THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics
Lecture 12: Independence of discrete random variables and Continuous random variables

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

Conditional PMF: summary

- Conditional PMF is exactly like conditional probabilities. Your new sample space is one where the conditioning event has taken place.
- By definition, we have $P(X = x|Y = y) = \dfrac{P(X = x, Y = y)}{P(Y = y)}$, where $P(Y = y) > 0$.
- $\displaystyle\sum_x P(X = x|Y = y) = 1$.
- The multiplication rule:
  - $P(X = x, Y = y) = P(X = x|Y = y)P(Y = y)$.
  - For 3 random variables, $P(X = x, Y = y, Z = z)$ equals $P(X = x|Y = y, Z = z)P(Y = y|Z = z)P(Z = z)$
- Total probability rule:
  $P(X = x) = \displaystyle\sum_y P(X = x, Y = y) = \sum_y P(X = x|Y = y)P(Y = y)$.

Conditional Expectation

- Recall the **expectation** of X. $E[X] = \sum_x xP(X = x)$.
- The **conditional expectation** of random variable X given event A with $P(A) > 0$ is defined as: $E[X|A] = \sum_x xP(X = x|A)$.
- For a function $g(X)$, $E[g(X)|A] = \sum_x g(x)P(X = x|A)$.
- The conditional expectation of X given $Y = y$ is given by $E[X|Y = y] = \sum_x xP(X = x|Y = y)$.
- How are $E[X|Y = y]$ related to $E[X]$?

  —**Total expectation theorem!**

Total expectation theorem

- We have $$E[X|Y=y]P(Y=y) = \sum_x xP(X=x|Y=y)P(Y=y)$$

Total expectation theorem

- We have $$E[X|Y=y]P(Y=y) = \sum_x xP(X=x|Y=y)P(Y=y)$$
$$\sum_y E[X|Y=y]P(Y=y) = \sum_y \sum_x xP(X=x|Y=y)P(Y=y)$$
$$\sum_y E[X|Y=y]P(Y=y) = \sum_y \sum_x x \underbrace{P(X=x|Y=y)P(Y=y)}_{P(X=x,Y=y)}$$

Total expectation theorem

- We have

$$E[X|Y=y]P(Y=y) = \sum_x xP(X=x|Y=y)P(Y=y)$$

$$\sum_y E[X|Y=y]P(Y=y) = \sum_y \sum_x xP(X=x|Y=y)P(Y=y)$$

$$\begin{aligned}\sum_y E[X|Y=y]P(Y=y) &= \sum_y \sum_x x \underbrace{P(X=x|Y=y)P(Y=y)}_{P(X=x,Y=y)} \\ &= \underbrace{\sum_x x \sum_y P(X=x|Y=y)P(Y=y)}_{P(X=x)} \\ &= \sum_x xP(X=x)\end{aligned}$$

Total expectation theorem

- We have $$E[X|Y=y]P(Y=y) = \sum_x xP(X=x|Y=y)P(Y=y)$$
$$\sum_y E[X|Y=y]P(Y=y) = \sum_y \sum_x xP(X=x|Y=y)P(Y=y)$$
$$\sum_y E[X|Y=y]P(Y=y) = \sum_y \sum_x x \underbrace{P(X=x|Y=y)P(Y=y)}_{P(X=x,Y=y)}$$
$$= \sum_x x \underbrace{\sum_y P(X=x|Y=y)P(Y=y)}_{P(X=x)}$$
$$= \sum_x xP(X=x)$$
$$= E[X]$$

- So $E[X]$ is just a weighted average of $E[X|Y=y]$, the weights being the probability of $Y=y$.

Roadmap - discrete r.v. - independence

- Independence of two events
- Independence of a random variable and an event
- Independence of two random variables–pairwise independence
  - Implications – expectation of the product of two independent r.v's
  - Implications – variance of the sum of two independent r.v.'s
- Generalization to multiple random variables
  - Implication – variance of a binomial
- Conditional independence

Independence of random variables

- Lets first consider two events $\{X = x\}$ and A. We know that these two events are independent if $P(\{X = x\}, A) = P(\{X = x\})P(A)$
- In other words if $P(A) > 0$, then $P(X = x|A) = P(X = x)$, i.e. knowing the occurrence of A does not change our belief about $\{X = x\}$.
- We will call the random variable X and event A to be independent if

$$P(X = x, A) = P(X = x)P(A) \qquad \textbf{For all } x$$

- Two random variables are said to be independent if

$$P(X = x, Y = y) = P(X = x)P(Y = y) \qquad \textbf{For all } x \textbf{ and } y$$

- To put it a bit differently,

$$P(X = x|Y = y) = P(X = x) \qquad \textbf{For all } x \textbf{ and } y \textbf{ such that } P(Y = y) > 0$$

A super important implication

We saw that $E[X+Y] = E[X] + E[Y]$ no matter whether X and Y are independent or not.

- If X and Y are independent, $E[XY] = E[X]E[Y]$
- $$E[XY] = \sum_{x,y} xyP(X=x, Y=y) = \sum_{x,y} xyP(X=x)P(Y=y)$$ $$= \left(\sum_{x} xP(X=x)\right)\left(\sum_{y} yP(Y=y)\right) = E[X]E[Y]$$
- In fact, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

Variance of sum of independent random variables

Let X and Y be two independent random variables. What is $\text{var}(X+Y)$?

- Remember! $\text{var}(X+Y) = E[(X+Y)^2] - (E[X+Y])^2$
- $$E[(X+Y)^2] = E[X^2 + Y^2 + 2XY] = E[X^2] + E[Y^2] + 2E[XY]$$
  $$= E[X^2] + E[Y^2] + 2E[X]E[Y]$$

Variance of sum of independent random variables

Let X and Y be two independent random variables. What is $\text{var}(X+Y)$?

- Remember! $\text{var}(X+Y) = E[(X+Y)^2] - (E[X+Y])^2$
- $$\begin{aligned} E[(X+Y)^2] &= E[X^2 + Y^2 + 2XY] = E[X^2] + E[Y^2] + 2E[XY] \\ &= E[X^2] + E[Y^2] + 2E[X]E[Y] \\ E[X+Y]^2 &= (E[X] + E[Y])^2 = E[X]^2 + E[Y]^2 + 2E[X]E[Y] \\ \text{var}(X+Y) &= E[(X+Y)^2] - (E[X+Y])^2 \end{aligned}$$

Variance of sum of independent random variables

Let X and Y be two independent random variables. What is $\text{var}(X+Y)$?

- Remember! $\text{var}(X+Y) = E[(X+Y)^2] - (E[X+Y])^2$
- $$\begin{aligned} E[(X+Y)^2] &= E[X^2 + Y^2 + 2XY] = E[X^2] + E[Y^2] + 2E[XY] \\ &= E[X^2] + E[Y^2] + 2E[X]E[Y] \\ E[X+Y]^2 &= (E[X] + E[Y])^2 = E[X]^2 + E[Y]^2 + 2E[X]E[Y] \\ \text{var}(X+Y) &= E[(X+Y)^2] - (E[X+Y])^2 \\ &= \underbrace{E[X^2] - E[X]^2}_{\text{var}(X)} + \underbrace{E[Y^2] - E[Y]^2}_{\text{var}(Y)} = \text{var}(X) + \text{var}(Y) \end{aligned}$$

Variance of sum of independent random variables

Let X and Y be two independent random variables. What is $\text{var}(X+Y)$?

- Remember! $\text{var}(X+Y) = E[(X+Y)^2] - (E[X+Y])^2$
- $$\begin{aligned} E[(X+Y)^2] &= E[X^2 + Y^2 + 2XY] = E[X^2] + E[Y^2] + 2E[XY] \\ &= E[X^2] + E[Y^2] + 2E[X]E[Y] \\ E[X+Y]^2 &= (E[X] + E[Y])^2 = E[X]^2 + E[Y]^2 + 2E[X]E[Y] \\ \text{var}(X+Y) &= E[(X+Y)^2] - (E[X+Y])^2 \\ &= \underbrace{E[X^2] - E[X]^2}_{\text{var}(X)} + \underbrace{E[Y^2] - E[Y]^2}_{\text{var}(Y)} = \text{var}(X) + \text{var}(Y) \end{aligned}$$
- Variance of sum of independent random variables equals the sum of the variances!

Independence of several random variables

- Three random variables X, Y and Z are said to be independent if

$$P(X = x, Y = y, Z = z) = P(X = x)P(Y = y)P(Z = z) \qquad \text{For all } x, y, z$$

- If X, Y, Z are independent, then so are $f(X)$, $g(Y)$ and $h(Z)$.
- Also, any random variable $f(X, Y)$ and $g(Z)$ are independent.
- Are $f(X, Y)$ and $g(Y, Z)$ independent?

Independence of several random variables

- Three random variables X, Y and Z are said to be independent if

$$P(X = x, Y = y, Z = z) = P(X = x)P(Y = y)P(Z = z) \qquad \text{For all } x, y, z$$

- If X, Y, Z are independent, then so are $f(X)$, $g(Y)$ and $h(Z)$.
- Also, any random variable $f(X, Y)$ and $g(Z)$ are independent.
- Are $f(X, Y)$ and $g(Y, Z)$ independent?
  - Not necessarily, both have Y in common.

Independence of several random variables

- Three random variables X, Y and Z are said to be independent if

$$P(X = x, Y = y, Z = z) = P(X = x)P(Y = y)P(Z = z) \qquad \text{For all } x, y, z$$

- If X, Y, Z are independent, then so are $f(X)$, $g(Y)$ and $h(Z)$.
- Also, any random variable $f(X, Y)$ and $g(Z)$ are independent.
- Are $f(X, Y)$ and $g(Y, Z)$ independent?
  - Not necessarily, both have Y in common.

- For n independent random variables, X_1, $X_2, \ldots, X_n$, we also have:

$$\text{var}(X_1 + X_2 + X_3 + \cdots + X_n) = \text{var}(X_1) + \text{var}(X_2) + \cdots + \text{var}(X_n)$$

Variance of a Binomial

Consider n independent Bernoulli variables $X_1, X_2, \ldots, X_n$, each with probability p of having value "1". The sum $Y = \sum_i X_i$ is a *Binomial*(n, p) random variable.

- We saw last time that $E[Y] = \sum_i E[X_i] = np$. What about the variance?
- Recall that $\text{var}(X_i) = p(1-p)$ for $i \in \{1, 2, \ldots, n\}$.
- $\text{var}(Y) = \text{var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \text{var}(X_i) = np(1-p)$.

Conditional Independence

- Very similar to conditional independence of events!
- X and Y are conditionally independent, given a positive probability event A if

$$P(X = x, Y = y|A) = P(X = x|A)P(Y = y|A) \qquad \text{For all } x \text{ and } y$$

- Same as saying $P(X = x|Y = y, A) = P(X = x|A)$, i.e.
- Once you know that A has occurred, knowing $\{Y = y\}$ has occurred does not give you any more information!
- Like we learned before, conditional independence does not imply unconditional independence.

Example-conditionally independent but not marginally

- I separately phone two students (Alice and Bob) and tell them the midterm grade.
- To each, I report the same grade, $G \in \{A+, A..., C\}$.
- The signal is bad and, Alice and Bob each independently make an educated guess of what I said.
- Let the grades guessed by Alice and Bob be X and Y.
- Are X and Y marginally independent?

Example-conditionally independent but not marginally

- I separately phone two students (Alice and Bob) and tell them the midterm grade.
- To each, I report the same grade, $G \in \{A+, A..., C\}$.
- The signal is bad and, Alice and Bob each independently make an educated guess of what I said.
- Let the grades guessed by Alice and Bob be X and Y.
- Are X and Y marginally independent?
  - NO. you would think, $P(X = A|Y = A) > P(X = A)$.

Example-conditionally independent but not marginally

- I separately phone two students (Alice and Bob) and tell them the midterm grade.
- To each, I report the same grade, $G \in \{A+, A..., C\}$.
- The signal is bad and, Alice and Bob each independently make an educated guess of what I said.
- Let the grades guessed by Alice and Bob be X and Y.
- Are X and Y marginally independent?
  - NO. you would think, $P(X = A|Y = A) > P(X = A)$.
- What if I tell you that $G = A-$?

Example-conditionally independent but not marginally

- I separately phone two students (Alice and Bob) and tell them the midterm grade.
- To each, I report the same grade, $G \in \{A+, A..., C\}$.
- The signal is bad and, Alice and Bob each independently make an educated guess of what I said.
- Let the grades guessed by Alice and Bob be X and Y.
- Are X and Y marginally independent?
  - NO. you would think, $P(X = A|Y = A) > P(X = A)$.
- What if I tell you that $G = A-$?
  - Are X and Y conditionally independent given $\{G = A\}$.

Example-conditionally independent but not marginally

- I separately phone two students (Alice and Bob) and tell them the midterm grade.
- To each, I report the same grade, $G \in \{A+, A..., C\}$.
- The signal is bad and, Alice and Bob each independently make an educated guess of what I said.
- Let the grades guessed by Alice and Bob be X and Y.
- Are X and Y marginally independent?
  - NO. you would think, $P(X = A|Y = A) > P(X = A)$.
- What if I tell you that $G = A-$?
  - Are X and Y conditionally independent given $\{G = A\}$.
  - YES! Because if we know the grade I actually said, the two variables are no longer dependent.

Example-marginally independent but not conditionally

- I toss two dice independently and X and Y are the readings on them.
- Are X and Y independent?
- Now I tell you that $X + Y = 12$. Are they still independent?

Cumulative distribution function

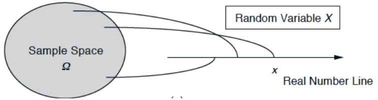
- I toss two dice independently and X and Y are the readings on them.
- Are X and Y independent?
- Now I tell you that $X + Y = 12$. Are they still independent?

Roadmap

- Discrete vs continuous random variables
- Probability mass function vs Probability density function
  - Properties of the pdf
- Cumulative distribution function
  - Properties of the cdf
- Expectation, variance and properties
- The normal distribution

Review: Random variables

A random variable is mapping from the sample space Ω into the real numbers.



So far, we've looked at **discrete random variables**, that can take a finite, or at most countably infinite, number of values, e.g.

- Bernoulli random variable – can take on values in $\{0, 1\}$.
- Binomial(n, p) random variable – can take on values in $\{0, 1, \ldots, n\}$.
- Geometric(p) random variable – can take on any positive integer.

Continuous random variable

A **continuous** random variable is a random variable that:

- Can take on an uncountably infinite range of values.
- For any specific value $X = x$, $P(X = x) = 0$.

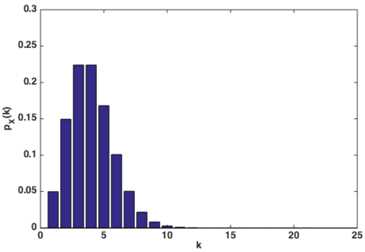
Examples might include:

- The time at which a bus arrives.
- The volume of water passing through a pipe over a given time period.
- The height of a randomly selected individual.

Probability mass function

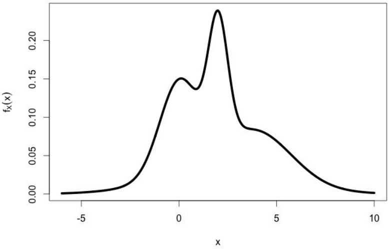
Remember for a discrete random variable X , we could describe the probability of X a particular value using the **probability mass function**.

- e.g. if $X \sim \text{Poisson}(\lambda)$, then the PMF of X is $p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$
- We can read off the probability of a specific value of k from the PMF.
- We can use the PMF to calculate the **expected value** and the **variance** of X.
- We can plot the PMF using a histogram



Probability density function

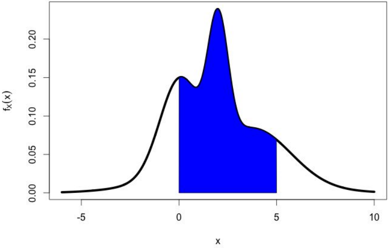
- For a continuous random variable, we cannot construct a PMF – each specific value has zero probability.
- Instead, we use a continuous, non-negative function $f_X(x)$ called the **probability density function**, or PDF, of X.



Probability density function

- For a continuous random variable, we cannot construct a PMF – each specific value has zero probability.
- Instead, we use a continuous, non-negative function $f_X(x)$ called the **probability density function**, or PDF, of X.
- The probability of X lying between two values x_1 and x_2 is simply the area under the PDF, i.e.

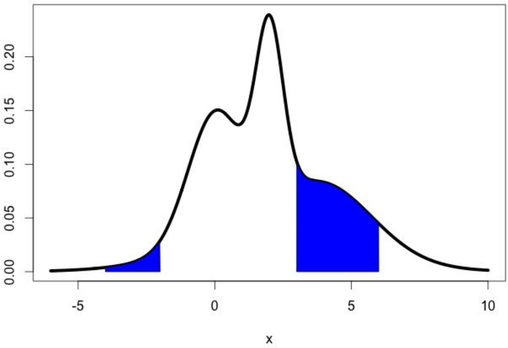
$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$



Probability density function

- More generally, for any subset B of the real line,

$$P(X \in B) = \int_B f_X(x)dx$$



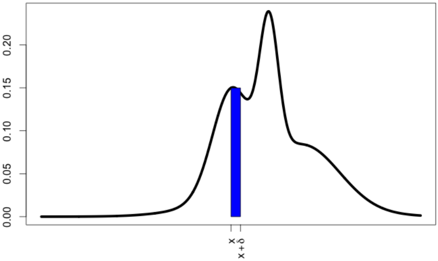
- Here, $B = (-4, -2) \cup (3, 6)$.

Properties of the pdf

- *Note that $f_X(a)$ is not $P(X = a)$!!*
- For any single value a, $P(X = a) = \int_a^a f_X(x)dx = 0$.
- This means that, for example, $P(X \leq a) = P(X < a) + P(X = a) = P(X < a)$.
- Recall that a valid probability law must satisfy $P(\Omega) = 1$ and $P(A) > 0$.
- f_X is non-negative, so $P(x \in B) = \int_{x \in B} f_X(x)dx \geq 0$ for all B
- To have normalization, we require,
  - $\int_{-\infty}^{\infty} f_X(x) = P(-\infty < X < \infty) = 1 \leftarrow$ total area under curve is 1.
- Note that $f_X(x)$ can be greater than 1 – even infinite! – for certain values of x, provided the integral over all x is 1.

Intuition

- We can think of the probability of our random variable lying in some small interval of length δ, $[x, x+\delta]$
- $P(X \in [x, x+\delta]) = \int_x^{x+\delta} f_X(t)dt \approx f_X(x) \cdot \delta$



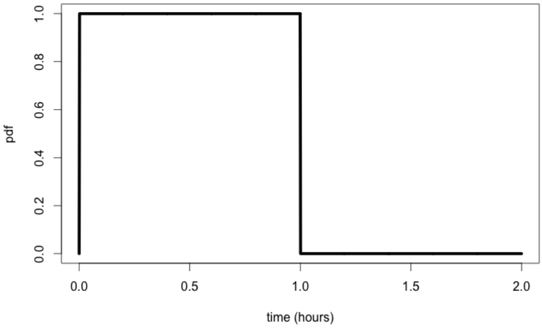
- Note however that $f_X(x)$ is **not** the probability at x.

Example: Continuous uniform random variable

- I know a bus is going to arrive some time in the next hour, but I don't know when. If I assume all times within that hour are equally likely, what will my PDF look like?

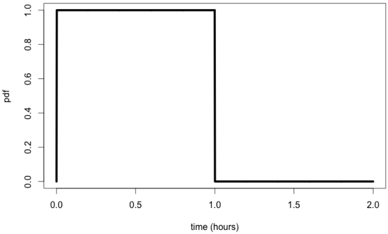
Example: Continuous uniform random variable

- I know a bus is going to arrive some time in the next hour, but I don't know when. If I assume all times within that hour are equally likely, what will my PDF look like?



$$f_X(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

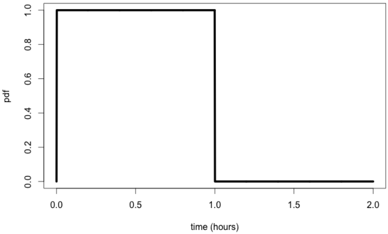
Example: Continuous uniform random variable



$$f_X(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

- What is $P(X > 0.5)$?
- What is $P(X > 1.5)$?
- What is $P(X = 0.7)$?

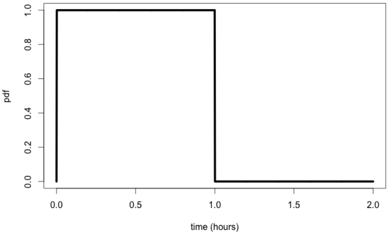
Example: Continuous uniform random variable



$$f_X(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

- What is $P(X > 0.5)$? 0.5
- What is $P(X > 1.5)$?
- What is $P(X = 0.7)$?

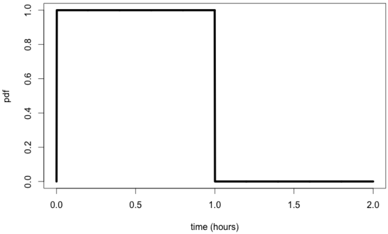
Example: Continuous uniform random variable



$$f_X(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

- What is $P(X > 0.5)$? 0.5
- What is $P(X > 1.5)$? 0
- What is $P(X = 0.7)$?

Example: Continuous uniform random variable



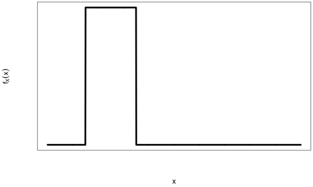
$$f_X(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

- What is $P(X > 0.5)$? 0.5
- What is $P(X > 1.5)$? 0
- What is $P(X = 0.7)$? 0

Continuous uniform random variable

- More generally, X is a **continuous uniform random variable** if it has PDF

$$f_X(x) = \begin{cases} c & \text{if } a \le x \le b \\ 0 & \text{otherwise.} \end{cases}$$

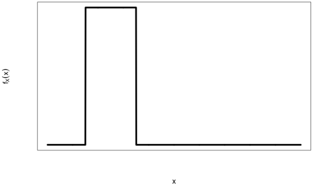


- What is c?

Continuous uniform random variable

- More generally, X is a **continuous uniform random variable** if it has PDF

$$f_X(x) = \begin{cases} c & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

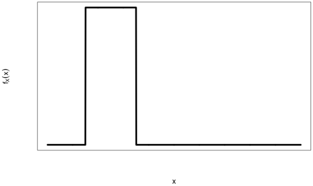


- What is c?
  - Well first lets see what $\int_a^b f_X(x)dx$ is!

Continuous uniform random variable

- More generally, X is a **continuous uniform random variable** if it has PDF

$$f_X(x) = \begin{cases} c & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

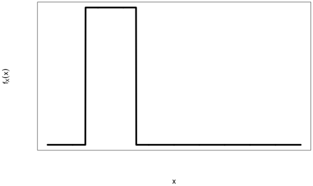


- What is c?
  - Well first lets see what $\int_a^b f_X(x)dx$ is!
  - This is just the area under the curve, i.e. $(b-a) \times c$...

Continuous uniform random variable

- More generally, X is a **continuous uniform random variable** if it has PDF

$$f_X(x) = \begin{cases} c & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

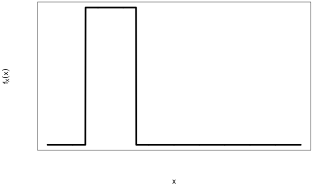


- What is c?
  - Well first lets see what $\int_a^b f_X(x)dx$ is!
  - This is just the area under the curve, i.e. $(b - a) \times c$...
  - But we want this to be 1. So c is

Continuous uniform random variable

- More generally, X is a **continuous uniform random variable** if it has PDF

$$f_X(x) = \begin{cases} c & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$



- What is c?
  - Well first lets see what $\int_a^b f_X(x)dx$ is!
  - This is just the area under the curve, i.e. $(b-a) \times c$...
  - But we want this to be 1. So c is $c = 1/(b-a)$

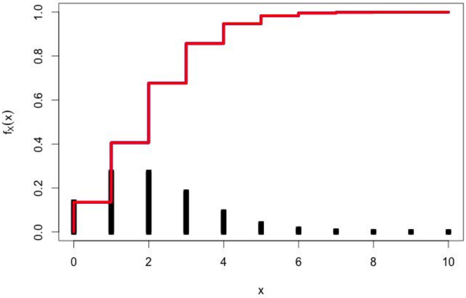
Cumulative distribution function

- Often we are interested in $P(X \leq x)$
- For example,
  - What is the probability that the bus arrives before 1:30?
  - What is the probability that a randomly selected person is under 5'7"?
  - What is the probability that this month's rainfall is less than 3in?
- We can get this from our PDF:

$$F_X(x) = P(X \leq x) = \begin{cases} \sum_{x' \leq x} p_X(x) & \text{if } X \text{ is a discrete r.v.} \\ \\ \int_{\infty}^{x} f_X(x')dx' & \text{if } X \text{ is a continuous r.v.} \end{cases}$$

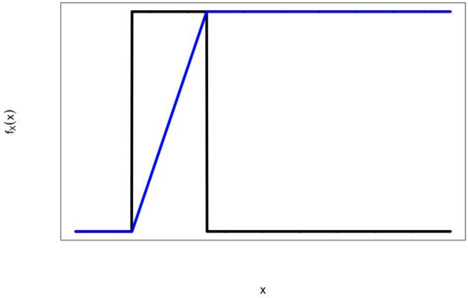
- This is called the **cumulative distribution function** (CDF) of X.
- Note: If we know $P(X \leq x)$, we also know $P(X > x)$

Cumulative distribution function



- If X is discrete, $F_X(x)$ is a piecewise-constant function of x.
- $F_X(x) = \sum_{x' \leq x} p_X(x')$

Cumulative distribution function

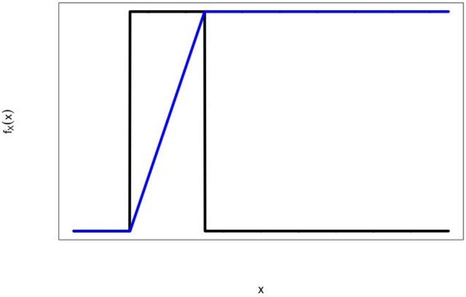


- The CDF is **monotonically non-decreasing**:

$$\text{if } x \le y, \text{ then } F_X(x) \le F_X(y)$$

- $F_X(x) \to 0$ as $x \to -\infty$
- $F_X(x) \to 1$ as $x \to \infty$

Cumulative distribution function



- If X is continuous, $F_X(x)$ is a continuous function of x
- $F_X(x) = \int_{t=-\infty}^{x} f_X(t)dt$

Expectation of a continuous random variable

- For discrete random variables, we found

$$E[X] = \sum_x x p_X(x)$$

- We can also think of the expectation of a continuous random variable – the number we would expect to get, on average, if we repeated our experiment infinitely many times.
- What do you think the expectation of a continuous random variable is?

Expectation of a continuous random variable

- For discrete random variables, we found

$$E[X] = \sum_x x p_X(x)$$

- We can also think of the expectation of a continuous random variable – the number we would expect to get, on average, if we repeated our experiment infinitely many times.
- What do you think the expectation of a continuous random variable is?
- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- *Similar to the discrete case... but we are integrating rather than summing*
- Just as in the discrete case, we can think of $E[X]$ as the "center of gravity" of the PDF.

Expectation of functions of a continuous random variable

- What do you think the expectation of a function $g(X)$ of a continuous random variable is?

Expectation of functions of a continuous random variable

- What do you think the expectation of a function $g(X)$ of a continuous random variable is?
- Again, similar to the discrete case...
- $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
- Note, $g(X)$ can be a continuous random variable, e.g. $g(X) = X^2$, or a discrete random variable, e.g.

$$g(X) = \begin{cases} 1 & \text{if } X \geq 0 \\ 0 & \text{if } X < 0 \end{cases}$$

Expectation of functions of a continuous random variable

- What do you think the expectation of a function $g(X)$ of a continuous random variable is?
- Again, similar to the discrete case...
- $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
- Note, $g(X)$ can be a continuous random variable, e.g. $g(X) = X^2$, or a discrete random variable, e.g.

$$g(X) = \begin{cases} 1 & \text{if } X \geq 0 \\ 0 & \text{if } X < 0 \end{cases}$$

- We can use the first and second moment to calculate the variance of X,

$$\text{var}[X] = E[X^2] - E[X]^2$$

Expectation of functions of a continuous random variable

- What do you think the expectation of a function $g(X)$ of a continuous random variable is?
- Again, similar to the discrete case...
- $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$
- Note, $g(X)$ can be a continuous random variable, e.g. $g(X) = X^2$, or a discrete random variable, e.g.

$$g(X) = \begin{cases} 1 & \text{if } X \geq 0 \\ 0 & \text{if } X < 0 \end{cases}$$

- We can use the first and second moment to calculate the variance of X,

$$\text{var}[X] = E[X^2] - E[X]^2$$

- We can also use our results for expectations and variances of linear functions:

$$E[aX + b] = aE[X] + b$$

$$\text{var}(aX + b) = a^2 \text{var}(X)$$

Mean of a uniform random variable

Let X be a uniform random variable over $[a, b]$. What is its expected value?

- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$

Mean of a uniform random variable

Let X be a uniform random variable over $[a, b]$. What is its expected value?

- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- $f_X(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$

Mean of a uniform random variable

Let X be a uniform random variable over $[a, b]$. What is its expected value?

- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- $f_X(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$
- So, $E[X] = \int_{-\infty}^{a} x \times 0 dx + \int_{a}^{b} \frac{x}{b-a} dx + \int_{b}^{\infty} x \times 0 dx$

Mean of a uniform random variable

Let X be a uniform random variable over $[a, b]$. What is its expected value?

- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- $f_X(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$
- So, $E[X] = \int_{-\infty}^{a} x \times 0 dx + \int_{a}^{b} \frac{x}{b-a} dx + \int_{b}^{\infty} x \times 0 dx$
  $= \int_{a}^{b} \frac{x}{b-a} dx$

Mean of a uniform random variable

Let X be a uniform random variable over $[a, b]$. What is its expected value?

- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- $f_X(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$
- So, $E[X] = \int_{-\infty}^{a} x \times 0 dx + \int_{a}^{b} \frac{x}{b-a} dx + \int_{b}^{\infty} x \times 0 dx$

$$= \int_{a}^{b} \frac{x}{b-a} dx$$

$$= \left[\frac{x^2}{2(b-a)} \right]_a^b$$

Mean of a uniform random variable

Let X be a uniform random variable over $[a, b]$. What is its expected value?

- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- $f_X(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$
- So, $E[X] = \int_{-\infty}^{a} x \times 0 dx + \int_{a}^{b} \frac{x}{b-a} dx + \int_{b}^{\infty} x \times 0 dx$

$$= \int_{a}^{b} \frac{x}{b-a} dx$$

$$= \left[\frac{x^2}{2(b-a)} \right]_a^b$$

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

Variance of a uniform random variable

To calculate the variance, we need to calculate the second moment:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$

Variance of a uniform random variable

To calculate the variance, we need to calculate the second moment:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$
$$= \int_a^b \frac{x^2}{b-a} dx$$

Variance of a uniform random variable

To calculate the variance, we need to calculate the second moment:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\ &= \int_a^b \frac{x^2}{b-a} dx \\ &= \left[\frac{x^3}{3(b-a)} \right]_a^b \end{aligned}$$

Variance of a uniform random variable

To calculate the variance, we need to calculate the second moment:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\ &= \int_a^b \frac{x^2}{b-a} dx \\ &= \left[\frac{x^3}{3(b-a)} \right]_a^b \\ &= \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3} \end{aligned}$$

Variance of a uniform random variable

To calculate the variance, we need to calculate the second moment:

$$
\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) dx \\
&= \int_a^b \frac{x^2}{b-a} dx \\
&= \left[\frac{x^3}{3(b-a)} \right]_a^b \\
&= \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}
\end{aligned}
$$

So, the variance is

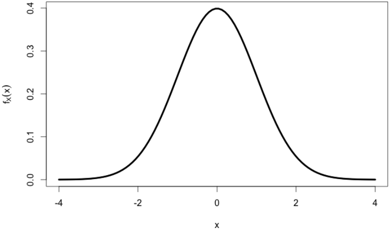
$$
\text{var}(X) = E[X^2] - E[X]^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}
$$

The normal distribution

- A normal, or Gaussian, random variable is a continuous random variable with PDF

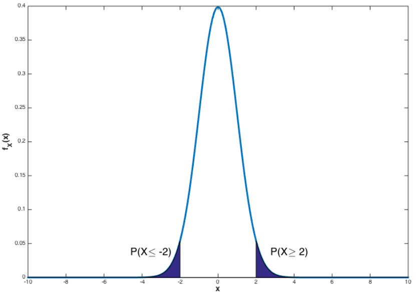
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

  where μ and σ are scalars, and $\sigma > 0$.
- We write $X \sim N(\mu, \sigma^2)$.
- The mean of X is μ, and the variance is σ^2 (how could we show this?)



The normal distribution

- The normal distribution is the classic “bell-shaped curve”.
- It is a good approximation for a wide range of real-life phenomena.
  - Stock returns.
  - Molecular velocities.
  - Locations of projectiles aimed at a target.



- Further, it has a number of nice properties that make it easy to work with. Like symmetry. In the above picture, $P(X \geq 2) = P(X \leq -2)$.

Linear transformations of normal distributions

- Let $X \sim N(\mu, \sigma^2)$
- Let $Y = aX + b$
- What are the mean and variance of Y?

Linear transformations of normal distributions

- Let $X \sim N(\mu, \sigma^2)$
- Let $Y = aX + b$
- What are the mean and variance of Y?
- $E[Y] = a\mu + b$
- $\text{var}[Y] = a^2\sigma^2$.

Linear transformations of normal distributions

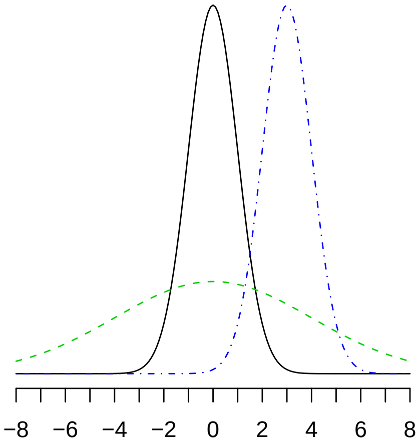
- Let $X \sim N(\mu, \sigma^2)$
- Let $Y = aX + b$
- What are the mean and variance of Y?
- $E[Y] = a\mu + b$
- $\text{var}[Y] = a^2\sigma^2$.

- In fact, if $Y = aX + b$, then Y is *also* a normal random variable, with mean $a\mu + b$ and variance $a^2\sigma^2$:

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

The normal distribution

- Example: Below are the pdfs of $X_1 \sim N(0, 1)$, $X_2 \sim N(3, 1)$, and $X_3 \sim N(0, 16)$.
- Which pdf goes with which X?



The standard normal

- I tell you that, if $X \sim N(0,1)$, then $P(X < -1) = 0.159$.
- If $Y \sim N(1,1)$, what is $P(Y < 0)$?
- Well we need to use the table of the **Standard Normal**.
- How do I transform Y such that it has the standard normal distribution?
- We know that a linear function of a normal random variable is also normally distributed!

The standard normal

- I tell you that, if $X \sim N(0,1)$, then $P(X < -1) = 0.159$.
- If $Y \sim N(1,1)$, what is $P(Y < 0)$?
- Well we need to use the table of the **Standard Normal**.
- How do I transform Y such that it has the standard normal distribution?
- We know that a linear function of a normal random variable is also normally distributed!
- Well $Z = Y - 1$ has mean zero and variance 1.

The standard normal

- I tell you that, if $X \sim N(0,1)$, then $P(X < -1) = 0.159$.
- If $Y \sim N(1,1)$, what is $P(Y < 0)$?
- Well we need to use the table of the **Standard Normal**.
- How do I transform Y such that it has the standard normal distribution?
- We know that a linear function of a normal random variable is also normally distributed!
- Well $Z = Y - 1$ has mean zero and variance 1.
- So $P(Y < 0) = P(Z - 1 < -1) = P(X < -1) = 0.159$.

The standard normal

- If $Y \sim N(0, 4)$, what value of y satisfies $P(Y < y) = 0.159$?
- The variance of Y is 4 times that of a standard normal random variable.
- Transform into a $N(0, 1)$ random variable!

The standard normal

- If $Y \sim N(0, 4)$, what value of y satisfies $P(Y < y) = 0.159$?
- The variance of Y is 4 times that of a standard normal random variable.
- Transform into a $N(0, 1)$ random variable!
- Use $Z = Y/2$...Now $Z \sim N(0, 1)$.

The standard normal

- If $Y \sim N(0, 4)$, what value of y satisfies $P(Y < y) = 0.159$?
- The variance of Y is 4 times that of a standard normal random variable.
- Transform into a $N(0, 1)$ random variable!
- Use $Z = Y/2$...Now $Z \sim N(0, 1)$.
- So, if $P(Y < y) = P(2Z < y) = P(Z < y/2)$.
- We want y such that $P(Z < y/2) = 0.159$. But we know that $P(Z < -1) = 0.159$, so?

The standard normal

- If $Y \sim N(0, 4)$, what value of y satisfies $P(Y < y) = 0.159$?
- The variance of Y is 4 times that of a standard normal random variable.
- Transform into a $N(0, 1)$ random variable!
- Use $Z = Y/2$...Now $Z \sim N(0, 1)$.
- So, if $P(Y < y) = P(2Z < y) = P(Z < y/2)$.
- We want y such that $P(Z < y/2) = 0.159$. But we know that $P(Z < -1) = 0.159$, so?
- So $y/2 = -1$ and as a result $y = -2$...!

The standard normal

- It is often helpful to map our normal distribution with mean μ and variance σ^2 onto a normal distribution with mean 0 and variance 1.
- This is known as the **standard normal**
- If we know probabilities associated with the standard normal, we can use these to calculate probabilities associated with normal random variables with arbitary mean and variance.
- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{x - \mu}{\sigma} \sim N(0, 1)$.
- (Note, we often use the letter Z for standard normal random variables)

The standard normal

- The CDF of the standard normal is denoted Φ:

$$\Phi(z) = P(Z \leq z) = P(Z < z) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

- We cannot calculate this analytically.
- The **standard normal table** lets us look up values of $\Phi(y)$.

	.00	.01	.02	0.03	0.04	⋯
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	⋯
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	⋯
0.2	0.5793	**0.5832**	0.5871	0.5910	0.5948	⋯
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	⋯
⋮	⋮	⋮	⋮	⋮	⋮	

$$P(Z < 0.21) = 0.5832$$

CDF of a normal random variable

If $X \sim N(3, 4)$, what is $P(X < 0)$?

- First we need to **standardize**:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 3}{2}$$

- So, a value of $x = 0$ corresponds to a value of $z = -1.5$
- Now, we can translate our question into the standard normal:

$$P(X < 0) = P(Z < -1.5) = P(Z \leq -1.5)$$

- Problem... our table only gives $\Phi(z) = P(Z \leq z)$ for $z \geq 0$.

CDF of a normal random variable

If $X \sim N(3, 4)$, what is $P(X < 0)$?

- First we need to **standardize**:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 3}{2}$$

- So, a value of $x = 0$ corresponds to a value of $z = -1.5$
- Now, we can translate our question into the standard normal:

$$P(X < 0) = P(Z < -1.5) = P(Z \leq -1.5)$$

- Problem... our table only gives $\Phi(z) = P(Z \leq z)$ for $z \geq 0$.
- But, $P(Z \leq -1.5) = P(Z \geq 1.5)$, due to symmetry.

CDF of a normal random variable

If $X \sim N(3, 4)$, what is $P(X < 0)$?

- First we need to **standardize**:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 3}{2}$$

- So, a value of $x = 0$ corresponds to a value of $z = -1.5$
- Now, we can translate our question into the standard normal:

$$P(X < 0) = P(Z < -1.5) = P(Z \leq -1.5)$$

- Problem... our table only gives $\Phi(z) = P(Z \leq z)$ for $z \geq 0$.
- But, $P(Z \leq -1.5) = P(Z \geq 1.5)$, due to symmetry.
- Our table only gives us "less than" values.

CDF of a normal random variable

If $X \sim N(3, 4)$, what is $P(X < 0)$?

- First we need to **standardize**:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 3}{2}$$

- So, a value of $x = 0$ corresponds to a value of $z = -1.5$
- Now, we can translate our question into the standard normal:

$$P(X < 0) = P(Z < -1.5) = P(Z \leq -1.5)$$

- Problem... our table only gives $\Phi(z) = P(Z \leq z)$ for $z \geq 0$.
- But, $P(Z \leq -1.5) = P(Z \geq 1.5)$, due to symmetry.
- Our table only gives us "less than" values.
- But, $P(Z \geq 1.5) = 1 - P(Z < 1.5) = 1 - P(Z \leq 1.5) = 1 - \Phi(1.5)$.

CDF of a normal random variable

If $X \sim N(3,4)$, what is $P(X < 0)$?

- First we need to **standardize**:

$$Z = \frac{X-\mu}{\sigma} = \frac{X-3}{2}$$

- So, a value of $x = 0$ corresponds to a value of $z = -1.5$
- Now, we can translate our question into the standard normal:

$$P(X < 0) = P(Z < -1.5) = P(Z \leq -1.5)$$

- Problem... our table only gives $\Phi(z) = P(Z \leq z)$ for $z \geq 0$.
- But, $P(Z \leq -1.5) = P(Z \geq 1.5)$, due to symmetry.
- Our table only gives us "less than" values.
- But, $P(Z \geq 1.5) = 1 - P(Z < 1.5) = 1 - P(Z \leq 1.5) = 1 - \Phi(1.5)$.
- And we're done!
  $P(X < 0) = 1 - \Phi(1.5) =$ (look at the table...)$1 - 0.9332 = 0.0668$

Recap

- With continuous random variables, any specific value of $X = x$ has zero probability.
- So, writing a function for $P(X = x)$ – like we did with discrete random variables – is pretty pointless.
- Instead, we work with **PDFs** $f_X(x)$ – functions that we can integrate over to get the probabilities we need.

$$P(X \in B) = \int_B f_X(x)dx$$

- We can think of the PDF $f_X(x)$ as the "probability mass per unit area" near x.
- We are often interested in the probability of $X \leq x$ for some x – we call this the cumulative distribution function $F_X(x) = P(X \leq x)$.
- Once we know $f_X(x)$, we can calculate expectations and variances of X.