Lec-17. 数理统计介绍、随机样本、统计量

主讲教师：吴利苏 (**wulisu@sdust.edu.cn**)
主　　页：wulisu.cn

本节课内容

1. 数理统计部分介绍

2. 统计量与常用统计量

3. 三个重要抽样分布
   - χ^2 分布, t 分布, F 分布

数理统计

- 数理统计是以概率论为基础，根据试验或观察数据，来研究随机现象，对研究对象的客观规律做出合理的估计和判断.

数理统计

- 数理统计是以概率论为基础，根据试验或观察数据，来研究随机现象，对研究对象的客观规律做出合理的估计和判断.
- 数理统计主要内容:

数理统计

- 数理统计是以概率论为基础，根据试验或观察数据，来研究随机现象，对研究对象的客观规律做出合理的估计和判断.
- 数理统计主要内容:
  - 数据收集（获取、预处理、数据清洗);

数理统计

- 数理统计是以概率论为基础，根据试验或观察数据，来研究随机现象，对研究对象的客观规律做出合理的估计和判断.
- 数理统计主要内容：
    - 数据收集（获取、预处理、数据清洗）;
    - 数据处理（聚类分析、特征分析、降维分析、主成分分析、特征提取）;

数理统计

- 数理统计是以概率论为基础，根据试验或观察数据，来研究随机现象，对研究对象的客观规律做出合理的估计和判断.
- 数理统计主要内容：
  - 数据收集（获取、预处理、数据清洗）;
  - 数据处理（聚类分析、特征分析、降维分析、主成分分析、特征提取）;
  - 结果分析（统计推断）.

概率论和数理统计

- 在概率论中，已知随机变量的分布的前提下，研究它的性质、特点和规律性. 例如，求数字特征 $E(X), D(X)$、求随机变量函数的分布 $F(x)$.
- 在数理统计中，随机变量的分布未知或者部分未知，通过数据对随机变量作出推断.

数理统计学习内容

Chap-6 总体、随机样本、统计量、常用的统计量和抽样分布

- χ^2 分布、t 分布、F 分布.

Chap-7 估计问题

- 点估计：分布函数已知，参数未知，估计未知参数.
  - ★ 矩估计、极大似然估计.
- 区间估计：对参数处在某个区间的可信程度的估计.

Chap-8 假设检验问题

- 分布函数未知，或分布函数只知道形式而参数未知，提出某些假设，进行检验推断.

基本概念

- 总体　试验的全部可能的观察值;

基本概念

- 总体　试验的全部可能的观察值;
- 个体　总体中的每个可能观察值;

基本概念

- 总体　试验的全部可能的观察值;
- 个体　总体中的每个可能观察值;
- 总体的容量　总体中所包含的个体数;

基本概念

- 总体　试验的全部可能的观察值;
- 个体　总体中的每个可能观察值;
- 总体的容量　总体中所包含的个体数;
- 有限总体　容量有限的总体;

基本概念

- 总体　试验的全部可能的观察值;
- 个体　总体中的每个可能观察值;
- 总体的容量　总体中所包含的个体数;
- 有限总体　容量有限的总体;
- 无限总体　容量无限的总体，通常将容量非常大的有限总体也按无限总体处理.

例

- 研究 2000 名学生的年龄, 这些学生的年龄的全体构成一个总体, 每个学生的年龄就是个体.
- 一湖泊中某种鱼的含汞量, 有限总体.
- 一城市空气质量，PM2.5 值, 无限总体.
- 考察全国正在使用的某种型号灯泡的寿命所形成的总体. 由于可能观察值的个数很多, 可认为是无限总体.

总体分布

- 总体 X 中的每一个个体有一个的取值, 这些取值构成一个分布, 因此 X 可以看成一个随机变量.
- X 的分布函数记为 $F(x)$, 称总体 X 具有分布 $F(x)$.

如何推断总体分布的未知参数（或分布）?

在实际中，往往总体的分布未知或总体的分布已知，但某些参数未知．因此要对总体进行推断，研究所有个体是不可能的，故须抽出部分个体进行研究．

如何推断总体分布的未知参数（或分布）?

在实际中，往往总体的分布未知或总体的分布已知，但某些参数未知. 因此要对总体进行推断，研究所有个体是不可能的，故须抽出部分个体进行研究.

- 样本　从总体中抽出的部分个体.

如何推断总体分布的未知参数（或分布）?

在实际中，往往总体的分布未知或总体的分布已知，但某些参数未知. 因此要对总体进行推断，研究所有个体是不可能的，故须抽出部分个体进行研究.

- 样本　从总体中抽出的部分个体.
- 样本容量　样本中所含个体的个数..

如何推断总体分布的未知参数（或分布）?

在实际中，往往总体的分布未知或总体的分布已知，但某些参数未知．因此要对总体进行推断，研究所有个体是不可能的，故须抽出部分个体进行研究．

- 样本　从总体中抽出的部分个体．
- 样本容量　样本中所含个体的个数..
- 简单随机样本　独立同分布的样本 $(X_1, ..., X_n)$ 称为容量是 n 的简单随机样本，简称样本．

如何推断总体分布的未知参数（或分布）?

在实际中，往往总体的分布未知或总体的分布已知，但某些参数未知. 因此要对总体进行推断，研究所有个体是不可能的，故须抽出部分个体进行研究.

- 样本　从总体中抽出的部分个体.
- 样本容量　样本中所含个体的个数..
- 简单随机样本　独立同分布的样本 $(X_1, ..., X_n)$ 称为容量是 n 的简单随机样本，简称样本.
- 样本值　$X_1, ..., X_n$ 的观察值 $x_1, ..., x_n$.

如何推断总体分布的未知参数（或分布）?

在实际中，往往总体的分布未知或总体的分布已知，但某些参数未知. 因此要对总体进行推断，研究所有个体是不可能的，故须抽出部分个体进行研究.

- 样本　从总体中抽出的部分个体.
- 样本容量　样本中所含个体的个数..
- 简单随机样本　独立同分布的样本 $(X_1, ..., X_n)$ 称为容量是 n 的简单随机样本，简称样本.
- 样本值　$X_1, ..., X_n$ 的观察值 $x_1, ..., x_n$.

注: 后面所有的样本均指简单随机样本.

(简单随机) 抽样

- 获得简单随机样本的抽样称为(简单随机) 抽样.

(简单随机) 抽样

- 获得简单随机样本的抽样称为(简单随机) 抽样.

如何进行 (简单随机) 抽样?

(简单随机) 抽样

- 获得简单随机样本的抽样称为(简单随机) 抽样.

如何进行 (简单随机) 抽样?

- 对于有限总体, 采用放回抽样.

(简单随机) 抽样

- 获得简单随机样本的抽样称为(简单随机) 抽样.

如何进行 (简单随机) 抽样?

- 对于有限总体, 采用放回抽样.
- 对于无限总体, 一般采取不放回抽样.

(简单随机) 抽样

- 获得简单随机样本的抽样称为(简单随机) 抽样.

如何进行 (简单随机) 抽样?

- 对于有限总体, 采用放回抽样.
- 对于无限总体, 一般采取不放回抽样.
- 但当总体容量很大的时候, 不放回抽样对结果影响较小, 因此通常采用不放回抽样, 并将所得到的样本近似当作 (简单随机) 样本来处理.

样本的分布函数和概率密度函数

若 $X_1, \ldots, X_n$ 是总体 X 的样本, X 的分布函数 $F(x)$. 则由独立性得 $X_1, \ldots, X_n$ 的联合分布函数

$$F^*(x_1, \ldots, x_n) = \prod F(x_i).$$

又若 X 具有概率密度 f, 则 $X_1, \ldots, X_n$ 的联合概率密度为

$$f^*(x_1, \ldots, x_n) = \prod f(x_i).$$

例

设一批灯泡的寿命 $X($ 小时$)$ 服从参数为 θ 的指数分布，θ 未知. 从该批灯泡中采用简单随机抽样抽取容量为 10 的样本 $X_1, ..., X_{10}$. 对样本实施观测，得到样本值为

6394	1105	4717	1399	7952
17424	3275	21639	2360	2896

写出样本的概率密度.

解：总体 $X \sim Exp(\theta)$, $X_1, ..., X_{10}$ 为来自总体 X 的一个样本, 则 $(X_1, ..., X_{10})$ 的联合概率密度为

$$f^*(x_1, ..., x_{10}) = \prod_{i=1}^{10} f(x_i)$$

$$= \begin{cases} \frac{1}{\theta^{10}} e^{-\frac{1}{\theta}\sum\limits_{i=1}^{10} x_i} & x_1 > 0, .., x_n > 0; \\ 0 & \text{其他}. \end{cases}$$

$\square$

解：总体 $X \sim Exp(\theta)$, $X_1, ..., X_{10}$ 为来自总体 X 的一个样本, 则 $(X_1, ..., X_{10})$ 的联合概率密度为

$$f^*(x_1, ..., x_{10}) = \prod_{i=1}^{10} f(x_i)$$

$$= \begin{cases} \frac{1}{\theta^{10}} e^{-\frac{1}{\theta}\sum\limits_{i=1}^{10} x_i} & x_1 > 0, .., x_n > 0; \\ 0 & \text{其他}. \end{cases}$$

$\square$

- 如何由已知样本值来估计未知参数 θ?

为了估计指数分布的参数 θ, 进行抽样观测，得到样本 $X_1, ..., X_{10}$ 的样本值

6394	1105	4717	1399	7952
17424	3275	21639	2360	2896

由 $E(X) = \theta$, 所以可以用样本的均值

$$\overline{X} = \frac{1}{10}\sum_{k=1}^{10} x_i$$

估计未知参数 θ.

构造统计量

从样本中提取有用的信息来研究总体的分布及各种数字特征—构造统计量.

- 统计量 设 $X_1, ..., X_n$ 是来自总体 X 的一个样本, $g(X_1, ..., X_n)$ 是 $X_1, ..., X_n$ 的函数, 若 g 中不含未知数, 则称 $g(X_1, ..., X_n)$ 是一个统计量.

注: $X_1, ..., X_n$ 是随机变量, 而统计量 $g(X_1, ..., X_n)$ 是随机变量的一个函数. 设 $x_1, ..., x_n$ 是相应于样本 $X_1, ..., X_n$ 的一个样本值, 则称 $g(x_1, ..., x_n)$ 是 $g(X_1, ..., X_n)$ 的观察值.

例

设 X_1, X_2, X_3 是来自总体 $N(\mu, \sigma^2)$ 的一个样本, 其中 μ 已知, σ^2 未知, 判断下列各式哪些是统计量.

$$T_1 = X_1, \qquad T_2 = X_1 + X_2 e^{X_3}$$

$$T_3 = \frac{1}{3}(X_1 + X_2 + X_3), \qquad T_4 = \max\{X_1, X_2, X_3\}$$

$$T_5 = X_1 + X_2 - 2\mu, \qquad T_6 = \frac{1}{\sigma^2}(X_1^2 + X_2^2 + X_3^2)$$

★常用的统计量★

设 $X_1, ..., X_n$ 是来自总体的一个样本,.

- 样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$;
- 样本方差

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right);$$

样本标准差

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

★常用的统计量★

- 样本 k 阶 (原点) 矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,\ k = 1, 2, \ldots;$
- 样本 k 阶中心矩 $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k.$

注: B_2 与 S^2 不一样. 样本方差 S^2 中, 除 n 会低估方差, 为保证无偏性, 故修正为除 $n-1$.（Chap7-3）

当总体数字特征未知时

- 用样本均值 $\overline{X}$ 估计总体均值 $\mu = E(X)$;

当总体数字特征未知时

- 用样本均值 $\overline{X}$ 估计总体均值 $\mu = E(X)$;
- 用样本方差 S^2 估计总体方差 $\sigma^2 = E(X-\mu)^2$;

当总体数字特征未知时

- 用样本均值 $\overline{X}$ 估计总体均值 $\mu = E(X)$;
- 用样本方差 S^2 估计总体方差 $\sigma^2 = E(X-\mu)^2$;
- 用样本矩 A_k 估计总体原点矩 $\mu_k = E(X^k)$;

当总体数字特征未知时

- 用样本均值 $\overline{X}$ 估计总体均值 $\mu = E(X)$;
- 用样本方差 S^2 估计总体方差 $\sigma^2 = E(X-\mu)^2$;
- 用样本矩 A_k 估计总体原点矩 $\mu_k = E(X^k)$;
- 用样本中心矩 B_k 估计总体中心矩 $\nu_k = E(X-\mu)^k$.

当总体数字特征未知时

- 用样本均值 $\overline{X}$ 估计总体均值 $\mu = E(X)$;
- 用样本方差 S^2 估计总体方差 $\sigma^2 = E(X-\mu)^2$;
- 用样本矩 A_k 估计总体原点矩 $\mu_k = E(X^k)$;
- 用样本中心矩 B_k 估计总体中心矩 $\nu_k = E(X-\mu)^k$.

当总体数字特征未知时

- 用样本均值 $\overline{X}$ 估计总体均值 $\mu = E(X)$;
- 用样本方差 S^2 估计总体方差 $\sigma^2 = E(X-\mu)^2$;
- 用样本矩 A_k 估计总体原点矩 $\mu_k = E(X^k)$;
- 用样本中心矩 B_k 估计总体中心矩 $\nu_k = E(X-\mu)^k$.

这些非常直观的想法，有什么理论依据吗？

当总体数字特征未知时

- 用样本均值 $\overline{X}$ 估计总体均值 $\mu = E(X)$;
- 用样本方差 S^2 估计总体方差 $\sigma^2 = E(X-\mu)^2$;
- 用样本矩 A_k 估计总体原点矩 $\mu_k = E(X^k)$;
- 用样本中心矩 B_k 估计总体中心矩 $\nu_k = E(X-\mu)^k$.

这些非常直观的想法，有什么理论依据吗？

- 样本矩 $\xrightarrow{\quad P \quad}$ 总体距.

性质

若总体 X 的 k 阶矩 $E(X^k) = \mu_k$ 存在，则当 $n \to \infty$ 时，有 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k$.

性质

若总体 X 的 k 阶矩 $E(X^k)=\mu_k$ 存在，则当 $n\to\infty$ 时，有 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\xrightarrow{P}\mu_k$.

证明：由于 $X_1,...,X_n$ 独立且与 X 同分布，

$$E(X_1^k)=E(X_2^k)=...=E(X_n^k)=\mu_k,$$

由辛钦大数定律知，$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\xrightarrow{P}E(A_k)=\mu_k$. □

进一步由依概率收敛的性质知

$$g(A_1,...,A_k)\xrightarrow{P}g(\mu_1,...,\mu_k),$$

其中 g 是连续函数. Chap7 矩估计的理论依据.

当总体分布未知时: 经验分布函数 $\xrightarrow{P=1}$ 总体分布

定义

设 $x_1, \ldots, x_n$ 是来自分布函数 $F(x)$ 的总体 X 的样本观察值. X 的**经验分布函数** $F_n(x)$ 定义为

$$F_n(x) = \frac{\sharp(x_i \leq x)}{n}, \quad -\infty < x < +\infty$$

其中 $\sharp(x_i \leq x)$ 表示 $x_1, \ldots, x_n$ 中小于或等于 x 的个数.

注: 由定义, 当给定样本观察值 $x_1, \ldots, x_n$ 时, $F_n(x)$ 满足分布函数的三个条件:

1. $F_n(x)$ 是 x 的不减函数.
2. $0 \leq F_n(x) \leq 1$, $F(-\infty) = 0$, $F(+\infty) = 1$.
3. $F_n(x)$ 是一个右连续函数.

故 $F_n(x)$ 是一个分布函数.

- 将 $x_1, ..., x_n$ 按自小到大的次序排序，重新编号为 $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$，则

$$F_n(x) = \begin{cases} 0 & x < x_{(1)}; \\ \frac{k}{n} & x_{(k)} \leq x < x_{(k+1)} \quad k = 1, 2, ..., n-1; \\ 1 & x \geq x_{(n)}. \end{cases}.$$

- 将 $x_1, ..., x_n$ 按自小到大的次序排序，重新编号为 $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$，则

$$F_n(x) = \begin{cases} 0 & x < x_{(1)}; \\ \frac{k}{n} & x_{(k)} \le x < x_{(k+1)} \quad k = 1, 2, ..., n-1;. \\ 1 & x \ge x_{(n)}. \end{cases}$$

- 当 $x_1, ..., x_n$ 各不同时，$F_n(x)$ 是以等概率 $\frac{1}{n}$ 取 $x_1, ..., x_n$ 的离散型随机变量的分布函数.

例

设 X 有样本观察值 $-1, 1, 2$, 则

$$F_3(x) = \begin{cases} 0 & x < -1; \\ \frac{1}{3} & -1 \leq x < 1; \\ \frac{2}{3} & 1 \leq x < 2; \\ 1 & x \geq 2. \end{cases}$$

当给定 x 时, $F_n(x)$ 是样本 $X_1, \ldots, X_n$ 的函数, 故它是一个统计量.

定理 (格里汶科定理)

设 $X_1, \ldots, X_n$ 是来自以 $F(x)$ 为分布函数的总体 X 的样本, $F_n(x)$ 是经验分布函数, 则有

$$P\left\{\lim_{n\to\infty}\sup_{-\infty<x<\infty}|F_n(x)-F(x)|=0\right\}=1.$$

上面定理表明 $F_n(x)$ 在整个实数轴上以概率 1 均匀收敛于 $F(x)$. 所以, 当 n 很大时, $F_n(x)$ 可以很好地近似总体分布函数 $F(x)$. 这是以样本推断总体的依据.

抽样分布

- 统计量的分布被称为抽样分布.

抽样分布

- 统计量的分布被称为抽样分布.
- 当总体 X 服从一般分布 (如指数分布、均匀分布等), 要得出统计量 (n 个随机变量的函数) 的分布是很困难的.

抽样分布

- 统计量的分布被称为抽样分布.
- 当总体 X 服从一般分布 (如指数分布、均匀分布等), 要得出统计量 (n 个随机变量的函数) 的分布是很困难的.
- 当总体 X 服从正态分布时，统计量 $\overline{X}, S^2$ 的分布是可以计算的，那么服从什么分布呢？

抽样分布

- 统计量的分布被称为抽样分布.
- 当总体 X 服从一般分布 (如指数分布、均匀分布等), 要得出统计量 (n 个随机变量的函数) 的分布是很困难的.
- 当总体 X 服从正态分布时，统计量 $\overline{X}, S^2$ 的分布是可以计算的，那么服从什么分布呢?
- 下面将介绍数理统计中三个重要的抽样分布—χ^2 分布, t 分布, F 分布.

1. χ^2 分布

定义

设 $X_1, ..., X_n$ 是来自总体 $N(0,1)$ 的样本, 则称统计量

$$\chi^2 = X_1^2 + ... + X_n^2$$

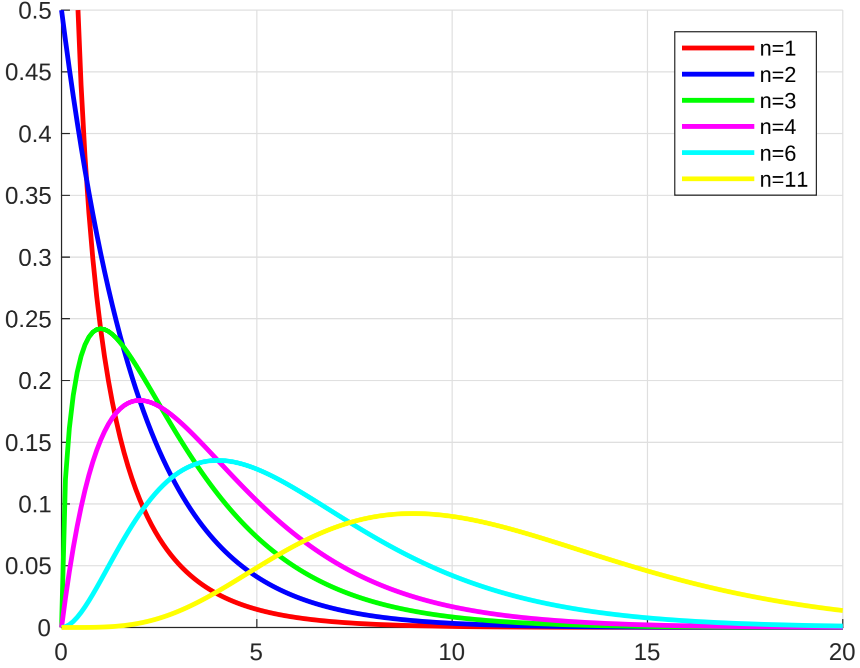
为服从自由度为 n 的χ^2 分布, 记为 $\chi^2 \sim \chi^2(n)$.

$\chi^2(n)$ 的概率密度和图像

$\chi^2(n)$ 的概率密度为

$$f(y)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}y^{\frac{n}{2}-1}e^{-\frac{y}{2}} & y>0;\\ 0 & \text{其他}.\end{cases}.$$

其中 $\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx.$


0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
n=1
n=2
n=3
n=4
n=6
n=11

χ^2 分布和 Γ 分布

$\chi^2(n)$ 的概率密度为

$$f(y) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} & y > 0; \\ 0 & \text{其他}. \end{cases}.$$

$\Gamma(\alpha, \theta)(\alpha > 0, \theta > 0)$ 的概率密度为（Page80）

$$f(x) = \begin{cases} \frac{1}{\theta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\theta}} & x > 0; \\ 0 & \text{其他}. \end{cases}$$

性质

- $\chi^2(1) \sim \Gamma(\frac{1}{2}, 2)$;
- $\chi^2(n) \sim \Gamma(\frac{n}{2}, 2)$.

证明：$X_i \sim N(0,1)$，所以 (Page-52, 80)

$$X_i^2 \sim \chi^2(1) \sim \Gamma(\frac{1}{2}, 2),$$

X_i^2 相互独立，由 Γ 分布的可加性，

$$\chi^2 = \sum_{i=1}^{n} X_i^2 = \Gamma(\frac{n}{2}, 2).$$

χ^2 分布的性质

性质（可加性）

设 $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, 且 χ_1^2, χ_2^2 相互独立, 则

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2).$$

- 一般地, 设 $\chi_i^2 \sim \chi^2(n_i)$, 且 $\chi_i^2(i = 1, ..., m)$ 相互独立, 则

$$\sum_{i=1}^{m} \chi_i^2 \sim \chi^2(n_1 + ... + n_m).$$

χ^2 分布的期望和方差

性质 (期望和方差)

若 $\chi^2 \sim \chi^2(n)$, 则有 $E(\chi^2) = n$, $D(\chi^2) = 2n$.

证: $X_i \sim N(0,1)$, $E(X_i^2) = D(X_i) = 1$,

$$D(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = 3 - 1 = 2.$$

则 $E(\chi^2) = E(\sum\limits_{i=1}^{n} X_2^2) = \sum\limits_{i=1}^{n} E(X_i^2) = n.$

$D(\chi^2) = D(\sum\limits_{i=1}^{n} X_i^2) = \sum\limits_{i=1}^{n} D(X_i^2) = 2n.$ $\square$

注: $X \sim N(0,1)$, 则矩为 $E(X^{2k}) = 1 \times 3 \times \cdots (2k-1)$, $E(X^{2k+1}) = 0$.

χ^2 分布的概率计算—上/下 α 分位数（Page-120）

任意给定随机变量 X, 分布函数 $F(x)$, 概率密度函数 $f(x)$,

- 下 α 分位数

$$P\{X \le \chi_{\underline{\alpha}}\} = F(\chi_{\underline{\alpha}}) = \int_{-\infty}^{\underline{\alpha}} f(x)\,dx = \alpha.$$

- 上 α 分位数

$$P\{X > \chi_{\alpha}\} = 1 - F(\chi_{\alpha}) = \int_{\chi_{\alpha}}^{+\infty} f(x)\,dx = \alpha.$$

χ^2 分布的上 α 分位数

定义 (上分位数)

给定 $0 < \alpha < 1$, 满足条件

$$P\{\chi^2 > \chi^2_\alpha(n)\} = \int_{\chi^2_\alpha(n)}^{\infty} f(y)\,dy = \alpha$$

的 $\chi^2_\alpha(n)$ 称为 χ^2 分布的上 α 分位数.

求 χ^2 分布的上分位数

- 查附录 5(Page-400, $n = 40$ 为止)
  $\alpha = 0.05$, $n = 20$, $\chi^2_{0.05}(20) = 31.410$.
  $\alpha = 0.1$, $n = 25$, $\chi^2_{0.1}(25) = 34.382$.

求 χ^2 分布的上分位数

- 查附录 5(Page-400, $n=40$ 为止)
  $\alpha=0.05$, $n=20$, $\chi^2_{0.05}(20)=31.410$.
  $\alpha=0.1$, $n=25$, $\chi^2_{0.1}(25)=34.382$.
- 当 n 充分大时, 费希尔证明

$$\chi^2_\alpha \approx \frac{1}{2}(z_\alpha+\sqrt{2n-1})^2.$$

其中 z_α 是标准正态分布上的上 α 分位数.

求 χ^2 分布的上分位数

- 查附录 5(Page-400, $n=40$ 为止)
  $\alpha=0.05$, $n=20$, $\chi^2_{0.05}(20)=31.410$.
  $\alpha=0.1$, $n=25$, $\chi^2_{0.1}(25)=34.382$.
- 当 n 充分大时, 费希尔证明

$$\chi^2_\alpha \approx \frac{1}{2}(z_\alpha+\sqrt{2n-1})^2.$$

其中 z_α 是标准正态分布上的上 α 分位数.

- 当 $n>40$ 时, 可用上式求,
  $\chi^2_{0.05}(50)\approx\frac{1}{2}(1.645+\sqrt{99})^2=67.221$

例

设总体 $X \sim N\left(\mu, \sigma^2\right), \mu, \sigma^2$ 已知. $(X_1, \cdots, X_n)$ 是取自总体 X 的样本. 求统计量

$$\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

的分布.

例

设总体 $X \sim N\left(\mu, \sigma^2\right), \mu, \sigma^2$ 已知. $(X_1, \cdots, X_n)$ 是取自总体 X 的样本. 求统计量

$$\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

的分布.

证明：令 $Y_i = \frac{X_i - \mu}{\sigma}$, 则 $Y_i \sim N(0, 1)$.

因此 $\chi^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \sum_{i=1}^{n} Y_i^2 \sim \chi^2(n)$. □

例

1. 已知 $X_1, X_2, \cdots, X_n$ 为总体 X 的 (简单随机) 样本, $E(X)=\mu, D(X)=\sigma^2$. 求 $E(\overline{X}), D(\overline{X}), E(S^2), E(B_2)$.

2. 对于正态总体 X, 求 $D(S^2)$.

3. 已知总体 X 的样本 $X_1, \cdots, X_n$ 和总体 Y 的样本 $Y_1, \cdots, Y_n$ 相互独立, 且 $D(X)=D(Y)=\sigma^2$,

$$S_1^2=\frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\overline{X})^2, \quad S_2^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2$$

令

$$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}.$$

求 $E(S_w^2), D(S_w^2)$.