Probability and Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

March 20, 2019

Topics

Objectives

Objectives:

- Obtain the cumulative distribution function from a probability density function.
- Use a cumulative distribution function to find probabilities.
- Obtain the probability density function from a cumulative distribution function.

Cumulative Distribution Functions: Continuous (4.2)

- The ***cumulative distribution function*** (or ***cdf***) of a random variable X, denoted $\boldsymbol{F(x)}$, is defined for all x as

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

Cumulative Distribution Functions: Continuous (4.2)

- The ***cumulative distribution function*** (or ***cdf***) of a random variable X, denoted $\boldsymbol{F(x)}$, is defined for all x as

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

  We'll focus on the case in which X is **continuous.**

Cumulative Distribution Functions: Continuous (4.2)

- The ***cumulative distribution function*** (or ***cdf***) of a random variable X, denoted $\boldsymbol{F(x)}$, is defined for all x as

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

We'll focus on the case in which X is **continuous.**

(But we'll also briefly look at the **discrete** case on the last few slides).

Cumulative Distribution Functions: Continuous (4.2)

- The ***cumulative distribution function*** (or ***cdf***) of a random variable X, denoted $\boldsymbol{F(x)}$, is defined for all x as

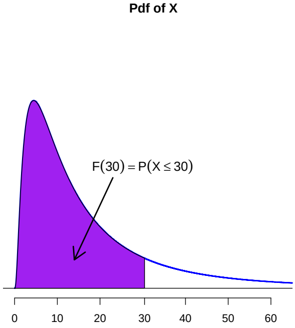
$$F(x) \;=\; P(X \leq x).$$

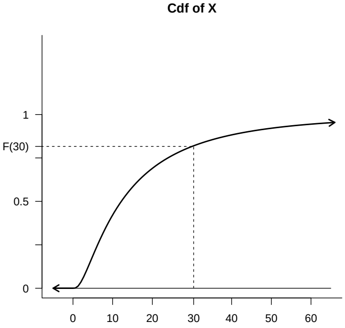
  We'll focus on the case in which X is **continuous.**

  (But we'll also briefly look at the **discrete** case on the last few slides).

- If X is **continuous** with **pdf** $f(x)$, then

$$F(x) \;=\; \int_{-\infty}^{x} f(y)\; dy.$$


Pdf of X
F(30) = P(X ≤ 30)
0
10
20
30
40
50
60


Cdf of X
1
F(30)
0.5
0
0
10
20
30
40
50
60

Example

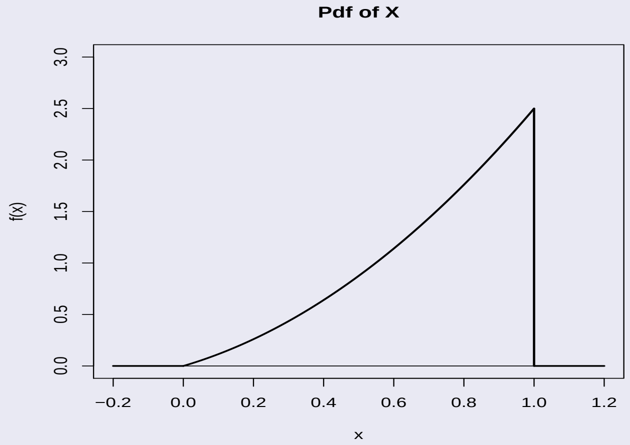
Suppose (again) that X is the gain in a certain investment, in thousands of dollars, and has **pdf**

$$f(x) = \begin{cases} \frac{1}{2}(3x^2 + 2x) & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Example

Suppose (again) that X is the gain in a certain investment, in thousands of dollars, and has **pdf**

$$f(x) = \begin{cases} \frac{1}{2}(3x^2 + 2x) & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$


Pdf of X
f(x)
x
0.0
0.5
1.0
1.5
2.0
2.5
3.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2

For $x < 0$, the **cdf** is

$$F(x) \;=\; 0$$

because there's no area under the **pdf** to the left of 0.

For $x < \mathbf{0}$, the **cdf** is

$$F(x) \;=\; 0$$

because there's no area under the **pdf** to the left of 0.

For $x > \mathbf{1}$,

$$F(x) \;=\; 1$$

because all of the area under the **pdf** is accumulated to the left 1.

For $0 \leq x \leq 1$, the **cdf** is

$$F(x) \quad = \quad \int_{-\infty}^{x} f(y)\,dy$$

For $\mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1}$, the **cdf** is

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(y)\,dy \\ &= \int_0^x \tfrac{1}{2}(3y^2 + 2y)\,dy \end{aligned}$$

For $\mathbf{0} \leq \boldsymbol{x} \leq \mathbf{1}$, the **cdf** is

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(y)\,dy \\ &= \int_{0}^{x} \frac{1}{2}(3y^2+2y)\,dy \\ &= \frac{1}{2}(y^3+y^2)\Big|_0^x \end{aligned}$$

For $0 \le x \le 1$, the **cdf** is

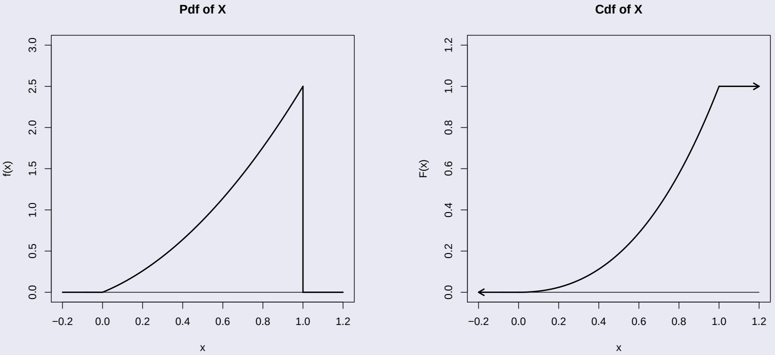
$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(y)\,dy \\
&= \int_{0}^{x} \tfrac{1}{2}(3y^2 + 2y)\,dy \\
&= \tfrac{1}{2}(y^3 + y^2)\Big|_0^x \\
&= \tfrac{1}{2}(x^3 + x^2).
\end{aligned}
$$

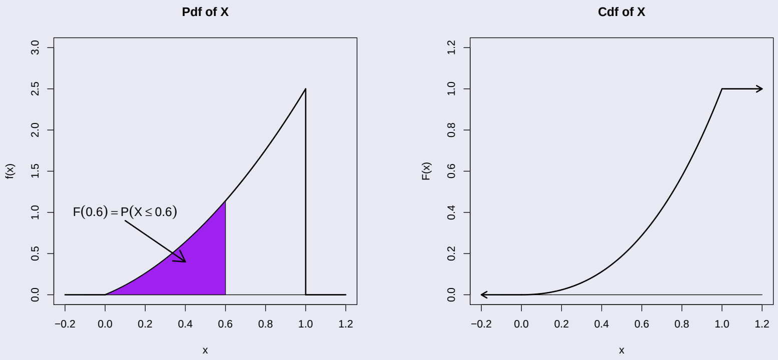
For $\mathbf{0 \leq x \leq 1}$, the **cdf** is

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(y)\, dy \\ &= \int_{0}^{x} \tfrac{1}{2}(3y^2 + 2y)\, dy \\ &= \tfrac{1}{2}(y^3 + y^2)\Big|_0^x \\ &= \tfrac{1}{2}(x^3 + x^2). \end{aligned}$$

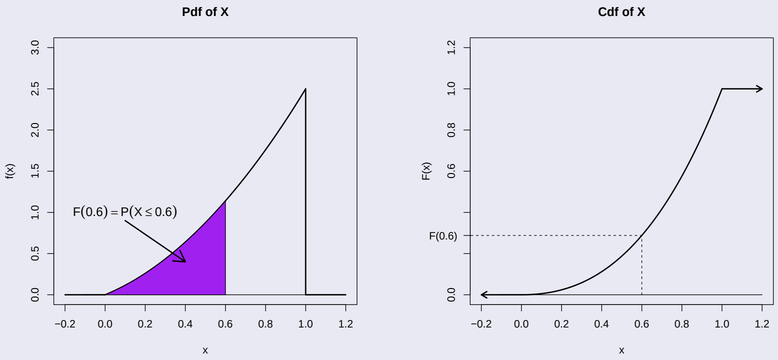
Thus the **cdf** of X is

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2}(x^3 + x^2) & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$$


Pdf of X
f(x)
x
0.0
0.5
1.0
1.5
2.0
2.5
3.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Cdf of X
F(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
1.2
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2



Pdf of X
f(x)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
F(0.6) = P(X ≤ 0.6)
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
x
Cdf of X
F(x)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
x



Pdf of X
f(x)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
F(0.6) = P(X ≤ 0.6)
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
x
Cdf of X
F(x)
0.0
F(0.6)
0.6
0.8
1.0
1.2
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2
x

The probability $P(X \le 0.6)$ that the investment gain will be **less than $600** is

$$P(X \le 0.6) \;=\; F(0.6)$$

The probability $P(X \leq 0.6)$ that the investment gain will be **less than $600** is

$$\begin{aligned} P(X \leq 0.6) &= F(0.6) \\ &= \tfrac{1}{2}(0.6^3 + 0.6^2) \end{aligned}$$

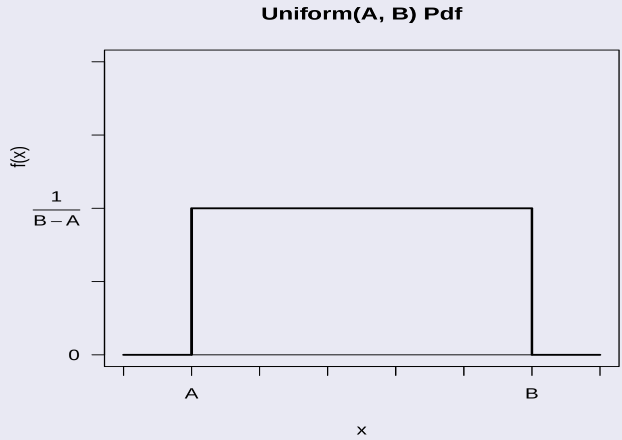
The probability $P(X \le 0.6)$ that the investment gain will be **less than \$600** is

$$\begin{aligned} P(X \le 0.6) &= F(0.6) \\ &= \frac{1}{2}(0.6^3 + 0.6^2) \\ &= \mathbf{0.288}. \end{aligned}$$

Example

Suppose X, the thickness of a certain metal sheet, follows a **uniform distribution** over the range A to B.

The **pdf** is below.


Uniform(A, B) Pdf
f(x)
1
B − A
0
A
B
x

For $x < A$, the **cdf** is

$$F(x) \;=\; 0$$

because there's no area under the **pdf** to the left of A.

For $x < A$, the **cdf** is

$$F(x) \;=\; 0$$

because there's no area under the **pdf** to the left of A.

For $x > B$,

$$F(x) \;=\; 1$$

because all of the area under the **pdf** is accumulated to the left B.

For $A \leq x \leq B$, the **cdf** is

$$F(x) \quad = \quad \int_{-\infty}^{x} f(y)\, dy$$

For $A \leq x \leq B$, the **cdf** is

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(y)\,dy \\
&= \int_{A}^{x} \frac{1}{B-A}\,dy
\end{aligned}
$$

For $A \leq x \leq B$, the **cdf** is

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(y)\,dy \\ &= \int_{A}^{x} \frac{1}{B-A}\,dy \\ &= \left.\frac{y}{B-A}\right|_{A}^{x} \end{aligned}$$

For $A \le x \le B$, the **cdf** is

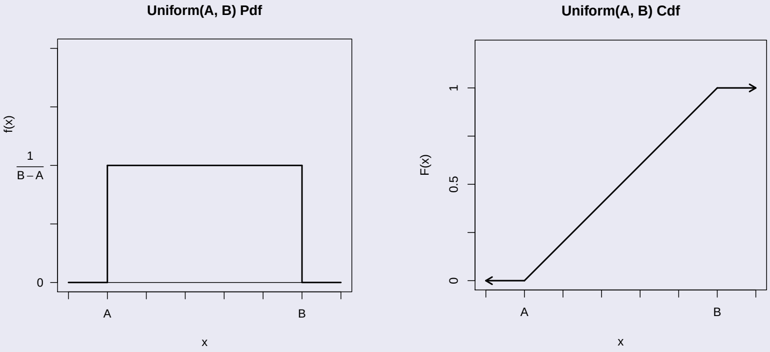
$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(y)\,dy \\ &= \int_{A}^{x} \frac{1}{B-A}\,dy \\ &= \left.\frac{y}{B-A}\right|_{A}^{x} \\ &= \frac{x-A}{B-A}. \end{aligned}$$

For $A \leq x \leq B$, the **cdf** is

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(y)\,dy \\ &= \int_{A}^{x} \tfrac{1}{B-A}\,dy \\ &= \tfrac{y}{B-A}\Big|_{A}^{x} \\ &= \tfrac{x-A}{B-A}. \end{aligned}$$

Thus the **cdf** of a **uniform** random variable is

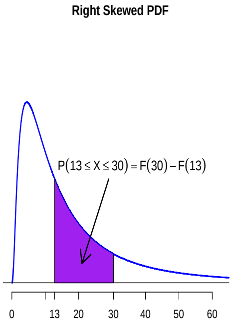
$$F(x) = \begin{cases} 0 & \text{for } x < A \\ \frac{x-A}{B-A} & \text{for } A \leq x \leq B \\ 1 & \text{for } x > B \end{cases}$$

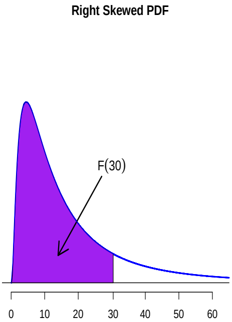

Uniform(A, B) Pdf
f(x)
1
B – A
0
A
B
x
Uniform(A, B) Cdf
F(x)
0
0.5
1
A
B
x

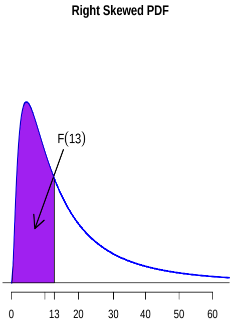
Proposition

Suppose X is a random variable with **cdf** $F(x)$. Then for any numbers a and b, with $a \leq b$,

1. $P(X > a) \; = \; 1 - F(a)$
2. $P(a < X \leq b) \; = \; F(b) - F(a)$


Right Skewed PDF
P(13 ≤ X ≤ 30) = F(30) − F(13)
0
13
20
30
40
50
60


Right Skewed PDF
F(30)
0
10
20
30
40
50
60


Right Skewed PDF
F(13)
0
13
20
30
40
50
60

Example

Suppose (again) that X is the gain in a certain investment, in thousands of dollars.

We found that the **cdf** of X is

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2}(x^3 + x^2) & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x > 1 \end{cases}$$

Example

Suppose (again) that X is the gain in a certain investment, in thousands of dollars.

We found that the **cdf** of X is

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{2}(x^3 + x^2) & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x > 1 \end{cases}$$

The probability $P(X > 0.6)$ that it will be **more than $600** is

$$P(X > 0.6) = 1 - F(0.6)$$

The probability $P(X > 0.6)$ that it will be **more than \$600** is

$$\begin{aligned} P(X > 0.6) &= 1 - F(0.6) \\ &= 1 - \frac{1}{2}(0.6^3 + 0.6^2) \end{aligned}$$

The probability $P(X > 0.6)$ that it will be **more than \$600** is

$$\begin{aligned} P(X > 0.6) &= 1 - F(0.6) \\ &= 1 - \tfrac{1}{2}(0.6^3 + 0.6^2) \\ &= 1 - 0.288 \end{aligned}$$

The probability $P(X > 0.6)$ that it will be **more than \$600** is

$$\begin{aligned} P(X > 0.6) &= 1 - F(0.6) \\ &= 1 - \tfrac{1}{2}(0.6^3 + 0.6^2) \\ &= 1 - 0.288 \\ &= \mathbf{0.712}. \end{aligned}$$

The probability $P(0.4 \leq X \leq 0.8)$ that the investment gain will be **between $400 and $800** is

$$P(0.4 \leq X \leq 0.8) = F(0.8) - F(0.4)$$

The probability $P(0.4 \leq X \leq 0.8)$ that the investment gain will be **between \$400 and \$800** is

$$\begin{aligned} P(0.4 \leq X \leq 0.8) &= F(0.8) - F(0.4) \\ &= \tfrac{1}{2}(0.8^3 + 0.8^2) - \tfrac{1}{2}(0.4^3 + 0.4^2) \end{aligned}$$

The probability $P(0.4 \le X \le 0.8)$ that the investment gain will be **between \$400 and \$800** is

$$\begin{aligned} P(0.4 \le X \le 0.8) &= F(0.8) - F(0.4) \\ &= \tfrac{1}{2}(0.8^3 + 0.8^2) - \tfrac{1}{2}(0.4^3 + 0.4^2) \\ &= \mathbf{0.464}. \end{aligned}$$

- We obtain the **pdf** $f(x)$ by taking the **derivative** of the **cdf** $F(x)$.

Proposition

Suppose X is a continuous random variable with **cdf** $F(x)$ and **pdf** $f(x)$. Then at every value x for which the derivative $F'(x)$ exists,

$$F'(x) = f(x).$$

- We obtain the **pdf** $f(x)$ by taking the **derivative** of the **cdf** $F(x)$.

Proposition

Suppose X is a continuous random variable with **cdf** $F(x)$ and **pdf** $f(x)$. Then at every value x for which the derivative $F'(x)$ exists,

$$F'(x) \;=\; f(x).$$

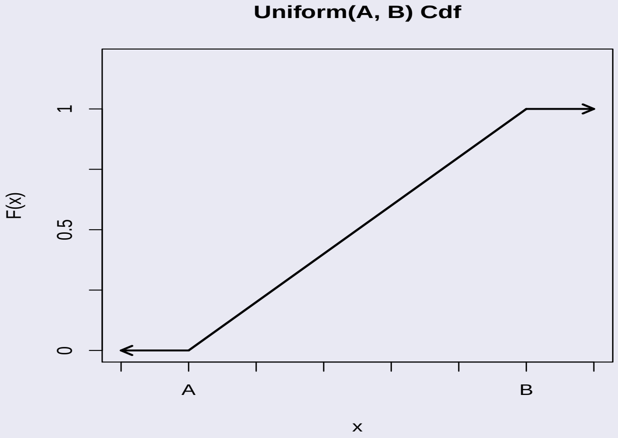
This is a consequence of the Fundamental Theorem of Calculus, which says

$$\frac{d}{dx}\int_{\infty}^{x} f(y)\;dy \;=\; f(x).$$

Example

Recall that the **cdf** of a **uniform** random variable over the range A to B is

$$F(x) = \begin{cases} 0 & \text{for } x < A \\ \frac{x-A}{B-A} & \text{for } A \le x \le B \\ 1 & \text{for } x > B \end{cases}$$


Uniform(A, B) Cdf
F(x)
0
0.5
1
A
B
x

The **pdf** is obtained from the **cdf** by taking its derivative:

$$f(x) \quad = \quad F'(x) \quad = \quad \begin{cases} \frac{d}{dx}(0) & \text{for } x < A \\ \frac{d}{dx}\left(\frac{x-A}{B-A}\right) & \text{for } A < x < B \\ \frac{d}{dx}(1) & \text{for } x > B \end{cases}$$

The **pdf** is obtained from the **cdf** by taking its derivative:

$$f(x) = F'(x) = \begin{cases} \frac{d}{dx}(0) & \text{for } x < A \\ \frac{d}{dx}\left(\frac{x-A}{B-A}\right) & \text{for } A < x < B \\ \frac{d}{dx}(1) & \text{for } x > B \end{cases}$$

$$= \begin{cases} \frac{1}{B-A} & \text{for } A < x < B \\ 0 & \text{otherwise} \end{cases}$$

Cumulative Distribution Functions: Discrete (3.2)

- Recall that the ***cumulative distribution function*** (or ***cdf***) of a random variable X, denoted $\boldsymbol{F(x)}$, is defined for all x as

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

- If X is **discrete** with **pmf** $p(x)$, then

$$F(x) \;=\; \sum_{y \leq x} p(y).$$

Cumulative Distribution Functions: Discrete (3.2)

- Recall that the ***cumulative distribution function*** (or ***cdf***) of a random variable X, denoted $\boldsymbol{F(x)}$, is defined for all x as

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

- If X is **discrete** with **pmf** $p(x)$, then

$$F(x) \;=\; \sum_{y \leq x} p(y).$$

- In the **discrete** case, $F(x)$ is a **step function**, with a step of size $p(x)$ at each of the possible values x of the random variable X.

Example

Suppose the **pmf** of a **discrete** random variable X is

x	1	1.5	2	4
$p(x)$	0.1	0.2	0.4	0.3

Example

Suppose the **pmf** of a **discrete** random variable X is

x	1	1.5	2	4
$p(x)$	0.1	0.2	0.4	0.3

Then the **cdf** $F(x)$ is a **step function**, with a step of size $p(x)$ at each of the values $x =$ 1, 1.5, 2, and 4.

Example

Suppose the **pmf** of a **discrete** random variable X is

x	1	1.5	2	4
$p(x)$	0.1	0.2	0.4	0.3

Then the **cdf** $F(x)$ is a **step function**, with a step of size $p(x)$ at each of the values $x = 1$, 1.5, 2, and 4. We can write this as

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ & \\ & \\ & \end{cases}$$

Example

Suppose the **pmf** of a **discrete** random variable X is

x	1	1.5	2	4
$p(x)$	0.1	0.2	0.4	0.3

Then the **cdf** $F(x)$ is a **step function**, with a step of size $p(x)$ at each of the values $x = 1$, 1.5, 2, and 4. We can write this as

$$F(x) = \begin{cases} 0 & \text{for } \ x < 1 \\ 0.1 & \text{for } \ 1 \leq x < 1.5 \\ & \\ & \end{cases}$$

Example

Suppose the **pmf** of a **discrete** random variable X is

x	1	1.5	2	4
$p(x)$	0.1	0.2	0.4	0.3

Then the **cdf** $F(x)$ is a **step function**, with a step of size $p(x)$ at each of the values $x = 1$, 1.5, 2, and 4. We can write this as

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ 0.1 & \text{for } 1 \leq x < 1.5 \\ 0.3 & \text{for } 1.5 \leq x < 2 \\ & \\ & \end{cases}$$

Example

Suppose the **pmf** of a **discrete** random variable X is

x	1	1.5	2	4
$p(x)$	0.1	0.2	0.4	0.3

Then the **cdf** $F(x)$ is a **step function**, with a step of size $p(x)$ at each of the values $x =$ 1, 1.5, 2, and 4. We can write this as

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ 0.1 & \text{for } 1 \leq x < 1.5 \\ 0.3 & \text{for } 1.5 \leq x < 2 \\ 0.7 & \text{for } 2 \leq x < 4 \\ & \end{cases}$$

Example

Suppose the **pmf** of a **discrete** random variable X is

x	1	1.5	2	4
$p(x)$	0.1	0.2	0.4	0.3

Then the **cdf** $F(x)$ is a **step function**, with a step of size $p(x)$ at each of the values $x =$ 1, 1.5, 2, and 4. We can write this as

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ 0.1 & \text{for } 1 \leq x < 1.5 \\ 0.3 & \text{for } 1.5 \leq x < 2 \\ 0.7 & \text{for } 2 \leq x < 4 \\ 1 & \text{for } x \geq 4 \end{cases}$$

Example

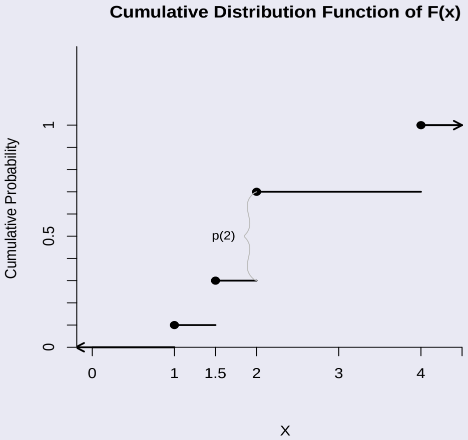
Suppose the **pmf** of a **discrete** random variable X is

x	1	1.5	2	4
$p(x)$	0.1	0.2	0.4	0.3

Then the **cdf** $F(x)$ is a **step function**, with a step of size $p(x)$ at each of the values $x = 1$, 1.5, 2, and 4. We can write this as

$$F(x) = \begin{cases} 0 & \text{for} \quad x < 1 \\ 0.1 & \text{for} \quad 1 \leq x < 1.5 \\ 0.3 & \text{for} \quad 1.5 \leq x < 2 \\ 0.7 & \text{for} \quad 2 \leq x < 4 \\ 1 & \text{for} \quad x \geq 4 \end{cases}$$

A graph of $F(x)$ is shown on the next slide.


Cumulative Distribution Function of F(x)
Cumulative Probability
0
0.5
1
p(2)
0
1
1.5
2
3
4
X