

概率论与数理统计

Week 5 10/10/2024

概率分布函数

Note

$F(x) = P(X < x)$ 定义为概率分布函数, 具有以下性质:

- 单调不减
- $F(-\infty) = 0, F(\infty) = 1$
- 右连续, 左极限不一定等于函数值, 但是有极限必然等于函数值

概率分布函数可以通过右极限减去左极限回到分布律, 即 $\lim_{x \rightarrow x^+} - \lim_{x \rightarrow x^-}$ 得到。

连续性随机变量及其概率密度

Note

连续性随机变量的分布函数定义为 $F(x) = \int_{-\infty}^x f(t)dt$

其中 $f(x)$ 定义为概率密度函数, 具有以下性质:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $\forall x_1, x_2 (x_2 > x_1), P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(t)dt$
- $F'(x) = f(x)$

均匀分布

Note

若密度函数满足 $f(x) = \begin{cases} \frac{1}{b-a} & x \in (a, b) \\ 0 & \text{others} \end{cases}$ 则称为均匀分布

指数分布

Note

若密度函数满足 $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$ 则称为指数分布

此时有 $F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$

当X服从参数 λ 的指数分布时, 可以记作 $X \sim E(\lambda)$

指数分布具有无记忆性, 即 $P\{x > t_0 + t | X > t_0\} = P\{X > t\}$

正态分布

Note

若密度函数满足 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, 则称为正态分布

记作 $X \sim N(\mu, \sigma)$

标准正态分布可以通过查表取得, 不标准的正态分布可以作变换 $\frac{x-\mu}{\sigma} = t$, 此时有 $P(x \leq b) = \Phi(\frac{x-\mu}{\sigma})$.

Week 6 10/17/2024

随机变量的函数分布

Note

若 $X \sim N(\mu, \sigma^2)$ $Y = g(X)$, 那么 Y 就是随机变量的一个函数分布。

若 X 是离散型随机变量, 那么求 Y 的分布律只需要将 X 的分布律的第一行变为 $g(X)$.

若 X 是连续性随机变量, 可以考虑以下方法:

• 等价法

- 求分布函数 $F(y) = P(\xi \leq y) = P(g(X) \leq y)$, 从而求得 $f_\xi(y) = F'_\xi(y)$
- 具体而言, 首先求 $P(x_1 \leq x \leq x_2), s.t. g(x_1) = g(x_2) = y$.
- 接着作积分 $F(y) = \int_{x_1}^{x_2} f_x(\xi) d\xi$
- 最后求导即得 Y 的密度函数
- e.g. $X \sim N(0, 1)$, 求 $Z = X^2$ 的概率密度 $f_z(Z)$

$$\text{首先考虑求 } Z \text{ 的分布函数 } F_z(Z) = \begin{cases} 0 & z \leq 0 \\ \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} & z > 0 \end{cases}$$

然后求导即可

• 反函数法

- 设 $X \sim f_X(x), -\infty < x < +\infty, g'(x) > 0$ (或 < 0), $Y = g(X)$, 则 Y 具有概率密度为:
- $f_Y(y) = \begin{cases} f_X(h(y)) \cdot |h'(y)| & \alpha < y < \beta \\ 0 & \text{others} \end{cases}$
- 其中 (α, β) 是 Y 的取值范围, $h(x) = g^{-1}(x)$

二维随机变量

Note

设 E 是一个随机实验, 样本空间 $S = e$; 设 $X = X(e)$ 和 $Y = Y(e)$ 是定义在 S 上的随机变量, 由他们构成的向量 (X, Y) 叫做二元随机变量或二维随机变量.

$$P(X = x_i, Y = y_j) = P_{ij}, i, j = 1, 2, \dots$$

离散型随机变量的联合概率分布律

$$P(X = i, Y = j) = P(X = i)P(Y = j|X = i)$$

离散型随机变量的边际分布律

$$P(Y = y_j) = P(X < +\infty, Y = y_j) = \sum_{i=1}^{+\infty} P_{ij} = P_{.j}, j = 1, 2, \dots$$

求某个变量的边际概率 $P(X = x_i)$, 只需要对某一行/某一列求和就可以了.

Week7 10/24/2024

二元随机变量的分布函数

① Note

$F(x, y) = P\{(X \leq x) \cap (Y \leq y)\}$ 记作二维随机变量的一个分布函数.

也就是从(0,0)到(x,y)的一个矩形的面积.显然 $F(x, y)$ 随着x,y的增加是单调不减的.

边际分布函数和条件分布函数

① Note

$$F_X(x) = F(x, +\infty) = P(X \leq x, Y \leq +\infty)$$

称作x的边际分布函数.

① Note

$$F_{X|Y}(x|y) = P(X \leq x | Y = y) = \frac{P(X \leq x, Y = y)}{P(Y = y)}$$

称作 $Y = y$ 条件下,X的条件分布函数.

联合概率密度

① Note

如果二元随机变量的分布函数 $F(X, Y)$,存在非负函数 $f(x, y)$,使得

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(u, v) du dv$$

显然的, $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(u, v) du dv = 1$

边际概率密度

① Note

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

称作X的边际概率密度函数.

条件概率密度

① Note

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$
 记作 $Y=y$ 条件下,X的条件概率密度函数.

二元均匀分布

① Note

也就是在某个二维区域当中概率密度相等

Week 8 10/31/2024

随机变量的独立性

① Note

若 $F(x, y) = F_X(x)F_Y(y)$ 总成立,那么我们就说随机变量 X, Y 是独立的.

此时有性质 $f(x, y) = f_X(x)f_Y(y)$,这样可以显著降低计算量.

- 判断离散型的随机变量是否独立,我们可以画这样的表格:

		Y		
		0	1	P(X=j)
X	1	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{2}$
	2	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{2}$
P(Y=i)		$\frac{1}{3}$	$\frac{2}{3}$	

如果我们发现 $P(x, y) = P(X = x)P(Y = y)$, 也就是说每个单元格等于该行乘该列, 那么我们就说这是独立的.

- 而处理离散型的随机变量, 就先按部就班地算出 $f_X(x)$ 和 $f_Y(y)$, 然后验证独立性就可以了

Note

独立随机变量密度函数的特征:

$$f(x, y) = g(x)h(y)$$

二元正态分布

Warning

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)\right)}$$

$$X \sim N(\mu_1, \sigma_1^2)$$

$$Y \sim N(\mu_2, \sigma_2^2)$$

X, Y 的分布都不依赖于参数 ρ

ρ (相关系数) 影响 X, Y 的独立性, $\rho = 0$ 时, 就有 X, Y 相互独立.

Week 10 11/21

切比雪夫大数定律

Note

若 $X_1, X_2, \dots$ 相互独立, 且具有相同的数学期望 μ 和相同的方差 σ^2 , 那么当 $n \rightarrow \infty$ 时, 有

$$\frac{1}{n} \sum X_k \xrightarrow{P} \mu$$

此时, 有概率估计式

$$P\left(\left|\frac{1}{n} \sum X_k - \mu\right| \geq \epsilon\right) \leq \frac{\frac{1}{n}\sigma^2}{\epsilon^2}$$

辛钦大数定律

① Note

设 $X_1, X_2, \dots$ 独立同分布, $E[X_i] = \mu$, 则当 $n \rightarrow \infty$ 时, 有

$$\frac{1}{n} \sum X_k \xrightarrow{P} \mu$$

构造平均数形式

$X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, g 在 (a, b) 连续, 则 $g(X_n, Y_n) \xrightarrow{P} g(a, b)$

这启示我们可以解决形如 $\frac{X_1+X_2+\dots+X_n}{X_1^2+X_2^2+\dots+X_n^2}$ 这样的题目, 因为上下都可以各自利用切比雪夫大数定律, 然后相除就可以得到结果.

处理 $\frac{1}{n} \sum X_i^2$ 问题时, 经常利用公式 $E(X^2) = D(x) + E^2(x)$

伯努利大数定律

① Note

设 n_A 为 n 重伯努利实验中事件 A 的发生次数, 并记事件 A 在每次实验中发生的概率为 p , 则有,

$$\frac{n_A}{n} \xrightarrow{P} p, \text{ 当 } n \rightarrow +\infty \text{ 时.}$$

独立同分布的中心极限定理

⚠ Warning

设 $X_1, X_2, \dots, X_n$ 独立同分布, 那么当 n 足够大时, 我们可以认为 X 趋近于 **正态分布**, 也就是说,

$$\sum X_i \rightarrow N(n\mu, n\sigma^2)$$

$$\frac{1}{n} \sum X_i \rightarrow N(\mu, \frac{\sigma^2}{n})$$

Week 11 11/28

样本的方差和均值

① Note

- 样本均值

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \implies E\bar{X} = EX, D(\bar{X}) = \frac{1}{n} DX$$

$$D(\bar{x}) = \frac{1}{n} D(x)$$

- 样本方差

$$S^2(x) = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

$$E(S^2(x)) = D(X)$$

- 样本矩

$$k\text{阶(原点)矩: } A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (k = 1, 2, \dots)$$

$$k\text{阶中心矩: } B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \quad (k = 1, 2, \dots)$$

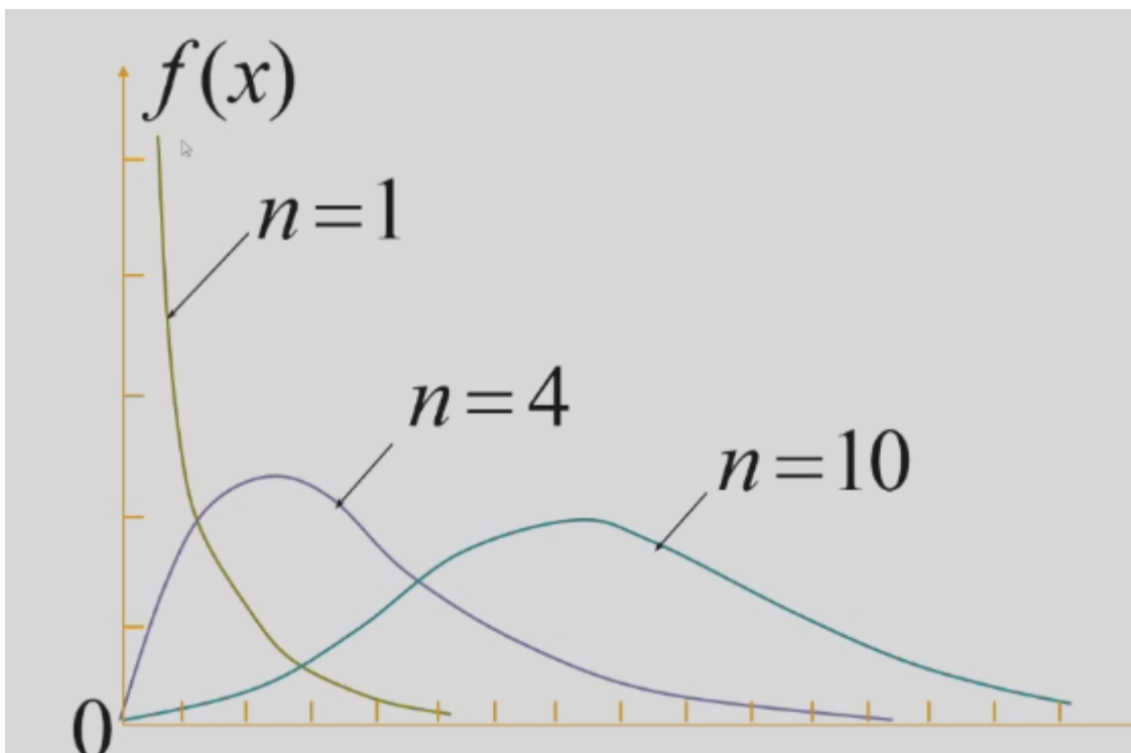
样本分布

① Note

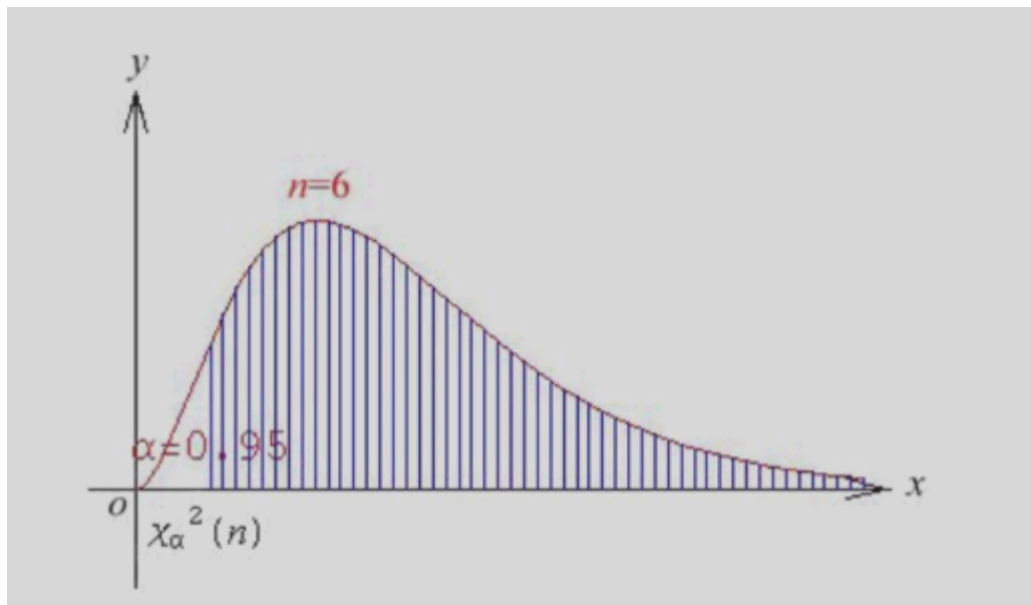
- χ^2 分布
 - 设随机变量 $X_1, X_2, \dots, X_n$ 相互独立, 且 $X_i \sim N(0, 1)$, 那么我们就称

$$\chi_n^2 = \sum_{i=1}^n X_i^2$$

为自由度为 n 的 χ^2 分布.



- 上图为不同自由度下的卡方分布密度函数.
- 卡方分布的重要性质:
 - 设 $\chi^2 \sim \chi^2(n)$, 则有 $E(\chi^2) = n, D(\chi^2) = 2n$
 - 设 $Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2)$, 且 Y_1, Y_2 互相独立, 则有 $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$



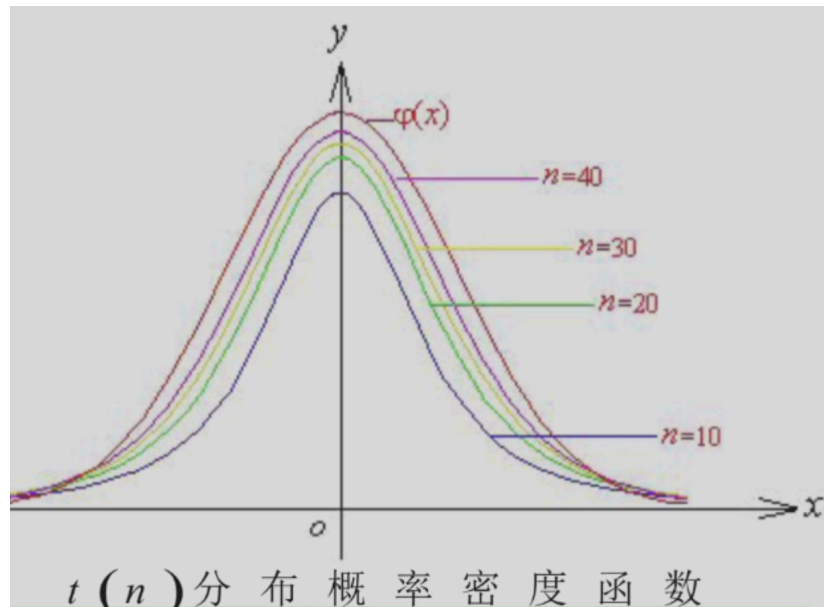
- 对给定的概率 α , $0 < \alpha < 1$, 称满足条件 $\int_{\chi^2_\alpha(n)}^{\infty} f_n(y) dy = \alpha$ 的点 $\chi^2_\alpha(n)$ 为 $\chi^2(n)$ 分布的上 α 分位数, 上 α 分位数 $\chi^2_\alpha(n)$ 的值可查 χ^2 分布表。

① Note

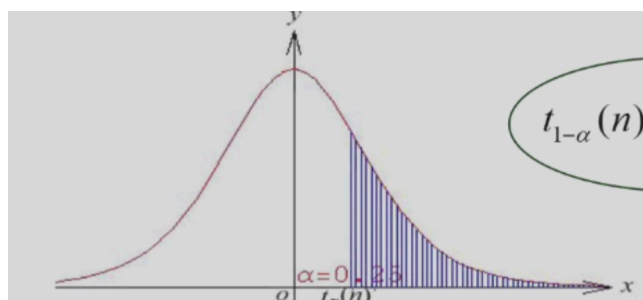
设 $X \sim N(0, 1)$, $Y \sim \chi^2(n)$, 并且假设 X, Y 相互独立,

则称随机变量 $T = \frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布。

记为 $T \sim t(n)$



对给定的 α , $0 < \alpha < 1$, 称满足条件 $\int_{t_\alpha(n)}^{\infty} f(t, n) dt = \alpha$ 的点 $t_\alpha(n)$ 为 $t(n)$ 分布的上 α 分位数。 t 分布的上 α 分位数可查 t 分布表。



① Note

- F 分布

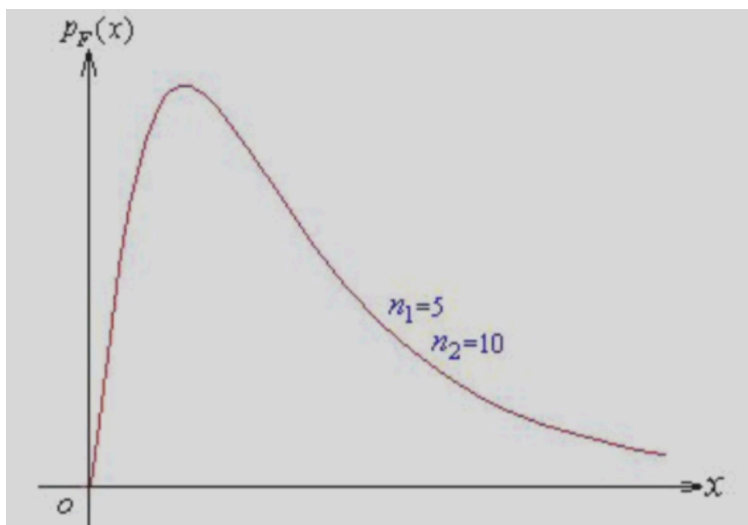
设 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, 且 X, Y 独立, 则

称随机变量 $F = \frac{X/n_1}{Y/n_2}$ 服从自由度 (n_1, n_2) 的 F 分布,

记为 $F \sim F(n_1, n_2)$,

其中 n_1 称为第一自由度, n_2 称为第二自由度。

重要性质: $F \sim F(n_1, n_2)$, 则 $F^{-1} \sim F(n_2, n_1)$



正态总体的抽样分布

设 $X_1, X_2, \dots, X_n$ 为来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本, $\bar{X}$ 是样本均值, S^2 是样本方差, 则有:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

设 $X_1, X_2, \dots, X_n$ 为来自正态总体 $N(\mu, \sigma^2)$ 的简单随机样本, $\bar{X}$ 是样本均值, S^2 是样本方差, 则有:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$\bar{X}$ 与 S^2 相互独立