Christopher Lum
lum@uw.edu

Lecture09b
Probability Distributions

The YouTube video entitled ‘TBD’ that covers this lecture is located at TBD.

Outline

-Random Variables. Probability Distributions
  -Discrete Random Variables and Distributions
  -Continuous Random Variables and Distributions
-Mean and Variance of a Distribution
  -Transformation of Mean and Variance
  -Expectations and Moments

Random Variables and Probability Distributions

We now investigate the concept of a **probability distribution**, or briefly, a **distribution**. This shows the probabilities of events in an experiment. The quantity that we observe in an experiment will be denoted by X and is called a **random variable** (or **stochastic variable**).

If the random values takes on discrete values (ie a die has 6 discrete values: 1, 2, 3, 4, 5, 6), we will obtain a **discrete random variable** and an associated **discrete distribution**.

If the random value can take on a continuous value (ie the voltage in a capacitor), we will obtain a **continuous random variable** and an associated **continuous distribution**.

In both cases, the distribution of X is determined by the **distribution function**

$$F(x) = P(X \leq x) \qquad \textbf{(Eq.1)}$$

This is the probability that in a trial, X will assume any value not exceeding x. This is also sometime referred to as the **cumulative distribution function**.

Definition: Random Variable

A random variable X is a function defined on the sample space S of an experiment. Its values are real numbers. For every number a, the probability $P(X = a)$ is defined and denotes the probability that X takes on the exact value a.

Similarly, for any interval I, the probability $P(X \in I)$ is defined and denotes the probability that X assumes any value in I.

From Eq.1, we obtain the fundamental formula for the probability corresponding to an interval $x \in (a, b]$

$$P(a < x \le b) = F(b) - F(a) \qquad \textbf{(Eq.2)}$$

Discrete Random Variables and Distributions

Consider now a discrete random variable X which can assume values $x_1, x_2, \ldots$. Each of these values has a positive probability associated with them which we denote p_i. For example
$p_1 = P(X = x_1)$, $p_2 = P(X = x_2)$, etc.

Furthermore, it is worthwhile to mention that the values of $x_1, x_2, \ldots$ do not need to be integer values. For example, suppose X is price of items in a grocery cart.

$x_1 = \$5.50$
$x_2 = \$17.43$
$x_3 = \$22.33$

$p_1 = P(X = x_1) = P(X = 5.50)$
$p_2 = P(X = x_2) = P(X = 17.43)$
$p_3 = P(X = x_3) = P(X = 22.33)$

Clearly, the discrete distribution of X is also determined by the **probability function** (aka **probability density function**, **probability distribution function**) , $f(x)$ of X, defined by

$$f(x) = \begin{cases} p_j = P(X = x_j) & \text{if } x = x_j \\ 0 & \text{otherwise} \end{cases} \qquad j = 1, 2, \ldots \qquad \textbf{(Eq.3)}$$

This function $f(x)$ returns the probability that the event x will occur.

From this, we get the values of the **distribution function** (aka **cumulative distribution function**) $F(x)$ by taking sums

$$F(x) = \sum_{x_j \le x} f(x_j) = \sum_{x_j \le x} p_j \qquad \textbf{(Eq.4)}$$

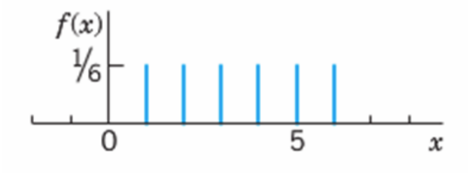
In Eq.4, for any given x, we sum all the probabilities p_j which are associated with values smaller than or equal to that of x. This is a step function with upward jumps of size p_j at the possible values x_j of X and is constant in between points. Make note that the inequality in Eq.4 is $\le$ as this is important for discrete distributions.

Example 1: Probability Function and Distribution Function

Consider a random variable

X = the value that a fair die turns up

We see that X has possible values of $x = 1, 2, 3, 4, 5, 6$ with probability of $p(x_j) = 1/6$, for $j = 1, 2, ..., 6$. Therefore, we can easily construct the probability function, $f(x)$ as shown below



Now, we can construct the distribution function, $F(x)$ using Eq.4

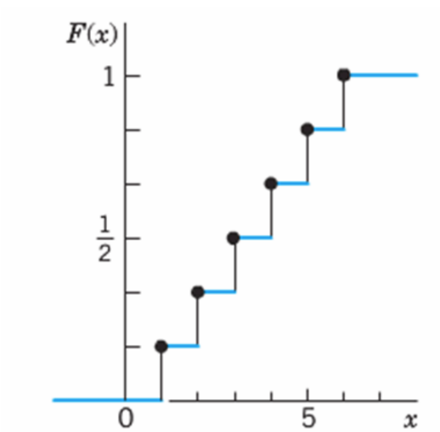
$$F(x_1) = \sum_{x_j \leq x_1} p_j = p_1 = \frac{1}{6}$$

$$F(x_2) = \sum_{x_j \leq x_2} p_j = p_1 + p_2 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$\vdots$$

$$F(x_6) = \sum_{x_j \leq x_6} p_j = p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

Graphically, we can represent $F(x)$ as shown below. Note that at points where the function value might be ambiguous (i.e. at the jumps), the black circles are used to denote the value of the function.



Example 2: Probability Function and Distribution Function

Now consider a random variable

X = the sum of two fair dice

We can enumerate all possible outcomes for this somewhat simple problem

	1	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

So we see that the sample space of this problem is

$$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\} = \{x_1, x_2, \ldots, x_{11}\}$$

However, we see that some values occur more often than others so not each of these values has an equal probability. It may be useful to enumerate all these probabilities.

$$p_1 = p(x_1) = P(X = 2) = \frac{1}{36}$$

$$p_2 = p(x_2) = P(X = 3) = \frac{2}{36} = \frac{1}{18}$$

$$p_3 = p(x_3) = P(X = 4) = \frac{3}{36} = \frac{1}{12}$$

$$p_4 = p(x_4) = P(X = 5) = \frac{4}{36} = \frac{1}{9}$$

$$p_5 = p(x_5) = P(X = 6) = \frac{5}{36}$$

$$p_6 = p(x_6) = P(X = 7) = \frac{6}{36} = \frac{1}{6}$$

$$p_7 = p(x_7) = P(X = 8) = \frac{5}{36}$$

$$p_8 = p(x_8) = P(X = 9) = \frac{4}{36} = \frac{1}{9}$$

$$p_9 = p(x_9) = P(X = 10) = \frac{3}{36} = \frac{1}{12}$$

$$p_{10} = p(x_{10}) = P(X = 11) = \frac{2}{26} = \frac{1}{18}$$

$$p_{11} = p(x_{11}) = P(X = 12) = \frac{1}{36}$$

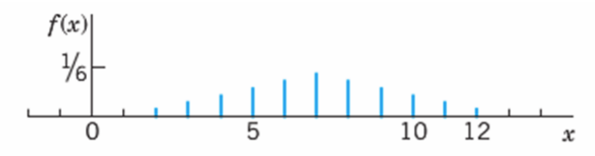
Note that $p_1 = p(x_1)$ is actually the probability that the sum is 2, not 1 (the indexing is off by one)

Math Joke: Obi-Wan Error

In software, being off by one in an index is sometimes referred to as the ‘Off-By-One Error’ → ‘OB1 Error’

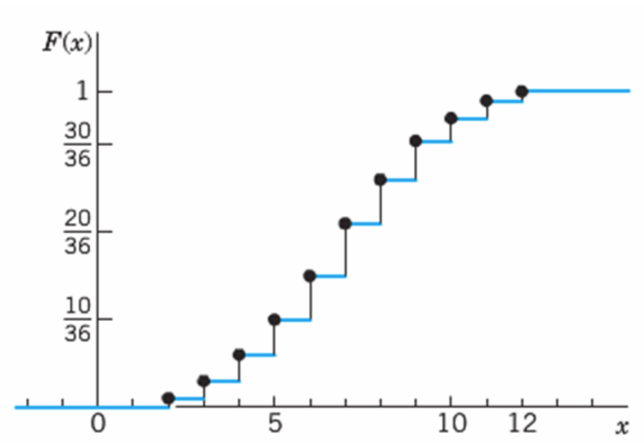
→ Obi-Wan Error

So the probability function is as shown below



Interestingly, note that even though the probability distribution functions for a single dice roll is uniform, when we sum the results of the two, we obtain a distribution that is not uniform, it is more bell shaped. We will revisit this when we discuss the Central Limit Theorem and Gaussian Distributions.

Graphically, we can represent the distribution function $F(x)$ as shown below



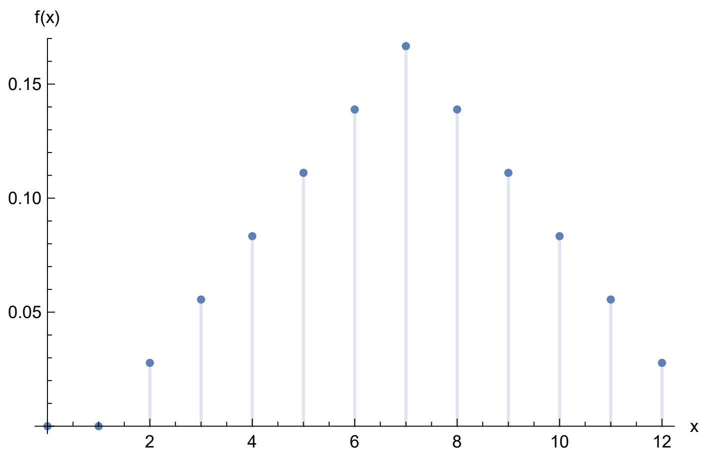
We can define this probability distribution function using a ‘Module’ in Mathematica.

If this was not covered in a previous lecture, go to Mathematica Module Tutorial (module_tutorial-.doc)

```
(*Define the probability distribution function*)
f[x_] := Module[
  (*Define local variables*)
  {PS, binIndex},

  (*Function body*)

  (*If x is not an integer or if x is outside range [2,12], return 0*)
  If[! Element[x, Integers],
   Return[0],
  ];

  If[x > 12,
   Return[0],
  ];

  If[x < 2,
   Return[0],
  ];

  (*x∈[2,12] and is an integer, compute the associated discrete probability*)
  PS = {1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36};

  binIndex = Floor[x] - 1;

  Return[PS[[binIndex]]]
 ]
```

We can now plot the probability distribution function, $f(x)$, using Mathematica's 'DiscretePlot' function.

```
xMin = 0;
xMax = 12;
dx = 1;
DiscretePlot[f[x], {x, xMin, xMax, dx},
 AxesLabel → {"x", "f(x)"}]

Clear[xMin, xMax, dx]
```

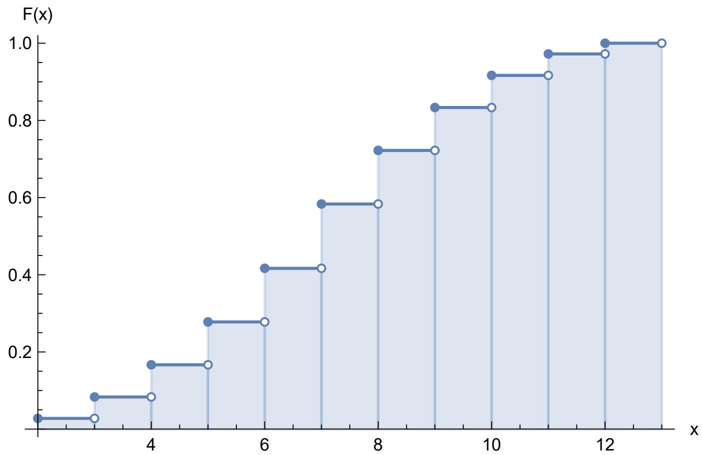


We can now define the cumulative distribution function

```
F[x_] := Module[
  (*Define local variables*)
  {},

  (*Function body*)

  (*If x is not an integer, return 0*)
  If[! Element[x, Integers],
   Return[0],
  ];

  (*Sum the probabilities up to this value*)
  Return[Sum[f[p], {p, 0, x}]]
 ]
```

We can plot this using ‘DiscretePlot’ along with some extra options to make it appear as in the textbook

```
xMin = 2;
xMax = 12;
dx = 1;
DiscretePlot[F[x], {x, xMin, xMax, dx},
 ExtentSize → Right, ExtentMarkers → {"Filled", "Empty"},
 AxesLabel → {"x", "F(x)"}]
```



Two useful formulas for discrete distributions are readily obtained as follows. For the probability corresponding to intervals we have from Eq.2 and Eq.4

$$P(a < X \leq b) = F(b) - F(a) = \sum_{a < x_j \leq b} p_j \qquad (X \text{ discrete}) \qquad \textbf{\textit{(Eq.5)}}$$

In Eq.5, one must exercise caution with the inequalities. Note that this is the probability that X is strictly greater than a but less than or equal to b.

From Eq.5 we can see that

$$\sum_j p_j = 1 \qquad \text{(sum of all probabilities)} \qquad \textbf{\textit{(Eq.6)}}$$

Example 3: Illustration of Eq.5

From Example 2 (summing the roll of two fair die), we can compute the probability of a sum of at least 4 and at most 8.

Solution: We need to be careful about the < vs. ≤ sign in Eq.5. To solve this problem, we seek

$$P(3 < X \leq 8) = F(8) - F(3)$$

We can compute $F(3)$ and $F(8)$ (taking special note to see that p_1 is actually $P(X = x_1) = P(X = 2)$)

$$F(x) = \sum_{x_j \leq x} f(x_j) = \sum_{x_j \leq x} p_j$$

$$F(3) = \sum_{x_j \leq 3} f(x_j)$$

$= \sum_{x_j \leq 3} p_j$ recall: this notation means we sum probabilities associated with values less than 3

$= P(X = 2) + P(X = 3)$

$= p_1 + p_2$

$= \frac{1}{36} + \frac{2}{36}$

$$\frac{1}{36} + \frac{2}{36}$$

F[3]

$$\frac{1}{12}$$

$$\frac{1}{12}$$

So we obtain

$F(3) = \frac{1}{12}$ = probability that $X \leq 3$

Similarly $F(8)$ is given by

$F(8) = \sum_{x_j \leq 8} f(x_j)$

$= \sum_{x_j \leq 8} p_j$ recall: this notation means we sum probabilities associated with values less than 8

$= P(X = 2) + P(x = 3) + P(x = 4) + P(x = 5) + P(x = 6) + P(x = 7) + P(x = 8)$

$= p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7$

$= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36}$

$$\frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36}$$

F[8]

$$\frac{13}{18}$$

$$\frac{13}{18}$$

So we obtain

$F(8) = \frac{13}{18}$ = probability that $X \leq 8$

So the total probability is

```
F[8] - F[3]
```

$\frac{23}{36}$

```
Clear[f, F]
```

Continuous Random Variables and Distributions

We can extend the idea of discrete random variables to continuous random variables. Recall that for both discrete and continuous random variables, we can define the cumulative distribution function as

$$F(x) = P(X \leq x)$$

We can use the cumulative distribution function to define the **probability density function**, $f(x)$, as

$$F(x) = \int_{-\infty}^{x} f(v)\, dv \qquad \textbf{\textit{(Eq.7)}}$$

Differentiation gives an alternative relation as

$$f(x) = F'(x) \qquad \textbf{\textit{(Eq.8)}}$$

Recall that Eq.2 defined how to determine the probability that the random variable is within an interval $(a, b]$ as $P(a < x \leq b]) = F(b) - F(a)$. So combining this with Eq.7 yields an important relationship between the probability density function and the probability of a continuous random variable occurring in an interval $(a, b]$

$$P(a < x \leq b]) = F(b) - F(a)$$

$$= \int_{-\infty}^{b} f(v)\, dv - \int_{-\infty}^{a} f(v)\, dv \qquad \text{recall: } b > a$$

$$P(a < x \leq b]) = \int_{a}^{b} f(v)\, dv \qquad \textbf{\textit{(Eq.9)}}$$

This is the continuous analog of Eq.5 (discrete random variable occurring in a range = $P(a < X \leq b) = F(b) - F(a)$)

From Eq.7 and the fact that $P(S) = 1$, we obtain

$$\int_{-\infty}^{\infty} f(v)\, dv = 1 \qquad \textbf{\textit{(Eq.10)}}$$

This shows that the probability density function must have the property that when integrated from -∞ to ∞, it equals 1.

This shows that the probability of x occurring in the interval $(a, b]$ is given by the area under the probability density function, $f(x)$, between a and b.

Note that continuous random variable is somewhat simpler than discrete random variables. This is because the four probabilities corresponding to $X \in (a, b]$, $X \in (a, b)$, $X \in [a, b)$, and $X \in [a, b]$ are all the same because we see that adding a single point to the to the integral does not change the integral.

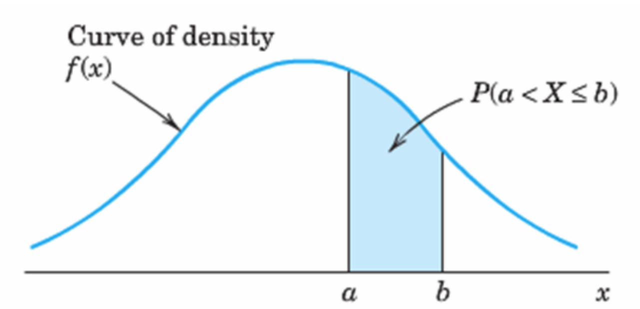


Fig. 515. Example illustrating formula (9)

Example 5: Continuous Distribution

Let X have a probability density function

$$f(x) = \begin{cases} f_1(x) = 0.75 \times (1 - x^2) & \text{if } x \in [-1, 1] \\ f_2(x) = 0 & \text{otherwise} \end{cases}$$

```
f1[x_] = 3/4 × (1 - x^2);
```

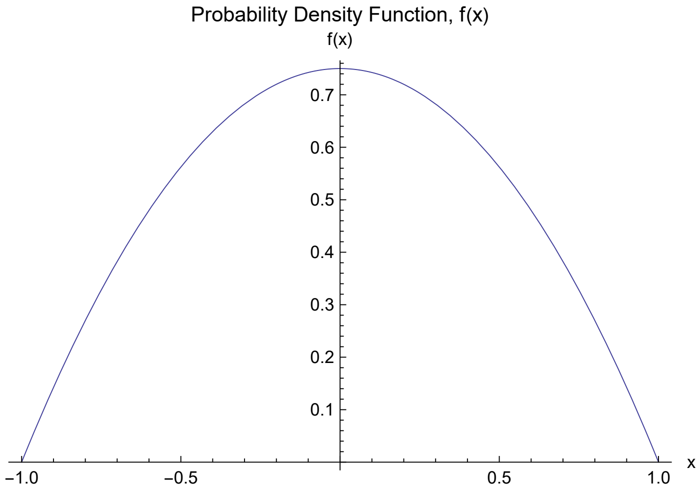
We can first verify that this is a valid probability density function by verifying that it satisfies Eq.10

$$\int_{-\infty}^{\infty} f(v)\, dv = 1$$

```
∫_{-1}^{1} f1[x] dx
```

```
1
```

```
Plot[f1[x], {x, -1, 1},
 AxesLabel → {"x", "f(x)"},
 PlotLabel → "Probability Density Function, f(x)"]
```



We can find the cumulative distribution function, $F(x)$, using Eq.7

$$F(x) = \int_{-\infty}^{x} f(v)\, dv$$

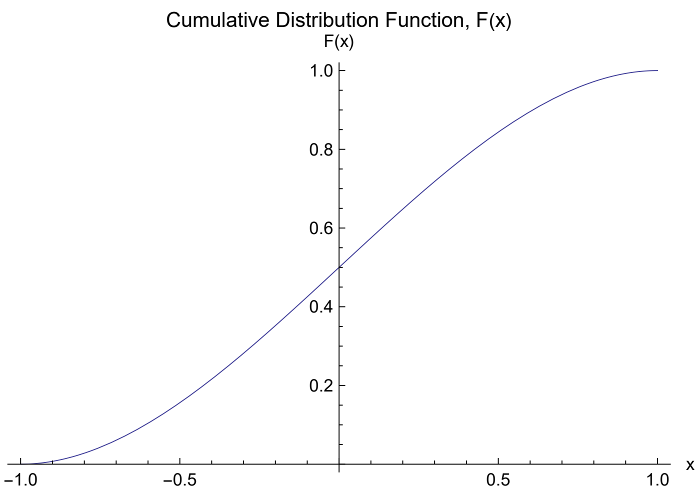
$$= \int_{-\infty}^{-1} f(v)\, dv + \int_{-1}^{x} f(v)\, dv \quad \text{let: } x \in [-1, 1]$$

$$= \int_{-\infty}^{-1} 0\, dv + \int_{-1}^{x} \frac{3}{4} \times \left(1 - v^2\right) dv$$

```
Print["F(x)"]
F1[x_] = ∫_{-1}^{x} f1[v] dv
Print[]

Plot[F1[x], {x, -1, 1},
 AxesLabel → {"x", "F(x)"},
 PlotLabel → "Cumulative Distribution Function, F(x)"]
```

F(x)

$\frac{1}{2}+\frac{3x}{4}-\frac{x^3}{4}$



So we obtain

$$F(x)=\begin{cases} F_1(x)=\frac{1}{2}+\frac{3}{4}x-\frac{1}{4}x^3 & x\in[-1,1] \\ F_2(x)=0 & \text{otherwise}\end{cases}$$

With this, can now compute quantities such as the probability that the random variable X takes values in between $-1/2$ and $+1/2$.

$$P\left(X\in\left[-\tfrac{1}{2},\tfrac{1}{2}\right]\right)=F\left(\tfrac{1}{2}\right)-F\left(-\tfrac{1}{2}\right)$$

```
F1[1/2] - F1[-1/2] // N
```

0.6875

Also, we can find x such that $P(X \leq x) = 0.95$ (95% of all occurrences fall in this range)

$$F(x)-F(-\infty)=0.95 \qquad \text{note: } F(-\infty)=0$$

$$\frac{1}{2} + \frac{3}{4} x - \frac{1}{4} x^3 = 0.95$$

```
Solve[F1[x] == 95/100, x] // N // Chop
```

```
{{x → 1.24814}, {x → 0.729299}, {x → -1.97744}}
```

We see that the only valid solution is

$$x \approx 0.73$$

```
Clear[F1, f1]
```

Mean and Variance of a Distribution

The mean, μ, and variance, σ^2, of a random variable X and of its distribution are the theoretical counterparts of the mean $\overline{x}$ and variance s^2 of a frequency distribution in Section 24.1 and serve a similar purpose. Indeed, the mean characterizes the central location and the variance measures the spread (the variability) of the distribution. The **mean**, μ, is defined by

$$\mu = \sum_j x_j f(x_j) \qquad \text{(discrete distribution)} \qquad \textbf{\textit{(Eq.1a)}}$$

$$\mu = \int_{-\infty}^{\infty} x f(x) \, dx \qquad \text{(continuous distribution)} \qquad \textbf{\textit{(Eq.1b)}}$$

The mean is also sometimes referred to as $E(X)$, the **expectation** of the random variable X because it gives the average value of X to be expected in many trials.

The **variance**, σ^2, is given by

$$\sigma^2 = \sum_j (x_j - \mu)^2 f(x_j) \qquad \text{(discrete distribution)} \qquad \textbf{\textit{(Eq.2a)}}$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx \qquad \text{(continuous distribution)} \qquad \textbf{\textit{(Eq.2b)}}$$

Quantities such as μ and σ^2 measure certain properties of a distribution and are called parameters. μ and σ^2 are two of the most important ones. From Eq.2, we see

$$\sigma^2 > 0 \qquad \textbf{\textit{(Eq.3)}}$$

Example 1: Mean and Variance

The random variable X = number of heads in a single toss of a fair coin has possible values $X = 0$ and $X = 1$ with probabilities $P(X = 0) = \frac{1}{2}$ and $P(X = 1) = \frac{1}{2}$. From Eq.1a, we thus obtain the mean

$$\mu = x_1 f(x_1) + x_2 f(x_2) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$$

From Eq.2a, we compute the variance as

$$\sigma^2 = \left(x_1 - \frac{1}{2}\right)^2 \cdot f(x_1) + \left(x_2 - \frac{1}{2}\right)^2 \cdot f(x_2) = \left(0 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} + \left(1 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{4}$$

Note that this takes physical meaning if we assume that we win 1 dollar for each head and 0 dollars for each tail. The expectation of X, μ, gives the expected winnings of this game (per game played) if played over a large number of trials.

We see that if the coin was unfair, we would not expect a mean of 1/2 and instead it would be “skewed” towards one side or the other. This is reflected in the non-uniform probabilities of each event.

Example 2: Uniform Distribution. Variance Measures Spread

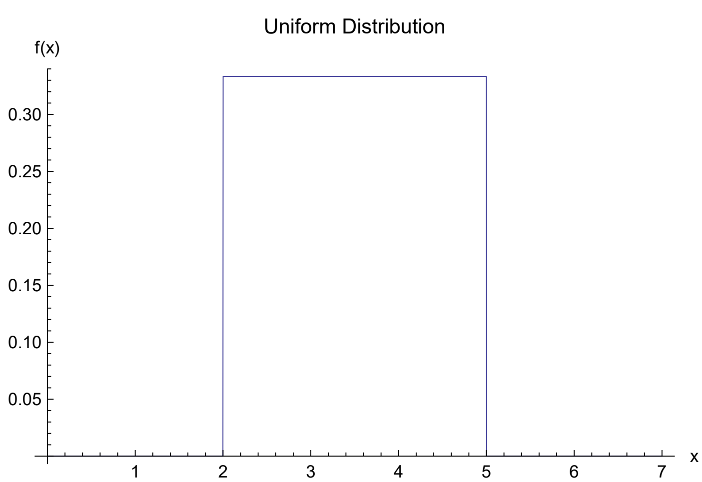
A distribution with the density

$$f(x) = \begin{cases} f_1(x) = \frac{1}{b-a} & a < x < b \\ f_2(x) = 0 & \text{otherwise} \end{cases}$$

Is called the uniform distribution on the interval $a < x < b$.

```
f[x_, a_, b_] := If[(a < x) && (x < b),
   1/(b - a),
   0]
```

```
Plot[f[x, 2, 5], {x, 0, 7},
 AxesLabel → {"x", "f(x)"},
 PlotLabel → "Uniform Distribution"]
```



From Eq.1b, we see the mean is given by

$$\mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

```
μ = ∫_a^b x 1/(b - a) ⅆx // Simplify
```

$$\frac{a+b}{2}$$

From Eq.2b, we calculate the variance as

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

```
variance = ∫_a^b (x - μ)^2 1/(b - a) ⅆx
```

$$\frac{1}{12}(a-b)^2$$

As we said earlier, the variance measures the spread and therefore, as expected, the variance grows with the spread of uniform distribution (i.e. as $|b-a|$ gets larger)

We note that different distributions can have the same mean but different variances as shown in Figure 516.

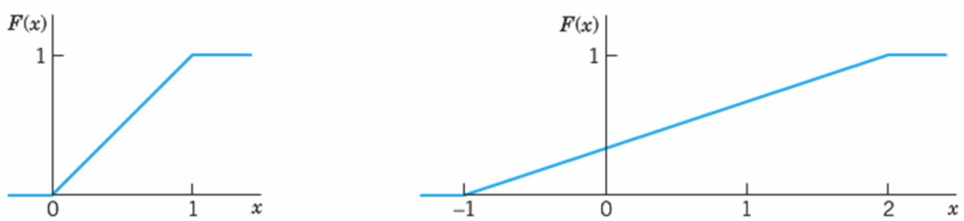


Fig. 516. Uniform distributions having the same mean (0.5) but different variances σ^2

Theorem 1: Mean of a Symmetric Distribution

If a distribution is symmetric with respect to $x = c$, that is $f(c-x) = f(c+x)$, then $\mu = c$.

Transformation of Mean and Variance

Given a random variable X with mean μ and variance σ^2, we want to calculate the mean and variance of $X^* = a_1 + a_2 X$ where a_1 and a_2 are given constants. This problem is important in statistics, where it often appears.

Theorem 2: Transformation of Mean and Variance

(a) If a random variable X has mean μ and variance σ^2, then the random variable

$$X^* = a_1 + a_2 X \qquad \textbf{(Eq.4)}$$

has mean μ^* and σ^{*2} where

$$\mu^* = a_1 + a_2 \mu \qquad \textbf{(Eq.5)}$$
$$\sigma^{*2} = a_2^2 \sigma^2$$

(b) In particular, the **standardized random variable** Z corresponding to X, given by

$$Z = \frac{X-\mu}{\sigma} \qquad \textbf{(Eq.6)}$$

has the mean of 0 and the variance of 1. We will make use of this fact later.

Expectations and Moments

Recall that Eq.1 defines the expectation of X, the value of X to be expected on the average, written $\mu = E(X)$. More generally, if $g(x)$ is non-constant and continuous for all x, then $g(X)$ is a random variable. Hence its mathematical expectation, or briefly, its expectation, $E(g(X))$ is the value of $g(X)$ to be expected on the average, defined similarly to Eq.1 by

$$E(g(X)) = \sum_j g(x_j)\, f(x_j) \quad \text{or} \quad E(g(X)) = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx \qquad \textbf{(Eq.7)}$$

We also define the ***k***

$$E\left(X^k\right) = \sum_j x_j^{\,k}\, f(x_j) \quad \text{or} \quad E\left(X^k\right) = \int_{-\infty}^{\infty} x^k\, f(x)\, dx \qquad \textbf{(Eq.8)}$$

and the $\boldsymbol{k}^{\textbf{th}}$ **central moment** of X (where $k = 1, 2, \ldots$)

$$E\left([X-\mu]^k\right) = \sum_j \left(x_j - \mu\right)^k f(x_j) \quad \text{or} \quad E\left([X-\mu]^k\right) = \int_{-\infty}^{\infty} (x-\mu)^k\, f(x)\, dx \qquad \textbf{(Eq.9)}$$

The mean of X is the first moment (ie Eq.8 with $k = 1$). The second central moment is the variance of X (ie Eq.9 with $k = 2$).