生物信息中的统计模型 (2015年春)

第1章 概率论基础

G. Casella “Statistical Inference”

Chapter 1-3

感谢清华大学自动化系江瑞教授提供PPT

Classical Probabilities

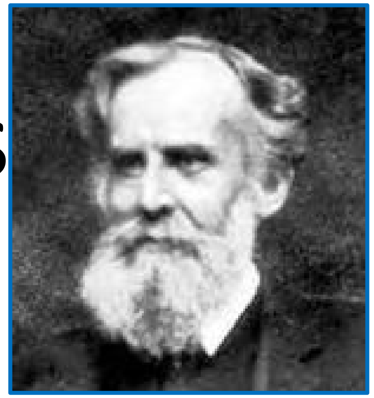
Pierre-Simon Laplace

Théorie analytique des probabilités

The probability of an event is **the ratio of the number of cases favorable to it, to the number of all cases possible** when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, **equally possible**.

http://en.wikipedia.org/wiki/Classical_definition_of_probability

Frequency Probabilities



John Venn

Frequency probabilities

Probabilities are related to well-defined **random experiments**. The set of all possible outcomes of a random experiment is called the **sample space** of the experiment. An **event** is defined as a particular subset of the sample space. The relative frequency of occurrence of an event, in a number of repetitions of the experiment, is a measure of the **probability** of that event.

http://en.wikipedia.org/wiki/Frequency_probability

Subjective Probabilities

Thomas Bayes

Bayesian probability

Probability is the degree to which a person (or community) believes that a proposition is true, **the degree of belief**.

http://en.wikipedia.org/wiki/Bayesian_probability

Axiomatic Definition

Kolmogorov

Probability axioms

Given a sample space $\mathcal{S}$ and an associated sigma algbra $\mathcal{B}$, a **probability function** is a function with domain $\mathcal{B}$ that satisfies

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$;
2. $P(\mathcal{S}) = 1$;
3. If $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint, then
$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

http://en.wikipedia.org/wiki/Kolmogorov_axioms

Random Experiments（随机试验）

Random experiments

A **random experiment** is an experiment for which the outcome cannot be predicted with certainty. The term "random experiment" is often simplified as "experiment."

- 随机试验在相同的条件下可以重复进行
- 随机试验的所有可能结果能够事先明确地指出来
- 某一次随机试验的结果不能在试验进行之前预料到

Sample Space（样本空间）

Sample space

The set, $\mathcal{S}$, of all possible outcomes of a particular random experiment is called the **sample space** for the experiment.

– 有限可列 (Finite countable)
– 无限可列 (Infinite countable)
– 无限不可列 (Infinite uncountable)

Examples of random experiments

- 随机试验
  - 掷一只骰子，观察朝上一面的点数
  - 在一批产品中，任取一件，观察是正品还是次品
  - 射击一目标，直到击中为止，记录射击的次数
  - 从一批灯泡中，任取一只，测其寿命

- 样本空间
  - 掷骰子试验（有限可列）： $\mathcal{S} = \{1,2,3,4,5,6\}$
  - 取一件产品（有限可列）： $\mathcal{S} = \{\text{正品，次品}\}$
  - 射击目标试验（无限可列）： $\mathcal{S} = \{1,2,3,\ldots\}$
  - 灯泡寿命试验（无限不可列）： $\mathcal{S} = \{t|t\geq 0\}$

Random Event （随机事件）

Event

An (random) **event** is any collection of possible outcomes of an experiment, that is, any subset of $\mathcal{S}$ (including $\mathcal{S}$ itself).

– 基本事件 vs. 复合事件
– 必然事件 vs. 不可能事件

Event Operations

- 包含
  Containment
  $$A \subset B \Leftrightarrow x \in A \Rightarrow x \in B$$
  $$A = B \Leftrightarrow A \subset B \text{ and } B \subset A$$
- 合集
  Union
  $$A \cup B = \{x : x \in A \text{ or } x \in B\}$$
- 交集
  Intersection
  $$A \cap B = \{x : x \in A \text{ and } x \in B\}$$
- 补集
  Complementation
  $$A^c = \{x : x \notin A\}$$
- 差集
  Theoretic difference
  $$A - B = \{x : x \in A \text{ and } x \notin B\}$$

Extension of Event Operations

Countable infinite collection of sets

$$\bigcup_{i=1}^{\infty} A_i = \{x \in \mathcal{S} : x \in A_i \text{ for some } i\}$$

$$\bigcap_{i=1}^{\infty} A_i = \{x \in \mathcal{S} : x \in A_i \text{ for all } i\}$$

Uncountable infinite collection of sets

$$\bigcup_{a \in \Gamma} A_a = \{x \in \mathcal{S} : x \in A_a \text{ for some } a\}$$

$$\bigcap_{a \in \Gamma} A_a = \{x \in \mathcal{S} : x \in A_a \text{ for all } a\}$$

$\Gamma :$ All possible real numbers. $A_a : (0, a]$.

Sigma Algebra

sigma algebra (Borel field)

A collection of subsets of $\mathcal{S}$ is called a **sigma algebra**, denoted by $\mathcal{B}$, if it satisfies the following three properties:

1. $\emptyset \in \mathcal{B}$
   (the empty set is an element of $\mathcal{B}$);
2. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$
   ($\mathcal{B}$ is closed under complementation);
3. If $A_1, A_2, \ldots \in \mathcal{B}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$
   ($\mathcal{B}$ is closed under countable unions).

Examples of Sigma Algebra-I

- $\mathcal{B}_1 = \{\emptyset \,, \mathcal{S}\}$ (the trivial sigma algebra)
  - $\emptyset \in \mathcal{B}_1$
  - $\mathcal{B}_1$ is closed under complementation
  - $\mathcal{B}_1$ is closed under countable unions

Examples of sigma algebra-II

- $\mathcal{B}_2 = \{\text{all subsets of } \mathcal{S}, \text{ including } \mathcal{S} \text{ itself}\}$
  - $\emptyset \in \mathcal{B}_2$
  - $\mathcal{B}_2$ is closed under complementation
  - $\mathcal{B}_2$ is closed under countable unions
- Example
  - $\mathcal{B} = \{ \emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$ for $\mathcal{S} = \{1,2,3\}$

Properties of a sigma algebra

- $\emptyset$ is always in a sigma algebra
  - By definition (1)
- $\mathcal{S}$ is always in a sigma algebra
  - By definitions (1) and (2)
- A sigma algebra is also closed under countable intersections
  - By definition (2), (3), and the DeMorgan's law

Kolmogorov Axioms

Kolmogorov axioms

Given a sample space $\mathcal{S}$ and an associated sigma algbra $\mathcal{B}$, a **probability function** is a function with domain $\mathcal{B}$ that satisfies

1. $P(A) \geq 0$ for all $A \in \mathcal{B}$;
2. $P(\mathcal{S}) = 1$;
3. If $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

可列可加性

Defining Probability Functions

定理：Let $S = \{s_1, \ldots, s_n\}$ be a finit set. Let $\mathcal{B}$ be any sigma algbra of subsets of S. Let $p_1, \ldots, p_n$ be nonnegative numbers that sum to 1. For any $A \in \mathcal{B}$, define $P(A)$ by

$$P(A) = \sum_{\{i: s_i \in A\}} p_i.$$

Then P is a probability funciton on $\mathcal{B}$. This remains true if $S = \{s_1, s_2, \ldots\}$ is a countable set.

Satisfaction of the Kolmogorov Axioms

Axiom 1 : Because $p_1, \ldots, p_n$ are nonnegative numbers,

$$P(A) = \sum_{\{i: s_i \in A\}} p_i \geq 0$$

Axiom 2 : Because $p_1, \ldots, p_n$ sum up to 1,

$$P(S) = \sum_{\{i: s_i \in S\}} p_i = \sum_{i=1}^{n} p_i = 1$$

Axiom 3 : Let $A_1, A_2, \ldots, A_k$ be pairwise disjoint events, then,

$$P\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{\{j: s_j \in \bigcup_{i=1}^{k} A_i\}} p_j = \sum_{i=1}^{k} \sum_{\{j: s_i \in A_i\}} p_j = \sum_{i=1}^{k} P(A_i),$$

because the same p_j's appear exactly once in the equality.

Classical Probabilities

- Sample space

  $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$ A finite countable sample space

- Define probability

  $P(s_i) = p_i = 1/n$ Equal probability

- Probability function

$$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} \frac{1}{n} = \frac{\#\{\text{elements in } A\}}{\#\{\text{elements in } \mathcal{S}\}}$$

where $A \in \mathcal{B} = \{\text{all subsets of } \mathcal{S}, \text{ including } \mathcal{S} \text{ itself}\}$

Dice

- Sample space
  $$\mathcal{S} = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$$
- Define probability
  $$P(s_i) = 1/6$$
- Probability function
  $$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} \frac{1}{n} = \frac{\#\{\text{elements in } A\}}{\#\{\text{elements in } \mathcal{S}\}}$$
- Calculation
  - $P(\text{观测到 3 的概率}) = 1/6$
  - $P(\text{观测到奇数点的概率}) = 3/6 = 1/2$
  - $P(\text{观测到大于等于 3 的概率}) = 4/6 = 2/3$

The Calculus of Probabilities

For two events

If P is a probability function and A and B are any two sets in $\mathcal{B}$, then

1. $P(A) = P(A \cap B) + P(A \cap B^c)$;
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
3. $P(A \cup B) \leq P(A) + P(B)$;
4. $P(A \cap B) \geq P(A) + P(B) - 1$ (Bonferroni's inequality);
5. If $A \subset B$, then $P(A) \leq P(B)$.

The calculus of probabilities

For countable events

If P is a probability function, then

1. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2, \ldots$;
2. $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$ for any sets $A_1, A_2, \ldots$.

Conditional Probability

Conditional probability

if A and B are events in $\mathcal{S}$, and $P(B) > 0$, then the **conditional probability** of A given B, written $P(A \mid B)$, is

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

– 样本空间由 $\mathcal{S}$ 变为 B

Satisfaction of the Axioms

(1) For any event $A \in \mathcal{S}$, $P(A \mid B) = P(A \cap B) / P(B) \geq 0$;

(2) For the sample space $\mathcal{S}$, $P(S \mid B) = P(S \cap B) / P(B) = P(B) / P(B) = 1$;

(3) If $A_1, A_2, \ldots$ are disjoint, then

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) &= P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right) / P(B) \\
&= P\left(\bigcup_{i=1}^{\infty} (A_i \cap B)\right) / P(B) \\
&= \sum_{i=1}^{\infty} P(A_i \cap B) / P(B) \\
&= \sum_{i=1}^{\infty} (P(A_i \cap B) / P(B)) \\
&= \sum_{i=1}^{\infty} P(A_i \mid B)
\end{aligned}
$$

Statistically Independent

Statistically independent

Two events, A and B, are said to be **statistically independent** if

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

A collection of events, $A_1, A_2, \dots A_n$ are **mutually independent** if for **any subcollection** $A_{i_1}, A_{i_2}, \dots A_{i_k}$

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Multiplication Rule

Multiplication rule

Let A and B be two events in $\mathcal{S}$. If $P(A) > 0$, then

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

if $P(B) > 0$, then

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

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

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

Multiplication Rule

Multiplication rule

Let A, B, and C be three events in $\mathcal{S}$, then

$$\begin{aligned} P(A \cap B \cap C) &= P((A \cap B) \cap C) \\ &= P(A \cap B)P(C \mid A \cap B) \\ &= P(A)P(B \mid A)P(C \mid A \cap B) \end{aligned}$$

$$P(ABC) = P(A)P(B \mid A)P(C \mid AB)$$

Chain Rule

Chain rule

Let $A_1, \ldots, A_k$ be k events in $\mathcal{S}$, then

$$\begin{aligned}
P\left(\bigcap_{i=1}^{k} A_i\right) &= P\left(\bigcap_{i=1}^{k-1} A_i\right) P\left(A_k \middle| \bigcap_{i=1}^{k-1} A_i\right) \\
&= P\left(\bigcap_{i=1}^{k-2} A_i\right) P\left(A_{k-1} \middle| \bigcap_{i=1}^{k-2} A_i\right) P\left(A_k \middle| \bigcap_{i=1}^{k-1} A_i\right) \\
&= \cdots \\
&= P(A_1 \cap A_2 \cap A_3) P(A_4 \mid A_1 \cap A_2 \cap A_3) \cdots P\left(A_k \middle| \bigcap_{i=1}^{k-1} A_i\right) \\
&= P(A_1 \cap A_2) P(A_3 \mid A_1 \cap A_2) P(A_4 \mid A_1 \cap A_2 \cap A_3) \cdots P\left(A_k \middle| \bigcap_{i=1}^{k-1} A_i\right) \\
&= P(A_1) p(A_2 \mid A_1) P(A_3 \mid A_1 \cap A_2) P(A_4 \mid A_1 \cap A_2 \cap A_3) \cdots P\left(A_k \middle| \bigcap_{i=1}^{k-1} A_i\right)
\end{aligned}$$

Law of Total Probability

Law of total probability

Let $A_1, A_2, \ldots$ be a partition of the sample space $\mathcal{S}$, then for any event B,

$$P(B) = \sum_{i=1}^{\infty} P(B \mid A_i) P(A_i).$$

$$P(B) = \sum_{i=1}^{\infty} P(B \cap A_i), \text{ and } P(B \cap A_i) = P(B \mid A_i) P(A_i)$$

Bayes' Rule

Bayes' rule

Let $A_1, A_2, \ldots$ be a partition of the sample space $\mathcal{S}$, and let B be any set. Then, for each $i = 1, 2, \ldots,$

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B \mid A_j)P(A_j)}.$$

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

$$P(A_i \cap B) = P(B \mid A_i)P(A_i), \text{ and}$$

$$P(B) = \sum_{i=1}^{\infty} P(B \mid A_i)P(A_i)$$

随机变量

Tossing coins

- 扔一枚硬币，观察到正面的概率
  - S = {H, T}
  - P(正面) = P({H}) = 1/2
- 扔一枚硬币三次，观察到两次正面的概率
  - S = {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}
  - P(两次正面) = P({HHT, HTH, THH}) = 3/8
- 扔一枚硬币一百次，观察到十次正面的概率
  - S = {2^{100} elements}
  - P(十次正面) = Unable to count!
- 实际上正面出现的次数仅有101种可能

It is much easier to deal with a summary variable than with the original probability structure.

How to Reduce the Sample Space?

- 定义计数函数
  - $X(s) = \#\{\mathrm{H}\}$
  - 定义域 $\mathcal{S}$ 包含 2^{100} 个元素
  - 值域 $[0, 100]$ 包含 101 个元素
- 观察到十次正面的次数
  - $P(\#\{\mathrm{H}\}=10)=P(X=10)=\mathrm{C}(100,10)\times 0.5^{10}\times 0.5^{90}\approx 1.37\times 10^{-17}$
- 扔任意硬币 n 次，观察到 x 次正面的次数
  - $P(X=x \mid n, p) = \mathrm{C}(n, k) \times p^k \times (1-p)^{n-k}$

Random Variables

Random variable

A **random variable** is a function from a sample space $\mathcal{S}$ into the real numbers.

- 随机变量是定义在样本空间上的实值函数
- 随机变量用大写字母表示，例如X, Y, Z
- 随机变量的取值用对应的小写字母表示，例如x, y, z

Change of the Sample Space

- 样本空间的转换
  - 在随机变量的定义域上 $\mathcal{S} = \{s_1, s_2, \ldots, s_n\}$
  - 在随机变量的值域上 $\mathcal{X} = \{x_1, x_2, \ldots, x_m\}$
  - 随机变量建立的映射 $X\colon \mathcal{S} \mapsto \mathcal{X}$
- 定义在随机变量定义域上的概率函数

$$\begin{aligned} P(s_j) &= p_j \\ P(A) &= \textstyle\sum_{s_j \in A} p_j \end{aligned}$$

- 定义在随机变量值域上的概率函数

$$P_X(X = x_i) = P(\{s_j \in \mathcal{S}\colon X(s_j) = x_i\})$$

Induce a Probability on the Range

Suppose that the range of X is also a finite set $\mathcal{X}$, we can then define

$$P_X(X = x_i) = P\left(\{s_j : s_j \in \mathcal{S}, X(s_j) = x_i\}\right)$$

Now, let the sigma algebra $\mathcal{B}$ be the collection of all subsets of $\mathcal{X}$,

Axiom 1 : for any set $A \in \mathcal{B}$,

$$\begin{aligned} P_X(A) \quad &= P\left(\bigcup_{x_i \in A} \{s_j : s_j \in \mathcal{S}, X(s_j) = x_i\}\right) \\ &= \sum_{x_i \in A} P\left(\{s_j : s_j \in \mathcal{S}, X(s_j) = x_i\}\right) \\ &\geq 0 \end{aligned}$$

Axiom 2 : for the entire sample space $\mathcal{X}$,

$$P_X(\mathcal{X}) \quad = P\left(\bigcup_{x_i \in \mathcal{X}} \{s_j : s_j \in \mathcal{S}, X(s_j) = x_i\}\right) = P(\mathcal{S}) = 1$$

Axiom 3 : for pairwise disjoint sets $A_i, A_2, \ldots,$

$$\begin{aligned} P_X\left(\bigcup_{k=1}^{\infty} A_k\right) &= P\left(\bigcup_{k=1}^{\infty} \left\{\bigcup_{x_i \in A_k} \{s_j : s_j \in \mathcal{S}, X(s_j) = x_i\}\right\}\right) \\ &= \sum_{k=1}^{\infty} P\left(\bigcup_{x_i \in A_k} \{s_j : s_j \in \mathcal{S}, X(s_j) = x_i\}\right) \\ &= \sum_{k=1}^{\infty} P_X(A_k) \end{aligned}$$

随机变量

- 随机变量的引入简化了研究的问题，体现了统计学中**数据简约**的思想
- 随机变量的取值很重要，但随机变量以什么概率取得这些值更重要

Distributions of Random Variables

- 随机变量的所有可能取值及取得每一个值的概率
- 扔一枚硬币三次，观察出现正面的次数
  - $\mathcal{S} = \{\text{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}\}$
  - $\mathcal{X} = \{0, 1, 2, 3\}$
  - X: $\mathcal{S} \mapsto \mathcal{X}$
    $X(\text{HHH})=3 \quad X(\text{HHT})=2 \quad X(\text{HTH})=2 \quad X(\text{THH})=2$
    $X(\text{TTH})=1 \quad X(\text{THT})=1 \quad X(\text{HTT})=1 \quad X(\text{TTT})=0$
  - $P(X=0)=1/8 \qquad P(X=1)=3/8$
    $P(X=2)=3/8 \qquad P(X=3)=1/8$

Cumulative distribution function (cdf)

Distribution function

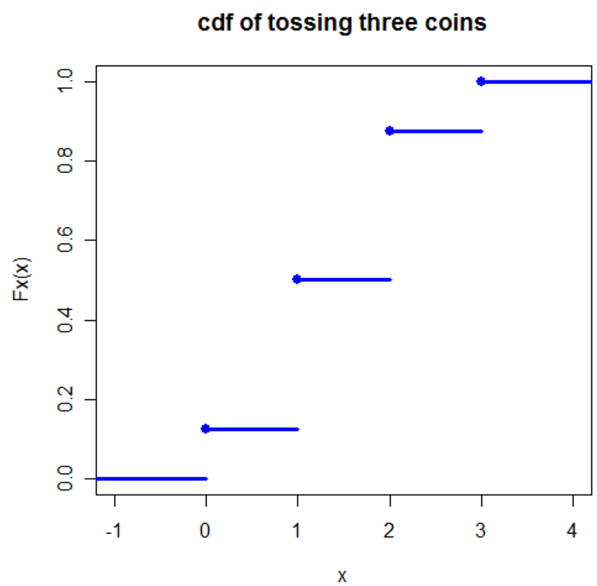
The **cumulative distribution function** (cdf) of a random variable X, denotd by $F_X(x)$, is defined by

$$F_X(x) = P_X(X \leq x), \text{ for all } x.$$

累积分布函数

- 扔一枚硬币三次，观察出现正面的次数
  - $\mathcal{X} = \{0, 1, 2, 3\}$
  - $P(X=0) = 1/8$ $\quad$ $P(X=1) = 3/8$
    $P(X=2) = 3/8$ $\quad$ $P(X=3) = 1/8$
  - 分布函数

$$F_X(x) = \begin{cases} 0 & \text{if} \quad -\infty < x < 0; \\ 1/8 & \text{if} \quad 0 \le x < 1; \\ 1/2 & \text{if} \quad 1 \le x < 2; \\ 7/8 & \text{if} \quad 2 \le x < 3; \\ 1 & \text{if} \quad 3 \le x < \infty. \end{cases}$$



Necessary and sufficient condition

Necessary and sufficient condition

The function $F(x)$ is a cdf if and only if the following **three conditions** hold:

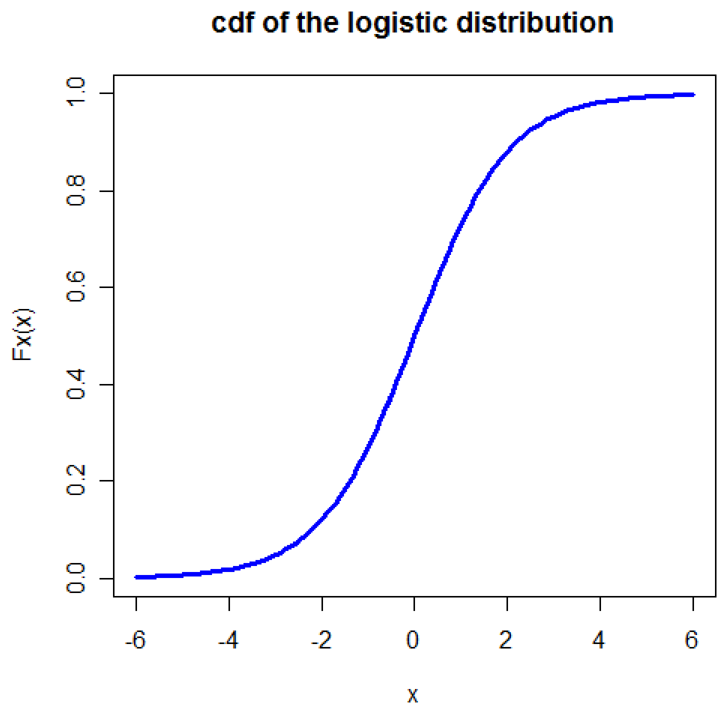
1. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$;
2. $F(x)$ is a nondescreasing function of x;
3. $F(x)$ is right-continuous; that is, for every number x_0, $\lim_{x \to x_0^+} F(x) = F(x_0)$.

Logistic cdf

- Logistic distribution

$$F_X(x) = \frac{1}{1+e^{-x}}$$

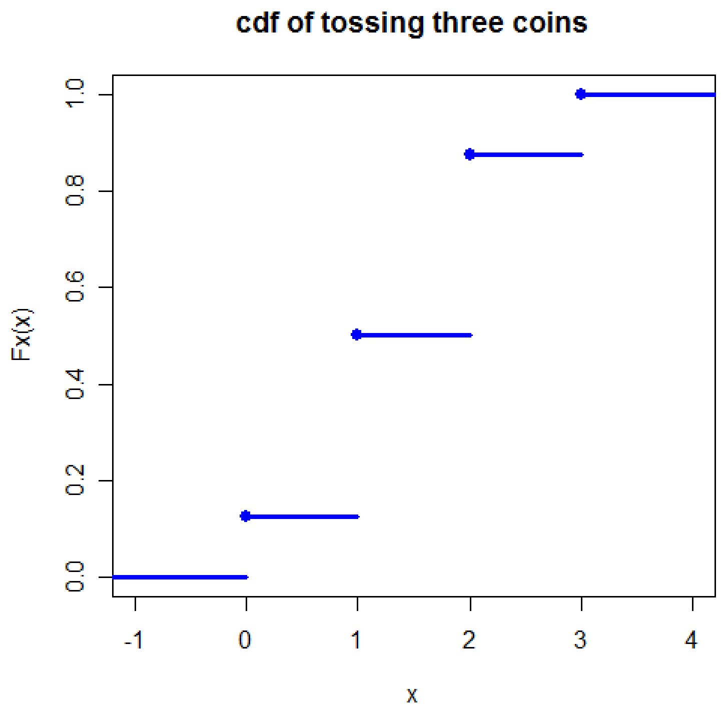
- 充要条件的满足性
  - 负无穷时为0
  - 正无穷时为1
  - 不减
  - 右连续



Discrete random variables

Discrete random variables

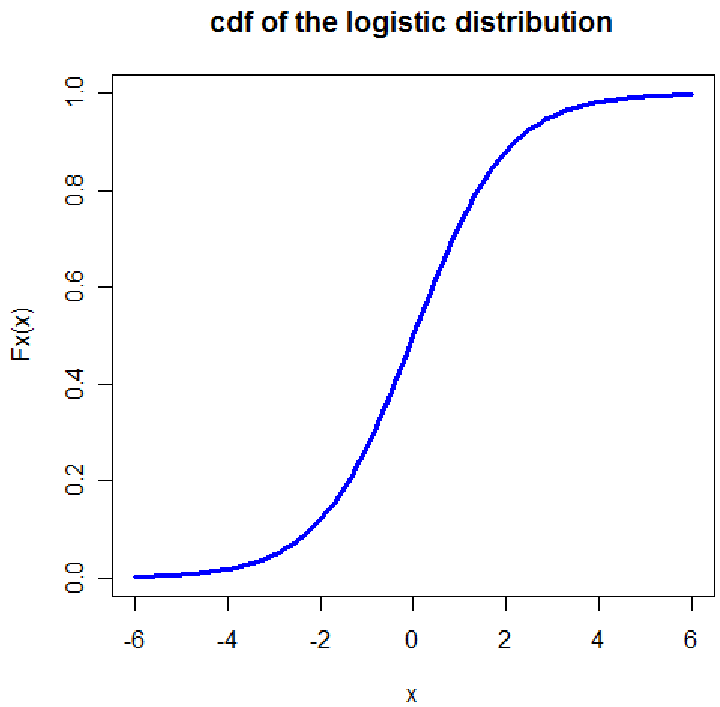
A random variable X is **discrete** if $F_X(x)$ is a step function of x.



Continuous random variables

Continuous random variables

A random variable X is **continuous** if $F_X(x)$ is a continuous function of x.



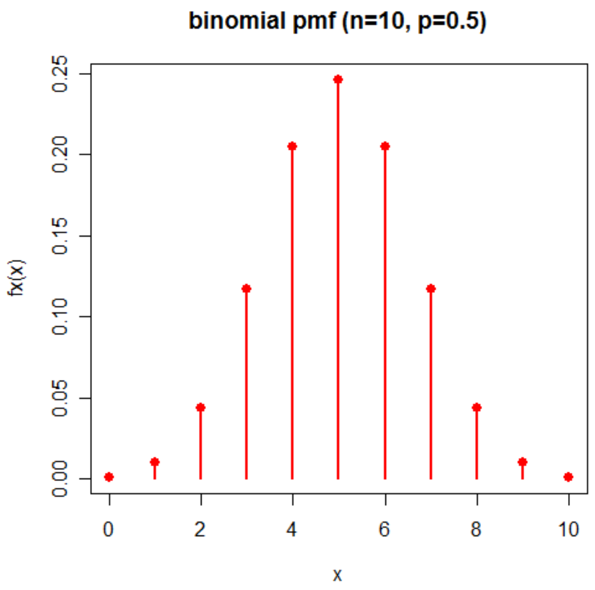
Probability mass functions

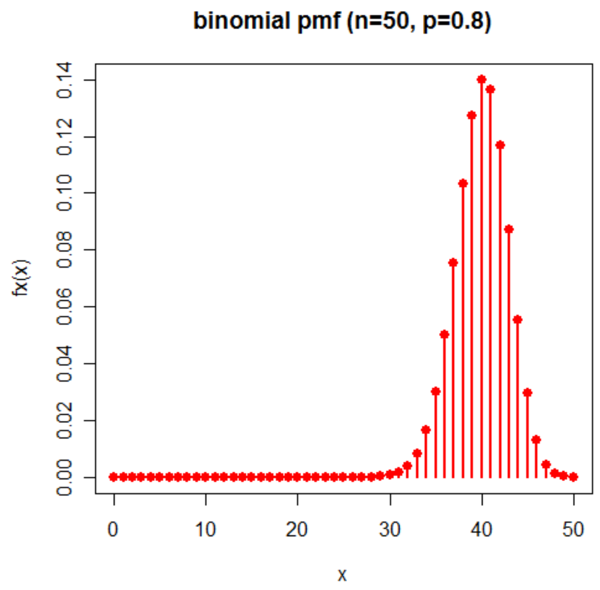
Probability mass function

The **probability mass function** (pmf) of a discrete random variable X, denoted by $f_X(x)$, is given by

$$f_X(x) = P_X(X = x), \text{ for all } x.$$

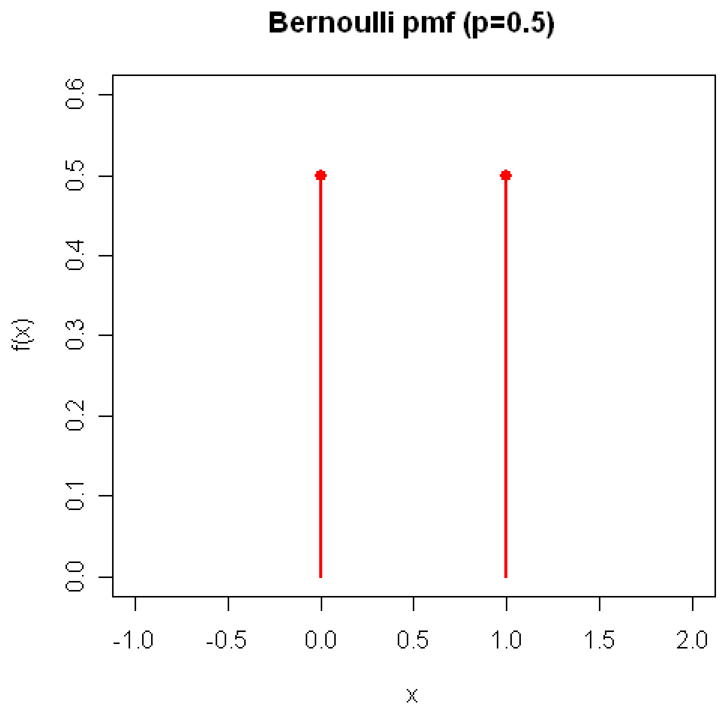
Exact





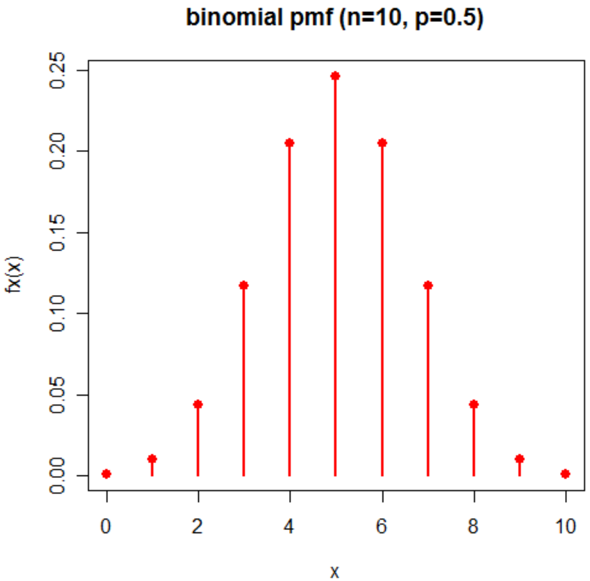
Bernoulli distribution

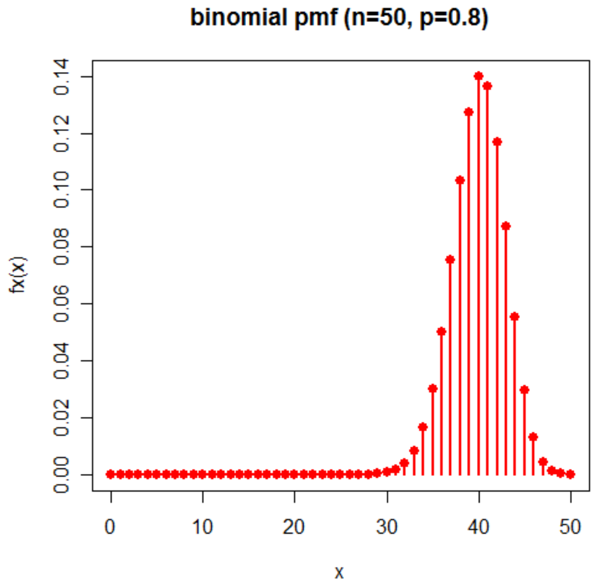
$$X = \begin{cases} 1 \text{ (success)} & \text{with probability } p \\ 0 \text{ (failure)} & \text{with probability } 1 - p \end{cases}$$

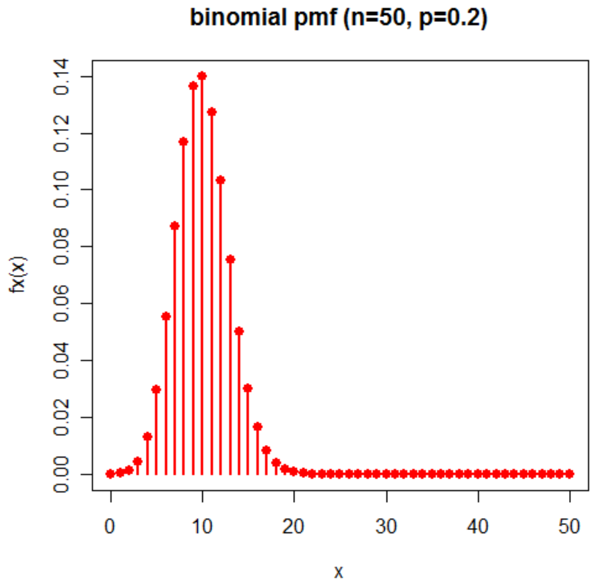


Binomial distribution

$$P(X = x \mid n, p) = \binom{n}{x} p^{x}(1-p)^{n-x}, x = 0, 1, \ldots, n$$







For a Continuous Random Variable

$$\boldsymbol{P(X{=}x) = ?}$$

- $\{X = x\} \subset \{x{-}\varepsilon < X \leq x\}$ for any x and ε
- $P\{X = x\} \leq P\{x{-}\varepsilon < X \leq x\}$
  $= P\{X \leq x \cap X > x{-}\varepsilon\}$
  $= P\{X \leq x \cap (X \leq x{-}\varepsilon)^{\mathrm{c}}\}$
  $= P\{X \leq x\} - P(X \leq x \cap X \leq x{-}\varepsilon\}$
  $= F_X(x) - F_X(x{-}\varepsilon)$
- $0 \leq P\{X = x\} \leq \lim_{\varepsilon\to 0} [F_X(x) - F_X(x{-}\varepsilon)] = 0$

$$\boldsymbol{P\{X{=}x\} = 0 \text{ for any } x}$$

$$P\{a{<}X{<}b\}{=}P\{a{<}X \leq b\}{=}P\{a{\leq}X{<}b\}{=}P\{a{\leq}X{\leq}b\} \text{ for any } x$$

Probability Density Functions

Probability density function

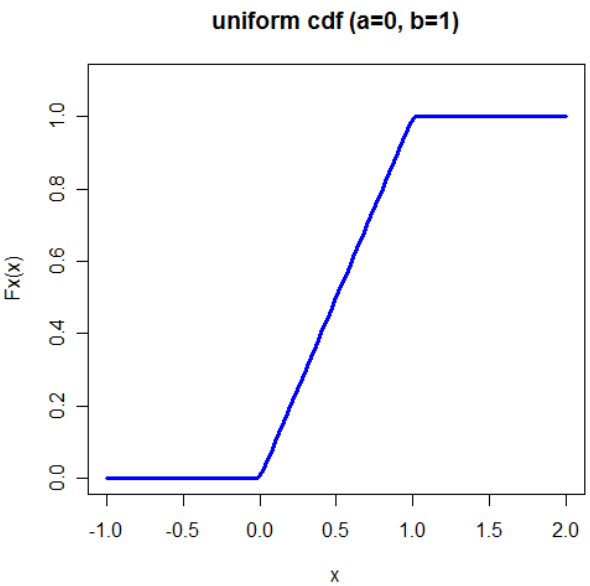
The **probability density function** (pdf), denoted by $f_X(x)$, of a continuous random variable X is the function that satisfies

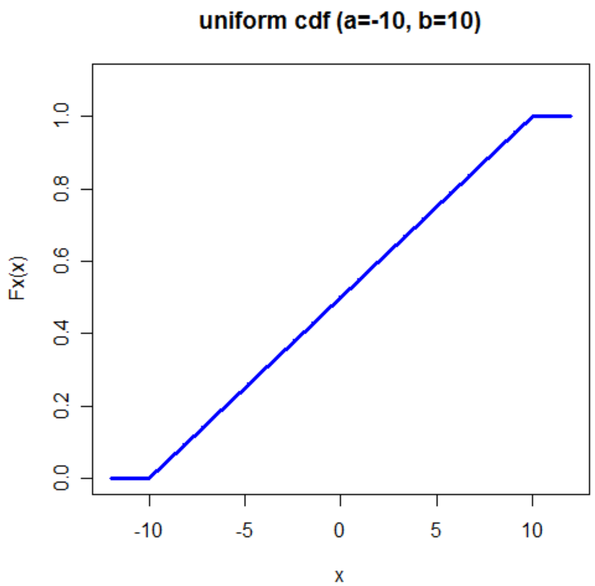
$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad \text{for all } x.$$

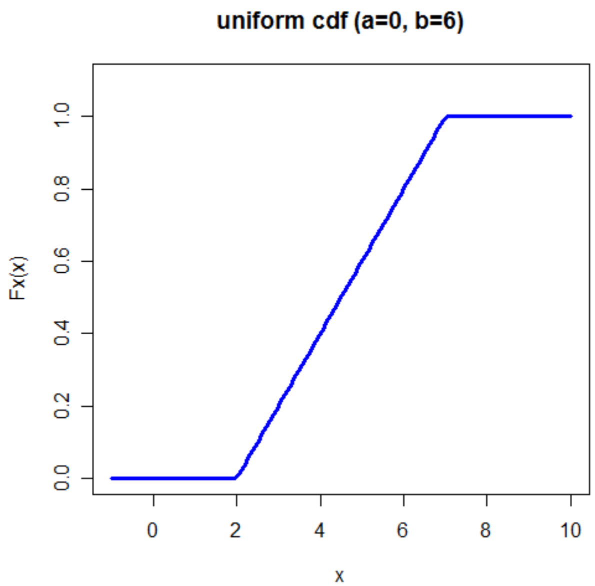
$$f_X(x) = \frac{d}{dx} F_X(x)$$

Uniform distribution

$$F_X(x \mid a,b) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$







Relation of cdfs and pdfs

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

$$P(X < x) = \int_{-\infty}^x f(t)dt$$

$$P(X > x) = \int_x^\infty f(t)dt = 1 - \int_{-\infty}^x f(t)dt$$

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt$$

Necessary and sufficient condition

Necessary and sufficient condition

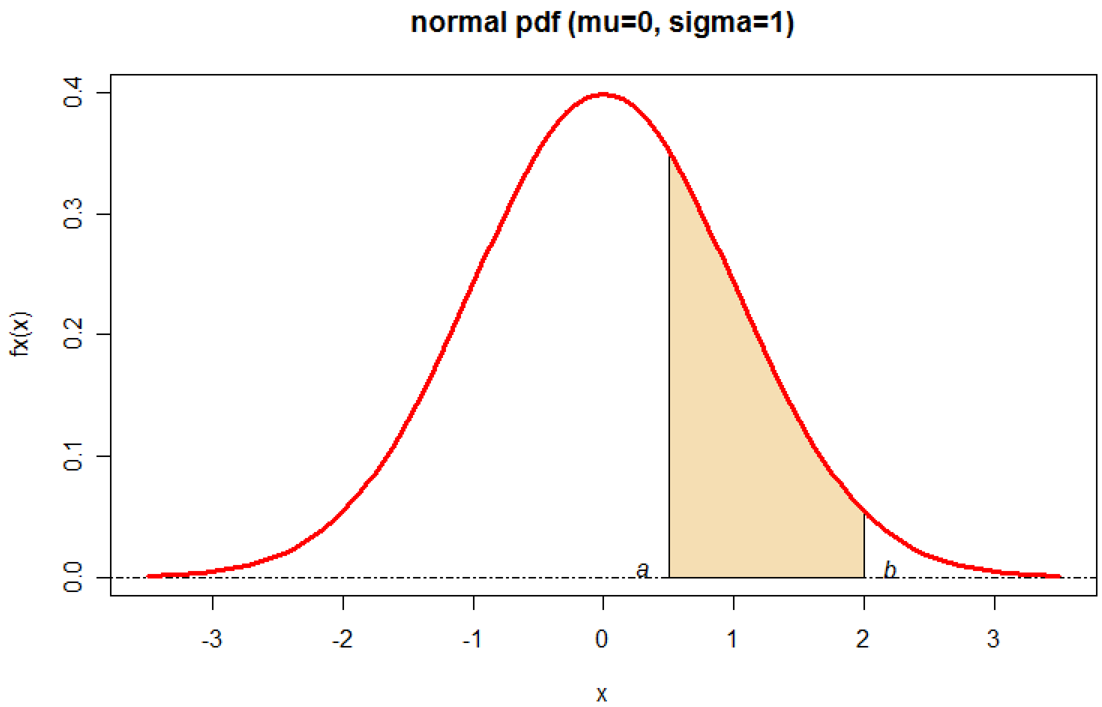
A function $f_X(x)$ is a pdf or pmf of a random variable X if and only if the following two conditions hold:

1. $f_X(x) \geq 0$ for all x;
2. $\sum_x f_X(x)=1$ (pmf) or

   $\int_{-\infty}^{\infty} f_X(x)=1$ (pdf).

Standard Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



$$\begin{aligned} &P(a \le x \le b) \\ &= P(a < x \le b) \\ &= P(a \le x < b) \\ &= P(a < x < b) \\ &= F_X(b) - F_X(a) \\ &= \int_a^b f_X(x)dx \end{aligned}$$

Mode

Mode

The **mode** of a random variable X is the value that occurs the most frequently in the probability distribution, corresponding to the maximum value in the pmf or pdf.

Median

Median

The **median** of a random variable X is a value m such that

$$P(X \le m) \ge \frac{1}{2} \text{ and } P(X \ge m) \ge \frac{1}{2}$$

For a continuous random variable X, the median m satisfies

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

Expectations

Expected value

The **expected value** or **mean** of a random variable $g(X)$, denoted by $\mathrm{E}g(x)$, is

$$\mathrm{E}g(x) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) & \text{if } X \text{ is descrete} \end{cases}$$

provided that the integral or sum exists. If $\mathrm{E}\left|g(X)\right| = \infty$, we say that $\mathrm{E}g(X)$ does not exist.

Normal Mode

Suppose X is a normal (μ, σ^2) random variable. Then,

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ \log f(x) &= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}(\sigma^2)(x-\mu)^2 \end{aligned}$$

Obviously, the maximum value is obtained at $x = \mu$.

Therefore,

The mode of a normal distribution is its location parameter.

Normal Median

Suppose X is a normal (μ, σ^2) random variable. Then,

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

$$\int_{-\infty}^{\mu} f(x)dx = \int_{-\infty}^{\mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \xrightleftharpoons{y=x-\mu} \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy$$

$$\int_{\mu}^{\infty} f(x)dx = \int_{\mu}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \xrightleftharpoons{y=-(x-\mu)} \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} dy$$

Obviously, these two integrals are equal

Therefore,

The median of a normal distribution is its location parameter.

Standard Normal Expectation

Suppose $$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty \leq x < \infty$$

that is, X has an **standard normal distribution** $N(0,1)$. Then,

$$\begin{aligned}
\mathrm{E}X &= \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} dx \\
&= -\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} d\left(-\frac{x^2}{2}\right) \\
&= e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} \\
&= 0
\end{aligned}$$

Properties of Expectation

Properties of expectation

Let X be a random variable and let $a, b,$ and c be constants. Then for any functions $g_1(X)$ and $g_2(X)$ whose expectations exists,

1. $\mathrm{E}[ag_1(X) + bg_2(X) + c] = a\mathrm{E}g_1(X) + b\mathrm{E}g_2(X) + c$;
2. If $g_1(X) \geq 0$ for all x, then $\mathrm{E}g_1(X) \geq 0$;
3. If $g_1(X) \geq g_2(X)$ for all x, then $\mathrm{E}g_1(X) \geq \mathrm{E}g_2(X)$;
4. If $a \leq g_1(X) \leq b$ for all x, then $a \leq \mathrm{E}g_1(X) \leq b$.

Moments of random variables

Moment

For eqch integer n, the $\boldsymbol{n}$**th moment** of a random variable X, μ'_n, is

$$\mu'_n = \mathrm{E}X^n.$$

The $\boldsymbol{n}$**th central moment** of X, μ_n, is

$$\mu_n = \mathrm{E}(X - \mu)^n,$$

where $\mu = \mu'_n = \mathrm{E}X.$

Mean and Variance

Variance

The **mean** of a random variable X is its first moment

$$\mu = \mathrm{E}X.$$

The **variance** of a random variable X is its second central moment

$$\mathrm{Var}X = \mathrm{E}(X - \mathrm{E}X)^2.$$

The positive sequre root of $\mathrm{Var}X$ is the **standard deviation** of X.

Properties of Variance

Properties of variances

If X is a random variable with finite variance, then for any constants a and b

$$\mathrm{Var}(aX+b) = a^2 \mathrm{Var} X.$$

$$\begin{aligned}
\text{since} \quad \mathrm{Var}(aX+b) &= \mathrm{E}((aX+b) - E(aX+b))^2 \\
&= \mathrm{E}(aX - aEX)^2 \\
&= a^2 \mathrm{E}(X - \mathrm{E}X)^2 \\
&= a^2 \mathrm{Var}(X).
\end{aligned}$$

$$\mathrm{Var} X = \mathrm{E}X^2 - (\mathrm{E}X)^2$$

$$\begin{aligned}
\text{since} \quad \mathrm{Var} X &= \mathrm{E}(X - \mathrm{E}X)^2 \\
&= \mathrm{E}(X^2 - 2X\mathrm{E}X + (\mathrm{E}X)^2) \\
&= \mathrm{E}X^2 - 2(\mathrm{E}X)^2 + (\mathrm{E}X)^2 \\
&= \mathrm{E}X^2 - (\mathrm{E}X)^2.
\end{aligned}$$

Standard Normal Variance

Suppose X is a standard normal random variable. Then,

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \\
\mathrm{E}X &= \int_{-\infty}^{\infty} x(2\pi)^{-1/2} \exp(-x^2/2)dx \\
&= -(2\pi)^{-1/2} \int_{-\infty}^{\infty} \exp(-x^2/2)d(-x^2/2) \\
&= -(2\pi)^{-1/2} \exp(-x^2/2)\Big|_{-\infty}^{\infty} \\
&= 0 \\
\mathrm{E}X^2 &= \int_{-\infty}^{\infty} x^2(2\pi)^{-1/2} \exp(-x^2/2)dx \\
&= -(2\pi)^{-1/2} \int_{-\infty}^{\infty} xd\exp(-x^2/2) \\
&= -(2\pi)^{-1/2} x\exp(-x^2/2)\Big|_{-\infty}^{\infty} + (2\pi)^{-1/2} \int_{-\infty}^{\infty} \exp(-x^2/2)dx \\
&= 1
\end{aligned}
$$

Therefore

$$\mathrm{Var}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = 1$$

$$\int_{-\infty}^{\infty} \exp(-p^2x^2 + qx)dx = \exp\left(\frac{q^2}{4p^2}\right)\frac{\sqrt{\pi}}{p} \quad (p > 0)$$

Moment Generating Function

Moment generating function

Let X be a random variable with cdf $F_X(x)$. The **moment generating function (mgf)** of X, $M_X(t)$, is

$$M_X(t) = \mathrm{E}e^{tX},$$

provided that the expectation exists for t in some neighborhood of 0.

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$$

$$M_X(t) = \sum_{x=-\infty}^{\infty} e^{tx} P(X = x)$$

Normal Moment Generation Function

$$
\begin{aligned}
M(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} + tx\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2 - 2\mu x + \mu^2 + 2\sigma^2 tx}{2\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{x^2 - 2(\mu+\sigma^2 t)x + (\mu+\sigma^2 t)^2 + \mu^2 - (\mu+\sigma^2 t)^2}{2\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2}\left(x - (\mu+\sigma^2 t)\right)^2 + \left(\mu t + \frac{1}{2}\sigma^2 t^2\right)\right] dx \\
&= \color{red}{\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{1}{2\sigma^2}\left(x - (\mu+\sigma^2 t)\right)^2\right] dx}_{\color{blue}{=1}}
\end{aligned}
$$

Deriving moments from mgf

Deriving moments

If X has mgf $M_X(t)$, then

$$\mathrm{E}X^n = M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}.$$

That is, the n-th moment is equal to the n-th derivative of $M_X(t)$, evaluated at $t = 0$.

Standard Normal Moments

Standard normal mgf is

$$M(t) = \exp\left(\frac{t^2}{2}\right)$$

$$\frac{d}{dx}M(t) = t\exp\left(\frac{t^2}{2}\right) \Rightarrow \mu_1' = 0 \Rightarrow \mu = 0$$

$$\frac{d^2}{dx^2}M(t) = (t^2+1)\exp\left(\frac{t^2}{2}\right) \Rightarrow \mu_2' = 1 \Rightarrow \sigma^2 = 1$$

$$\frac{d^3}{dx^3}M(t) = (t^3+3t)\exp\left(\frac{t^2}{2}\right) \Rightarrow \mu_3' = 0 \Rightarrow \beta_s = 0$$

$$\frac{d^4}{dx^4}M(t) = (t^4+6t^2+3)\exp\left(\frac{t^2}{2}\right) \Rightarrow \mu_4' = 3 \Rightarrow \beta_k = 3$$