第七讲
概率论
Lecture 7
Probability

明玉瑞 Yurui Ming

yrming@gmail.com

声明
Disclaimer

- 本讲义在准备过程中由于时间所限，所用材料来源并未规范标示引用来源。所引材料仅用于教学所用，作者无意侵犯原著者之知识版权，所引材料之知识版权均归原著者所有；若原著者介意之，请联系作者更正及删除。

  The time limit during the preparation of these slides incurs the situation that not all the sources of the used materials (texts or images) are properly referenced or clearly manifested. However, all materials in these slides are solely for teaching and the author is with no intention to infringe the copyright bestowed on the original authors or manufacturers. All credits go to corresponding IP holders. Please address the author for any concern for remedy including deletion.

概率
Probability

- 概率是对某个事件发生的可能性的度量，准确的说，事件发生的可能性大小的度量。该值表示期望发生的情况与所有可能发生的情况的总数的比率。如果把事件记为A，我们用$P(A)$来表示该值，显然$P(A) \in [0,1]$。若$P(A) = 0$，则A称为不可能事件；若$P(A) = 1$，则称A为必然事件。

  Probability is the measure of how likely it is some event will occur, or in a more precise way, the amount of how likely an event occurs. It is a value that expresses the ratio of favorable cases to the whole number of cases possible. If we denote the event as A, we use $P(A)$ to represent this value. It is manifest that $P(A) \in [0,1]$, and two extremes are interpreted as the probability of the impossible event: $P(A) = 0$, and the probability of the sure event: $P(A) = 1$.

- 例如，当掷硬币时，正面朝上的概率是 50%。这是因为一枚硬币有两个面，所以硬币正面朝上的概率为 50%，反面朝上的概率为 50%。

  For example, the chance of landing on heads is 50% when toss a coin. This is because a coin has two sides, so there is a 50% chance that the coin will land on heads and 50% that a coin will land on tails.

基本定义
Basic Definitions

- 试验：是指实验或随机过程的单次执行或发生。由于事件的不确定性，只有通过实验，才能观察到不确定事件的结果。简而言之，试验代表实验的一次尝试或迭代，是实验的基本单位。

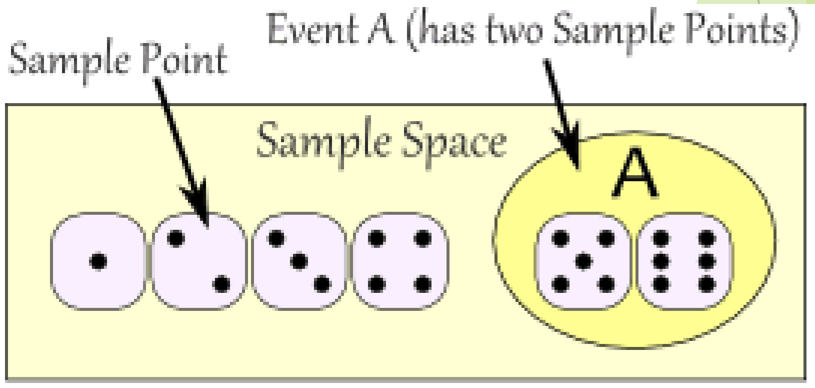
  Trial: it refers to a single performance or occurrence of an experiment or random process. Due to the uncertainty nature of an event, an outcome can only be observed after the trial. In simple terms, a trial represents one attempt or iteration of the experiment and is the basic unit of the experiment.

- 样本空间：统计实验的样本空间是所有可能结果的集合。

  The SAMPLE SPACE of a statistical experiment is the set of all possible outcomes.

- 事件：样本空间的子集。

  An EVENT is a subset of the sample space.



统计实验
Statistical Experiment

- 在上述定义中我们提到了统计实验。统计实验也称可重复实验，是一种可能具有多个不同结果的实验，所有可能出现的结果是预知的，但由于涉及随机因素，将发生的具体结果无法事先确定。抛硬币或从一副牌中随机抽取一张牌都是统计实验。

  We refer to statistical experiment in the above definitions. A statistical experiment is also called a repeatable experiment, is an action or occurrence that can have multiple different outcomes, all of which can be specified in advance, but where the particular outcome that will occur cannot be specified in advance because it depends on random chance. Flipping a coin or choosing a card from a deck at random are both statistical experiments.

- 针对可重复的实验，我们可以理解概率如下：假设我们无限次重复这个实验，特定结果的概率是在无限次试验中，我们获得该结果的次数占总试验次数的比例。

  In the context of a repeatable experiment, one way to think about probability is as follows. Suppose we repeat the experiment an infinite number of times. The probability of a particular outcome is the proportion of times we get that outcome out of our infinite number of trials.

统计实验
Statistical Experiment

- 在上面的定义中，如果我们的实验不可重复怎么办？例如，如果我们想知道明天下雨的概率是多少？这个结果有一定的概率，但无法通过可重复的实验来确定（甚至无法近似确定）。所以，另外一种理解概率的方式是将其视为描述和处理信念程度的方式。如果我说“明天下雨的概率是75%”，这表明我对明天会下雨的信心较强，如果没下雨我会有些惊讶。另一方面，如果我说“明天下雨的概率是2%”，我对不会下雨有非常强烈的信念，如果真下雨我会非常惊讶。这种视角会和我们后面讲到的贝叶斯学派有关联。

  In the above definition, what if our experiment is not repeatable? For example, what if we want to know the probability that it will rain tomorrow? There is certainly some probability associated with that outcome, but no way to determine it (or even approximate it) by a repeatable experiment. So, another way to think about probability is as a way of talking about and mathematically working with degrees of belief. If I say “the probability of rain tomorrow is 75%”, that is a way of expressing the fact that my belief in rain tomorrow is fairly strong, and I would be somewhat surprised if it did not rain. On the other hand, if I say “the probability of rain tomorrow is 2%”, I have a very strong belief that it will not rain, and I would be very surprised if it did. This viewpoint is associated with the Bayesians we will talk about later.

基本运算
Basic Operations

- 令A、B为样本空间$\mathcal{S}$中两个事件，则我们有：
  - $\emptyset$：表示不可能事件；
  - $\bar{A}$：表示A不发生；
  - $A\cup B$：表示A或B至少有一发生；
  - A,B：表示A和B同时发生；
  - $A\backslash B$：表示A发生且B不发生；
- Let A and B be two events of the sample space $\mathcal{S}$. We will denote:
  - "the impossible event" by $\emptyset$;
  - "A does not occur" by $\bar{A}$;
  - "either A or B occur" by $A\cup B$;
  - "both A and B occur" by A,B;
  - "A occurs and B does not" by $A\backslash B$.
- 如果$A\cup B=\mathcal{S}$，则称事件A和事件B是互为穷集的，换句话说，我们确信A或B中的一个必然会发生。特别地，$A\cup\bar{A}=\mathcal{S}$。如果如果$A,B=\emptyset$，则称事件A和事件B是互斥的，在这种情况下，我们确定如果A发生，则B不可能发生。显然，我们有$A,\bar{A}=\emptyset$。

  The events A and B are exhaustive if $A\cup B=\mathcal{S}$, in other words we are sure that either A or B will occur. Thus, in particular $A\cup\bar{A}=\mathcal{S}$. The events A and B are exclusive if $A,B=\emptyset$. In this case, we are sure that if A occurs then B cannot. Clearly, we have $A,\bar{A}=\emptyset$.

基本定义
Basic Definitions

- 现在我们考虑一个事件的概率。根据定义，不可能事件的概率为零，而必然事件的概率为一。更有意义的情况是既非不可能事件，也非必然事件的概率。如果我们假设样本空间$\mathcal{S}$中的所有结果是等可能发生的，那么一个事件E的概率，我们用$P(E)$来表示，就是事件E中结果的数量除以样本空间$\mathcal{S}$中结果的数量，即一个事件的概率是样本空间中既是该事件结果的数量在样本空间结果数量中所占的比例，如下式所示：

  Let’s consider the probability of an event. By definition, an impossible event has probability zero, and a certain event has probability one. The more meaningful cases are events that are neither impossible nor certain. For the moment, let’s assume that all outcomes in the sample space $\mathcal{S}$ are equally likely. If that is the case, then the probability of an event E, which we write as $P(E)$, is simply the number of outcomes in E divided by number of outcomes in $\mathcal{S}$, that is, the probability of an event is the proportion of outcomes in the sample space that are also outcomes in that event, as shown below:

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

概率定义
Definition of Probability

- 上面的对概率的定义更像一个直观的描述，实际上，当推广到更抽象的空间时，可以引出以集合子集的度量为基础，概率的公理化定义。该定义要求集合包含空集，且在补集和可数并集的情况下是封闭的（即$\bigcup_{i=1}^{\infty} A_i \in \mathcal{S}$）：

  The above definition of probability is more like an intuitive description. Actually, when generalized to more abstract spaces, it leads to an axiomatic definition of probability in terms of measure over a collection of subsets. This collection is assumed to contain the empty set, and to be closed under the complementation and countable union (i.e. $\bigcup_{i=1}^{\infty} A_i \in \mathcal{S}$) :

- 定理：令$\mathcal{S}$表示样本空间，定义在$\mathcal{S}$上的集合函数$P(\cdot)$若满足下面条件，则称为概率函数：
  - 对$\mathcal{S}$中任一事件A，有$P(A) \geq 0$;
  - $P(\mathcal{S}) = 1$;
  - 如果A_1，A_2，…，两两互斥，即对任意的i, j，$A_i, A_j = \emptyset$，则$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

- Let $\mathcal{S}$ denote the sample space. A set function $P(\cdot)$ defined in $\mathcal{S}$ is a probability function if:
  - For any event A in $\mathcal{S}$, then $P(A) \geq 0$;
  - $P(\mathcal{S}) = 1$;
  - If A_1, A_2, . . . are exclusive events in $\mathcal{S}$ and hence $A_i, A_j = \emptyset$ for all i, j, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

概率函数
Probability Function

- 令$P(\cdot)$表示样本空间$\mathcal{S}$上的集合函数$P(\cdot)$，则$P(\cdot)$满足下面性质：
  - $P(\emptyset)=0$；
  - $P(\cdot)$满足可数可加性：令A_1，A_2，…，A_n为$\mathcal{S}$中的事件序列，满足对所有的$i\neq j$，有$A_i, A_j=\emptyset$，则$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)$。若这些事件是$\mathcal{S}$的一个划分，即$\bigcup_{i=1}^{n}A_i=\mathcal{S}$，则我们有$P(\bigcup_{i=1}^{n}A_i)=1$；
  - 注意到$P(A\cup\bar{A})=P(A)+P(\bar{A})=1$，则对任意的$A\in\mathcal{S}$，有$P(A\cup\bar{A})\leq 1$；
  - 若$A\subset B$，则$P(B\backslash A)=P(B)-P(A)$；
  - 一般的，我们有$P(A\cup B)=P(A)+P(B)-P(A,B)$。

- Let $P(\cdot)$ be a probability function defined over the sample space $\mathcal{S}$. Then $P(\cdot)$ satisfies the following properties:
  - $P(\emptyset)=0$;
  - $P(\cdot)$ is finitely additive: if A_1, A_2, …, A_n are events in $\mathcal{S}$, such that $A_i, A_j=\emptyset$ for all $i\neq j$, then $P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)$. If these events form a partition of $\mathcal{S}$, i.e. they are such that $\bigcup_{i=1}^{n}A_i=\mathcal{S}$, then $P(\bigcup_{i=1}^{n}A_i)=1$;
  - $P(A\cup\bar{A})=P(A)+P(\bar{A})=1$, so that $P(A)\leq 1$ for any A in $\mathcal{S}$;
  - If $A\subset B$ then $P(B\backslash A)=P(B)-P(A)$;
  - Generally, we have $P(A\cup B)=P(A)+P(B)-P(A,B)$.

条件概率
Conditional Probability

- 到目前为止，所考虑的概率都是无条件概率。然而，在某些情况下，我们可能对在某个事件发生的条件下另一个事件发生的概率感兴趣。这在形式上是通过考虑事件B在事件A发生的条件下的条件概率来完成的。这个事件用$B|A$表示。

  The probabilities considered so far are unconditional probabilities. In some situations, however, we may be interested in the probability of an event given the occurrence of some other event. This is formally done by considering the conditional probability of B given that A occurs. This event is denoted by $B|A$.

- 条件概率：令A与B为样本空间$\mathcal{S}$中的事件，且$P(A) > 0$。则在事件A发生的情况下事件B发生的条件概率由下面式子定义。在该定义中，有时为了强调$P(A)$为无条件概率，有时又称之为边缘概率。

  Conditional Probability: Let A and B events in $\mathcal{S}$, and suppose that $P(A) > 0$. The conditional probability of B given A is as follows. Notably, to emphasize that $P(A)$ is unconditional, $P(A)$ is called marginal probability.

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

条件概率

Conditional Probability

- 从条件概率的定义中，我们可以推导出两个事件的交集的概率，即它们的联合概率，用条件概率和边际概率表示：$P(A,B)=P(B|A)P(A)$。这个规则可以应用于更多的事件，并产生乘法规则或分解规则。

  From the definition of conditional probability, we can derive the probability of the intersection of two events, called their joint probability in terms of conditional and marginal probabilities: p $P(A,B) = P(B|A)P(A)$. This rule can be applied to a larger number of events and produces the multiplication rule or factorization rule.

- 乘法法则：事件$A_1, A_2, \ldots, A_n$的联合分布可以表示为：

  (Multiplication Rule) The joint probability of a set of events $A_1, A_2, \ldots, A_n$ can be expressed as:

$$P(A_1, A_2, \cdots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)\cdots P(A_n|A_1, A_2, \cdots, A_{n-1})$$

- 考虑事件A和B。由于事件A和$\bar{A}$构成了$\mathcal{S}$的一个划分，我们可以将事件B分解为两个互斥事件A, B和$\bar{A}, B$的并集，因此我们有以下全概率公式：

  Consider the events A and B. Since the events A and $\bar{A}$ form a partition of $\mathcal{S}$, so that we can decompose B into the union of the two exclusive events A, B and $\bar{A}, B$, so we have the following Total Probability Formula:

$$P(B) = P(A,B) + P(\bar{A},B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$$

采样
Sampling

- 当样本空间的所有元素具有相同的概率时，计算概率通常归结为计数某种选择方式的数量。具体来说，我们通常对从n个元素或样本的集合中选择r个事物的方法感兴趣。这被称为从n个元素的集合中抽样或采样。这个问题的答案取决于我们对选择的具体含义：顺序是否重要，是否允许重复。

  When all elements of the sample space are equally likely, calculating probability often boils down to counting the number of ways of making some selection. Specially, we are often interested in finding how many ways there are of choosing r things from an n element set. This is called sampling from an n element set. The answer to this question depends on exactly what we mean by selection: is the order important and is repetition allowed.

- 采样或抽样的根本目的是进行概率的计算。当我们对一些概率模型或随机过程有很好的了解时，我们可以通过这种形式的计算来得到概率值。否则，特别是当实验可重复时，我们不得不进行实际的试验，从试验的结果来推断概率值。而试验及产生结果的过程，我们也称采样。

  The purpose of sampling is to perform probability calculations. When we have a good understanding of the probability models or random processes, we can obtain the probability values via direct calculation. Otherwise, especially when the experiment is a repeatable one, we have to conduct actual experiments to infer the probability values from the results of the trials. The process of conducting experiments and observing the outcomes is also called sampling.

采样
Sampling

- 从具有n个元素的集合中，或者说n个基本事件的样本空间中，选择（或采样）出r个样本的方式总共有：
  - 按顺序序取且允许放回与重复：n^r；
  - 按顺序序取且不允许放回与重复：$\frac{n!}{(n-r)!}$；
  - 不按顺序序取且不允许放回与重复：$\binom{n}{r}$。
- The number of ways of selecting (sampling) r things from an n element set is:
  - ordered with replacement/repetition allowed: n^r;
  - ordered without replacement/no repetition: $\frac{n!}{(n-r)!}$;
  - unordered without replacement/no repetition: $\binom{n}{r}$.
- 在回答涉及采样的问题时，仔细阅读问题并确定所涉及的采样类型非常重要。具体来说，需要考虑选择了多少个事物，从哪个集合进行选择，顺序是否重要，是否允许重复。有时候可能会有多种方法可以使用，但必须保持一致。

  It is important when answering questions involving sampling that you read the question carefully and decide what sort of sampling is involved. Specifically, how many things are you selecting, what set are you selecting from, does the order matter, and is repetition allowed or not. Sometimes more than one sort can be used but you must be consistent.

预测与推理
Prediction and Inference

- 在上面对概率的基本概念介绍的基础上，我们来概括概率的基本用途。概括地讲，概率概率可以用于以下两类问题：

  Based on the introduction of probability above, let's summarize the mainly usage of it. Broadly speaking, probabilities can be used for the following two types of problems:

- 预测：从原因到结果的推理。假设已知一组原因及它们之间的相互作用方式，那么可能/不可能的结果是什么？许多早期发展概率论的工作是为了回答此类问题，以便能够预测某些事件发生的频率。

  Prediction: reasoning from causes to effects. Given a known set of causes and knowledge about how they interact, what are likely/unlikely outcomes? Much of the early work developing probability theory was motivated by answering questions like these in order to be able to predict how often certain events would occur.

- 推断：从结果到原因的推理。假设已知可能的原因及它们之间的相互作用方式，并且观察到一些结果，那么哪些原因是可能/不可能的？例如，观察一幅图像的特征，以确定它是一只猫还是一只狗。

  Inference: reasoning from effects to causes. Given knowledge about possible causes and how they interact, as well as some observed outcomes, which causes are likely/unlikely? For examples, observe features of an image, to determine if it is a cat or dog.

预测与推理
Prediction and Inference

- 我们说概率论主要用于预测和推理方面。在推理中，我们可能会使用到最重要的工具是贝叶斯定理，也称为贝叶斯规则。通过用$P(A|B)P(B)$（使用乘法规则）替代条件概率的定义$P(A,B)$，我们得到了贝叶斯规则：

  We say that probability theory is useful for prediction and inference. Perhaps the most important tool we'll use for inference is BAYES' RULE, also known as BAYES' THEOREM. By replacing the $P(A,B)$ in the definition of conditional probability with$P(A|B)P(B)$ (using the product rule), we derive the Bayes' Rule:

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

- 通常，对于推理问题，我们需要通过关于可以直接观察到的事物（例如测试结果）和两者之间的概率关系的信息来推断无法直接观察到的事物。

  Usually, for an inference problem, we need to infer something that can't be observed directly using information about something that can be observed directly (a test result) and the probabilistic relationship between the two.

独立性
Independence

- 独立性和条件独立性：当我们应用贝叶斯定理时，我们假设事件A的发生会改变事件B的概率。如果另一方面，事件A的发生不会改变事件B的概率，那么事件B和A是独立事件。由定义$P(B) = P(B|A) = P(A,B)/P(A)$，最后我们推得$P(A)P(B) = P(A,B)$。

  Independence and Conditional Independence: When we apply Bayes' Theorem, we assume that the occurrence of an event A changes the probability of B. If, on the other hand, occurrence of A does not change the probability of B, then A and B are independent events. Thus, by definition we have $P(B) = P(B|A) = P(A,B)/P(A)$, from which we derive that $P(A)P(B) = P(A,B)$.

- 独立性：对两个事件A与B，且$P(A) > 0$，$P(B) > 0$，若$P(A)P(B) = P(A,B)$，则称事件A与B相对独立，即两个独立事件的联合概率可分解为边缘概率的乘积。若对事件A_1，A_2，…，A_n，如果对所有可能的组合都满足$P(A_i)P(A_j) = P(A_i, A_j)$，则称它们两两独立。

  Independence: Two events A and B, with $P(A) > 0$ and $P(B) > 0$ are independent if $P(A)P(B) = P(A,B)$. In words, the joint probability of two independent events factorizes into the product of the marginal probabilities. The events A_1, A_2, ..., A_n are called pair-wise independent, if $P(A_i)P(A_j) = P(A_i, A_j)$ for all possible pairs.

独立性
Independence

- 进一步地，当联合概率的因式分解规则适用于所有事件子集的交集时，A_1，A_2，...，A_n被称为互相独立。两两独立通常并不意味着互相独立。例如，假设A_1、A_2、A_3是两两独立的，$P(A_i) > 0$对所有i成立，并且A_1、A_2、A_3的交集为空集，则$P(A_1, A_2, A_3) = 0$，因而因式分解规则不适用。

  In addition, when the factorization rule applies for all intersection of subsets of events, then A_1, A_2, ..., A_n are called mutually independent. Pair-wise independence does not generally imply mutual independence. For example, suppose that A_1, A_2, A_3 are pair-wise independent, $P(A_i) > 0$ for all i and $A_1, A_2, A_3 = \emptyset$, then $P(A_1, A_2, A_3) = 0$ and the factorization rule does not apply.

- 条件独立性：对于两个事件A和B，假设$P(A) > 0$且$P(B) > 0$，在给定事件C的条件下，如果$P(A, B|C) = P(A|C)P(B|C)$，那么它们是在事件C给定下独立的。独立并不意味着条件独立，因此可能出现这样的情况，两个事件A和B在给定事件C的条件下是独立的，但是当单独考虑它们时它们不是独立的。

  Conditional Independence: Two events A and B, with $P(A) > 0$ and $P(B) > 0$ are independent given C, if $P(A, B|C) = P(A|C)P(B|C)$. Independence does not imply conditional independence, so that it may happen that two events A and B are independent given C, but they are not independent when considered alone.

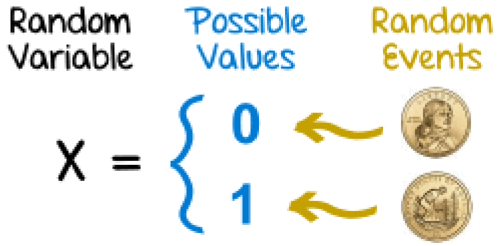
随机变量
Random Variable

- 随机变量（或RV）是代表样本空间或其某个分区中可能的事件的变量。换句话说，随机变量的可能的取值是样本空间或分区中的事件，并且这些值覆盖了样本空间或分区中的所有事件。我们将用大写字母表示随机变量，用小写字母或数字表示可能的取值。

  A RANDOM VARIABLE (or RV) is a variable that represents the possible events in the sample space or some partition of it. Put another way, an RV has possible values of each being one event in the sample space or a partition, and where the values cover all events in the sample space or the partition. We will write random variables with uppercase letters, and their possible values with lowercase letters or numbers.

- 随机变量本身的取值并不是重点，重点是取特定值的可能性的数值度量，即概率。

  The key is not the rule of assigning values to the variable itself, it is the values measure the values upon experiment.



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

概率分布
Probability Distribution

- 因此，有人采用以下的定义作为随机变量的定义：随机变量x是一种由机会决定的变量，它对实验的每个结果都有一个单一的数值。

  Therefore, there is also an adaptation of the following definition as RV: A random variable X is a variable that has a single numerical value，determined by chance，for each outcome of an experiment.

- 事件的概率表示了一次试验中某一个结果发生的可能性大小。若要全面了解试验，则必须知道试验的全部可能结果及各种可能结果发生的概率，即随机试验的概率分布。随机试验的概率分布就是随机变量X的概率分布，即随机变量的可能取值及取得对应值的概率。

  The probability of an event represents the likelihood of a particular outcome occurring in a single experiment. To fully understand the experiment, one must know all the possible outcomes and the probabilities of each outcome occurring, which is referred to as the probability distribution of the random experiment. The probability distribution of the random experiment is the probability distribution of the random variable X, which includes all possible values of the random variable and the corresponding probabilities of obtaining those values.

离散型随机变量

Discrete R.V.

- 随机变量分为离散型随机变量与连续型随机变量。根据随机变量所属类型的不同，概率分布取不同的表现形式。

  Random variables are categorized into discreate random variable and continuous random variable. Depending on the type of random variable, the probability distribution takes different forms.

- 取有限（或可数）个状态的随机变量被称为离散随机变量。对于离散型变量X，假设其可取值为x_1，x_2，x_3，…，且X取x_k的概率为$P(X=x_k)=f(x_k)$，则一般地，我们称满足如下条件的$f(x)\overset{\text{def}}{=}P(X=x)$为概率函数或概率分布：

  Random variables that take a finite (or countable) number of states are referred to as discrete random variables. For a discrete random variable X, suppose that the possible values that it can assume are given by x_1, x_2, x_3 , …, and with probabilities given by $P(X=x_k)=f(x_k)$. Then $f(x)$ given by $f(x)\overset{\text{def}}{=}P(X=x)$ is which satisfies the following conditions is called the probability function, also referred to as probability distribution:

$$f(x)\geq 0$$

$$\sum_x f(x)=1$$

概率分布
Probability Distribution

- 对于离散型随机变量，我们求得概率分布的目的是为了得到该随机变量在某一离散值的概率。下表是常见的离散型概率分布的情况：

- For discrete random variables, the purpose of obtaining the probability distribution is to get the probability of the random variable taking on a specific discrete value. The following table shows the commonly used discrete probability distribution:

Distribution	$f(x)$	Support	Mean	Variance
Bernoulli (p)	$p^x(1-p)^{1-x}$	$0, 1$	p	$p(1-p)$
binomial(n, p)	$\binom{n}{x} p^x(1-p)^{n-x}$	$0, 1, \ldots, n$	np	$np(1-p)$
geometric(p)	$p(1-p)^x$	$0, 1, 2, \ldots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
neg. binom.(r, p)	$\binom{x+r-1}{r-1} p^r(1-p)^x$	$0, 1, 2, \ldots$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Poisson(λ)	$\frac{\lambda^x e^{-\lambda}}{x!}$	$0, 1, 2, \ldots$	λ	λ
hypergeom.(m, N, n)	$\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$0, 1, \ldots, n$	$\frac{nm}{N}$	$\frac{nm(N-n)(N-m)}{N^2(N-1)}$

累积分布函数
Cumulative Distribution Function

- 累积分布函数（c.d.f）：设X是一个离散随机变量，其累积分布函数被定义为$F_X(x) = P(X \le x)$，对于所有$x \in (-\infty, -\infty)$。

  Cumulative Distribution Function (c.d.f): Let X be a discrete random variable. The cumulative distribution function is defined as $F_X(x) = P(X \le x)$, for all $x \in (-\infty, -\infty)$.

- 实际上，任何累计分布函数必满足下面性质，满足下面性质的函数，必然可作为累计分布函数：
  1. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$;
  2. $F(x)$ is a non-decreasing function of x;
  3. $F(x)$ is right-continuous, that is the right-limit $\lim_{x\downarrow x_i} F(x) = F(x_i)$.

- Actually, any cumulative distribution function satisfies the following properties. And any function satisfying the the following properties is eligible for an cumulative distribution function:
  1. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$;
  2. $F(x)$ is a non-decreasing function of x;
  3. $F(x)$ is right-continuous, that is the right-limit $\lim_{x\downarrow x_i} F(x) = F(x_i)$.

连续型随机变量
Continuous R.V.

- 随机变量X在$\mathcal{R}$及其子集上连续取值时，称为连续型随机变量。

  If RV X assumes value from $\mathcal{R}$ or its subset in a continuous way, then is called a continuous RV.

- 对于连续型随机变量X，由于没有办法得知其在每一点的概率，我们比较关心其在一段区间如$[a,b]$上的概率。这可通过累计分布函数表示：$P([a,b]) = F_X(a \leq x \leq b)$。实际上，我们对满足如下性质的$f(x)$更为关心，而通常将其称为概率密度函数（pdf）：

  For a continuous random variable X, as there is no way to know its probability at each point, we are more concerned about its probability within an interval such as $[a,b]$. This can be expressed using the cumulative distribution function: $P([a,b]) = F_X(a \leq x \leq b)$. In fact, we are more interested in $f(x)$ that satisfies the following properties, which is usually referred to as the probability density function (pdf):

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

$$f(x) \geq 0, x \in (-\infty, +\infty)$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

正态分布
Normal Distribution

- 正态分布：在区间$(-\infty,+\infty)$上定义的连续型随机变量X，若其具有如下概率密度函数，则称X为具有正态分布或高斯分布的随机变量，其中参数μ和σ^2分别代表分布的对称点和分布关于μ的“散布”。当$\mu=1$且$\sigma^2=1$时，该分布被称为标准正态分布。正态分布具有很多良好的性质，经常用于模拟自然发生的现象，是统计学、概率论和各种科学学科中的基本工具。

  Normal Distribution: A continuous random variable X defined in $(-\infty,+\infty)$ and with p.d.f. as below, then it is said to have a Normal distribution — or Gaussian distribution — with parameters μ and σ^2. The parameters μ and σ^2 identify, respectively, the point of symmetry of the distribution — the p.d.f. is symmetrical about μ — and the “spread” of the p.d.f. about μ. When $\mu=1$ and $\sigma^2=1$ the distribution is called Normal Standard. The normal distribution has several nice features and can model naturally occurring phenomena, which makes it a fundamental tool in statistics, probability theory, and various scientific disciplines.

- 简而言之，概率衡量了一个事件出现的可能性；概率分布刻画了所有可能出现的事件与其出现可能性大小的关系

  Probability calculates the likelihood of one event; Probability distribution depicts the relationship between each possible outcome for a random variable and their probabilities.

means and variances. In the Table, $\Gamma(k)$ is the gamma function, equal to $(k-1)!$ when k is an integer.

概率密度函数 Probability Density Functions

- 下表给出了常见的连续型随机变量的概率密度函数：

  The table below shows the probability density functions of common continuous random variables:

Distribution	Probability density function f (variable; parameters)	Characteristic function $\phi(u)$	Mean	Variance σ^2
Uniform	$f(x;a,b) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{ibu}-e^{iau}}{(b-a)iu}$	$\overline{x} = \dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$
Binomial	$f(r;n,p) = \dfrac{n!}{r!(n-r)!}\, p^r q^{n-r}$ $r = 0,1,2,\ldots,n\,; \quad 0 \le p \le 1\,; \quad q = 1-p$	$(q+pe^{iu})^n$	$\overline{r} = np$	npq
Poisson	$f(r;\mu) = \dfrac{\mu^r e^{-\mu}}{r!}\,; \quad r = 0,1,2,\ldots\,; \quad \mu > 0$	$\exp[\mu(e^{iu}-1)]$	$\overline{r} = \mu$	μ
Normal (Gaussian)	$f(x;\mu,\sigma^2) = \dfrac{1}{\sigma\sqrt{2\pi}}\exp(-(x-\mu)^2/2\sigma^2)$ $-\infty < x < \infty\,; \quad -\infty < \mu < \infty\,; \quad \sigma > 0$	$\exp(i\mu u - \frac{1}{2}\sigma^2 u^2)$	$\overline{x} = \mu$	σ^2
Multivariate Gaussian	$f(\boldsymbol{x};\boldsymbol{\mu},S) = \dfrac{1}{(2\pi)^{n/2}\sqrt{\lvert S\rvert}}$ $\times \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T S^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$ $-\infty < x_j < \infty; \quad -\infty < \mu_j < \infty; \quad \det S > 0$	$\exp\left[i\boldsymbol{\mu}\cdot\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^T S\boldsymbol{u}\right]$	$\boldsymbol{\mu}$	S_{jk}
χ^2	$f(z;n) = \dfrac{z^{n/2-1}e^{-z/2}}{2^{n/2}\Gamma(n/2)}\,; \quad z \ge 0$	$(1-2iu)^{-n/2}$	$\overline{z} = n$	$2n$
Student's t	$f(t;n) = \dfrac{1}{\sqrt{n\pi}}\dfrac{\Gamma[(n+1)/2]}{\Gamma(n/2)}\left(1+\dfrac{t^2}{n}\right)^{-(n+1)/2}$ $-\infty < t < \infty\,; \quad n$ not required to be integer	—	$\overline{t} = 0$ for $n \ge 2$	$n/(n-2)$ for $n \ge 3$
Gamma	$f(x;\lambda,k) = \dfrac{x^{k-1}\lambda^k e^{-\lambda x}}{\Gamma(k)}\,; \quad 0 < x < \infty\,;$ k not required to be integer	$(1-iu/\lambda)^{-k}$	$\overline{x} = k/\lambda$	k/λ^2