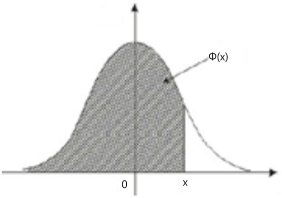
- 设$X \sim N(\mu, \sigma^2)$，$\mu \in \mathbb{R}, \sigma^2 > 0$，其概率密度函数(p.d.f.)

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, x \in \mathbb{R}.$$

- 标准正态$N(0,1)$ 累积分布函数(c.d.f.)

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$



**Theorem (2.1)**

*[【0】定理2.2.1]* 设随机变量$X_1,\dots,X_n$相互独立，且$X_k\sim N(a_k,\sigma_k^2)$，$k=1,\dots,n$。令$c_1,\dots,c_n$为常数，记$T=\sum_{k=1}^n c_kX_k$，则

$$T\sim N(\mu,\tau^2),$$

其中$\mu=\sum_{k=1}^n c_ka_k$，$\tau^2=\sum_{k=1}^n c_k^2\sigma_k^2$。

- 证明参考【0】的证明过程，证明特征函数
$$\phi_T(t)=\mathbb{E}[e^{itT}]=\exp\{it\mu-\tfrac{1}{2}\tau^2t^2\}。$$
- **推论2.1** [【0】推论2.2.1] 在上述定理中，若$a_1=\cdots=a_n=a$，$\sigma_1^2=\cdots=\sigma_n^2=\sigma^2$，则有
$$T\sim N(a\textstyle\sum_{k=1}^n c_k,\sigma^2\sum_{k=1}^n c_k^2).$$
- **推论2.2** [【0】推论2.2.1] 在推论2.1中若取$c_1=\cdots=c_n=1/n$，即$X_1,\dots,X_n\quad i.i.d.\sim N(a,\sigma^2)$，$T=\overline{X}$，则有
$$T\sim N(a,\sigma^2/n).$$

**Theorem (2.2)**

*[【0】定理2.2.2]* 设随机变量$X_1, \dots, X_n \quad i.i.d. \sim N(a, \sigma^2)$，$\vec{X} = (X_1, \dots, X_n)^T$，$\vec{Y} = (Y_1, \dots, Y_n)^T$，$A = (a_{ij})$为$n \times n$常数矩阵，记$\vec{Y} = A\vec{X}$，则

*(1)* $\vec{Y} \sim N_n(\vec{\mu}, \Sigma)$，其中$\vec{\mu} = \begin{pmatrix} a\sum_{k=1}^n a_{1k} \\ \vdots \\ a\sum_{k=1}^n a_{nk} \end{pmatrix}$，$\Sigma = \sigma^2 AA^T$；

*(2)* 当$A$为正交阵时，$Y_1, \dots, Y_n$ 相互独立，且结论*(1)*中$\Sigma = \sigma^2 \mathbf{I}_n$；

*(3)* 在条件*(2)*的基础上，若进一步假设$a = 0$，则

$$Y_1, \dots, Y_n \quad i.i.d. \sim N(0, \sigma^2).$$

- **证明参考【0】的证明过程。**

# 抽样分布之一：$\mathcal{X}^2$分布

- 概率密度函数（p.d.f.）

$$f(x|n) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2}, \quad x > 0, n \in \mathbb{N}^+,$$

- 记$\mathbb{P}(\mathcal{X}_n^2 > \mathcal{X}_n^2(\alpha)) = \alpha$，如下图像参考【0】图2.4.1~2.4.2。

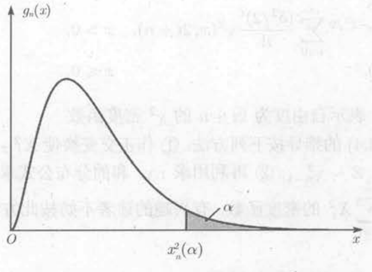



图 2.4.2　$\chi_n^2$ 的上侧 $\alpha$ 分位数

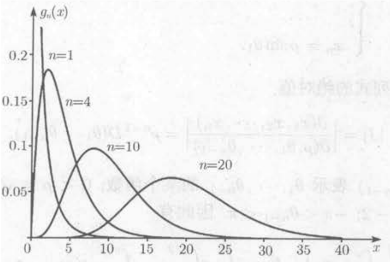



图 2.4.1　$\chi_n^2$ 的密度函数 $g_n(x)$ 形状图

**应用** $\mathcal{X}^2$不常直接用于模型假设中，而常用于假设检验中，尤其是独立性检验，拟合优度检验，生存分析中的秩检验等，也是定义假设检验另两大分布$t$分布和$F$分布的元素。

# $\mathcal{X}^2$分布基本性质

设$X \sim \mathcal{X}_n^2$，则

1. 如果$Z \sim N(0,1)$，则$Z^2 \sim \mathcal{X}_1^2$;
2. 如果$X_1, \ldots, X_p$独立，且$X_i \sim \mathcal{X}_{n_i}^2, i = 1, \ldots, p$，则
$$X_1 + \cdots + X_p \sim \mathcal{X}^2_{n_1+\cdots+n_p};$$
3. $X \sim Gamma(\frac{n}{2}, \frac{1}{2})$;
4. $\mathbb{E}X = n, \quad Var(X) = 2n$;
5. 特征函数 $\phi_X(t) = (1-2it)^{-n/2}$ ;
6. 设$X_1, \ldots, X_n \quad i.i.d. \sim Exp(\lambda)$，则 $2\lambda n\overline{X} = 2\lambda \sum_{i=1}^n X_i \sim \mathcal{X}_{2n}^2$,（参考【0】推论2.4.5）.

# 总体正态的样本均值与样本方差的分布

**Theorem (2.3)**

*[【0】定理2.2.3]* 设$X_1,\ldots,X_n \quad i.i.d. \sim N(\mu,\sigma^2)$，则

(1) $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$;

(2) $\frac{n-1}{\sigma^2}S_X^2 \sim \mathcal{X}_{n-1}^2$;

(3) $\overline{X}$与$S_X^2$相互独立。

- **证明参考【0】的证明过程。**

注1 $\overline{X}$与$S_X^2$的联合分布即为各自分布的乘积;

注2 定理2.3(3)独立性的证明另一方法：证明

1. $\overline{X}$与$X_i-\overline{X}, i=1,\ldots,n$ 均独立;
2. $(n-1)S_X^2 = [\sum_{i=2}^n (X_i-\overline{X})]^2 + \sum_{i=2}^n (X_i-\overline{X})^2$.

注3 定理2.3(3)独立性的证明亦可参考利用本章第5节中Basu定理。

- **定义2.1** 设随机变量$X \sim N(0,1)$，$Y \sim \mathcal{X}_n^2$ ,$n \in \mathbb{N}^+$，且$X$和$Y$相互独立，则称

$$T = \frac{X}{\sqrt{Y/n}}$$

是自由度为$n$的 $t$分布(Student's t distribution with $n$ degrees of freedom)，记为$T \sim t_n$，概率密度函数（p.d.f.）

$$f(x|n) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty.$$

- 记$\mathbb{P}(T > t_n(\alpha)) = \alpha$，则（图像参考【0】图2.4.3~2.4.4）

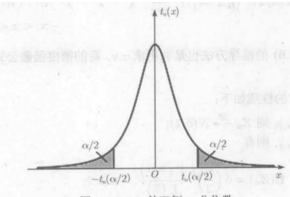



图 2.4.4 $t_n$ 的双侧 $\alpha$ 分位数

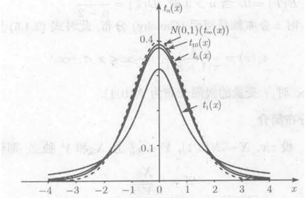



图 2.4.3 $t_n$ 的密度函数 $t_n(x)$ 形状图

# 抽样分布之二：$t$分布

- **$t$分布的基本性质** 若$T \sim t_n$，则
  1. 当$n=1$时，$t_1 = Cauchy(0,1)$;
  2. 当$n \geq 2$时，$\mathbb{E}T = 0$；当$n \geq 3$时，$Var(T) = \frac{n}{n-2}$;
  3. 当$n \to \infty$时，$t_n \to N(0,1)$。参考【0】p38
- **研究动机**：对于简单样本$X_1, \ldots, X_n \quad i.i.d. \sim N(\mu, \sigma^2)$,
  1. 若$\sigma^2$已知，则$\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$;
  2. 若$\sigma^2$未知，则
- **推论2.1** [【0】推论2.4.2]
  设$X_1, \ldots, X_n \quad i.i.d. \sim N(\mu, \sigma^2)$，则

$$\frac{\sqrt{n}(\overline{X}-\mu)}{S_X} \sim t_{n-1}.$$

# 抽样分布之二：$t$分布

注1 推论2.1可用于在总体均值$\mu$和$\sigma^2$均未知的情形下，求$\mu$的区间估计以及检验样本均值$\overline{X}$和假设均值$\mu$ 之间是否存在显著区别。

注2 如果是检验两样本的总体均值是否存在显著区别，可采用如下结论：

- **推论2.2**[【0】推论2.4.3]
设$X_1,\dots,X_m \;\; i.i.d. \sim N(\mu_1,\sigma^2)$, $Y_1,\dots,Y_n \;\; i.i.d. \sim N(\mu_2,\sigma^2)$, 且样本$X_1,\dots,X_m$和$Y_1,\dots,Y_n$独立，则

$$T = \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w}\cdot\sqrt{\frac{mn}{m+n}} \sim t_{m+n-2},$$

其中$(m+n-2)S_w^2 = (m-1)S_X^2+(n-1)S_Y^2$.

# 抽样分布之三：$F$分布

- **定义2.2** 设随机变量$X \sim \mathcal{X}_m^2$, $Y \sim \mathcal{X}_n^2$ , $m, n \in \mathbb{N}^+$, 且$X$和$Y$独立，则称

$$F = \frac{X/m}{Y/n}$$

是自由度为$m$和$n$的$F$分布(Snedecor's F distribution with m and n degrees of freedom), 简记$F \sim F_{m,n}$, 概率密度函数（p.d.f.）

$$f(x|m,n) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2}\frac{x^{\frac{m}{2}-1}}{(1+\frac{m}{n}x)^{(m+n)/2}}, \quad x > 0.$$

- 记$\mathbb{P}(F > F_{m,n}(\alpha)) = \alpha$，则（图像参考【0】图2.4.5∼2.4.6）

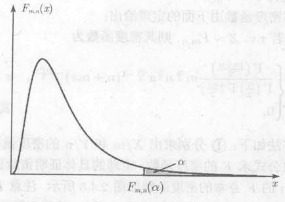



图 2.4.6 $F_{m,n}$ 的上侧 $\alpha$ 分位数

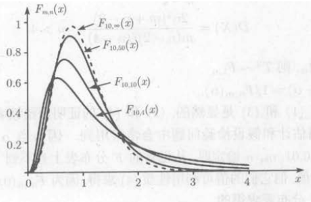



图 2.4.5 $F_{m,n}$ 的密度函数 $f_{m,n}(x)$ 形状图

- **研究动机：**用于检验两样本其总体方差是否存在显著区别，具体见如下结论：
- **推论2.3**[【0】推论2.4.4]
  设$X_1,\ldots,X_m \quad i.i.d. \sim N(\mu_1,\sigma_1^2)$，$Y_1,\ldots,Y_n \quad i.i.d. \sim N(\mu_2,\sigma_2^2)$，且样本$X_1,\ldots,X_m$和$Y_1,\ldots,Y_n$独立，则

$$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F_{m-1,n-1}.$$

- **$F$分布的基本性质：**
  1. 若$Z \sim F_{m,n}$，则 $\frac{1}{Z} \sim F_{n,m}$;
  2. 若$T \sim t_n$，则 $T^2 \sim F_{1,n}$;
  3. 若$Z \sim F_{m,n}$，则 $\mathbb{E}Z = \frac{n}{n-2}, n > 2$;
  4. 若$Z \sim F_{m,n}$，则 $\frac{(m/n)Z}{1+(m/n)Z} \sim Beta(m/2,n/2)$;
  5. 临界值$F_{m,n}(1-\alpha) = \frac{1}{F_{n,m}(\alpha)}$。

  性质4,5分别参考【1】定理5.3.8以及【0】p40

# 作业

- 习题2: Ex. 1, 11, 12, 16, 17, 19.

注 变异系数 $\nu = \frac{\sqrt{Var(X)}}{\mathbb{E}X}$；峰度 $\beta_2 = \frac{\mathbb{E}[X-\mathbb{E}X]^4}{[Var(X)]^2} - 3$。参考【0】p20.