(e^t p + 1 - p)^{n-1}pe^t \Rightarrow \phi'(0) = \mathbb{E}\{X\} = np$$

$$\phi''(t) = n(n-1)(e^t p + 1 - p)^{n-2}(pe^t)^2 + n(e^t p + 1 - p)^{n-1}pe^t$$
$$\Rightarrow \phi''(0) = \mathbb{E}\{X^2\} = n(n-1)p^2 + np$$

From (2.23):

$$\text{var}(X) = \mathbb{E}\{X^2\} - (\mathbb{E}\{X\})^2 = \sigma^2 = n(n-1)p^2 + np - n^2 p^2 = np(1-p)$$

References:

1. S. M. Ross, *Introduction to Probability Models*, Academic Press, 2014
2. D. G. Childers, *Probability and Random Processes: Using MATLAB with Applications to Continuous and Discrete Time Systems*, Irwin, 1997
3. D. Morin, *Probability: For the Enthusiastic Beginner*, CreateSpace Independent Publishing Platform, 2016
4. R. D. Yates and D. J. Goodman, *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineer*, Wiley, 2014
5. S. Kay, *Intuitive Probability and Random Processes using MATLAB*, Springer, 2006
6. H. Pishro-Nik, *Introduction to Probability, Statistics, and Random Processes*, Kappa Research LLC, 2014
   https://www.probabilitycourse.com