CPE 486/586: Machine Learning for Engineers

04 Statistics and Probability

Fall 2025

Rahul Bhadani

Outline

Probability

Probability Theory

1. Probability is a branch of mathematics that deals with uncertainty.
2. Probability is the likelihood or chance that something will occur.
3. Probability describes things whose outcomes are uncertain or random.

Sample Space

Definition

A sample space $\mathcal{S}$ is the possible outcome of an experiment. A point s in $\mathcal{S}$ is called sample outcome, realization, or outcome. Subsets of $\mathcal{S}$ are called events.
Example:
In an experiment of tossing two coins, our sample space is $\mathcal{S} = \{HH, HT, TH, TT\}$. An event that at least one of the coins is heads is $\mathcal{E} = \{HH, HT, TH\}$.

Probability

Definition

For each event $\mathcal{E}$ in the sample space $\mathcal{S}$, the probability is a function which associates with $\mathcal{E}$ a number between 0 and 1, i.e, $P(\mathcal{E}) \in [0, 1]$.

Probability

Definition

It is technically hard to describe the probability of each single event. In that case, we consider a good collection $\mathcal{B}$ of events which is large enough to contain all the useful events including $\emptyset$, and $\mathcal{S}$, and is closed under all possible countable set operations. This collection $\mathcal{B}$ is called a sigma-algebra or Borel field. Probability is a set function defined only on this collection, i.e.,

$$P : \mathcal{B} \to [0, 1]$$

Probability: Example

Example: Tossing a fair die.

1. One possible σ-algebra $\mathcal{B}_1$ is $\{\emptyset, \mathcal{S}\}$.
2. Another σ-algebra is the power set of $\mathcal{S}$.
3. Another σ-algebra is $\mathcal{B} = \{\emptyset, \mathcal{S}, \{0\}, \{2, 3, 4\}\}$.

Conditional Probability

Definition

The probability of an event A given B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

given $P(B) > 0$.

Conditional Probability

Some facts

1. $P(B|B) = 1$
2. $P(A|B) = \dfrac{P(A)}{P(B)}$ for $A \subset B$
3. $P(B|A) = 1$ for $A \subset B$
4. Multiplication formula is useful when it is easier to obtain conditional probability: $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$.
5. $P(A) = P(A \cap B) + P(A \cap B^c) = P(B)P(A|B) + P(B^c)P(A|B^c)$, where B^c is the complement of B.

Conditional Probability

Example 1

Let box 1 contain 2 red balls and 3 green balls and box 2 contain 5 red balls and 4 green balls. One box is chosen at random and one ball is drawn randomly from the chosen box. What is the probability of getting a red ball?

Conditional Probability

Example 1 Solution

Consider event A = choosing a red ball, even B = choosing the first box. Then $P(B) = P(B^c) = \dfrac{1}{2}$ (B^c is the complement of B). Then, the probability of drawing a red ball given the first box can be written as $P(A|B) = \dfrac{2}{5}$. $P(A|B^c) = \dfrac{5}{9}$. Then the probability of getting a red ball is $P(A) = \dfrac{1}{2} \times \dfrac{2}{5} + \dfrac{1}{2} \times \dfrac{5}{9} = \dfrac{43}{90}$.

Bayes' Rule as Inversion of Conditional Probability

Assume that we can partition sample space $\mathcal{S}$ into B_1, B_2, B_3. The conditional probabilities can be $P(A|B_i), i = 1, 2, \cdots$. Then, the posterior probabilities can be written as

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)} \tag{2}$$

Bayes' Rule as Inversion of Conditional Probability

Example 1 Extension

Consider the example 1 from the previous section. Given that we obtained a red ball, what is the probability that box 1 was selected?

Bayes' Rule as Inversion of Conditional Probability

Example 1 Extension Solution

We are interested in determining $P(B|A)$. From the Bayes' Rule,

$$P(B|A) = \frac{\frac{1}{2} \times \frac{2}{5}}{\frac{1}{2} \times \frac{2}{5} + \frac{1}{2} \times \frac{5}{9}} = \frac{\dfrac{1}{5}}{\dfrac{43}{90}} = \frac{18}{43} \tag{3}$$

Bayes' Rule as Inversion of Conditional Probability

Example 2: Rare disease probability

Consider that the probability of occurrence of a disease is 1/1000. A test accurately predicts the occurrence with 99% accuracy and negates the disease with 98% accuracy. Given that the test has shown positive, what is the probability of actually having the disease?

Bayes' Rule as Inversion of Conditional Probability

Example 2 Solution: Rare disease probability

Let's consider B_1 = event of the occurrence of disease, $B_2 =$ event that the patient is healthy. A = event that a patient has been tested positive.
From the question, we have $P(B_1) = 0.001$. $P(B_2) = 1 - P(B_1) = 0.999$
$P(A|B_1) = 0.99$, $P(A|B_2) = 0.02$, then fron Bayes' rule

$$P(B_1|A) = \frac{0.001 \times 0.99}{0.001 \times 0.99 + 0.999 \times 0.02} = 0.047210 \tag{4}$$

Independence

Definition

Two events are independent if the occurrence of one of the events gives us no information about whether or not the other event will occur. In other words, events do not influence each other. Formally, we write

$$P(A|B) = P(A) \tag{5}$$

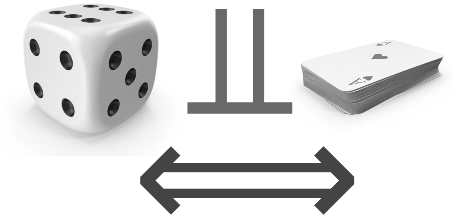
which is still applicable even if $P(A)$ or $P(B)$ is 0.

Independence

$P(A|B) = P(A)$ is equivalent to $P(A \cap B) = P(A)P(B)$, that is, the probability that they both occur is equal to the product of the probabilities of the two individual events.

Independence: Symmetric Relationship and Unrelated Events

Symmetric relationship: A is independent of B implies B is independent of A.

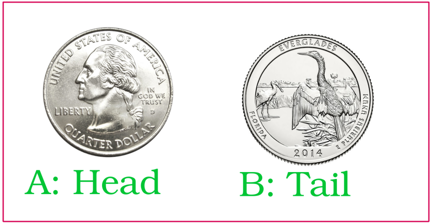


Unrelated events must be independent.

Independence vs. Mutual Exclusivity

Independent and mutually exclusive events are not the same. If $P(A) > 0$, and $P(B) > 0$, then:

1. If A and B are mutually exclusive, they cannot be independent.
2. If A and B are independent, they cannot be mutually exclusive.



$$A \cap B = \varnothing$$

Mutually exclusive but not independent

Independence: Complementary Events

If A and B are independent, so are A, B^c ; A^c, B; and A^c, B^c.

Independence of Multiple Events

The idea of independence can be extended to more than two events. For example, A, B, and C are independent if:

1. A and B are independent,
2. A and C are independent, and
3. B and C are independent; and
4. $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Independence

Example 3

Let's say that a man and a woman each have a pack of 52 playing cards. Each draws a card from his/her pack. Find the probability that they each draw the ace of clubs ♣.

Independence

Example 3 Solution

As they are drawing their cards independently of each other, the required probability is $\dfrac{1}{52} \times \dfrac{1}{52}$.

Independence

Example 4

What is the chance of getting at least one six in 4 throws of a dice?

Independence

Example 4 Solution

We first calculate the probability of no six in 4 throws of a dice which is given by $(5/6)^4$ as each throw is independent of one another. Thus the probability of getting at least one six in 4 throws of a dice is $1 - (5/6)^4$.

Statistics

Random Variable

Random Variable

Definition

A random variable is a real-valued function defined on the sample space $\mathcal{S}$, that is a rule which assigns a number to each outcome. Random variables can be thought of as a stochastic function. Random variables are denoted by capital letters such as X, Y, Z. The outcome of random variables is represented with corresponding small letters such as x, y, z.

Random Variable

Example

For example, the probability that a random variable takes the value of 2 would be expressed as $P(X = 2)$ where $x = 2$.

- A random variable is a function $\mathcal{S} \to \mathbb{R}$.

• A random variable is called discrete if it takes only a finite or countably many values.
• A random variable is called continuous if it can assume all the values in an interval.

Probability Mass Function

Definition

At this point, we can transition from probability to probability law which will help us in writing computer programs at a later stage.
The probability law of a discrete random variable can be described by the function defined as

$$p(x) = P(X = x) \tag{6}$$

Probability Mass Function

Let $x_1, x_2, \cdots$ be the points where p gives positive masses (or values) that we can call $p_1, p_2, \cdots$. Thus $P(X = x_j) = p_j, j = 1, 2, \cdots$

- $p_j \geq 0$, and $\sum_{j=1}^{\infty} p_j = 1$.

Probability Mass Function: Example

Example 5

If a coin is tossed twice, the sample space $\mathcal{S} = \{HH, HT, TH, TT\}$. Let X be the random variable that denotes the number of heads that can come up. With each sample point, we can associate a number for X, as shown below:

	HH	HT	TH	TT
X	2	1	1	0

Note: We can also define some other random variable on the same sample space, such as the square of the number of heads, i.e. $Y = X^2$ or the number heads minus the number of tails $Z = X - W$, where W is the random variable the denotes the number of tails that can come up.

Probability Mass Function: Example

Example 6

$$p_k = P(X = k) = q^{k-1}p, \text{ where } 0 < p < 1, q = 1 - p, \quad k = 1, 2, \cdots \tag{7}$$

This is pmf as it satisfies $p_k \geq 0$, and $\sum_{k=0}^{\infty} p_k = 1$.

Probability Density Function

When it comes to the continuous random variable X that takes real values, then it is not possible to define the probability for a single point, as they are real values. At what real value we should define a probability – there can be a possibility of an infinite number after a decimal point!!!. Hence, we definite the probability over a range. In this case, we get the probability density function (PDF) $p(x)$ such that

$$P(x_0 < x < x_1) = \int_{x_0}^{x_1} p(x)dx. \tag{8}$$

Probability Density Function

Endpoints x_0 and x_1 may not be included. They do not matter for continuous random variable X. In this case, $p(x)$ satisfies the following property:

- $p(x) \geq 0 \quad \forall x$
- $\int_{-\infty}^{\infty} p(x)dx = 1$

Probability Distribution

Cumulative Distribution Function (CDF)

Definition

For any $x \in \mathbb{R}$, we define cumulative distribution function as

$$F(x) = P(X \leq x) \tag{9}$$

They are sometimes to referred just as **distribution** or **probability distribution**. Shorthand notation for CDF is F_X.

CDF:Example

Example 7: CDF of a Discrete Random Variable

Consider a discrete random variable X such that $P(X = 0) = 1/3$, $P(X = 1) = 1/2$, $P(X = 2)1/6$. Then pmf of X, $p(x)$ is written as $p(0) = 1/3$, $p(1) = 1/2$, $p(2) = 1/6$ and $p(x) = 0$ for all other x.
The CDF of X, $F(X)$ is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ 1/3, & 0 \leq x \leq 1 \\ 5/6, & 1 \leq x < 2 \\ 1, & x \geq 2 \end{cases} \tag{10}$$

In this case, the graph of F is a step function, having jumps at the point where X has mass.

CDF: Example

What would be the CDF from Example 6?

$$
\begin{aligned}
F(x) &= P(X \leq k) \\
&= q^0 + q^1 p + q^2 p + \cdots + q^{k-1} p \\
&= p(q^0 + q^1 + q^2 + \cdots + q^{k-1})
\end{aligned} \tag{11}
$$

We apply geometric progression formula $S_{GP} = \dfrac{a(1 - r^n)}{1 - r}$ where a is the first term in GP and r is the common ratio. Hence,

$$
F(x) = p \frac{q^k - 1}{q - 1} \tag{12}
$$

CDF of a Continuous Random Variable (RV)

For a continuous RV, we can define the CDF as

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{13}$$

for the probability density function (PDF) $f(x)$. We also write f_X to denote that f is the PDF of the random variable X. Here, t is the dummy variable of the integration.

1. If $F(x)$ is differentiable at a given point, then $f(x) = F'(x)$.
2. f must satisfy

$$f(t) \geq 0, \quad \int_{-\infty}^{\infty} f(t)dt = 1. \tag{14}$$

Identically Distributed

If X and Y are two random variables with $P(X \in A) = P(X \in A)$ for all sets A. Then, they are called identically distributed. This is equivalent to $F_X(x) = F_Y(x)$ for all x, i.e. two CDFs are equal.

Independent and Identically Distributed (IID)

In addition to identically distributed, if two random variables are mutually independent, then they are called independent and identically distributed or (IID). It means sample items are independent events; the knowledge of the value of one variable will not give any information about the value of the other and vice versa.

Example

Example 8

Hit a dartboard of radius R randomly. Let X be the random variable denoting the distance of the chosen point from the center. In this case, probabilities are proportional to area, so $F(x) = P(X \leq x) = \frac{\pi x^2}{\pi R^2} = (x/R)^2, \quad 0 \leq x \leq R$.

Expected Values or Expectations

Consider a random variable X with pdf or pmf $f(x)$. If we have a function $g(X)$ of a random variable, then we can define the expectation as

$$\mathbb{E}(g(X)) = \begin{cases} \int_{-\infty}^{\infty} g(x)f(x)dx, & \text{if } X \text{ is continuous} \\ \sum_{x\in\mathcal{X}} g(x)f(x), & \text{if } X \text{ is discrete} \end{cases} \tag{15}$$

provided that $|\mathbb{E}(g(X)| < \infty$. If $|\mathbb{E}(g(X)| = \infty$, we say that the expectation doesn't exist.
Here $\mathcal{X}$ is the sample space of the random variable X.
Some literature also write E instead of $\mathbb{E}$ for the expectation.
Expectation is also referred to as mean or average.

Median

If X is a continuous random variable and has CDF F_X. Its median m is the value that satisfies $F_X(m) = 1/2$, that is,

$$\int_{-\infty}^{m} f_X(x)dx = \int_{m}^{\infty} f_X(x)dx = \frac{1}{2} \tag{16}$$

Equivalently, $m = F_X^{-1}(1/2)$.

Variance and Standard Deviation

The variance of a random variable X can be defined as

$$var(X) = \mathbb{E}(X - \mathbb{E}(X))^2 \tag{17}$$

We also denote $\sigma^2 = var(X)$. The standard deviation of X is the square root of $var(X)$,i.e.,

$$\sigma = \sqrt{var(X)} \tag{18}$$

Variance and standard deviation measure the degree of spread of a distribution around its mean $\mathbb{E}(X)$.

Statistical Moments

Statistical Moments

- Statistical moments describe the shape and characteristics of probability distributions
- The k-th moment about the origin is defined as:

$$\mu'_k = E[X^k] = \int_{-\infty}^{\infty} x^k f(x) dx$$

- The k-th central moment is:

$$\mu_k = E[(X - \mu)^k] = \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx$$

- First four moments provide crucial distribution information:
  - $\mu'_1 = \mu$ (mean)
  - $\mu_2 = \sigma^2$ (variance)
  - μ_3 relates to skewness
  - μ_4 relates to kurtosis

Skewness: Measuring Asymmetry

Definition

Skewness measures the asymmetry of a probability distribution around its mean.

- **Population Skewness:**

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E\left[(X-\mu)^3\right]}{\sigma^3}$$

- **Sample Skewness (Pearson's moment coefficient):**

$$g_1 = \frac{m_3}{s^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{s^3}$$

- **Adjusted Sample Skewness:**

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} \cdot g_1$$

Skewness: Measuring Asymmetry

Definition

Skewness measures the asymmetry of a probability distribution around its mean.

- **Population Skewness:**

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E\left[(X-\mu)^3\right]}{\sigma^3}$$

- **Sample Skewness (Pearson's moment coefficient):**

$$g_1 = \frac{m_3}{s^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{s^3}$$

- **Adjusted Sample Skewness:**

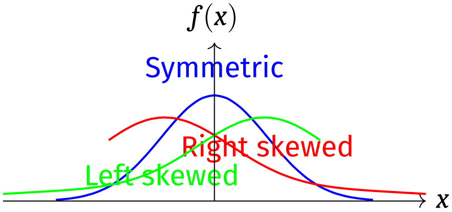
$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} \cdot g_1$$

Interpreting Skewness Values

- $\gamma_1 = 0$: Perfectly symmetric
- $\gamma_1 > 0$: Right-skewed (positive skew)
- $\gamma_1 < 0$: Left-skewed (negative skew)

Rule of thumb:

- $|\gamma_1| < 0.5$: Approximately symmetric
- $0.5 \leq |\gamma_1| < 1$: Moderately skewed
- $|\gamma_1| \geq 1$: Highly skewed



Kurtosis: Measuring Tail Behavior

Definition

Kurtosis measures the ”tailedness” and peakedness of a probability distribution.

- **Population Kurtosis:**

$$\gamma_2 = \frac{\mu_4}{\sigma^4} = \frac{E[(X-\mu)^4]}{\sigma^4}$$

- **Sample Kurtosis:**

$$g_2 = \frac{m_4}{s^4} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4}$$

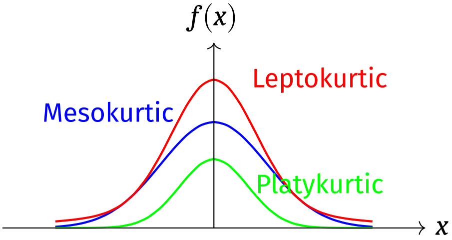
- **Excess Kurtosis:**

$$\text{Excess Kurtosis} = \gamma_2 - 3$$

(Normal distribution has kurtosis = 3, so excess kurtosis = 0)

Types of Kurtosis

- **Mesokurtic:** $\gamma_2 = 3$ (Normal distribution)
- **Leptokurtic:** $\gamma_2 > 3$ (Heavy tails, sharp peak)
- **Platykurtic:** $\gamma_2 < 3$ (Light tails, flat peak)



Excess Kurtosis Interpretation:

- Excess $= 0$: Normal-like tails
- Excess > 0: Heavier tails than normal
- Excess < 0: Lighter tails than normal

Common Families of Distribution: Discrete Distribution

Discrete Uniform Distribution

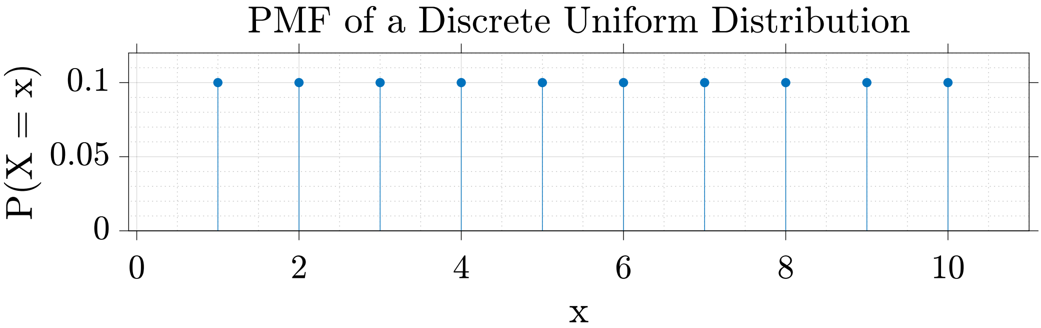
Each distribution is specified by some parameters that is a design time parameter. One of the goal of machine learning algorithms is to estimate those parameters from data.

1. Discrete Uniform

X: possible values $1, 2, 3, \cdots, N$. Here N is the parameter.

$$\begin{aligned} P(X = x|N) &= \frac{1}{N}, \quad x = 1, 2, \cdots, N \\ \mathbb{E}(X) &= \frac{1+N}{2} \\ var(X) &= \frac{(N+1)(N-1)}{12} \end{aligned} \tag{19}$$

Discrete Uniform Distribution



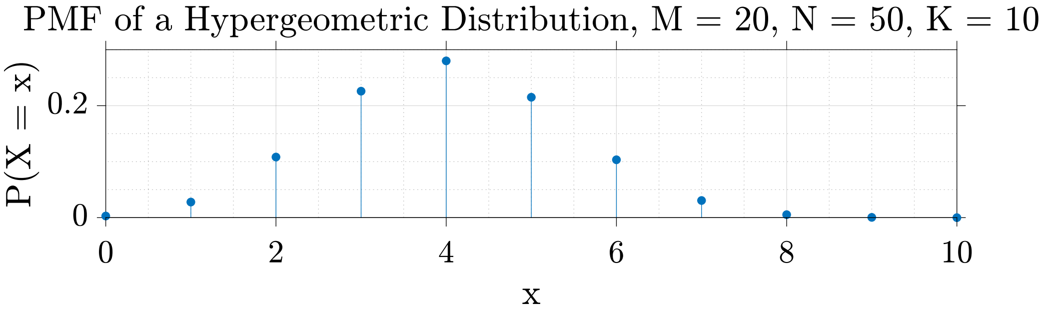
Hypergeometric Distribution

2. Hypergeometric

$$\begin{aligned} P(X = x | M, N, K) &= \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} \\ \mathbb{E}(X) &= \frac{KM}{N} \\ var(X) &= K\frac{M}{N}\frac{(N-M)(M-K)}{N(N-1)} \end{aligned} \tag{20}$$

The following problem exhibits Hypergeometric distribution: there is a large urn filled with N balls, M red, and $N - M$ green balls. Draw K balls at random without replacement.

Hypergeometric Distribution



Binomial Distribution

3. Binomial

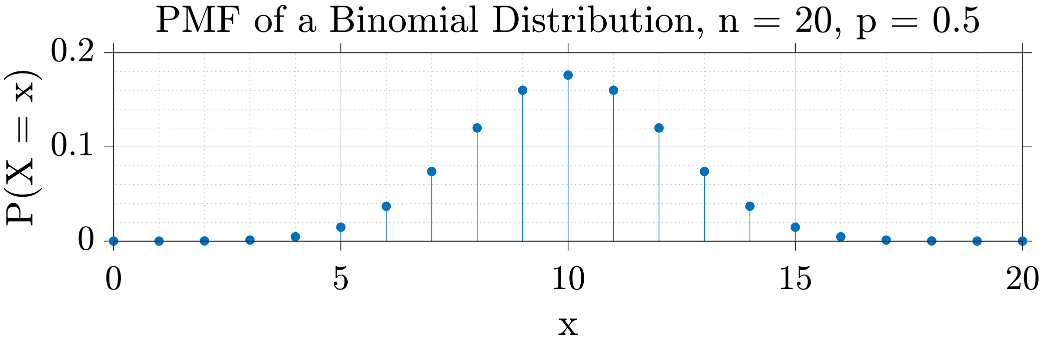
Repeat a random experiment n times that satisfies the following conditions

1. Only two possible outcomes: success and failure.
2. The probability of success p is the same for each trial.
3. The experiments are independent of each other.
4. $X =$ the number of total number of success in n trials.

We call these Bernoulli trials. We can then write as $X \sim Bin(n, p)$ with pmf:

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots n \\ \mathbb{E}(X) &= np \\ var(X) &= np(1-p) \end{aligned} \tag{21}$$

Binomial Distribution



Poisson Distribution

4. Poisson

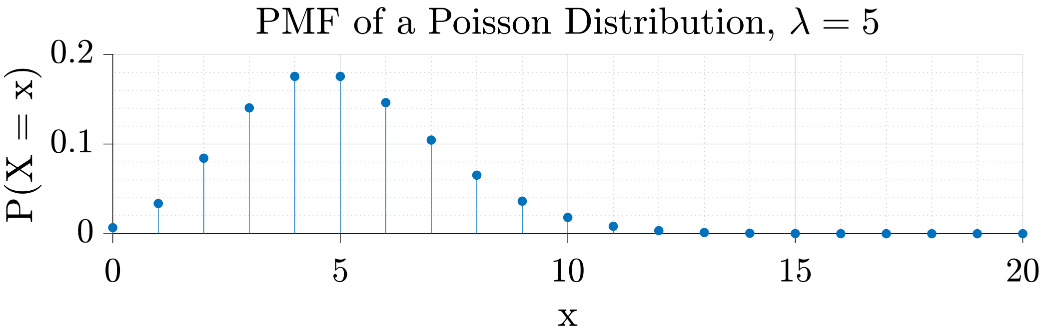
The random variable X takes non-negative integer values such that

$$
\begin{aligned}
P(X = x) &= \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \cdots \\
\mathbb{E}(X) &= \lambda \\
var(X) &= \lambda
\end{aligned}
\tag{22}
$$

Poisson Distribution

Poisson distribution is often used for describing the number of occurrences of a certain event in a very large number of observations, the probability for the event to occur in each observation being very small. Some examples: (i) Nuclear decay of atoms; (ii) Mutation of DNA; (iii) Photon counting by a photodetector.

Poisson Distribution



Common Families of Distribution: Continuous Distribution

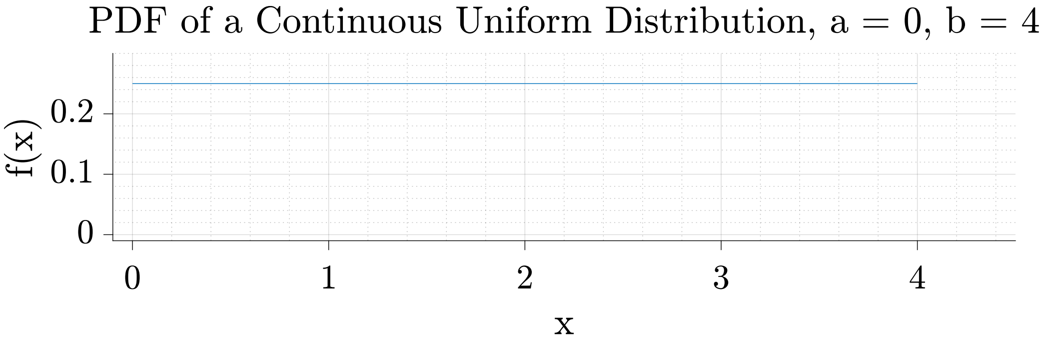
Continuous Uniform Distribution

5. Continuous Uniform

We say $X \sim Unif(a, b)$ if it has PDF

$$
\begin{aligned}
f(x|a, b) &= \frac{1}{b-a}, \quad a < x < b \\
\mathbb{E}(X) &= \frac{1}{a+b} \\
var(X) &= \frac{1}{12}(b-a)^2
\end{aligned}
\tag{23}
$$

Continuous Uniform Distribution



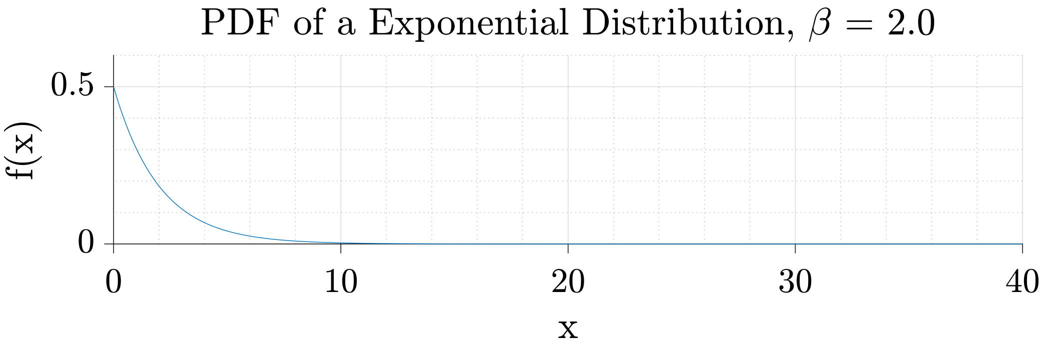
Exponential Family Distribution I

6. Exponential Family

We say $X \sim Exp(\beta)$ if it has PDF

$$\begin{aligned} f(x|\beta) &= \frac{1}{\beta} e^{-x/\beta}, \quad x > 0 \\ \mathbb{E}(X) &= \beta \\ var(X) &= \beta^2 \end{aligned} \tag{24}$$

Exponential Family Distribution



Gamma Family Distribution

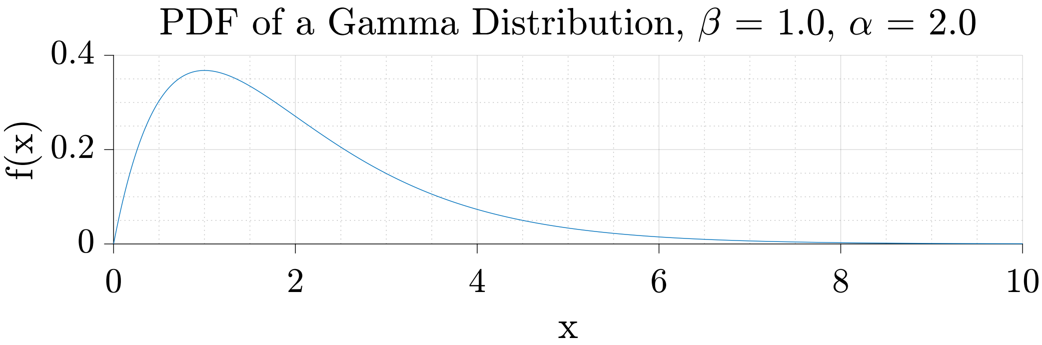
6. Gamma Family

For a random variable X, Gamma distribution is characterized by two parameters: α and β. Its PDF is given by

$$f(x|\alpha, \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} e^{-x/\beta} x^{\alpha-1} \tag{25}$$

where gamma function $\Gamma(\alpha)$ is given by $\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1} dx$. In the gamma distribution, α is a shape parameter that defines the peakedness of the distribution while β is a scale parameter that influences the spread of the distribution. The mean of the gamma distribution, $\mathbb{E}(X) = \alpha\beta$ while the variance $var(X) = \alpha\beta^2$.

Gamma Family Distribution



Weibull Distribution I

7. Weibull

If we have a random variable $X \sim Exp(\beta)$, and another random variable $Y \sim X^{1/\gamma}$, then in this case the Y follows a distribution called as Weibull distribution. The PDF of Weibull distribution is given by

$$f(y|\beta,\gamma) = \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^{\gamma}/\beta}, \quad y > 0 \tag{26}$$

Weibull Distribution I

For the Weibull distribution, the mean and variance are given by

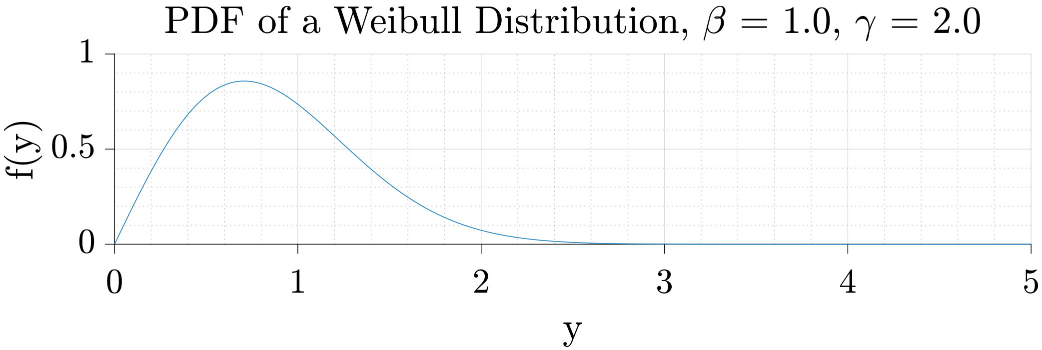
$$\begin{aligned} \mathbb{E}(Y) &= \beta^{1/\gamma}\Gamma(1 + \frac{1}{\gamma}) \\ var(Y) &= \beta^{2/\gamma}\left[\Gamma(1 + \tfrac{2}{\gamma}) - \left(\Gamma(1 + \tfrac{2}{\gamma})\right)^2\right] \end{aligned} \tag{27}$$

Weibull Distribution I

Remark

The difference between the Weibull and Gamma distributions is that in the Gamma distribution, y has a linear term in the exponential while in the Weibull distribution, there is y to the power γ.

Weibull Distribution I



Normal Distribution/Gaussian Distribution I

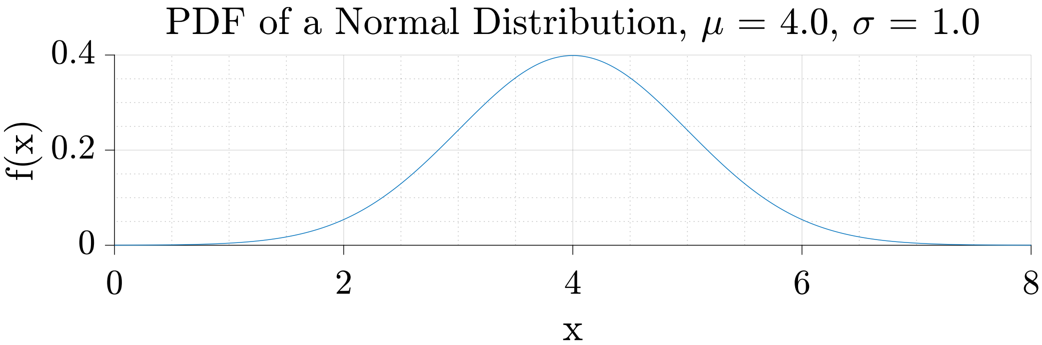
8. Gaussian

Gaussian distribution is the most common distribution everyone is familiar with. For a random variable X following a Gaussian distribution, we write $X \sim N(\mu, \sigma^2)$. Its PDF is given by

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{28}$$

Its expectations and variance are given by $\mathbb{E}(X) = \mu$, and $var(X) = \sigma^2$ respectively. A standard normal distribution has a mean of 0 and a variance of 1 denoted as $X \sim N(0, 1)$.

Normal Distribution/Gaussian Distribution I



Laplace Distribution I

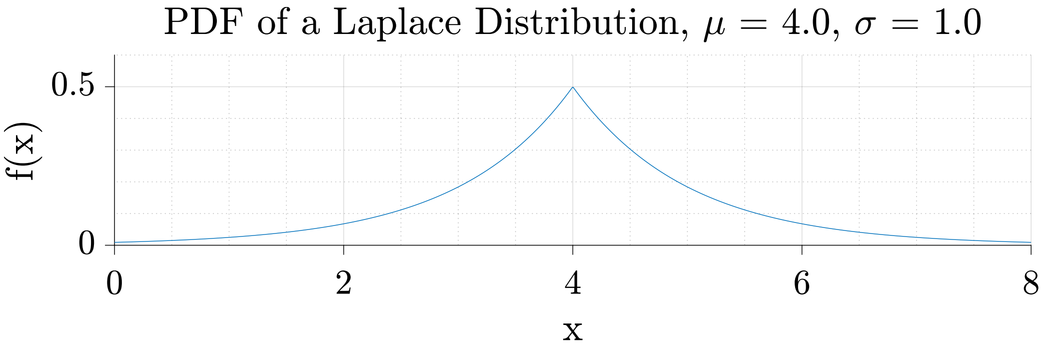
9. Laplace

A random variable X following the Laplace distribution, also known as the double exponential distribution has the following PDF:

$$f(x|\mu, \sigma) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, \quad -\infty < x < \infty \tag{29}$$

Its mean and variance are given by $\mathbb{E}(X) = \mu$, $var(X) = 2\sigma^2$ respectively.

Laplace Distribution I



Synthetic Data Generation from Common Families of Distribution

Generating Data from Discrete Distribution

`scipy.stats.rv_discrete`

Synthetic datasets from distributions discussed above can be generated using `scipy.stats.rv_discrete`. Below is the code snippet that requires Python 3.8 or above followed by a stem plot showing the PMF of each distribution.

Generating Data from Discrete Distribution

rv_discrete is a base class to construct specific distribution classes and instances for discrete random variables. It can also be used to construct an arbitrary distribution defined by a list of support points and corresponding probabilities.
It allows you to sample a random number of that particular distribution you are specifying, thereby generating synthetic datasets that follow the given distribution.

Generating Data from Hypergeometric Distribution

```
# hypergeometric
N = 100
M = 50
K = 10
x2k = np.arange(min(M,K))
y2k = np.zeros(min(M,K))
for i, x in enumerate(x2k):
    y2k[i] = (math.comb(M, x)*math.comb(N-M, K-x))/math.comb(N, K)
# because of numerical round off and how many samples
# we choose, sum of y2k is less than 1. So just normalize it
y2k =y2k/sum(y2k)

# Create a probability Mass Function
pmf2 = rv_discrete(name='hypergeometric', values=(x2k, y2k))
```

Plotting Probability Mass Functions

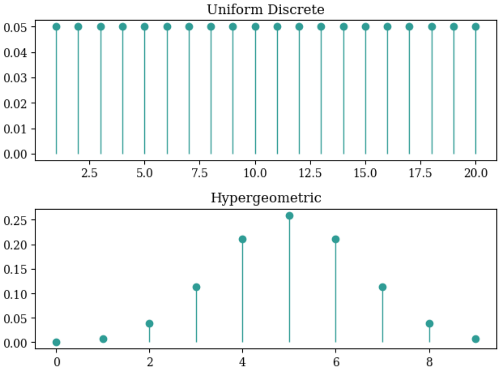
```
import matplotlib.pyplot as plt
fig, ax = plt.subplots(2, 1)
ax=np.ravel(ax)

ax[0].plot(x1k, pmf1.pmf(x1k), 'o', ms=6, mec='#2F9C95', markerfacecolor="#2F9C95")
ax[0].vlines(x1k, 0, pmf1.pmf(x1k), colors='#2F9C95', lw=1)
ax[0].set_title('Uniform Discrete')

ax[1].plot(x2k, pmf2.pmf(x2k), 'o', ms=6, mec='#2F9C95', markerfacecolor="#2F9C95")
ax[1].vlines(x2k, 0, pmf2.pmf(x2k), colors='#2F9C95', lw=1)
ax[1].set_title('Hypergeometric')

plt.tight_layout()
plt.show()
```

Plotting Probability Mass Functions



Generating Data from Continuous Distribution

scipy.stats.rv_continuous

Synthetic datasets from continuous distributions discussed in the lecture can be generated using `scipy.stats.rv_continuous`. Below is the code snippet that requires Python 3.8. Code is followed by a histogram plot of the probability density function of some distributions discussed here.

Generating Data from Exponential Distribution

```
class Exponential(rv_continuous):
    "Exponential"
    def __init__(self, beta, **kwargs):
        super().__init__(**kwargs)
        self.beta = beta
    def _pdf(self, x):
        if(x <=0):
            return 0
        y = (1.0/self.beta)*np.exp(-x/self.beta)
        return y
P2 = Exponential(name='Exponential', beta = 2.0)
# Sample 1000 numbers
B2 = P2.rvs(size = 1000)
```

Generating Data from Gamma Distribution

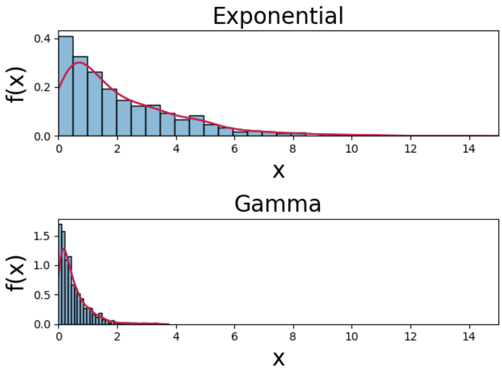
```
class Gamma(rv_continuous):
    "Gamma"
    def __init__(self, alpha, beta, **kwargs):
        super().__init__(**kwargs)
        self.alpha = alpha
        self.beta = beta
    def _pdf(self, x):
        if(x <=0):
            return 0
        y = (1.0/((self.beta**self.alpha)*gamma(self.alpha)))*
                        np.exp(-x/self.beta)*(x**(self.alpha-1))
        return y
P3 = Gamma(name='Gamma', alpha = 1, beta = 0.5)
B3 = P3.rvs(size = 1000)
```

Plotting Probability Density Functions

Plotting PDFs

```
fig, ax = plt.subplots(2, 1)
ax=np.ravel(ax)
a= s.histplot(B2, ax = ax[0], stat = 'density', kde=True)
s.kdeplot(B2, color='crimson', ax=a)
ax[0].set_xlim([0, 15])
ax[0].set_xlabel('x',  fontsize = 20)
ax[0].set_ylabel('f(x)', fontsize =20)
ax[0].set_title('Exponential', fontsize =20)
a = s.histplot(B3, ax = ax[1], stat = 'density', kde=True)
s.kdeplot(B3, color='crimson', ax=a)
ax[1].set_xlim([0, 15])
ax[1].set_xlabel('x',  fontsize = 20)
ax[1].set_ylabel('f(x)', fontsize =20)
ax[1].set_title('Gamma', fontsize =20)
plt.tight_layout()
```

Plotting Probability Density Functions



Python Notebook for Synthetic Data Generation

https://github.com/rahulbhadani/CPE486586_FA25/blob/main/Code/CPE486586_Ch04_SyntheticDataGeneration.ipynb

The End