概率论与数理统计（8）

清华大学

2020 年春季学期

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词：

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词：
  - 数据挖掘 (Data Mining);

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词：
    - 数据挖掘 (Data Mining);
    - 机器学习（Machine learning), 深度学习（deep learning);

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词：
  - 数据挖掘 (Data Mining);
  - 机器学习（Machine learning), 深度学习（deep learning);
  - 大数据科学（Big Data);

数理统计

- 统计学是一门研究如何有效地收集和分析（随机性）数据的学科。
- 应用极其广泛：物理，生物，医学，销售，金融，保险，机械制造，体育，博彩业。。。
- 几个时髦的名词：
    - 数据挖掘 (Data Mining)；
    - 机器学习（Machine learning), 深度学习（deep learning);
    - 大数据科学（Big Data);
- 这门课只涉及基本概念和方法。只涉及数据分析，不涉及数据收集。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。
- 我们想了解总体的性质，如数学期望，方差，某些特殊性质所占的比例等等。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。
- 我们想了解总体的性质，如数学期望，方差，某些特殊性质所占的比例等等。
- 可以把所有的个体都研究一遍。没有任何遗漏，错误。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。
- 我们想了解总体的性质，如数学期望，方差，某些特殊性质所占的比例等等。
- 可以把所有的个体都研究一遍。没有任何遗漏，错误。当总体的数量太大时，不大实际。

总体和样本

- 总体：研究对象的全体：如全班期中考试的成绩，班上所有学生的年龄，一批货物中所有货物各自的重量，价格，等等。
- 总体是一大堆数据，可以看作一个分布。
- 个体是单个研究对象，如说某某的身高，成绩之类的。
- 我们想了解总体的性质，如数学期望，方差，某些特殊性质所占的比例等等。
- 可以把所有的个体都研究一遍。没有任何遗漏，错误。当总体的数量太大时，不大实际。
- 随机抽样方法。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。
- 当然样本容量越大越接近总体。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。
- 当然样本容量越大越接近总体。
- 样本要具有随机性：每个样品的分布应该与总体的相同。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。
- 当然样本容量越大越接近总体。
- 样本要具有随机性：每个样品的分布应该与总体的相同。
- 样本要有独立性：即 $x_1, \ldots, x_n$ 相互独立。

简单随机抽样

- 随机地从总体中抽出 n 个个体，记为 $x_1, \ldots, x_n$。它们就是总体的一个样本，n 称为样本容量，样本中的个体称为样品。
- 当然样本容量越大越接近总体。
- 样本要具有随机性：每个样品的分布应该与总体的相同。
- 样本要有独立性：即 $x_1, \ldots, x_n$ 相互独立。若总体的分布函数为 $F(x)$，则样本容量为 n 的样本联合分布函数为

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{n} F(x_i).$$

- 经验分布函数：设 $x_1, \ldots, x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)}, \ldots, x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_k \leqslant x < x_{(k+1)}, \quad k = 1, \ldots, n-1, \\ 1, & x \geqslant x_{(n)}, \end{cases}$$

为经验分布函数。

- 经验分布函数：设 $x_1, \ldots, x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)}, \ldots, x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_k \leqslant x < x_{(k+1)}, \quad k = 1, \ldots, n-1, \\ 1, & x \geqslant x_{(n)}, \end{cases}$$

为经验分布函数。

- 若定义 $I_i(x) = \begin{cases} 1, & x_i \leqslant x, \\ 0, & x_i > x, \end{cases}$,

- 经验分布函数：设 $x_1, \ldots, x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)}, \ldots, x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_k \leqslant x < x_{(k+1)}, \quad k = 1, \ldots, n-1, \\ 1, & x \geqslant x_{(n)}, \end{cases}$$

为经验分布函数。

- 若定义 $I_i(x) = \begin{cases} 1, & x_i \leqslant x, \\ 0, & x_i > x, \end{cases}$，则 $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_i(x)$

- 经验分布函数：设 $x_1,\ldots,x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)},\ldots,x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x)=\begin{cases}0, & x<x_{(1)},\\ \frac{k}{n}, & x_k\leqslant x<x_{(k+1)}, \quad k=1,\ldots,n-1,\\ 1, & x\geqslant x_{(n)},\end{cases}$$

为经验分布函数。

- 若定义 $I_i(x)=\begin{cases}1, & x_i\leqslant x,\\ 0, & x_i>x,\end{cases}$，则 $F_n(x)=\frac{1}{n}\sum_{i=1}^{n}I_i(x)\to F(x)$.

- 经验分布函数：设 $x_1, \ldots, x_n$ 是取自总体分布函数为 $F(x)$ 的样本，如果将样本观测值由小到大排列，为 $x_{(1)}, \ldots, x_{(n)}$，则称其为有序样本，用有序样本定义的函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \frac{k}{n}, & x_k \leqslant x < x_{(k+1)}, \quad k = 1, \ldots, n-1, \\ 1, & x \geqslant x_{(n)}, \end{cases}$$

为经验分布函数。

- 若定义 $I_i(x) = \begin{cases} 1, & x_i \leqslant x, \\ 0, & x_i > x, \end{cases}$，则 $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_i(x) \to F(x)$.
- (格里纹科定理)：$P(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0) = 1$.

样本数据的整理和显示

- 频数频率表
- 直方图
- 茎叶图

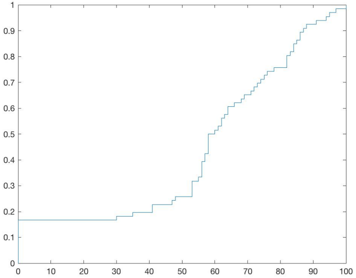
- 频数频率表
- 直方图
- 茎叶图

```
>> stemleafplot(V)
  0 | 0 0 0 0 0 0 0 0 0 0 0
  1 |
  2 |
  3 | 0 5
  4 | 1 1 7 8
  5 | 3 3 3 3 5 6 6 6 6 7 7 8 8 8 8 8
  6 | 0 1 2 2 3 4 4 6 8
  7 | 1 2 3 4 5 6 8
  8 | 2 2 2 3 4 4 5 6 6 7 8
  9 | 1 4 5 7
 10 | 0
```

样本数据的整理和显示

- 频数频率表
- 直方图
- 茎叶图

```
>> stemleafplot(V)
  0 | 0 0 0 0 0 0 0 0 0 0 0
  1 |
  2 |
  3 | 0 5
  4 | 1 1 7 8
  5 | 3 3 3 3 5 6 6 6 6 7 7 8 8
  6 | 0 1 2 2 3 4 4 6 8
  7 | 1 2 3 4 5 6 8
  8 | 2 2 2 3 4 4 5 6 6 7 8
  9 | 1 4 5 7
 10 | 0
```



统计量及其分布

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。

统计量及其分布

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。
- $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_i$.

统计量及其分布

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。
- $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_i$.
- $\frac{1}{n}\sum_{k=1}^{n}(x_i - \mu)^2$，其中 μ 是总体分布的数学期望（已知）。

统计量及其分布

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。
- $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_i$.
- $\frac{1}{n}\sum_{k=1}^{n}(x_i - \mu)^2$，其中 μ 是总体分布的数学期望 (已知)。
- $\frac{1}{n}\sum_{k=1}^{n}(x_i - \bar{x})^2$.

- 设 $x_1, \ldots, x_n$ 为某总体的样本，若样本函数 $T = T(x_1, \ldots, x_n)$ 中不含有任何未知参数，则称 T 为统计量。统计量的分布为抽样分布。
- $\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_i$.
- $\frac{1}{n} \sum_{k=1}^{n} (x_i - \mu)^2$，其中 μ 是总体分布的数学期望 (已知)。
- $\frac{1}{n} \sum_{k=1}^{n} (x_i - \bar{x})^2$.
- $\frac{1}{n-1} \sum_{k=1}^{n} (x - \bar{x})^2$

样本均值

- 设 $x_1, \ldots, x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x} = \frac{x_1 + \cdots + x_n}{n}$.

样本均值

- 设 $x_1, \ldots, x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x} = \frac{x_1 + \cdots + x_n}{n}$.
- 若把样本中的数据与样本均值之差称为偏差，则样本所有偏差之和为 0，即 $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

样本均值

- 设 $x_1, \ldots, x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x} = \frac{x_1+\cdots+x_n}{n}$.
- 若把样本中的数据与样本均值之差称为偏差，则样本所有偏差之和为 0，即 $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = 0.$$

样本均值

- 设 $x_1, \ldots, x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x} = \frac{x_1+\cdots+x_n}{n}$.
- 若把样本中的数据与样本均值之差称为偏差，则样本所有偏差之和为 0，即 $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = 0.$$

- 数据观测值与均值的偏差平方和最小，即若考虑 $f(c) = \sum_{i=1}^{n}(x_i - c)^2$，则 $c = \bar{x}$ 取到最小值。

样本均值

- 设 $x_1,\ldots,x_n$ 为某总体的样本，则其算数平均称为样本均值，一般用 $\bar{x}$ 表示，即 $\bar{x}=\frac{x_1+\cdots+x_n}{n}$.
- 若把样本中的数据与样本均值之差称为偏差，则样本所有偏差之和为 0，即 $\sum_{i=1}^n(x_i-\bar{x})=0$.

$$\sum_{i=1}^n(x_i-\bar{x})=\sum_{i=1}^n x_i-n\bar{x}=0.$$

- 数据观测值与均值的偏差平方和最小，即若考虑 $f(c)=\sum_{i=1}^n(x_i-c)^2$，则 $c=\bar{x}$ 取到最小值。

$$\begin{aligned}\sum_{i=1}^n(x_i-c)^2&=\sum_{i=1}^n(x_i-\bar{x}+\bar{x}-c)^2\\&=\sum_{i=1}^n(x_i-\bar{x})^2+n(\bar{x}-c)^2+2\sum_{i=1}^n(x_i-\bar{x})(\bar{x}-c)\end{aligned}$$

样本均值的抽样分布

- 若总体分布为 $N(\mu, \sigma^2)$, 则 $\bar{x} \sim N(\mu, \sigma^2/n)$.

样本均值的抽样分布

- 若总体分布为 $N(\mu, \sigma^2)$, 则 $\bar{x} \sim N(\mu, \sigma^2/n)$.

$$x_1 + \cdots + x_n \sim N(n\mu, n\sigma^2), \quad \bar{x} \sim N(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}).$$

样本均值的抽样分布

- 若总体分布为 $N(\mu, \sigma^2)$, 则 $\bar{x} \sim N(\mu, \sigma^2/n)$.

$$x_1 + \cdots + x_n \sim N(n\mu, n\sigma^2), \quad \bar{x} \sim N(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x) = \mu$, $Var(x) = \sigma^2$, 则 $E(\bar{x}) = \mu$, $Var(\bar{x}) = \frac{\sigma^2}{n}$.

样本均值的抽样分布

- 若总体分布为 $N(\mu,\sigma^2)$, 则 $\bar{x} \sim N(\mu,\sigma^2/n)$.

$$x_1 + \cdots + x_n \sim N(n\mu, n\sigma^2), \quad \bar{x} \sim N(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x) = \mu$, $Var(x) = \sigma^2$, 则 $E(\bar{x}) = \mu$, $Var(\bar{x}) = \frac{\sigma^2}{n}$. 当 n 充分大时，近似地 $\bar{x} \sim N(\mu,\sigma^2/n)$..

样本均值的抽样分布

- 若总体分布为 $N(\mu,\sigma^2)$, 则 $\bar{x} \sim N(\mu,\sigma^2/n)$.

$$x_1 + \cdots + x_n \sim N(n\mu, n\sigma^2), \quad \bar{x} \sim N(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x) = \mu$, $Var(x) = \sigma^2$, 则 $E(\bar{x}) = \mu$, $Var(\bar{x}) = \frac{\sigma^2}{n}$. 当 n 充分大时，近似地 $\bar{x} \sim N(\mu,\sigma^2/n)$..

$$\frac{\sum_{i=1}^n (x_i - \mu)}{\sqrt{n}\sigma} \tilde{\sim} N(0,1), \quad \bar{x} = \frac{\sigma}{\sqrt{n}} \frac{\sum_{i=1}^n (x_i - \mu)}{\sqrt{n}\sigma} + \mu \tilde{\sim} N(\mu, \frac{\sigma^2}{n}).$$

样本均值的抽样分布

- 若总体分布为 $N(\mu,\sigma^2)$, 则 $\bar{x} \sim N(\mu,\sigma^2/n)$.

$$x_1 + \cdots + x_n \sim N(n\mu, n\sigma^2), \quad \bar{x} \sim N(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x) = \mu$, $Var(x) = \sigma^2$, 则 $E(\bar{x}) = \mu$, $Var(\bar{x}) = \frac{\sigma^2}{n}$. 当 n 充分大时，近似地 $\bar{x} \sim N(\mu,\sigma^2/n)$..

$$\frac{\sum_{i=1}^n (x_i - \mu)}{\sqrt{n}\sigma} \tilde{\sim} N(0,1), \quad \bar{x} = \frac{\sigma}{\sqrt{n}} \frac{\sum_{i=1}^n (x_i - \mu)}{\sqrt{n}\sigma} + \mu \tilde{\sim} N(\mu, \frac{\sigma^2}{n}).$$

- 我们学过的分布中，还有哪些是可以精确知道 $\bar{x}$ 的分布的?

样本均值的抽样分布

- 若总体分布为 $N(\mu,\sigma^2)$，则 $\bar{x} \sim N(\mu,\sigma^2/n)$.

$$x_1+\cdots+x_n \sim N(n\mu, n\sigma^2), \quad \bar{x} \sim N(\frac{n\mu}{n}, \frac{n\sigma^2}{n^2}).$$

- 对于一般的分布，若 $E(x)=\mu$，$Var(x)=\sigma^2$，则 $E(\bar{x})=\mu$，$Var(\bar{x})=\frac{\sigma^2}{n}$. 当 n 充分大时，近似地 $\bar{x} \sim N(\mu,\sigma^2/n)$..

$$\frac{\sum_{i=1}^n (x_i-\mu)}{\sqrt{n}\sigma} \tilde{\sim} N(0,1), \quad \bar{x} = \frac{\sigma}{\sqrt{n}} \frac{\sum_{i=1}^n (x_i-\mu)}{\sqrt{n}\sigma} + \mu \tilde{\sim} N(\mu, \frac{\sigma^2}{n}).$$

- 我们学过的分布中，还有哪些是可以精确知道 $\bar{x}$ 的分布的？指数分布，伽玛分布，$\chi^2(n)$ 分布。（具有独立可加性）。

样本方差与样本标准差

- 设 $x_1, \ldots, x_n$ 为取自某总体的样本，则它关于均值平均偏差和为

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

称为样本方差，其算数根 $s_n = \sqrt{s_n^2}$ 称为样本标准差。

样本方差与样本标准差

- 设 $x_1, \ldots, x_n$ 为取自某总体的样本，则它关于均值平均偏差和为

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

称为样本方差，其算数根 $s_n = \sqrt{s_n^2}$ 称为样本标准差。当 n 不大时，一般会用

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

s^2 为无偏样本方差。对应的 $s = \sqrt{s^2}$ 代替 s_n.

样本方差

- 设总体 X 有 $E(X)=\mu$, $Var(X)=\sigma^2<\infty$, $x_1,\ldots,x_n$ 为从该总体取出的样本, 则

$$E(\bar{x})=\mu,\quad Var(\bar{x})=\sigma^2/n,\quad E(s^2)=\sigma^2.$$

样本方差

- 设总体 X 有 $E(X)=\mu$, $Var(X)=\sigma^2<\infty$, $x_1,\ldots,x_n$ 为从该总体取出的样本, 则

$$E(\bar{x})=\mu, \quad Var(\bar{x})=\sigma^2/n, \quad E(s^2)=\sigma^2.$$

-

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}^2)=\sum_{i=1}^{n}x_i^2-n\bar{x}^2\\ &= n(\sigma^2+\mu^2)+n(\mu^2-\frac{\sigma^2}{n})=(n-1)\sigma^2.\end{aligned}$$

- 设总体 X 有 $E(X)=\mu$, $Var(X)=\sigma^2<\infty$, $x_1,\ldots,x_n$ 为从该总体取出的样本, 则

$$E(\bar{x})=\mu,\quad Var(\bar{x})=\sigma^2/n,\quad E(s^2)=\sigma^2.$$

-

$$\sum_{i=1}^n(x_i-\bar{x})^2=\sum_{i=1}^n(x_i^2-2x_i\bar{x}+\bar{x}^2)=\sum_{i=1}^n x_i^2-n\bar{\bar{x}}^2$$

$$=n(\sigma^2+\mu^2)+n(\mu^2-\frac{\sigma^2}{n})=(n-1)\sigma^2.$$

$$E(s_n^2)=\frac{n-1}{n}\sigma^2,\quad E(s^2)=\sigma^2.$$

次序统计量及其抽样分布

- 设 $x_1, \ldots, x_n$ 是取自总体 X 的样本，$x_{(i)}$ 称为该样本的第 i 个次序统计量，它的取值是将样本观测值有小到大排列后的第 i 个观测值，其中 $x_{(1)} = \min\{x_1, \ldots, x_n\}$ 为该样本的最小次序统计量，$x_{(n)} = \max\{x_1, \ldots, x_n\}$ 为该样本的最大次序统计量。$(x_{(1)}, \ldots, x_{(n)})$ 为该样本的次序统计量。

- 设 $x_1, \ldots, x_n$ 是取自总体 X 的样本，$x_{(i)}$ 称为该样本的第 i 个次序统计量，它的取值是将样本观测值有小到大排列后的第 i 个观测值，其中 $x_{(1)} = \min\{x_1, \ldots, x_n\}$ 为该样本的最小次序统计量，$x_{(n)} = \max\{x_1, \ldots, x_n\}$ 为该样本的最大次序统计量。$(x_{(1)}, \ldots, x_{(n)})$ 为该样本的次序统计量。
- 一般情况下，简单样本的样品是独立同分布的，而次序统计量既不独立也不同分布。

次序统计量及其抽样分布

- 设 $x_1, \ldots, x_n$ 是取自总体 X 的样本，$x_{(i)}$ 称为该样本的第 i 个次序统计量，它的取值是将样本观测值有小到大排列后的第 i 个观测值，其中 $x_{(1)} = \min\{x_1, \ldots, x_n\}$ 为该样本的最小次序统计量，$x_{(n)} = \max\{x_1, \ldots, x_n\}$ 为该样本的最大次序统计量。$(x_{(1)}, \ldots, x_{(n)})$ 为该样本的次序统计量。
- 一般情况下，简单样本的样品是独立同分布的，而次序统计量既不独立也不同分布。
- 如，若总体为均匀分布 $[a, b]$，想知道 a，b，可以考虑 $Y = \min\{x_1, \ldots, x_n\}$ 和 $Y' = \max\{x_1, \ldots, x_n\}$。如，极差 $R = x_{(n)} - x_{(1)}$.

单个次序统计量的抽样分布

- 设总体 X 的分布函数为 $F(x)$, 密度函数为 $p(x)$, 则 $x_{(k)}$ 的密度函数为

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}p(x).$$

- 设总体 X 的分布函数为 $F(x)$, 密度函数为 $p(x)$, 则 $x_{(k)}$ 的密度函数为

$$p_k(x)=\frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}p(x).$$

- 第 k 个值落在区间 $(x,x+\Delta]$ 内，即

$$\begin{cases} k-1\text{个小于}x, \\ 1\text{个落入}(x,x+\Delta], \\ n-k\text{个大于}x+\Delta. \end{cases}$$

一共有 $\binom{n}{k}\times k=\frac{n!}{(n-k)!(k-1)!}$ 种可能性，所有

$$F_k(x+\Delta)-F_k(x)\approx\frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}(p(x)\Delta).$$

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1} p(x).$$

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1}p(x).$$

- 最大次序统计量 $x_{(n)}$ 的密度函数为

$$p_n(x) = n(F(x))^{n-1}p(x).$$

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1} p(x).$$

- 最大次序统计量 $x_{(n)}$ 的密度函数为

$$p_n(x) = n(F(x))^{n-1} p(x).$$

- 若总体分布为均匀分布 $U(0,1)$, 则第 k 个次序统计量的密度函数为

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, \quad 0 < x < 1.$$

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1} p(x).$$

- 最大次序统计量 $x_{(n)}$ 的密度函数为

$$p_n(x) = n(F(x))^{n-1} p(x).$$

- 若总体分布为均匀分布 $U(0,1)$, 则第 k 个次序统计量的密度函数为

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, \quad 0 < x < 1.$$

这是贝塔分布 $Be(k, n-k+1)$。

单个次序统计量的抽样分布

- 最小次序统计量 $x_{(1)}$ 的密度函数为

$$p_1(x) = n(1 - F(x))^{n-1} p(x).$$

- 最大次序统计量 $x_{(n)}$ 的密度函数为

$$p_n(x) = n(F(x))^{n-1} p(x).$$

- 若总体分布为均匀分布 $U(0,1)$, 则第 k 个次序统计量的密度函数为

$$p_k(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, \quad 0 < x < 1.$$

这是贝塔分布 $Be(k, n-k+1)$。一般贝塔分布的密度函数为

$$p(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad E(Be) = \frac{a}{a+b}.$$

多个次序统计量及其抽样分布

- 次序统计量 $(x_{(i)}, x_{(j)})$, $i < j$, 的联合分布密度函数为

$$
\begin{aligned}
p_{i,j}(y,z) = & \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(y)]^{i-1}[F(z)-F(y)]^{j-i-1} \\
& \times [1-F(z)]^{n-j} p(y) p(z), \quad y \leqslant z.
\end{aligned}
$$

多个次序统计量及其抽样分布

- 次序统计量 $(x_{(i)}, x_{(j)})$, $i < j$, 的联合分布密度函数为

$$
\begin{aligned}
p_{i,j}(y,z) = & \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(y)]^{i-1}[F(z)-F(y)]^{j-i-1} \\
& \times [1-F(z)]^{n-j} p(y)p(z), \quad y \leqslant z.
\end{aligned}
$$

- 对于 $y < z$, 考虑事件 $\{x_{(i)} \in (y, y+\Delta y],\ x_{(j)} \in (z, z+\Delta z]\}$, 则有

$$
\begin{cases}
i-1\text{小于}y, \\
1\text{个落到}(y, y+\Delta y], \\
j-i-1\text{个落到}(y+\Delta y, z], \\
1\text{个落到}(z, z+\Delta z], \\
n-j\text{大于}z+\Delta z.
\end{cases}
$$

有 $\binom{n}{i} \times i \times \binom{n-i}{j-i} \times (j-i)$ 中可能的组合。

样本极差的抽样分布

- 设总体的分布为均匀分布 $U(0,1)$，则 $(x_{(1)}, x_{(n)})$ 的联合分布密度为

$$p(y,z) = n(n-1)(z-y)^{n-2}.$$

样本极差的抽样分布

- 设总体的分布为均匀分布 $U(0,1)$，则 $(x_{(1)}, x_{(n)})$ 的联合分布密度为

$$p(y,z) = n(n-1)(z-y)^{n-2}.$$

- 样本极差为 $R = x_{(n)} - x_{(1)}$，则给定 $x_{(1)} = y$,

$$p(y,r) = n(n-1)r^{n-2}, \quad y > 0, r > 0, y + r < 1.$$

则

$$p_R(r) = \int_0^{1-r} n(n-1)r^{n-2}\,dy = n(n-1)r^{n-2}(1-r).$$

样本极差的抽样分布

- 设总体的分布为均匀分布 $U(0,1)$，则 $(x_{(1)}, x_{(n)})$ 的联合分布密度为

$$p(y,z) = n(n-1)(z-y)^{n-2}.$$

- 样本极差为 $R = x_{(n)} - x_{(1)}$，则给定 $x_{(1)} = y$，

$$p(y,r) = n(n-1)r^{n-2}, \quad y > 0, r > 0, y + r < 1.$$

则

$$p_R(r) = \int_0^{1-r} n(n-1)r^{n-2}\,dy = n(n-1)r^{n-2}(1-r).$$

- 一般分布方法类似，不一定能用初等函数表示。

样本分位数与样本中位数

- 分布函数 $F(x)$ 的 p-分位数为 x_p 满足 $F(x_p) = p$.

样本分位数与样本中位数

- 分布函数 $F(x)$ 的 p-分位数为 x_p 满足 $F(x_p)=p$.
- 样本中位数 $m_{0.5}$ 定义为

$$m_{0.5}=\begin{cases} x_{(\frac{n+1}{2})}, & \text{n 为奇数}, \\ \frac{1}{2}\left(x_{(n/2)}+x_{(n/2+1)}\right), & \text{n 为偶数}. \end{cases}.$$

样本分位数与样本中位数

- 分布函数 $F(x)$ 的 p-分位数为 x_p 满足 $F(x_p) = p$.
- 样本中位数 $m_{0.5}$ 定义为

$$m_{0.5} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{n 为奇数}, \\ \frac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right), & \text{n 为偶数}. \end{cases}.$$

- 样本 p 分位数 m_p 定义为

$$m_p = \begin{cases} x_{([np+1])}, & \text{np 不为整数}, \\ \frac{1}{2}\left(x_{(np)} + x_{(np+1)}\right), & \text{np 为整数}. \end{cases}$$

样本分位数的抽样分布

- 设总体密度函数为 $p(x)$, x_q 为其 q 分位数，$p(x)$ 在 x_q 处连续且 $p(x_q) > 0$, 则当 $n \to \infty$ 时样本 q 分位数 m_q 的渐进分布为

$$m_q \dot{\sim} N(x_q, \frac{q(1-q)}{np^2(x_q)}).$$

特别地，对于样本中位数，

$$m_{0.5} \dot{\sim} N(x_{0.5}, \frac{1}{4np^2(x_q)}).$$

样本分位数的抽样分布

- （大概原因）假设密度函数为正，即 $\forall x \in \mathbb{R}$, $p(x) > 0$. 考虑 $Y_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leqslant x) = \frac{1}{n}\sum_{i=1}^{n} I_i(x)$.

样本分位数的抽样分布

- （大概原因）假设密度函数为正，即 $\forall x \in \mathbb{R}$, $p(x) > 0$. 考虑 $Y_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leqslant x) = \frac{1}{n}\sum_{i=1}^n I_i(x)$. $E(I_i(x)) = F_X(x)$, $Var(I_i(x)) = (1 - F_X(x))F_X(x)$.

- (大概原因)假设密度函数为正，即 $\forall x \in \mathbb{R}$, $p(x) > 0$. 考虑 $Y_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leqslant x) = \frac{1}{n}\sum_{i=1}^{n} I_i(x)$.$E(I_i(x)) = F_X(x)$, $Var(I_i(x)) = (1 - F_X(x))F_X(x)$. 所以有

$$\sqrt{n}(Y_n(x) - F(x)) \to A \sim N(0, F_X(x)(1 - F_X(x))).$$

考虑 $g(t) = F_X^{-1}(t)$, $0 < t < 1$. 则 $g'(t) = \frac{1}{p(F_X^{-1}(t))}$. 则,

$$\sqrt{n}(F_X^{-1}(Y_n(x)) - F_X^{-1}(F_X(x))) \to B \sim N(0, \frac{F_X(x)(1 - F_X(x))}{(p(F_X^{-1}(F_X(x))))^2}),$$

令 $x = X_{(nq)}$, 则 $F_X^{-1}(F_X(X_{(nq)})) = X_{(nq)}$ 而且 $|x_q - F_X^{-1}(Y_n(X_{(nq)}))| = O(\frac{1}{n}) \to 0$, 于是 $\sqrt{n}(X_{(nq)} - x_q) \to C \sim N(0, \frac{q(1-q)}{p^2(x_q)})$.

- (大概原因) 假设密度函数为正，即 $\forall x \in \mathbb{R}$, $p(x) > 0$. 考虑 $Y_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \leqslant x) = \frac{1}{n}\sum_{i=1}^n I_i(x)$. $E(I_i(x)) = F_X(x)$, $Var(I_i(x)) = (1 - F_X(x))F_X(x)$. 所以有

$$\sqrt{n}(Y_n(x) - F(x)) \to A \sim N(0, F_X(x)(1 - F_X(x))).$$

考虑 $g(t) = F_X^{-1}(t)$, $0 < t < 1$. 则 $g'(t) = \frac{1}{p(F_X^{-1}(t))}$. 则，

$$\sqrt{n}(F_X^{-1}(Y_n(x)) - F_X^{-1}(F_X(x))) \to B \sim N(0, \frac{F_X(x)(1 - F_X(x))}{(p(F_X^{-1}(F_X(x))))^2}),$$

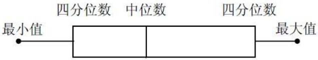
令 $x = X_{(nq)}$，则 $F_X^{-1}(F_X(X_{(nq)})) = X_{(nq)}$ 而且 $|x_q - F_X^{-1}(Y_n(X_{(nq)}))| = O(\frac{1}{n}) \to 0$，于是 $\sqrt{n}(X_{(nq)} - x_q) \to C \sim N(0, \frac{q(1-q)}{p^2(x_q)})$. 本页内容不做考试要求

五线概括与箱线图

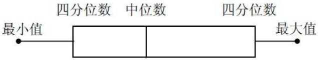
- 考虑 $x_{(1)}$，$m_{0.25}$，$m_{0.5}$，$m_{0.75}$，$x_{(n)}$. $m_{0.25}$ 与 $m_{0.75}$ 有称四分位数。

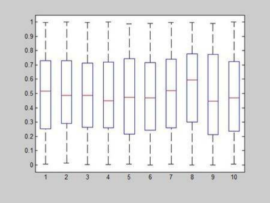
五线概括与箱线图

- 考虑 $x_{(1)}$，$m_{0.25}$，$m_{0.5}$，$m_{0.75}$，$x_{(n)}$. $m_{0.25}$ 与 $m_{0.75}$ 有称四分位数。
- 箱线图：



- 考虑 $x_{(1)}$，$m_{0.25}$，$m_{0.5}$，$m_{0.75}$，$x_{(n)}$. $m_{0.25}$ 与 $m_{0.75}$ 有称四分位数。
- 箱线图：





三大抽样分布

- 卡方 (χ^2) 分布：设 $X_1, \ldots, X_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $\chi^2 = X_1^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$.

三大抽样分布

- 卡方 (χ^2) 分布：设 $X_1, \ldots, X_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $\chi^2 = X_1^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$.
- $X_i^2 \sim Ga(1/2, 1/2)$, $\chi^2 \sim Ga(n/2, 1/2)$.

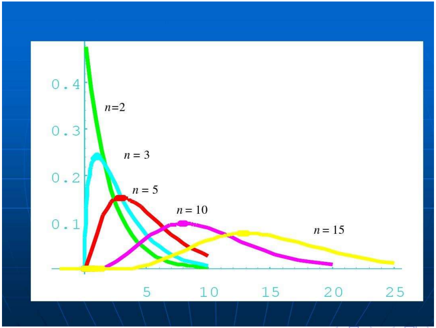
- 卡方 (χ^2) 分布：设 $X_1, \ldots, X_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $\chi^2 = X_1^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$.
- $X_i^2 \sim Ga(1/2, 1/2)$, $\chi^2 \sim Ga(n/2, 1/2)$.

$$p(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} y^{n/2-1} e^{-y/2}.$$

三大抽样分布

- 卡方 (χ^2) 分布：设 $X_1, \ldots, X_n$ 独立同分布于标准正态分布 $N(0,1)$，则 $\chi^2 = X_1^2 + \cdots + X_n^2$ 的分布称为自由度为 n 的 χ^2 分布，记为 $\chi^2(n)$.
- $X_i^2 \sim Ga(1/2, 1/2)$, $\chi^2 \sim Ga(n/2, 1/2)$.

$$p(y) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} y^{n/2-1} e^{-y/2}.$$

- 

卡方分布

- $x_1, \ldots, x_n$ 时来自正态分布 $N(\mu, \sigma^2)$ 的一个样本，其中 μ 时已知量。考虑统计量
$$T = \sum_{i=1}^{n} (x_i - \mu)^2.$$
- $$\frac{T}{\sigma^2} = \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma})^2 \sim \chi^2(n).$$

卡方分布

- $x_1,\dots,x_n$ 时来自正态分布 $N(\mu,\sigma^2)$ 的一个样本，其中 μ 时已知量。考虑统计量
$$T=\sum_{i=1}^{n}(x_i-\mu)^2.$$
- $$\frac{T}{\sigma^2}=\sum_{i=1}^{n}(\frac{x_i-\mu}{\sigma})^2\sim\chi^2(n).$$
- $\chi^2(n)$ 分布的特征函数为 $(1-2it)^{-n/2}$，则 $(1-i2\sigma^2t)^{-n/2}$ 为 T 的特征函数。

卡方分布

- $x_1, \ldots, x_n$ 时来自正态分布 $N(\mu, \sigma^2)$ 的一个样本，其中 μ 时已知量。考虑统计量
$$T = \sum_{i=1}^{n} (x_i - \mu)^2.$$
- $$\frac{T}{\sigma^2} = \sum_{i=1}^{n} (\frac{x_i - \mu}{\sigma})^2 \sim \chi^2(n).$$
- $\chi^2(n)$ 分布的特征函数为 $(1 - 2it)^{-n/2}$，则 $(1 - i2\sigma^2 t)^{-n/2}$ 为 T 的特征函数。
- $Ga(\alpha, \lambda)$： $p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$，$(1 - \frac{it}{\lambda})^{-\alpha}$.

卡方分布

- $x_1,\ldots,x_n$ 时来自正态分布 $N(\mu,\sigma^2)$ 的一个样本，其中 μ 时已知量。考虑统计量
$$T=\sum_{i=1}^n (x_i-\mu)^2.$$
- $$\frac{T}{\sigma^2}=\sum_{i=1}^n (\frac{x_i-\mu}{\sigma})^2 \sim \chi^2(n).$$
- $\chi^2(n)$ 分布的特征函数为 $(1-2it)^{-n/2}$，则 $(1-i2\sigma^2 t)^{-n/2}$ 为 T 的特征函数。
- $Ga(\alpha,\lambda)$： $p(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$，$(1-\frac{it}{\lambda})^{-\alpha}$.
- $T\sim Ga(n/2,\frac{1}{2\sigma^2})$. $p_T(t)=\frac{1}{(2\sigma^2)^{n/2}\Gamma(n/2)}e^{-\frac{t}{2\sigma^2}}t^{n/2-1}$.

卡方分布

- 设 $x_1, \ldots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2.$$

卡方分布

- 设 $x_1, \ldots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2.$$

- $\bar{x}$ 与 s^2 相互独立。

卡方分布

- 设 $x_1, \ldots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2.$$

- $\bar{x}$ 与 s^2 相互独立。
- $\bar{x} \sim N(\mu, \sigma^2/n)$.

卡方分布

- 设 $x_1, \dots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2.$$

- $\bar{x}$ 与 s^2 相互独立。
- $\bar{x} \sim N(\mu, \sigma^2/n)$.
- $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$.

卡方分布

- 设 $x_1, \ldots, x_n$ 是来自正态分布 $N(\mu, \sigma^2)$ 的样本，其样本均值和样本方差分别为

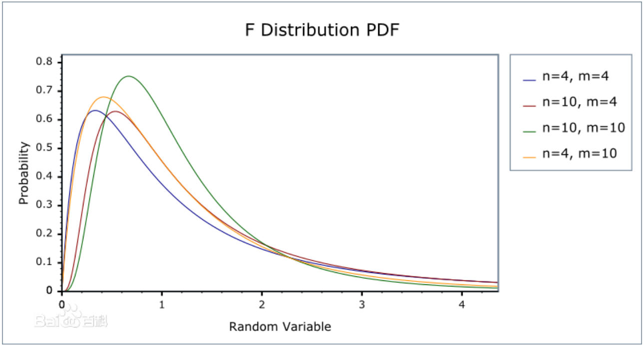
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_i - \bar{x})^2.$$

- $\bar{x}$ 与 s^2 相互独立。
- $\bar{x} \sim N(\mu, \sigma^2/n)$.
- $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$.
- 记号：$\chi^2_{1-\alpha}(n)$ 为自由度为 n 的 χ^2 分布的 $1-\alpha$ 分位数：$P(\chi^2 \leqslant \chi^2_{1-\alpha}(n)) = 1-\alpha$.

F 分布

- 若 $X_1 \sim \chi^2(m)$, $X_2 \sim \chi^2(n)$, 且 X_1 与 X_2 相互独立，则 $F = \frac{X_1/m}{X_2/n}$ 的分布为自由度为 m 和 n 的 F 分布，记为 $F \sim F(m, n)$.

F 分布

- 若 $X_1 \sim \chi^2(m)$, $X_2 \sim \chi^2(n)$, 且 X_1 与 X_2 相互独立，则 $F = \frac{X_1/m}{X_2/n}$ 的分布为自由度为 m 和 n 的 F 分布，记为 $F \sim F(m,n)$.
- 其密度函数为 $\frac{\Gamma(\frac{m+n}{2})(\frac{m}{n})^{m/2-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} y^{m/2-1}(1+\frac{m}{n}y)^{-\frac{m+n}{2}}$
- 

F 分布

- 记 $F_{1-\alpha}(m,n)$ 为自由度为 m, n 的 F 分布的 $1-\alpha$ 分位数。则

$$F_\alpha(n,m) = \frac{1}{F_{1-\alpha}(m,n)}.$$

F 分布

- 记 $F_{1-\alpha}(m,n)$ 为自由度为 m, n 的 F 分布的 $1-\alpha$ 分位数。则

$$F_\alpha(n,m)=\frac{1}{F_{1-\alpha}(m,n)}.$$

$$F\sim F(m,n),\quad \frac{1}{F}\sim F(n,m),$$

F 分布

- 记 $F_{1-\alpha}(m,n)$ 为自由度为 m, n 的 F 分布的 $1-\alpha$ 分位数。则

$$F_\alpha(n,m) = \frac{1}{F_{1-\alpha}(m,n)}.$$

$$F \sim F(m,n), \quad \frac{1}{F} \sim F(n,m),$$

$$\alpha = P(\frac{1}{F} \leqslant F_\alpha(n,m)) = P(F \geqslant \frac{1}{F_\alpha(n,m)})$$

F 分布

- 记 $F_{1-\alpha}(m,n)$ 为自由度为 m, n 的 F 分布的 $1-\alpha$ 分位数。则

$$F_\alpha(n,m)=\frac{1}{F_{1-\alpha}(m,n)}.$$

$$F\sim F(m,n),\quad \frac{1}{F}\sim F(n,m),$$

$$\alpha=P(\frac{1}{F}\leqslant F_\alpha(n,m))=P(F\geqslant\frac{1}{F_\alpha(n,m)})$$

$$P(F\leqslant\frac{1}{F_\alpha(n,m)})=1-\alpha.$$

F 分布

- 设 $x_1,\ldots,x_m$ 为来自 $N(\mu_1,\sigma_1^2)$ 的样本, $y_1,\ldots,y_n$ 为来自 $N(\mu_2,\sigma_2)$ 的样本，而且这两个样本相互独立，s_x^2 和 s_y^2 为它们的样本方差，则

$$F=\frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2}\sim F(m-1,n-1).$$

F 分布

- 设 $x_1, \ldots, x_m$ 为来自 $N(\mu_1, \sigma_1^2)$ 的样本, $y_1, \ldots, y_n$ 为来自 $N(\mu_2, \sigma_2)$ 的样本，而且这两个样本相互独立，s_x^2 和 s_y^2 为它们的样本方差，则

$$F = \frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2} \sim F(m-1, n-1).$$

- $s_x^2/\sigma_1^2 \sim \chi^2(m-1)$, $s_y^2/\sigma_2^2 \sim \chi^2(n-1)$.

F 分布

- 设 $x_1,\ldots,x_m$ 为来自 $N(\mu_1,\sigma_1^2)$ 的样本, $y_1,\ldots,y_n$ 为来自 $N(\mu_2,\sigma_2)$ 的样本，而且这两个样本相互独立，s_x^2 和 s_y^2 为它们的样本方差，则

$$F=\frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2}\sim F(m-1,n-1).$$

- $s_x^2/\sigma_1^2\sim\chi^2(m-1)$, $s_y^2/\sigma_2^2\sim\chi^2(n-1)$.
- $F\sim F(m-1,n-1)$.

F 分布-例子

- 设 $x_1, \ldots, x_{15}$ 是总体 $N(0, \sigma^2)$ 的一个样本，求
$$y = \frac{x_1^2 + \cdots + x_{10}^2}{2(x_{11}^2 + \cdots + x_{15}^2)}$$
的分布。

F 分布-例子

- 设 $x_1, \dots, x_{15}$ 是总体 $N(0, \sigma^2)$ 的一个样本，求

$$y = \frac{x_1^2 + \cdots + x_{10}^2}{2(x_{11}^2 + \cdots + x_{15}^2)}$$

  的分布。
- $(x_1^2 + \cdots + x_{10}^2)/\sigma^2 \sim \chi^2(10)$.

F 分布-例子

- 设 $x_1, \ldots, x_{15}$ 是总体 $N(0, \sigma^2)$ 的一个样本，求

$$y = \frac{x_1^2 + \cdots + x_{10}^2}{2(x_{11}^2 + \cdots + x_{15}^2)}$$

  的分布。
- $(x_1^2 + \cdots + x_{10}^2)/\sigma^2 \sim \chi^2(10)$.
- $(x_{11}^1 + \cdots + x_{15}^2)/\sigma^2 \sim \chi^2(5)$.

F 分布-例子

- 设 $x_1, \ldots, x_{15}$ 是总体 $N(0, \sigma^2)$ 的一个样本，求
$$y = \frac{x_1^2 + \cdots + x_{10}^2}{2(x_{11}^2 + \cdots + x_{15}^2)}$$
  的分布。
- $(x_1^2 + \cdots + x_{10}^2)/\sigma^2 \sim \chi^2(10)$.
- $(x_{11}^1 + \cdots + x_{15}^2)/\sigma^2 \sim \chi^2(5)$.
- $y = \frac{(x_1^2+\cdots+x_{10}^2)/\sigma^2/10}{(x_{11}^1+\cdots+x_{15}^2)/\sigma^2/5} \sim F(10, 5)$.

t 分布

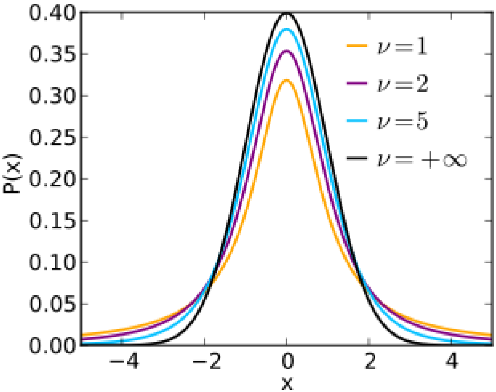
- 设随机变量 X 与 Y 相互独立且 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 则 $t = \frac{X_1}{\sqrt{Y/n}}$ 的分布为自由度为 n 的 t 分布。记 $t \sim t(n)$.

t 分布

- 设随机变量 X 与 Y 相互独立且 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 则 $t = \frac{X_1}{\sqrt{Y/n}}$ 的分布为自由度为 n 的 t 分布。记 $t \sim t(n)$.
- 其密度函数为 $\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{y^2}{n})^{-\frac{n+1}{2}}$, $y \in \mathbb{R}$.

t 分布

- 设随机变量 X 与 Y 相互独立且 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 则 $t = \frac{X_1}{\sqrt{Y/n}}$ 的分布为自由度为 n 的 t 分布。记 $t \sim t(n)$.
- 其密度函数为 $\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{y^2}{n})^{-\frac{n+1}{2}}$, $y \in \mathbb{R}$.



-

t 分布

- 自由度为 1 的 t 分布是柯西分布。

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n>1$，期望存在，均为 0，$n>2$ 时方差存在，为 $n/(n-2)$.

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n>1$，期望存在，均为 0，$n>2$ 时方差存在，为 $n/(n-2)$.
- $n\geqslant 30$ 时，可用标准正态分布逼近。

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n > 1$，期望存在，均为 0，$n > 2$ 时方差存在，为 $n/(n-2)$.
- $n \geqslant 30$ 时，可用标准正态分布逼近。
- $t_{1-\alpha} = -t_\alpha$.
- $x_1, \dots, x_n$ 为正态分布 $N(\mu, \sigma^2)$ 的一个样本，则 $t = \frac{\sqrt{n}(\bar{x}-\mu)}{s} \sim t(n-1)$.

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n > 1$，期望存在，均为 0，$n > 2$ 时方差存在，为 $n/(n-2)$.
- $n \geqslant 30$ 时，可用标准正态分布逼近。
- $t_{1-\alpha} = -t_\alpha$.
- $x_1, \ldots, x_n$ 为正态分布 $N(\mu, \sigma^2)$ 的一个样本，则
  $t = \frac{\sqrt{n}(\bar{x}-\mu)}{s} \sim t(n-1)$.

$$\frac{\sqrt{n}(\bar{x}-\mu)}{s} = \frac{(\bar{x}-\mu)/(\sigma/\sqrt{n})}{\sqrt{(n-1)s^2/\sigma^2/(n-1)}}.$$

t 分布

- 自由度为 1 的 t 分布是柯西分布。
- $n>1$，期望存在，均为 0，$n>2$ 时方差存在，为 $n/(n-2)$.
- $n \geqslant 30$ 时，可用标准正态分布逼近。
- $t_{1-\alpha} = -t_\alpha$.
- $x_1,\ldots,x_n$ 为正态分布 $N(\mu,\sigma^2)$ 的一个样本，则
  $t = \frac{\sqrt{n}(\bar{x}-\mu)}{s} \sim t(n-1)$.

$$\frac{\sqrt{n}(\bar{x}-\mu)}{s} = \frac{(\bar{x}-\mu)/(\sigma/\sqrt{n})}{\sqrt{(n-1)s^2/\sigma^2/(n-1)}}.$$

- $y_1,\ldots,y_m$ 为 $N(\mu',\sigma^2)$ 的样本，则若记 $s_w^2 = \frac{(n-1)s_x^2+(m-1)s_y^2}{m+n-2}$，则

$$\frac{(\bar{x}-\bar{y})-(\mu-\mu')}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2).$$