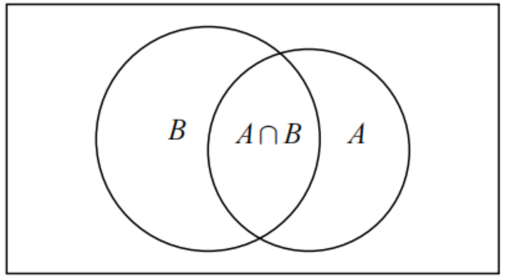
Part 2 数学回顾

概率论与数理统计回顾

- ●概率和结果
  - 结果(outcome)：一个随机过程可能发生的互斥的后果。
  - 概率(probability)：结果出现的概率是指长期观测的结果发生的次数占总样本的比例，也为结果的发生概率。
- ●样本空间和事件
  - 样本空间(sample space)：所有可能结果的集合。
  - 事件(event)：样本空间的子集，即，事件是一个或多个结果的集合。
- ●随机变量(random variable)：一个随机结果的数值概括。
  - 离散型(discrete)：只取离散值，如1,2,...
  - 连续型(continuous)：取一系列可能的连续值，如[0, ∞)

概率与条件概率



条件概率的示意图

- **例** 明天下雨的概率?
- 记事件"下雨"为 A，其发生的"概率"(probability)为P(A)。

- **例** 已知明天会出太阳，则下雨的概率有多大?
- 记事件"出太阳"为 B，则在出太阳的前提条件下，降雨的条件概率(conditional probability)为

$$P(A|B) \equiv \frac{P(AB)}{P(B)}$$

- AB表示事件A 与B同时发生(即交集，也记为$A \cap B$)，故$P(AB)$为 "太阳雨"的概率。
- **例** 股市崩盘的可能性为无条件概率；在已知经济已陷入严重衰退的情况下，股市崩盘的可能性则为条件概率。

独立事件与全概率公式

● 独立事件

- 如果条件概率等于无条件概率， $P(A|B)=P(A)$，即B是否发生不影响A的发生，则称A，B为相互独立的随机事件。
- 此时，$P(A|B)\equiv\frac{P(AB)}{P(B)}=P(A)$，故$P(AB)=P(A)P(B)$

● 全概率公式

- 如果事件组 $\{B_1,B_2,\ldots,B_n\}(n\geq 2)$ 两两互不相容，但必有一件事发生，且每件事的发生概率均为正数，则对任何事件A (无论 A 与$\{B_1,B_2,\ldots,B_n\}$是否有任何关系），都有

$$P(A)=\sum_{i=1}^{n}P(B_i)P(A|B_i)$$

- 全概率公式把世界分成了n个可能得情形$\{B_1,B_2,\ldots,B_n\}$，再把每种情况下的条件概率$P(A|B_i)$ “加权平均”而汇总成无条件概率（权重为每种情形发生的概率$P(B_i)$）。

分布与条件分布

●离散型概率分布

- 假设随机变量X的可能取值为$\{x_1, x_2, \ldots, x_k, \ldots\}$，其对应概率为$\{p_1, p_2, \ldots, p_k, \ldots\}$，即$p_k \equiv P(X = x_k)$，则称$X$为离散型随机变量，其分布律可表示为

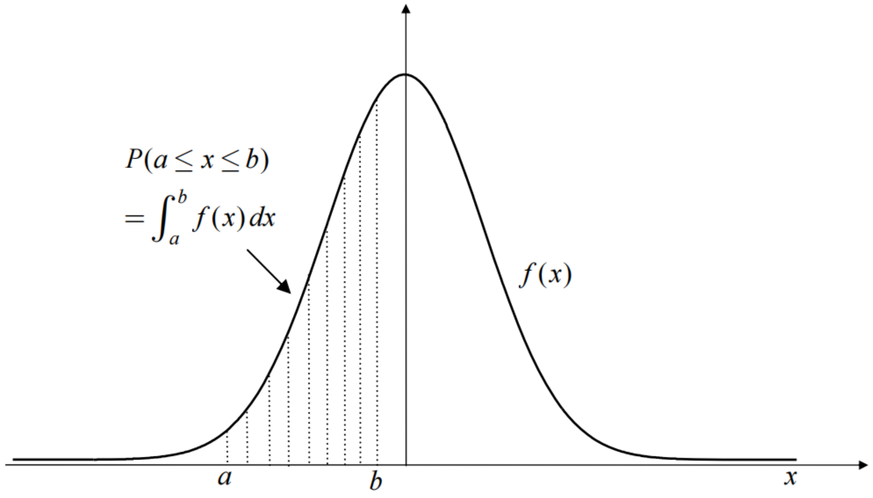
$$
\begin{array}{cccccc}
X & x_1 & x_2 & \cdots & x_k & \cdots \\
p & p_1 & p_2 & \cdots & p_k & \cdots
\end{array}
$$

- 其中，$p_k \geq 0$，$\sum_k p_k = 1$
- 常见的离散分布有两点分布(Bernoulli)、二项分布(Binomial)、泊松分布(Poisson)等。

分布与条件分布

●连续型概率分布

- 连续型随机变量 X可以取任意实数，其**概率密度函数**(probability density function，简记 pdf) $f(x)$满足，
- (1) $f(x) \geq 0$， $\forall x$
- (2) $\int_{-\infty}^{+\infty} f(x)dx = 1$
- (3) X落入区间$[a, b]$的概率为 $P(a \leq X \leq b) = \int_a^b f(x)dx$

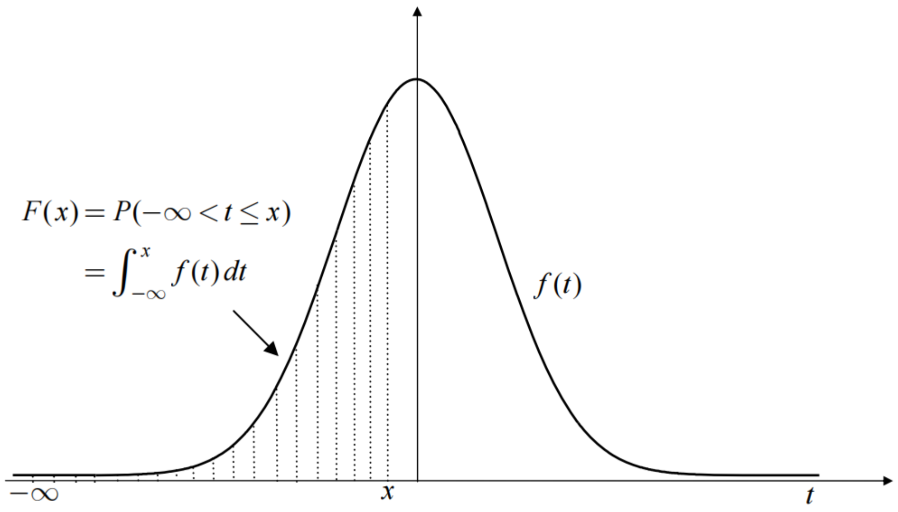


概率密度函数的示意图

- 定义**累积分布函数**(cumulative distribution function，简记 cdf)：

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

- 其中，t为积分变量。 $F(x)$ 度量的是，从$-\infty$至x为止，概率密度函数$f(t)$曲线下的面积。



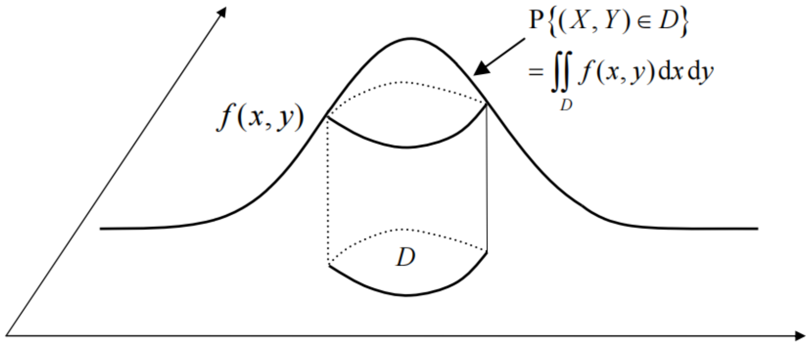
累积分布函数的示意图

●多维随机变量向量的概率分布

- 为研究变量间关系，常同时考虑两个或多个随机变量，即**随机向量**(random vector)。二维连续型随机向量(X, Y) 的**联合密度函数** (joint pdf) $f(x, y)$满足：
- (1) $f(x, y) \geq 0$，$\forall x, y$
- (2) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dxdy = 1$
- (3) (X, Y)落入平面某区域D的概率为$P\{(X, Y) \in D\} = \iint_D f(x, y) dxdy$
- 二维随机向量的联合密度函数就像倒扣的草帽。落入平面某区域*D*的概率就是此草帽下在区域*D* 之上的体积。

- n维连续型随机变量$(X_1, X_2, \dots, X_n)$

可由联合密度函数$f(x_1, x_2, \dots, x_n)$来描述



二维联合密度函数的示意图

- 从二维联合密度 $f(x,y)$，可计算X的（一维）边缘密度函数(marginal pdf)：

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

即给定$\mathrm{X} = x$，把所有Y取值的可能性都“加总”起来（积分的本质就是加总）。

- 类似地，可以计算Y的（一维）边缘密度函数：

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

即给定$\mathrm{Y} = y$，把所有X取值的可能性都“加总”起来。

- 定义二维随机向量(X,Y) 的累积分布函数为：

$$F(x,y) \equiv P(-\infty < X \leq x; -\infty < \mathrm{Y} \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{x} f(t,s)dtds$$

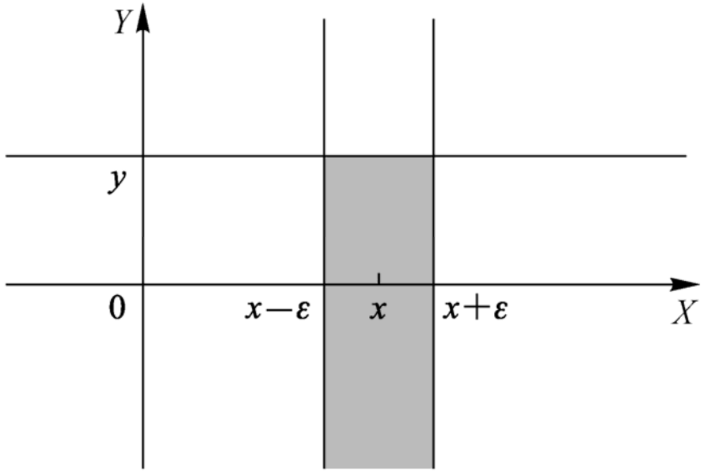
条件分布

- 条件分布(conditional distribution)的概念对于计量至关重要。
- 考虑在$X = x$条件下Y的条件分布，记为$Y|X = x$或$Y|x$
- 对于连续型分布，此条件分布相当于在“草帽”(联合密度函数)上$X = x$的位置垂直地切一刀所得的截面。
- 由于X为连续型随机变量，事件$\{X = x\}$发生的概率为0。

- 如何计算$Y|X=x$的条件概率密度(conditional pdf)?
- 考虑x附近的小邻域$[x-\varepsilon, x+\varepsilon]$
- 计算在$X \epsilon [x-\varepsilon, x+\varepsilon]$条件下$Y$的累积分布函数，即$P\{Y \le y|X \in [x-\varepsilon, x+\varepsilon]\}$，然后让$\varepsilon \to 0^+$，则可证明条件密度函数为，

$$f(y|x) = \frac{f(x,y)}{f_x(x)}$$

- 直观上，此公式与条件概率公式$P(A|B) \equiv \frac{P(AB)}{P(B)}$类似。



条件密度函数的计算

随机变量的数字特征

- **定义** 对于分布律为$p_k \equiv P(X = x_k)$的离散型随机变量X，其**期望**(expectation)为

$$E(X) \equiv \mu \equiv \sum_{k=1}^{\infty} x_k p_k$$

期望的直观含义就是对x_k进行加权平均，而权重为概率p_k

- **定义** 对于概率密度函数为$f(x)$的连续型随机变量X，其期望为

$$E(X) \equiv \mu \equiv \int_{-\infty}^{+\infty} x f(x) dx$$

上式也是对x进行加权平均，权重为概率密度$f(x)$

随机变量的数字特征

- 有时称求期望这种运算为**期望算子**(expectation operator)。
- 期望算子满足**线性性**(linearity)，即对于任意常数k

$$E(X+Y)=E(X)+E(Y),\quad E(kX)=kE(X)$$

- 定义 随机变量X的**方差**(variance)为

$$Var(X) \equiv \sigma^2 \equiv E[X - E(X)]^2$$

- 方差越大，则随机变量取值的波动幅度越大。
- 称方差的平方根为**标准差**(standard deviation)，记为σ。
- 在计算方差时，常利用以下简便公式：

$$Var(X) = E(X^2) - [E(X)]^2$$

- 常需要考虑两个变量之间的相关性，即一个随机变量的取值会对另一随机变量的取值有多大影响。

- 常需要考虑两个变量之间的相关性，即一个随机变量的取值会对另一随机变量的取值有多大影响。
- 定义 随机变量 X与 Y的**协方差**(covariance)为

$$Cov(X,Y) \equiv \sigma_{XY} \equiv E[(X-E(X))(Y-E(Y))]$$

- 如果当随机变量X的取值大于(小于)其期望$E(X)$时，随机变量Y的取值也倾向于大于(小于)其期望值$E(Y)$，则$Cov(X,Y) > 0$ ，二者存在正相关；
- 反之，如果当随机变量X的取值大于(小于)其期望$E(X)$时，随机变量Y的取值反而倾向于小于(大于)其期望值$E(Y)$)，则$Cov(X,Y) < 0$ ，二者存在负相关。
- 如果$Cov(X,Y) = 0$ ，则说明二者线性不相关(uncorrelated)，但不一定相互独立(independent)，因为还可能存在非线性的相关关系。
- 在计算协方差时，常使用以下简便公式：

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

- 协方差的运算也满足线性性，可以证明：

$$Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$$

- 定义 随机变量X 与 Y 的相关系数(correlation)为

$$\rho \equiv Corr(X,Y) \equiv \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

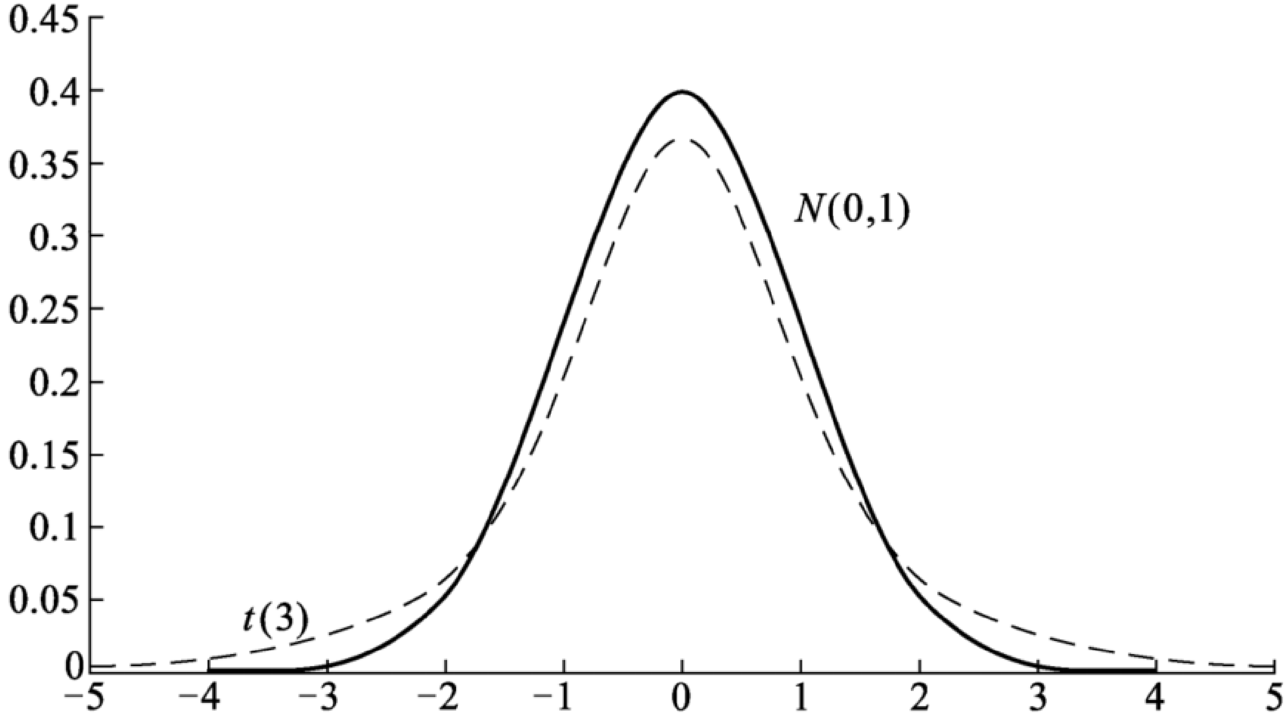
- 相关系数一定介于-1与1之间，即$-1 \leq \rho \leq 1$

- 如果以上各定义式中的积分不收敛，则随机变量的数字特征可能不存在。
- 比如，自由度为1的t分布变量，其期望与方差都不存在。

- 更一般地，对于随机变量X，可以定义一系列的数字特征，即各种矩(moment)的概念。
- **定义** 一阶原点矩为$E(X)$（即期望），二阶原点矩为$E(X^2)$，三阶原点矩为$E(X^3)$，四阶原点矩为$E(X^4)$，等等。
- **定义** 二阶中心矩为$E[X-E(X)]^2$（即方差），三阶中心矩为$E[X-E(X)]^3$，四阶中心矩为$E[X-E(X)]^4$，等等。
- 一阶原点矩(期望)表示随机变量的平均值。
- 二阶中心矩(方差)表示随机变量的波动程度。
- 三阶中心矩表示随机变量密度函数的不对称性(偏度)。
- 四阶中心矩表示随机变量密度函数的最高处(山峰)有多“尖”及尾部有多“厚”(峰度)。

- 三、四阶中心矩还取决于变量的单位。为此，首先将变量“标准化”(即减去期望μ，再除以标准差σ)，并引入以下定义。
- **定义** 随机变量X的**偏度**(skewness)为$E[(X-\mu)/\sigma]^3$
- 如果随机变量为对称分布(比如，正态分布)，则其偏度为 0；因为根据微积分，奇函数在关于原点对称的区间上积分为 0 。

- **定义** 随机变量X 的**峰度**(kurtosis)为$E[(X-\mu)/\sigma]^4$
- 对于正态分布，其峰度为 3。如果随机变量X 的峰度大于 3(比如t 分布)，则其密度函数的最高处(山峰)比正态分布更“尖”，而两侧尾部则更“厚”，称为“厚尾”(fat tails)。存在厚尾的概率分布更容易在尾部取值，称为极端值(outlier)。



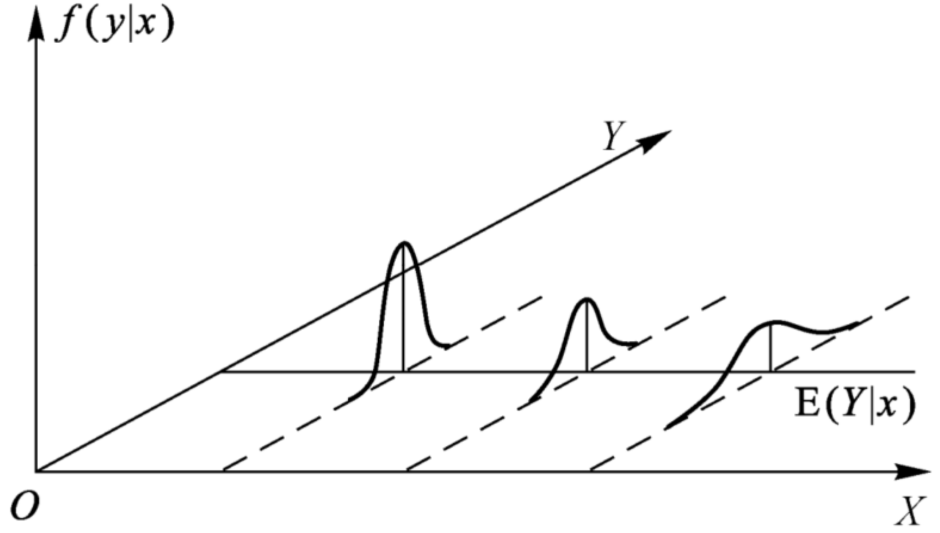
$N(0,1)$与$t(3)$的概率密度

- 定义 随机变量X 的**超额峰度** (excess kurtosis) 为$E[(X-\mu)/\sigma]^4-3$
- 由于正态分布的偏度为 0，峰度为 3，故可使用正态分布的偏度与峰度性质来检验某个分布是否为正态分布。
- 更一般地，对于随机变量 X 与任意函数$g(\cdot)$，称随机变量函数$g(X)$的期望$E[g(X)]=\int_{-\infty}^{+\infty} g(x)f(x)dx$为矩(moment)。

- 定义 **条件期望**(conditional expectation)就是条件分布$Y|x$的期望，即

$$E(Y|X=x)\equiv E(Y|x)=\int_{-\infty}^{+\infty} y f(y|x)\,dy$$

- 由于y已被积分积掉，故$E(Y|x)$只是x的函数。



条件期望与条件方差示意图

- 定义 **条件方差**(conditional variance)就是条件分布$Y|x$的方差

$$Var(Y|X=x) \equiv Var(Y|x) = \int_{-\infty}^{+\infty} \left[y - E[Y|x]\right]^2 f(y|x)dy$$

- 由于y已被积分积掉，故$Var(Y|x)$也只是x的函数。

- 定义 设$\boldsymbol{X} = (X_1 X_2 \ldots X_n)'$为$n$维随机向量，则其**期望**为$n$维列向量：

$$E(\boldsymbol{X}) \equiv E\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \equiv \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

• 定义 设$X=(X_1X_2\cdots X_n)'$为n维随机向量，则其**协方差矩阵** (covariance matrix)为n ×n对称矩阵：

$$
\begin{aligned}
\mathrm{Var}(\boldsymbol{X}) &\equiv \mathrm{E}\left[\left(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X})\right)\left(\boldsymbol{X}-\mathrm{E}(\boldsymbol{X})\right)'\right] \\
&= \mathrm{E}\left[\begin{pmatrix} X_1-\mathrm{E}(X_1) \\ \vdots \\ X_n-\mathrm{E}(X_n) \end{pmatrix}\begin{pmatrix} X_1-\mathrm{E}(X_1) & \cdots & X_n-\mathrm{E}(X_n) \end{pmatrix}\right] \\
&= \mathrm{E}\begin{pmatrix} [X_1-\mathrm{E}(X_1)]^2 & \cdots & [X_1-\mathrm{E}(X_1)][X_n-\mathrm{E}(X_n)] \\ \vdots & \ddots & \vdots \\ [X_1-\mathrm{E}(X_1)][X_n-\mathrm{E}(X_n)] & \cdots & [X_n-\mathrm{E}(X_n)]^2 \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}
\end{aligned}
$$

- 主对角线元素$\sigma_{ii} \equiv Var(X_i)$，非主对角线元素$\sigma_{\mathrm{ij}} \equiv Cov(X_i, X_j)$
- 协方差矩阵必然为半正定矩阵(positive semidefinite)。在一维情况下，这意味着随机变量的方差必然为非负。
- 对于随机向量**X**的期望与协方差矩阵的运算，有如下法则，假设$\boldsymbol{A}$为$m \times n$常数矩阵（不含随机变量），可以证明：
- （1）$E(\boldsymbol{AX}) = \boldsymbol{A}E(\boldsymbol{X})$（期望算子的线性性）
- （2）$Var(\boldsymbol{X}) = E(\boldsymbol{XX}') - E(\boldsymbol{X})[E(\boldsymbol{X})]'$ (一维公式的推广)
- （3）$Var(\boldsymbol{AX}) = \boldsymbol{A}Var(\boldsymbol{X})\boldsymbol{A}'$ (夹心估计量)
- 如果**A**为对称矩阵，则$Var(\boldsymbol{AX}) = \boldsymbol{A}Var(\boldsymbol{X})\boldsymbol{A}$，称为夹心估计量(sandwich estimator)，其中两边的$\boldsymbol{A}$为“面包”，而夹在中间的$Var(\boldsymbol{X})$为“菜”，在形式上类似于三明治。

- 以上随机变量的数字特征都可视为“总体矩”(population moments)。
- 在抽取随机样本后，可用样本数据计算相应的“样本矩”(sample moments)，作为相应总体矩的估计值。
- 即以“求样本平均值运算”$\left(\frac{1}{n}\sum_{i=1}^{n}(\cdot)\right)$来替代总体矩表达式的期望算子$E(\cdot)$
- 比如，可用样本均值(sample mean) $\left(\frac{1}{n}\sum_{i=1}^{n}(X_i)\right)$来估计总体均值(population mean)或期望$E(X)$。

迭代期望定律

- 定理 对于条件期望的运算，有以下重要的**迭代期望定律**(Law of iterated expectation)，
$$E(Y) = E_X[E(Y|x)]$$
- 无条件期望$E(Y)$等于，给定$X = x$情况下Y的条件期望$E(Y|x)$（仍为x 的函数），再对X求期望。
- 如果X为离散随机变量，则根据期望定义，上式可写为：
$$E(Y) = \sum_i P(X = x_i)E(Y|x_i)$$
- 无条件期望等于条件期望之加权平均，而权重为条件“$\mathrm{X} = x$”的概率(取值可能性)。

- 以数据集，验证迭代期望定律，即

$$E(lnw) = E_{rns}[E(lnw|rns)]$$

- 其中， rns为美国南方居民的虚拟变量，取值为0或1。
- 首先，计算$rns = 0$情况下， lnw的条件期望：

Variable	Obs	Mean	Std. Dev.	Min	Max
lnw	554	5.725644	.4129207	4.605	7.051

- 北方居民有554位，其条件期望$E(lnw|rns = 0) =$ 5.725644
- 其次，计算$rns = 1$情况下， lnw的条件期望：

Variable	Obs	Mean	Std. Dev.	Min	Max
lnw	204	5.581083	.4542189	4.718	6.844

- 南方居民有204 位，其条件期望$E(lnw|rns = 1) =$ 5.581083，故美国南方居民的工资略低于北方。

- 以北方与南方居民所占比重作为权重，将北方与南方居民的平均工资对数进行加权平均。
- 5.725644*(554/(554+204))+5.581083*(204/(554+204))=5.6867384
- 最后，用命令直接计算无条件期望。

Variable	Obs	Mean	Std. Dev.	Min	Max
lnw	758	5.686739	.4289494	4.605	7.051

- 二者的结果完全相等，验证迭代期望定律成立。

- 以离散型变量为例，证明等式。
- 假设X的可能取值为$x_1, x_2, \ldots, x_i, \ldots$，而$Y$的可能取值为$y_1, y, \ldots, y_j, \ldots$。记$p_i \equiv P(X = x_i)$，$q_j \equiv P(Y = y_i)$，而$p_{\mathrm{ij}} \equiv P(X = x_i, Y = y_j)$。
- 证明：从迭代期望定律的等式右边开始证明。

$$\mathrm{E}_X\left[\mathrm{E}(Y \mid x)\right] = \sum_i P(X = x_i)\,\mathrm{E}(Y \mid x_i) \qquad \text{(期望的定义式)}$$

$$= \sum_i P(X = x_i)\left[\sum_j P(Y = y_j \mid x_i) \cdot y_j\right] \qquad \text{(条件期望的定义式)}$$

$$= \sum_i \cancel{P(X = x_i)}\left[\sum_j \frac{P(X = x_i, Y = y_j)}{\cancel{P(X = x_i)}} \cdot y_j\right] \qquad \text{(条件概率的定义式)}$$

$$= \sum_i \left[\sum_j P(X = x_i, Y = y_j) \cdot y_j \right] \quad (消去 P(X = x_i))$$

$$= \sum_j \left[\sum_i P(X = x_i, Y = y_j) \cdot y_j \right] \quad (交换加总的次序)$$

$$= \sum_j \left[y_j \sum_i P(X = x_i, Y = y_j) \right] \quad (y_j 与 i 无关，可提出)$$

$$= \sum_j y_j P(Y = y_j) \quad (边缘概率与联合概率的关系)$$

$$= \mathrm{E}(Y) \quad (期望的定义式)$$

- 迭代期望定律很像全概率公式：无条件期望等于条件期望之加权平均，权重为条件概率密度。
- 将迭代期望定律推而广之，对于任意函数$g(\cdot)$，可得

$$E[g(Y)] = E_X E[g(Y)|X]$$

- 有时期望算子E_X的下标被省去，需注意对什么变量求期望。

概率论与数理统计回顾

- 联合、边缘和条件概率/分布
- 例如：

		X	
		0	1
Y	0	1/4	1/3
	1	1/3	1/12

概率论与数理统计回顾

- X的边际分布：
  - $prob(X = 0) = Prob(Y = 0, X = 0) + Prob(Y = 1, X = 0) = \frac{1}{4} + \frac{1}{3} = \frac{7}{12}$
  - $prob(X = 1) = Prob(Y = 0, X = 1) + Prob(Y = 1, X = 1) = \frac{1}{3} + \frac{1}{12} = \frac{5}{12}$

概率论与数理统计回顾

- Y的条件分布，例如$Prob(Y|X=0)$和$prob(Y|X=1)$
  - $prob(Y=0|X=0)=\frac{Prob(Y=0,X=0)}{Prob(X=0)}=\frac{1/4}{7/12}=\frac{3}{7}$
  - $prob(Y=1|X=0)=\frac{Prob(Y=1,X=0)}{Prob(X=0)}=\frac{1/3}{7/12}=\frac{4}{7}$
  - $prob(Y=0|X=1)=\frac{Prob(Y=0,X=1)}{Prob(X=1)}=\frac{1/3}{5/12}=\frac{4}{5}$
  - $prob(Y=1|X=1)=\frac{Prob(Y=0,X=1)}{Prob(X=1)}=1-prob(Y=0|X=1)=1/5$

概率论与数理统计回顾

- Y的条件期望，例如E$(Y|X)$
  - $E(Y|X=0)=\Sigma_y y\cdot Prob(Y=y|X=0)$

    $=0\times Prob(Y=0|X=0)+1\times Prob(Y=1|X=0)=1\times\frac{4}{7}=\frac{4}{7}$
  - $E(Y|X=1)=\Sigma_y y\cdot Prob(Y=y|X=1)$

    $=0\times Prob(Y=0|X=1)+1\times Prob(Y=1|X=1)=1\times\frac{1}{5}=\frac{1}{5}$
  - $E(Y|X)=\begin{cases}\frac{4}{7},\text{概率为}\,\frac{7}{12}\leftarrow Prob(X=0)\\ \frac{1}{5},\text{概率为}\,\frac{5}{12}\leftarrow Prob(X=1)\end{cases}$

  - 我们有$E\big(E(Y|X)\big)=\frac{4}{7}\times\frac{7}{12}+\frac{1}{5}*\frac{5}{12}=\frac{5}{12}$

概率论与数理统计回顾

- Y的无条件期望，例如$\mathrm{E}(Y|X)$
  - $E(Y) = \Sigma_y y \cdot Prob(Y = y)$

    $= 0 \times Prob(Y = 0) + 1 \times Prob(Y = 1) = 0 \times \frac{7}{12} + 1 \times \frac{5}{12} = \frac{5}{12}$

  - 我们有$E\big(E(Y|X)\big) = \frac{4}{7} \times \frac{7}{12} + \frac{1}{5} * \frac{5}{12} = \frac{5}{12}$

  - 注释：$E\big(E(Y|X)\big) = E(Y)$ ←迭代期望定律（LIE）

随机变量无关的三个层次概念

- **定义** 对于连续型随机变量X与Y，如果其联合密度等于边缘密度的乘积，即$f(x,y)=f_x(x)f_y(y)$，则称X与Y**相互独立**(independent)。
- 如果X与Y相互独立，则X与Y没有任何关系，故X的取值不对Y的取值产生任何影响，反之亦然。
- “相互独立”是有关随机变量“无关”的最强概念。
- 线性不相关的概念则更弱，仅要求协方差为0，即$Cov(X,Y)=0$
- “相互独立”意味着“线性不相关”，但反之不然。

- 在二者之间还有一个中间层次的无关概念，即“均值独立”(mean-independence)，在计量中很有用。
- **定义** 假设条件期望$E(Y|x)$存在。如果$E(Y|x)$ 不依赖于X，则称Y 均值独立于X (Y is mean-independent of X)。
- 均值独立不是一种对称的关系，即“Y 均值独立于X”并不意味着“X 均值独立于Y ” 。
- **命题** Y均值独立于X，当且仅当$E(Y|x) = E(Y)$(条件期望等于无条件期望) 。
- **证明**： （1）假设Y均值独立于X，则$E(Y|x)$不依赖于X，故$E_X[E(Y|x)] = E(Y|x)$。根据迭代期望定律， $E(Y) = E_X[E(Y|x)] = E(Y|x)$.
- （2）假设$E(Y|x) = E(Y)$，则显然$E(Y|x)$不依赖于X，故Y均值独立于X。

- **命题** 如果X与Y相互独立，则Y均值独立于X，且X均值独立于Y。
- **定理** 如果Y均值独立于X，或X均值独立于Y，则$Cov(X, Y) = 0$

证明：$\mathrm{Cov}(X,Y)=\mathrm{E}\left[\left(X-\mathrm{E}(X)\right)\left(Y-\mathrm{E}(Y)\right)\right]$ (协方差的定义)

$=\mathrm{E}_X\,\mathrm{E}_Y\left[\left(X-\mathrm{E}(X)\right)\left(Y-\mathrm{E}(Y)\right)\middle|x\right]$ (迭代期望定律)

$=\mathrm{E}_X\left[\left(X-\mathrm{E}(X)\right)\mathrm{E}_Y\left(Y-\mathrm{E}(Y)\middle|x\right)\right]$ $((X-\mathrm{E}(X))$视为常数提出)

$=\mathrm{E}_X\left[\left(X-\mathrm{E}(X)\right)\left(\mathrm{E}(Y\,|\,x)-\mathrm{E}(Y)\right)\right]$ (期望算子的线性性)

$=\mathrm{E}_X\left[\left(X-\mathrm{E}(X)\right)\cdot 0\right]=0$ (均值独立的性质)

- “相互独立”→“均值独立” →“线性不相关”；反之不然。

常用连续型统计分布

- 1.**正态分布：**如果随机变量X的概率密度函数为

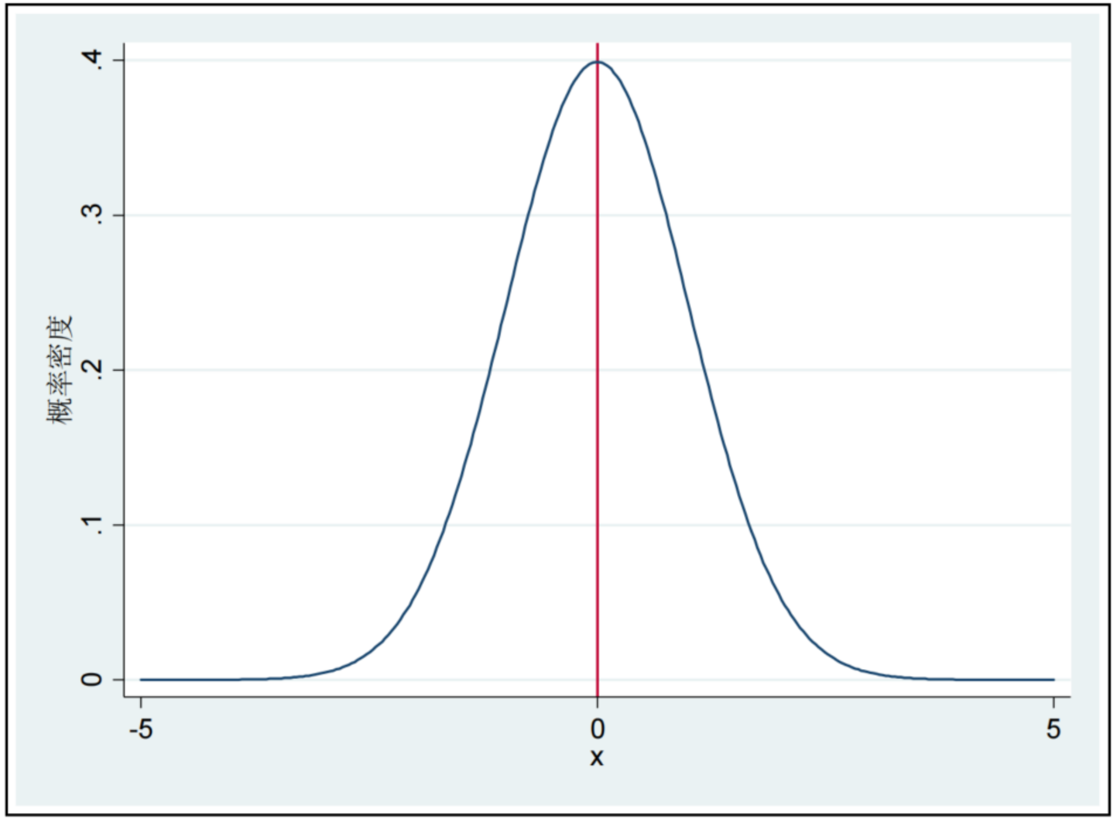
$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$$

则称X服从正态分布(normal distribution)，记为$X\sim N(\mu,\sigma^2)$，其中，μ为期望，而σ^2为方差。

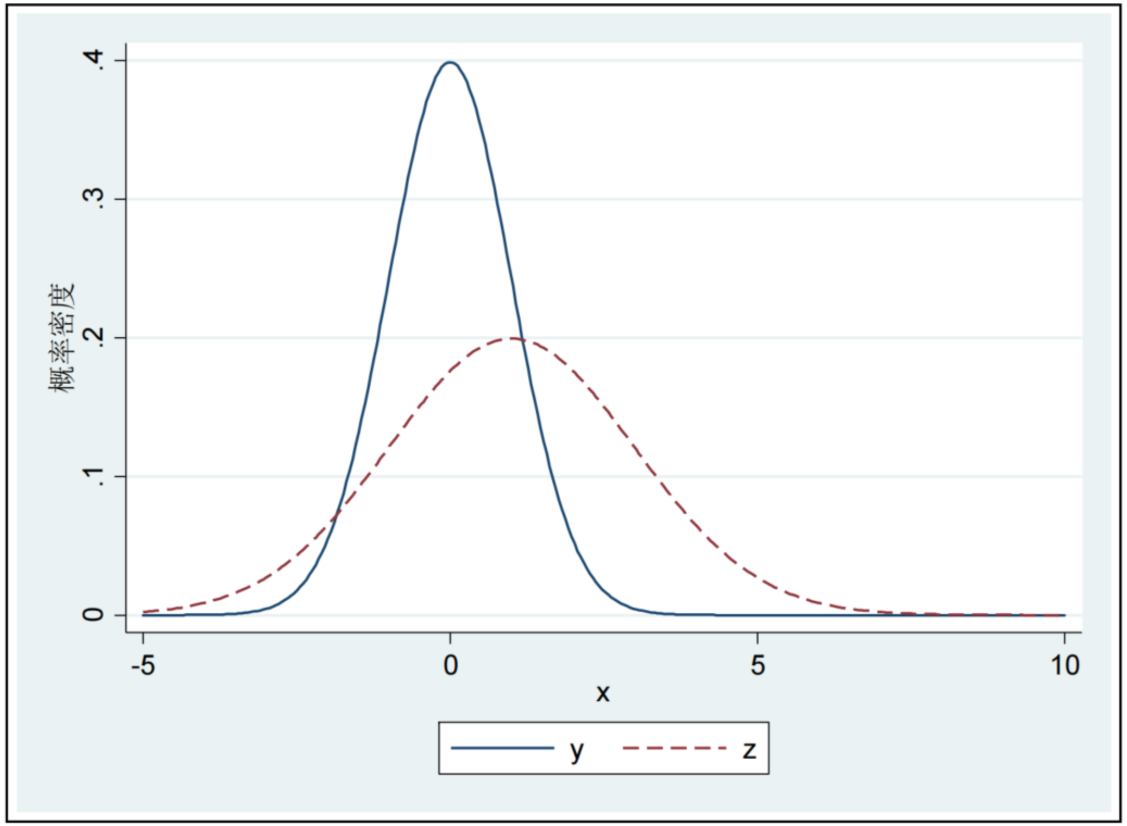
将X进行标准化，定义$Z\equiv\frac{X-\mu}{\sigma}$，则$Z$服从标准正态分布(standard normal distribution)，记为$Z\sim N(0,1)$，其概率密度函数为

$$\phi(x)=\frac{1}{\sqrt{2\pi}}exp\{-x^2/2\}$$

- 标准正态分布的概率密度以原点为对称，呈钟形通常记为$\phi(x)$；其累积分布函数记为$\Phi(x)$



标准正态的概率密度



$N(0,1)$与$N(1,4)$的密度函数

- **多维正态分布**：如果n维随机向量$X=(X_1X_2\dots X_n)'$的联合密度函数为

$$f(x_1,\dots,x_n)=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}exp\left\{-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right\}$$

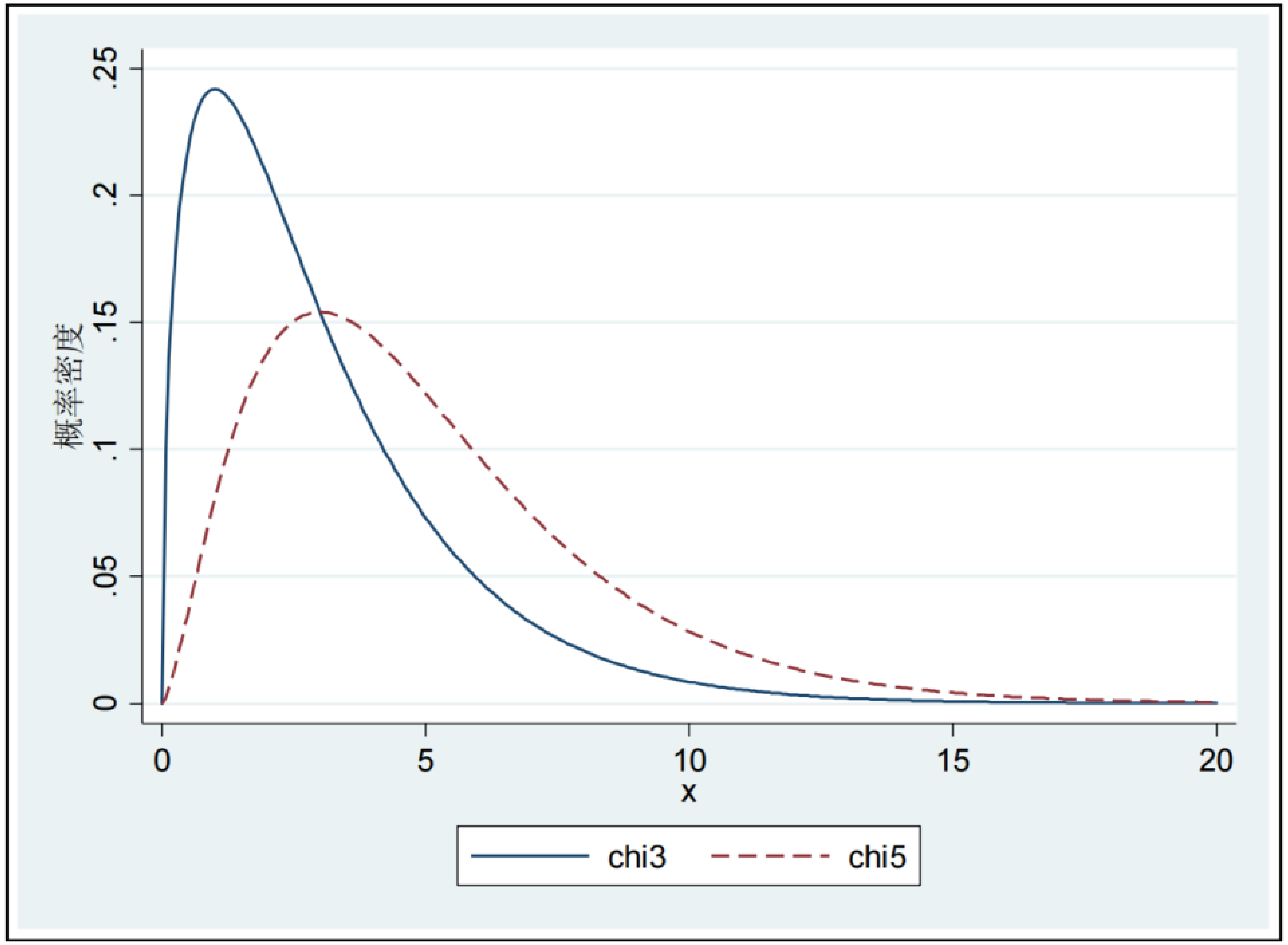
- 则称$\boldsymbol{X}$服从期望为$\boldsymbol{\mu}$、协方差矩阵为$\boldsymbol{\Sigma}$的n维正态分布，记为$\boldsymbol{X}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$。
- 其中，$(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$的二次型，其二次型矩阵为协方差矩阵的逆矩阵$\boldsymbol{\Sigma}^{-1}$；$|\boldsymbol{\Sigma}|$为协方差矩阵$\boldsymbol{\Sigma}$的行列式。
- 多维正态分布具有良好的性质。比如，多维正态的每个分量都是正态，其分量之任意线性组合仍然是正态。反之，每个分量均为一维正态并不足以保证其联合分布也是多维正态的。
- 如果$(X_1X_2\dots X_n)$服从n维正态分布，则“$X_1,X_2,\dots,X_n$相互独立”与“$X_1,X_2,\dots,X_n$两两不相关”是等价的。利用此性质，有时容易证明正态变量的独立性。

χ^2分布（卡方分布，Chi-square ）

- 如果$Z \sim N(0,1)$，则$Z^2 \sim \chi^2(1)$，即自由度为1的χ^2分布。
- 如果$\{Z_1, \ldots, Z_k\}$为独立同分布的标准正态，则其平方和服从自由度为k的卡方分布，记为

$$\sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$$

- 参数k为自由度(degree of freedom)，因为$\sum_{i=1}^{k} Z_i^2$由k个相互独立（自由）的随机变量所构成。
- χ^2分布来自标准正态的平方和，故取值为正。可以证明，$\chi^2(k)$分布的期望为k，而方差为$2k$。



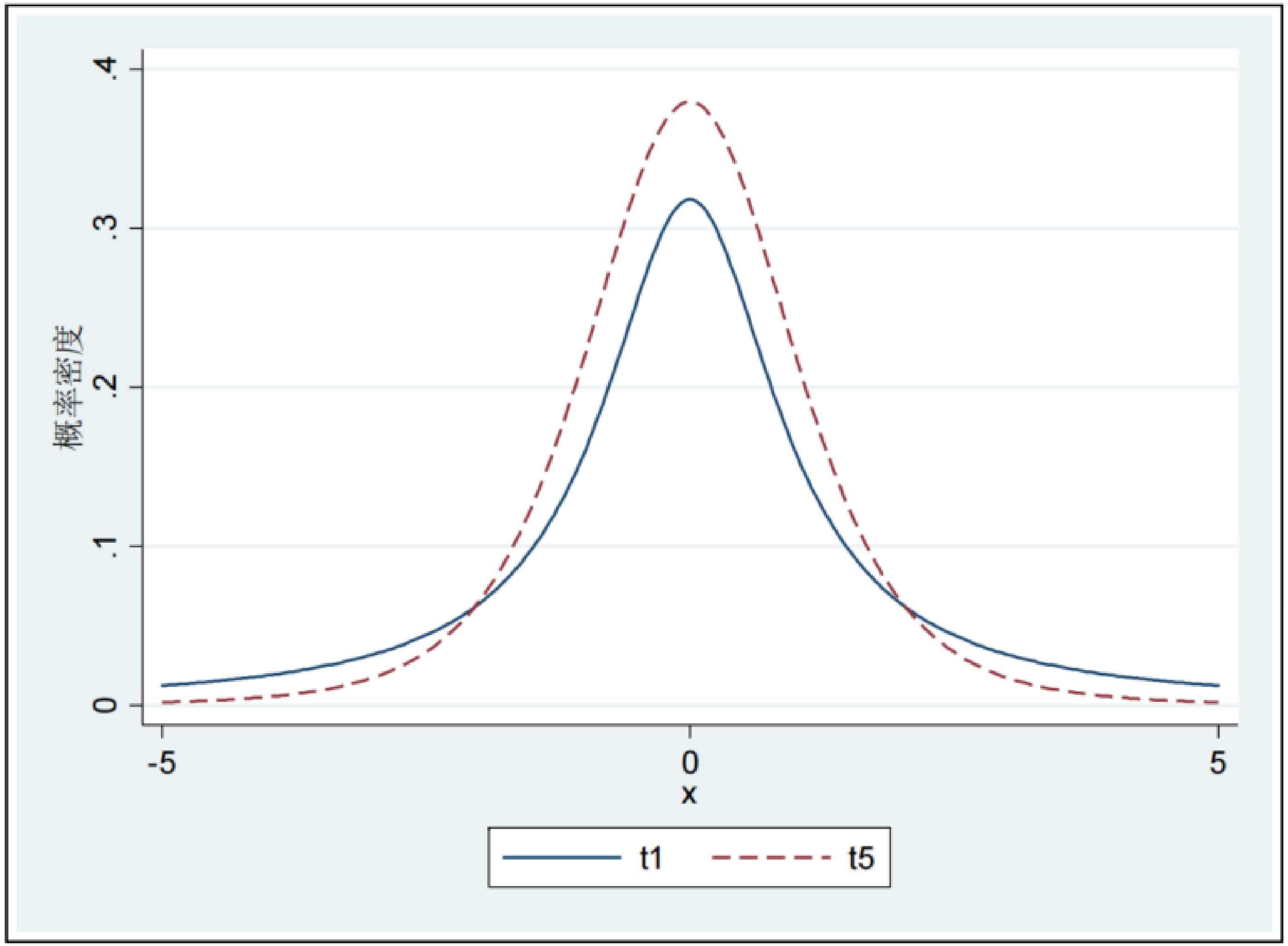
$\chi^2(3)$与$\chi^2(5)$的概率密度

t分布

- 假设$Z \sim N(0,1)$， $Y^2 \sim \chi^2(k)$，且Z与Y相互独立，则$\frac{Z}{\sqrt{Y/k}}$服从自由度为k的t分布，记为

$$\frac{Z}{\sqrt{Y/k}} \sim t(k)$$

- k为自由度。如果表达式中的分子与分母不互相独立，则一般不服从t分布。
- t分布也以原点为对称，但与标准正态分布相比，中间的“山峰”更低（但更尖），而两侧有“厚尾”（fat tails）。
- 当自由度$k \to \infty$时，t分布收敛于标准正态分布。



$t(1)$与$t(5)$的密度函数

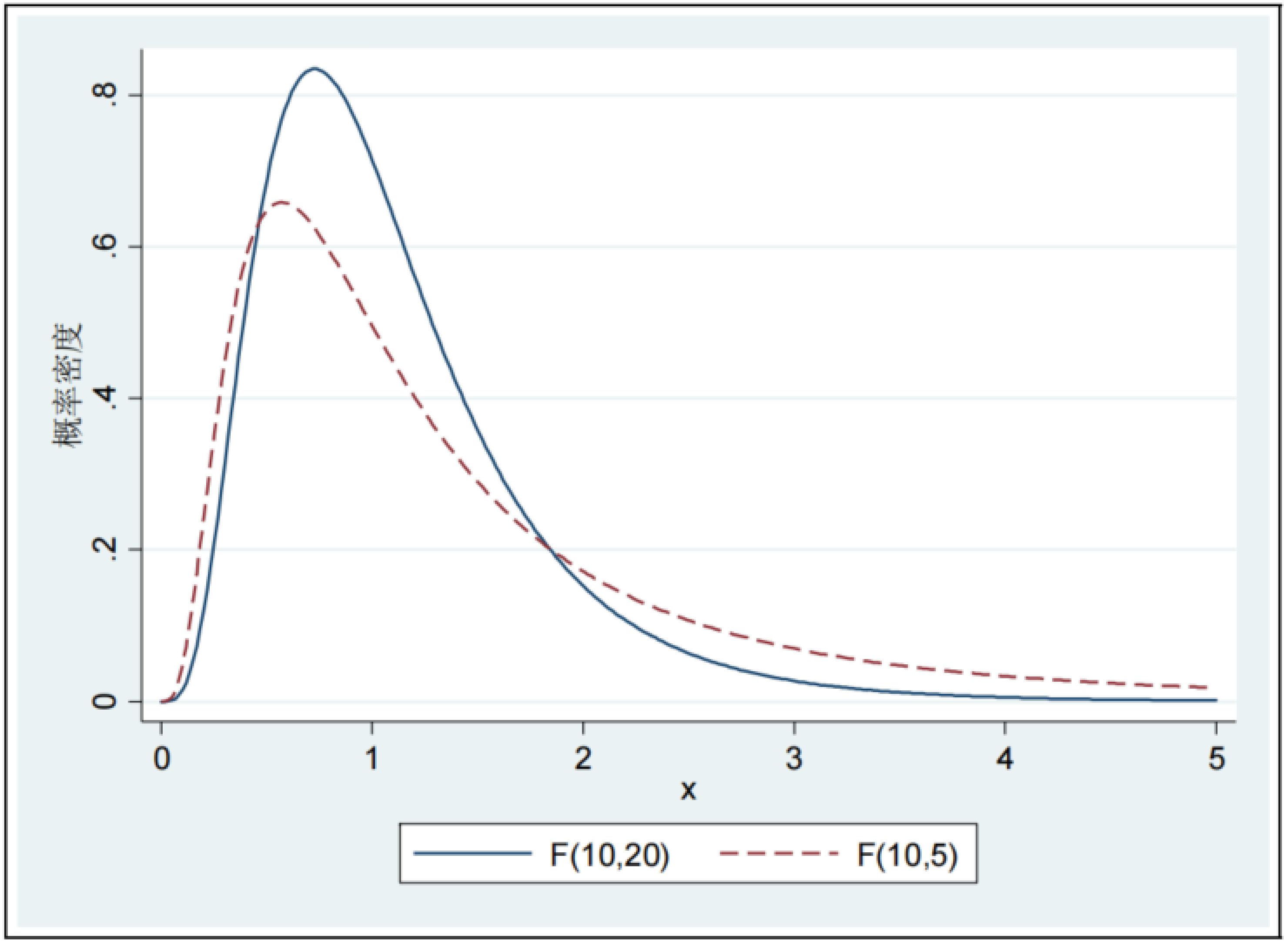
F分布

- 假设$Y_1 \sim \chi^2(k_1)$， $Y_2 \sim \chi^2(k_2)$，且Y_1， Y_2相互独立，则$\frac{Y_1/k_1}{Y_2/k_2}$服从自由度为k_1， k_2的$\boldsymbol{F}$分布，记为

$$\frac{Y_1/k_1}{Y_2/k_2} \sim F(k_1, k_2)$$

其中， k_1, k_2为自由度。

- F分布的取值也只能为正数，其概率密度形状与χ^2分布相似、
- 如上式中分子与分母不独立，则一般不服从F分布。



$F(10,20)$与$F(10,5)$的概率密度

- **命题** 如果$X\sim t(k)$，则$X^2\sim F(1,k)$
- **证明** 由于$X\sim t(k)$，故根据t分布的定义，可将X写为$X=\frac{Z}{\sqrt{Y/k}}\sim t(k)$，其中Z$\sim N(0,1)$，$Y\sim \chi^2(k)$，且$Z$与$Y$相互独立。
- 因此，

$$X^2=\left(\frac{Z}{\sqrt{Y/k}}\right)^2=\frac{Z^2/1}{Y/k}\sim F(1,k)$$

- 其中，由于$Z\sim N(0,1)$，估$Z^2\sim \chi^2(1)$；而且，由于Z与Y相互独立，故Z^2也与Y相互独立。根据F分布的定义，X^2服从自由度为$(1,k)$

的F分布。