Computer Science

CSC380: Principles of Data Science

Midterm review

Chicheng Zhang

Probability

Classical probability model

Special case

Assume each outcome is equally likely, and sample space is finite, then the probability of event is:

$$P(E) = \frac{|E|}{|S|}$$

Number of elements in event set

Number of possible outcomes (e.g. 36)

This is called classical probability model

Example: dice roll

- Suppose we roll two fair dice. A = {First die shows up 1}, B = {two dices summing up to 5}

- Find P(A), P(B), P(A and B)

- All 36 outcomes equally likely
- $P(A) = \frac{6}{36}$
- $P(B) = \frac{4}{36}$
- $P(A, B) = \frac{1}{36}$

$$A = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)\}$$

$$B = \{(1,4), (2,3), (3,2), (4,1)\}$$

$$A \cap B = \{(1,4)\}$$

Another attempt

- Suppose we roll two fair dice. B = {two dices summing up to 5}. Find P(B).

- Two dices can sum to any number in 2, .., 12

- Define the sample space to be

  {two dices summing up to 2, …, two dices summing up to 12}

- All outcomes are equally likely, so P(B) = $\frac{1}{11}$

- What’s wrong with this solution?
  - With the outcomes defined this way, It is incorrect that all outcomes are equally likely!

Rules of probability

- To recap and summarize:

Rules of Probability

1. **Non-negativity:** All probabilities are between 0 and 1 (inclusive)
2. **Unity of the sample space:** $P(S) = 1$
3. **Complement Rule:** $P(E^C) = 1 - P(E)$
4. **Probability of Unions:**
   (a) *In general,* $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
   (b) If E and F are disjoint, then $P(E \cup F) = P(E) + P(F)$

Example: dice roll

- Suppose we roll two fair dice. A = {First die shows up 1}, B = {two dices summing up to 5}. Find P(A or B).

- Two ways to solve this question:
  1. Find the event $A \cup B$
  2. Use inclusion-exclusion rule

$$P(A \cup B) = P(A) + P(B) - P(A, B) = \frac{9}{36} = \frac{1}{4}$$

- Equivalence between operations of sets and operations of propositions
  - And $\leftrightarrow \cap$, Or $\leftrightarrow \cup$, Not $\leftrightarrow C$
  - Set operation identities apply (e.g. De Morgan's Law)

Independence of events

- Independence of two events

Independence (version 2)

If A and B are independent events, then

$$P(A \cap B) = P(A)P(B)$$

Fact: if A, B are independent, then:

- A^C and B are independent $\quad P(A^C, B) = P(A^C)\, P(B)$
- A and B^C are independent
- A^C and B^C are independent

Independence of events

- Independence of multiple events

- Events $A_1, \ldots, A_n$ are independent if for any subsets $A_{i_1}, \ldots, A_{i_j}$,

$$P\left(A_{i_1}, \ldots, A_{i_j}\right) = P\left(A_{i_1}\right) \cdot \ldots \cdot P(A_{i_j})$$

Important consequence:

$$P(A_1, \ldots, A_n) = P(A_1) \cdot \ldots \cdot P(A_n)$$

Example: lightbulbs

- Suppose we have three lightbulbs, each is on with probability 0.4, independently. What is the probability that at least one lightbulb is on?

- E_1: light 1 is on; same for E_2, E_3
- We are asked to find $P(E_1 \cup E_2 \cup E_3)$

- Note: $P(E_1 \cap E_2 \cap E_3)$ is easy to find, but not quite what we want

Example: lightbulbs

- Suppose we have three lightbulbs, each is on with probability 0.4, independently. What is the probability that at least one lightbulb is on?

- We can use De Morgan's Law:
- $P\left((E_1 \cup E_2 \cup E_3)^C\right) = P(E_1^C \cap E_2^C \cap E_3^C)$

$$= P(E_1^C)P(E_2^C)P(E_3^C)$$

$$= 0.6^3$$

Thus, $P(E_1 \cup E_2 \cup E_3) = 1 - 0.6^3$

Repeated independent trials

- Suppose we have three lightbulbs, each is on with probability 0.4, independently. What is the probability that exactly 2 lightbulbs are on?

- $0.4^2 0.6^1$?
- $\binom{3}{2} 0.4^2 0.6^1$

- Why? Three outcomes: (on, on, off), (on, off, on), (off, on, on)
- This is also the basis of binomial distributions (Galton board)

Example: basketball

- Imagine a basketball player who takes three-point shots. Suppose they successfully make a three-pointer with a probability of 30% ($p = 0.3$), and each shot is independent of the others.
- What is the probability that exactly 10 shots are taken to make **5 successful three-pointers?**

Possible shot history:

SSSSFFFFFS

SSSSSFFFFF?

Last trial must be a ‘S’

Example: basketball

The probability we are looking for is

$$P(\text{ 4 successes in first 9 shots and success in } 10th \text{ shot})$$

$$= P(\text{4 successes in first 9 shots}) \times P(\text{success in } 10th \text{ shot})$$

$$= \binom{9}{4} p^4 (1-p)^5 \times p$$

Side note: the number of shots until exact 5 successes is known to follow from a *negative binomial distribution with parameter (r=5,p=0.3)*

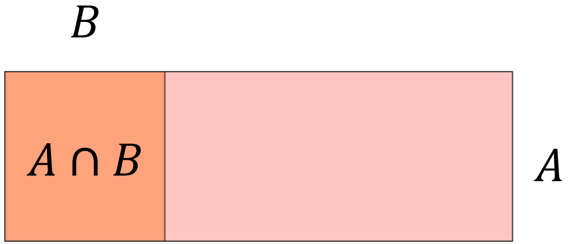
- *generalizes geometric distributions*

Conditional probability; probabilistic reasoning

Conditional Probability

- Consider the ways B can occur in the context of A (i.e., $A \cap B$), out of all the ways A can occur:

$$P(\,B \mid A\,) = \frac{P(A \cap B)}{P(A)}$$

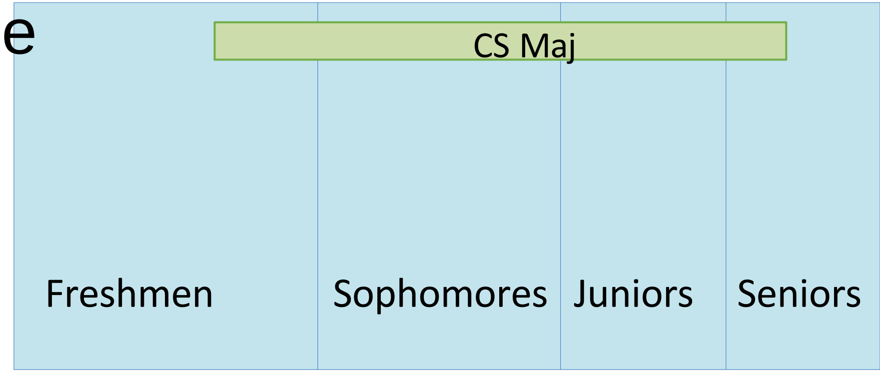


- This allows us to do probabilistic reasoning:
  - Compare $P(\,B \mid A\,)$ with $P(B)$
  - Suppose I am tested positive. Does this information increase my likelihood of getting COVID?

Basic probability facts for probabilistic reasoning

Law of Total Probability Suppose $B_1, \dots, B_n$ *form a partition* of the sample space S. Then,

$$P(A) = P(A, B_1) + \cdots + P(A, B_n)$$



Bayes rule

Prior probability Support of evidence

$$P(A \mid B) = \frac{P(A) \cdot P(B \mid A)}{P(B)}$$

Posterior probability

Conditional Probability

- Useful tools to reason about conditional probability:

1. Two-way tables

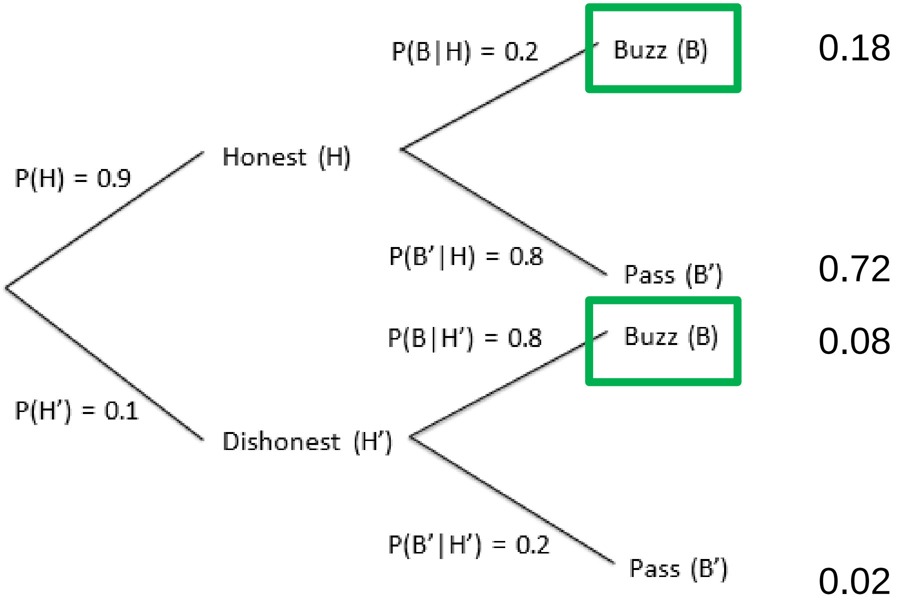
		Antigen B Absent	Antigen B Present	Marginal
Antigen A	Absent	0.44	0.10	0.54
	Present	0.42	0.04	0.46
	Marginal	0.86	0.14	1.00

Table: Probability Estimates for U.S. Blood Types

Law of total probability: summing over rows / columns

Conditional Probability

2. Probability trees



Law of total probability: summing over relevant branches

Conditional probability: weight of a branch relative to all relevant branches

Example: two boxes

- Select a box randomly and select a ball from it randomly. Probability that the selected ball is red?

- B_1: Box 1 selected; B_2: Box 2 selected; A: a red ball selected

- $P(A) = \sum_{i=1}^{2} P(B_i)\, P(A \mid B_i) = \frac{1}{2} \times \frac{60}{100} + \frac{1}{2} \times \frac{10}{30} = \frac{7}{15}$

Example: two boxes

- Select a box randomly and select a ball from it randomly. We are told it is red. Probability that box 1 was selected?

- $P(B_1 \mid A) = \frac{P(A,B_1)}{P(A)}$

$$= \frac{P(B_1)P(A|B_1)}{\sum_i P(B_i)P(A|B_i)} = \frac{\frac{1}{2}\times\frac{60}{100}}{\frac{1}{2}\times\frac{60}{100}+\frac{1}{2}\times\frac{10}{30}} = \frac{9}{14}$$

Independence revisited

Independent Events

We say that event A is **independent** of event B if conditioning on B does not change the probability of A, that is if

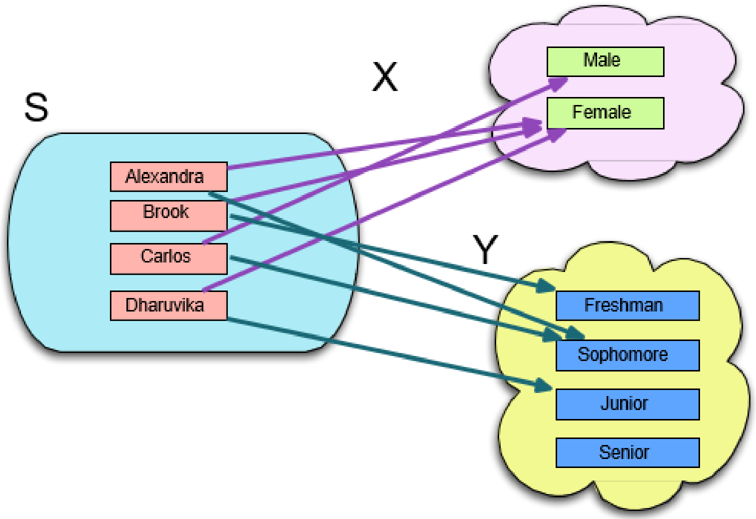
$$P(A \mid B) = P(A)$$

- Is disjointness equivalent to independence?
  - No, they are kind of opposite!

Random variables

Random variables

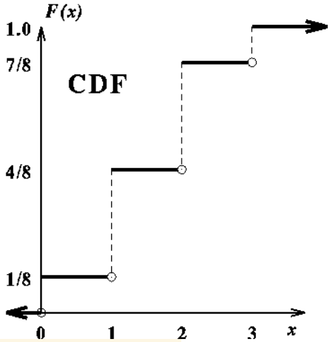
- Random variables: variables whose values are not deterministic but *random*



Discrete RVs

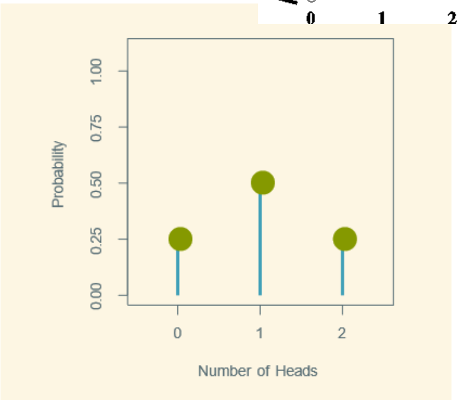
- The stochastic properties of RVs X are summarized by their probability distribution laws, represented by

  - Cumulative distribution function F



  - Or, probability mass function f

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



Probability and odd (HW3)

- In gambling, odds of 4:1 means for every 1 unit you bet, you could win 4 unit in profit if the event occurs, while risking 1 unit in loss if it doesn't

- Suppose you believe that A happens with probability p. Should you make the bet?

- You believe that your winning X is distributed as

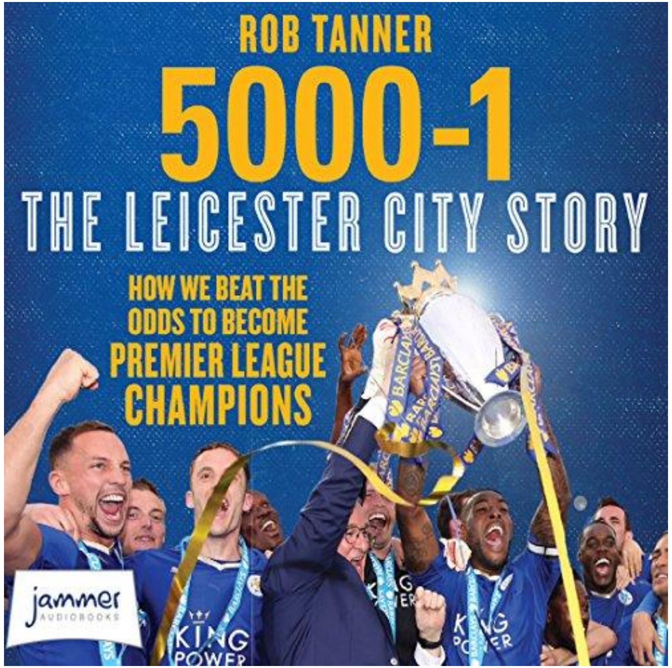
x	4	-1
$P(X = x)$	p	$1 - p$

- You should make the bet when $\mathrm{E}[\mathrm{X}] = p \cdot 4 + (-1)(1 - p) > 0$,
  - i.e, $p > 0.2$

Probability and odd (HW3)

- In 2015/2016, English Premier Soccer Team Leicester City won the championship, even though that they were listed with 5,000-to-1 odds to win the league

- What is the predicted probability of they winning the league?



Discrete RVs

- Converting X's CDFs to PMF and the other way around
- Given X's PMF, find its expectation $\mathrm{E}[X]$
- Given X's PMF, find $f(X)$'s PMF
- Given X's PMF, find $\mathrm{E}[f(X)]$
  - The expectation formula (the rule of the lazy statistician)
  - Find $\mathrm{Var}[X]$
    - Alternative formula: $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$

An example

- Suppose $X \sim \text{Geom}\left(\frac{1}{2}\right)$

x	1	2	3	…
$P(X = x)$	0.5	0.25	0.125	

- i.e., $P(X = x) = \frac{1}{2^x}$, for integer x
- What is $Y = 2^X$'s *PMF?*

y	2	4	8	…
$P(Y = y)$	0.5	0.25	0.125	

- What is $\mathrm{E}[Y]$?
  - $\sum_x 2^x \cdot \frac{1}{2^x} = \sum_x 1 = +\infty$

An example

- $\mathrm{E}[Y] = +\infty$

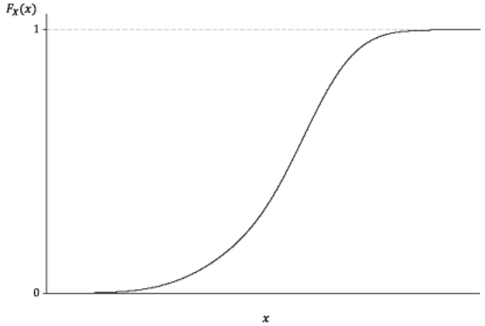
y	2	4	8	…
$P(Y = y)$	0.5	0.25	0.125	

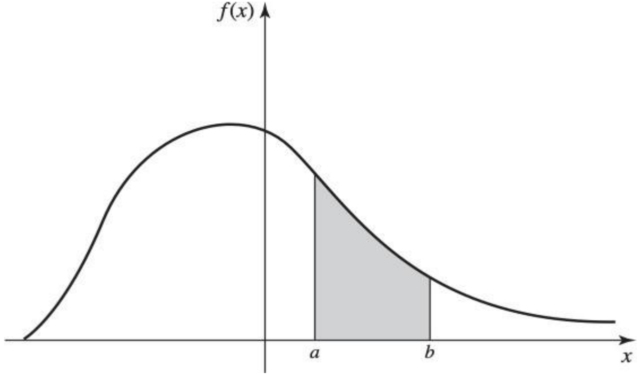
- This is called the *St. Petersburg's Paradox*
  - journal of the Imperial Academy of Sciences in St. Petersburg in 1738, by Daniel Bernoulli
  - Even though the expected winnings is infinite, in reality, most people would not pay a large amount to play this game

Continuous RVs

- For continuous RVs X, PMF is irrelevant. Probability distribution laws represented by:
  - Cumulative distribution function F
  - Or, probability density function f

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





Continuous RVs

- Converting X's CDFs to PDFs and the other way around
  - We may ask you to do simple integration (review examples & antiderivatives covered in class)
  - Keep in mind the "area under curve" interpretation of integration

- Given X's PMF, find its expectation $\mathrm{E}[X]$

- Given X's CDF, find $f(X)$'s CDF
  - Recommend: go through the examples we had in class about $f(X) = X + a$ and $f(X) = aX + b$

- Given X's PMF, find $\mathrm{E}[f(X)]$
  - The expectation formula (the rule of the lazy statistician)
  - Find $\mathrm{Var}[X]$
    - Alternative formula: $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$

Relationship between RVs and their transformations

- What are the relationships between X's PMF / PDF and those of $aX + b$?

- What's a general expression of $\mathrm{E}[aX + b]$ using $\mathrm{E}[X]$?

- What's a general expression of $\mathrm{Var}[aX + b]$ using $\mathrm{Var}[X]$?

Notable discrete RVs

I expect you to be comfortable with writing down the PMFs of

- Uniform distribution over a set
- Binomial distribution
- Geometric distribution

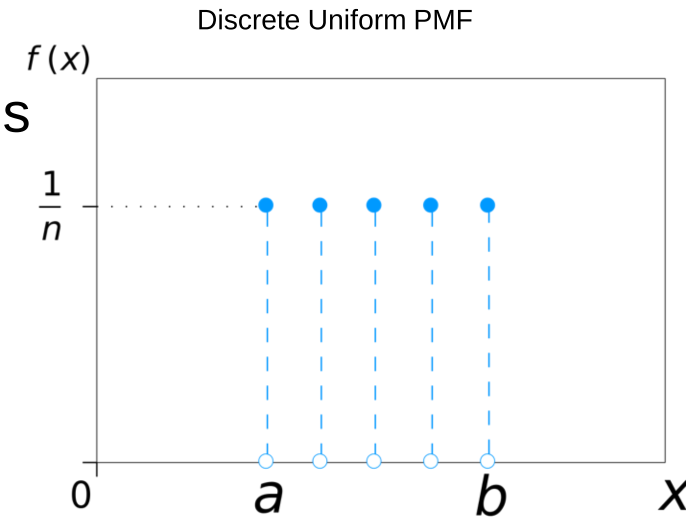
Additionally: understand the meaning of library functions

  - [scipy.stats.binom.pmf(x, n, p), scipy.stats.binom.cdf(x, n, p)]
  - [scipy.stats.geom.pdf(x, p), scipy.stats.geom.cdf(x, p)]

Uniform distribution over a set

More generally, consider $S = \{v_1, v_2, \dots, v_N\}$; X is drawn from the uniform distribution of S, then

$$P(X = k) = \begin{cases} \frac{1}{N} & \text{if } k \in \{v_1, v_2, \dots, v_N\} \\ 0 & \text{otherwise} \end{cases}$$



Useful connection to 'data' lecture:

- Mean / variance of X = Mean / variance of the dataset S
- PMF of $X \approx$ histogram of S

Example question

- Suppose that a baseball hitter has a probability of success p = 0.7, What is the probability that she hits more than 6 times (inclusively) out of a total of 15 throws?

- You may use the following outputs:

binom.pmf(5, 15, 0.7) = 0.003

binom.cdf(5, 15, 0.7) = 0.004

binom.pmf(6, 15, 0.7) = 0.012

binom.cdf(6, 15, 0.7) = 0.015

Example question

- Suppose that a baseball hitter has a probability of success p = 0.7, What is the probability that she hits more than 6 times (inclusively) out of a total of 15 throws?

- $X \sim \mathrm{Bin}(n = 15, p = 0.7)$
- We'd like to find $P(X \geq 6)$
  - Perhaps we can express it as CDF of X?
- $P(X \geq 6) = 1 - P(X \leq 5)$

  $= 1 -$ binom.cdf(5, 15, 0.7) $= 0.996$

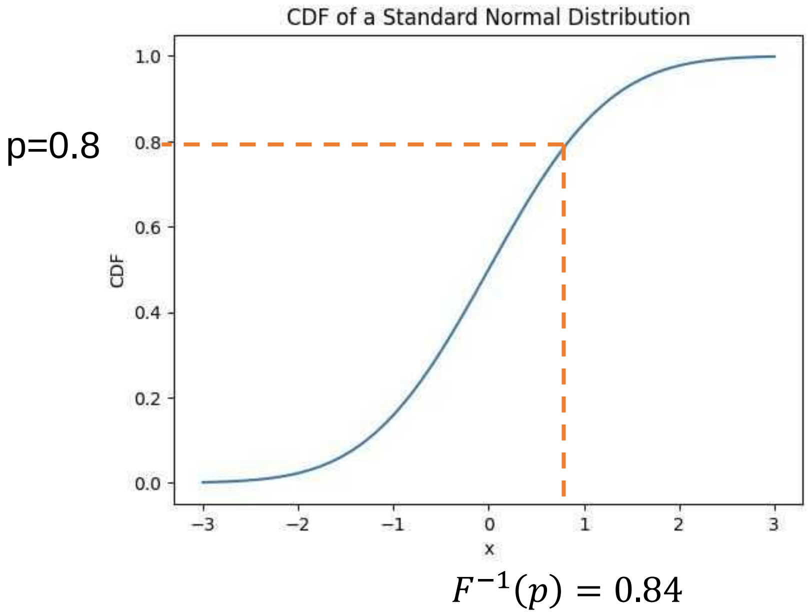
Notable continuous RVs

- I expect you to be comfortable with:
- Writing down the PDF and CDF of continuous uniform distributions
- Writing down the PDF and CDF of normal distributions
  - Have a good understanding on scaling and shifting properties of normal RVs
  - Write $P(a \leq X \leq b)$ in terms of standard norm CDFs
  - Understand the meaning of library function [scipy.stats.norm.cdf(x)]() and [scipy.stats.norm.pdf(x)]() – they are standard normal CDFs and PDFs

Percentage Point Function (HW4)

- The percentage point function (PPF), also known as the quantile function F^{-1}, is the inverse of the CDF F

- $F^{-1}(p)$ = threshold x, such that $P(X \leq x) = p$
- $F^{-1}\left(\frac{1}{2}\right)$: median
- $F^{-1}\left(\frac{1}{4}\right)$: 1-st quartile



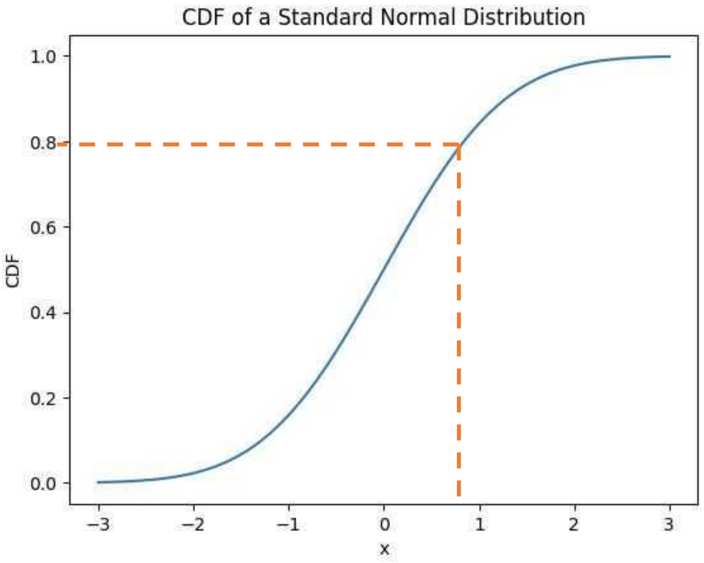
- Standard normal PPF: scipy.stats.norm.ppf

Percentage Point Function (HW4)

- Fun fact: to draw samples from a distribution with CDF F, we can do it in two steps:

- First, draw $U \sim \mathrm{Uniform}[0,1]$
- Second, return $F^{-1}(U)$

p=0.8



$F^{-1}(p) = 0.84$

- $X = F^{-1}(U)$ has CDF F (Exercise)

Multivariate RVs

- Understand the meaning of joint distribution, marginal distribution of a pair of RVs (X, Y)

- How to obtain marginal distributions from joint?
  - Marginalization
    - Summation for discrete (X,Y)
    - Integration for continuous (X,Y)

Best of luck!

“Exams are a tool for learning, not the purpose of education.”