

第 2 章 概率统计回顾

2.1 概率与条件概率

1. 概率

“概率”为在大量重复实验下, 事件发生频率趋向的某个稳定值。

记事件“下雨”为 A , 其发生的“概率”(probability)为 $P(A)$ 。

【例】天气预报明天 70% 概率下雨的含义?

2. 条件概率

【例】 已知明天会出太阳，下雨的概率有多大？

记事件“出太阳”为 B ，则在出太阳条件下降雨的“条件概率” (conditional probability) 为

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

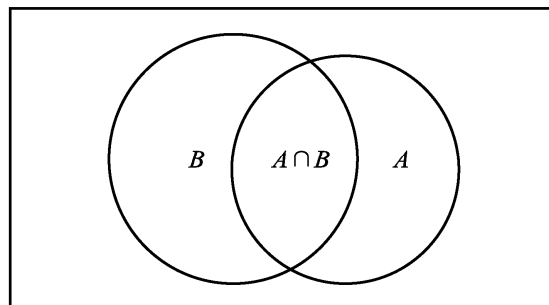


图 2.1 条件概率示意图

【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，股市崩盘的可能性为条件概率。

3. 独立事件

如果条件概率等于无条件概率，即 $P(A|B) = P(A)$ ，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， $P(A|B) \equiv \frac{P(A \cap B)}{P(B)} = P(A)$ ，故

$$P(A \cap B) = P(A)P(B)$$

4. 全概率公式

如果事件组 $\{B_1, B_2, \dots, B_n\}$ ($n \geq 2$) 两两互不相容, $P(B_i) > 0$ ($\forall i = 1, \dots, n$), 且 $B_1 \cup B_2 \cup \dots \cup B_n$ 为必然事件, 则对任何事件 A 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 个可能情形, 再把每种情况下的条件概率“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)。

该公式有助于理解迭代期望定律。

2.2 分布与条件分布

1. 离散型概率分布

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

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

其中， $p_k \geq 0$ ， $\sum_k p_k = 1$ 。

2. 连续型概率分布

连续型随机变量可以取任意实数, 其“概率密度函数”(probability density function, 简记 pdf) $f(x)$ 满足,

(i) $f(x) \geq 0, \forall x$;

(ii) $\int_{-\infty}^{+\infty} f(x) dx = 1$;

(iii) X 落入区间 $[a, b]$ 的概率为 $P(a \leq X \leq b) = \int_a^b f(x) dx$ 。

定义“累积分布函数”(cumulative distribution function, 简记 cdf):

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

3. 多维随机向量的概率分布

为研究变量间关系，常考虑“随机向量”(random vector)。

二维连续型随机向量 (X, Y) 的“联合密度函数” $f(x, y)$ 满足，

(i) $f(x, y) \geq 0, \forall x, y$;

(ii) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$;

(iii) (X, Y) 落入平面某区域 D 的概率为

$$P\{(X, Y) \in D\} = \iint_D f(x, y) dx dy。$$

n 维连续型随机向量 (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。

从二维联合密度 $f(x, y)$, 可计算 X 的(一维)边缘密度函数:

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

类似地, 可计算 Y 的(一维)边缘密度函数:

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

定义二维随机向量 (X, Y) 的累积分布函数为:

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

4. 条件分布

考虑在 $X = x$ 条件下 Y 的条件分布(conditional distribution), 记为 $Y|X = x$ 。如果 X 为连续型随机变量, 事件 $\{X = x\}$ 发生的概率为 0, 该如何计算 $Y|X = x$ 的“条件概率密度”?

考虑 $X \in [x - \varepsilon, x + \varepsilon]$, 然后让 $\varepsilon \rightarrow 0^+$, 可证明条件密度函数为:

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

此公式与条件概率公式类似。

2.3 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)

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

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**

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

期望算子(expectation operator)满足“线性性” (linearity), 即 $E(X + Y) = E(X) + E(Y)$, $E(kX) = k E(X)$, k 为任意常数。

定义 随机变量 X 的方差(variance)

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

称方差的平方根为标准差(standard deviation), 记为 σ 。

命题 $\text{Var}(X) = E[X^2] - [E(X)]^2$

证明:
$$\begin{aligned}\text{Var}(X) &\equiv E[X - E(X)]^2 = E\{X^2 - 2E(X)X + [E(X)]^2\} \\ &= E[X^2] - 2[E(X)]^2 + [E(X)]^2 = E[X^2] - [E(X)]^2\end{aligned}$$

定义 随机变量 X 与 Y 的协方差(covariance)

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

如果 $\text{Cov}(X, Y) > 0$ ，二者正相关；反之，负相关。

如果 $\text{Cov}(X, Y) = 0$ ，则线性不相关，但不一定相互独立。

计算协方差的简便公式：

$$\begin{aligned}\text{Cov}(X, Y) &\equiv E[(X - E(X))(Y - E(Y))] \\ &= E[XY - X E(Y) - E(X)Y + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

协方差的运算也满足线性性：

$$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

将协方差标准化：

定义 随机变量 X 与 Y 的相关系数(correlation)

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

可以证明， $-1 \leq \rho \leq 1$ 。

对于随机变量 X ，可定义一系列的数字特征，即各阶“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)。

将三、四阶中心矩标准化：

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 (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 $E[(X - \mu)/\sigma]^4$ 。

正态分布的峰度为 3。

定义 随机变量 X 的超额峰度 (excess kurtosis) 为 $E[(X - \mu)/\sigma]^4 - 3$ 。

如果随机变量 X 的峰度大于 3 (比如 t 分布), 则其密度函数的两侧尾部更“厚”(fat tails)。

可使用正态分布的偏度与峰度性质来检验某分布是否为正态。

更一般地, 对于任意函数 $g(\cdot)$, 称随机变量函数的期望 $E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx$ 为矩(moment)。

定义 条件期望(conditional expectation)即条件分布 $Y|x$ 的期望:

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

由于 y 已被积分掉, 故 $E(Y|x)$ 只是 x 的函数。

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

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

由于 y 已被积分掉, 故 $\text{Var}(Y|x)$ 只是 x 的函数。

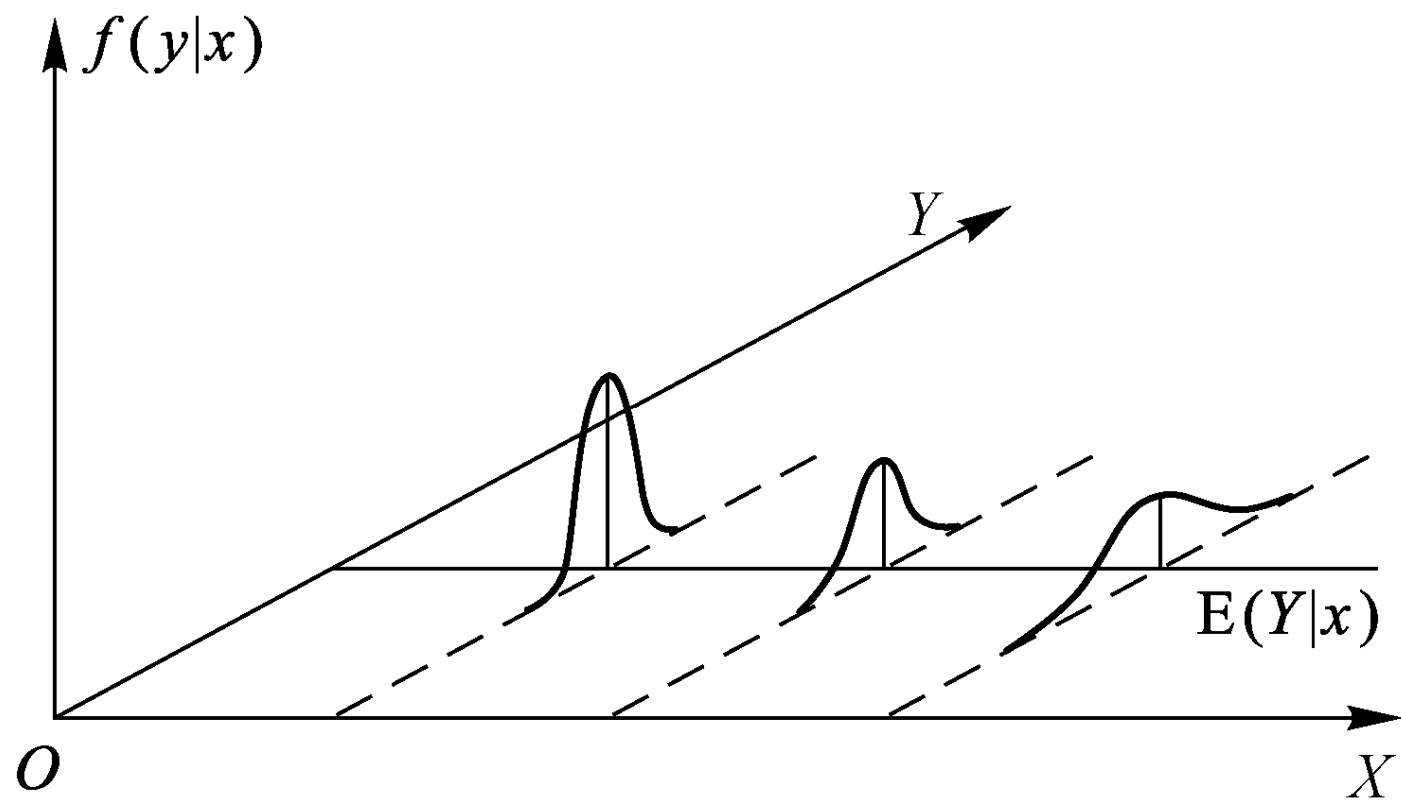


图 2.2 条件期望与条件方差示意图

为引入随机向量的数字特征，回顾矩阵的相关概念。

定义 对于 $n \times n$ 对称矩阵 A (symmetric)，如果对于任意 n 维非零列向量 c ，都有二次型 $c'Ac \geq 0$ ，则称 A 为半正定矩阵。

定义 对于 $n \times n$ 对称矩阵 A ，如果对于任意 n 维非零列向量 c ，都有二次型 $c'Ac > 0$ ，则称 A 为正定矩阵。

从几何上看，对于正定矩阵，可通过坐标变换变为一个主对角线上元素全部为正数的对角矩阵(特征值全部为正)。

故正定矩阵的行列式一定不等于 0，故其逆矩阵一定存在。

在一维情形下，正定矩阵就相当于正数。

类似地，可定义半负定与负定矩阵。

命题 对于任意矩阵 D ， $D'D$ 为半正定矩阵。

证明： 由于 $D'D = (D'D)'$ ，故 $D'D$ 为对称矩阵。

不失一般性，假设 $D'D$ 为 n 级矩阵。对于任意 n 维非零列向量 c ，
二次型

$$c'(D'D)c = (c'D')(Dc) = \underbrace{(Dc)' Dc}_{\text{平方和}} \geq 0$$

故 $D'D$ 为半正定矩阵。

定义 设 $\mathbf{X} = (X_1 \ X_2 \ \cdots \ X_n)'$ 为 n 维随机向量, 则其“协方差矩阵” (covariance matrix) 为 $n \times n$ 的对称半正定矩阵:

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

主对角线元素 $\sigma_{ii} \equiv \text{Var}(X_i)$ ，非主对角线元素 $\sigma_{ij} \equiv \text{Cov}(X_i, X_j)$ 。

假设 \mathbf{A} 为 $m \times n$ 常数矩阵(不含随机变量)，可证明：

(i) $\mathbf{E}(\mathbf{A}\mathbf{X}) = \mathbf{A}\mathbf{E}(\mathbf{X})$ ；

(ii) $\text{Var}(\mathbf{X}) = \mathbf{E}(\mathbf{X}\mathbf{X}') - \mathbf{E}(\mathbf{X})[\mathbf{E}(\mathbf{X})]'$ ；

(iii) $\text{Var}(\mathbf{A}\mathbf{X}) = \mathbf{A}\text{Var}(\mathbf{X})\mathbf{A}'$ 。

命题 n 维随机向量 \mathbf{X} 的协方差矩阵 $\text{Var}(\mathbf{X})$ 为半正定矩阵。

证明： 根据协方差矩阵的定义， $\text{Var}(\mathbf{X})$ 为 $n \times n$ 对称矩阵。

对于 n 维非零列向量 \mathbf{c} , 随机变量 $\mathbf{c}'\mathbf{X}$ 的方差必然大于或等于 0。
因此,

$$\text{Var}(\mathbf{c}'\mathbf{X}) = \mathbf{c}' \text{Var}(\mathbf{X}) \mathbf{c} \geq 0$$

根据定义, $\text{Var}(\mathbf{X})$ 为半正定矩阵。

考虑两个随机向量之间的协方差矩阵。

定义 设 $\mathbf{X} = (X_1 \ X_2 \ \cdots \ X_n)'$ 为 n 维随机向量, $\mathbf{Y} = (Y_1 \ Y_2 \ \cdots \ Y_m)'$ 为 m 维随机向量, 则这两个随机向量之间的协方差矩阵为

$$\text{Cov}(\mathbf{X}, \mathbf{Y}) \equiv \text{E} \left[(\mathbf{X} - \text{E}(\mathbf{X})) (\mathbf{Y} - \text{E}(\mathbf{Y}))' \right] = \text{E}(\mathbf{X}\mathbf{Y}') - \text{E}(\mathbf{X})\text{E}(\mathbf{Y}')$$

2.4 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),

$$E(Y) = E_X [E(Y | x)]$$

以连续型变量为例来证明。

证明: $\text{LHS} = E(Y) = \int_{-\infty}^{+\infty} y f_y(y) dy$

$$\begin{aligned} \text{RHS} &= E_X \left[\int_{-\infty}^{+\infty} y \frac{f(x, y)}{f_x(x)} dy \right] = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} y \frac{f(x, y)}{f_x(x)} dy \right] f_x(x) dx \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dx dy = \int_{-\infty}^{+\infty} y \left[\int_{-\infty}^{+\infty} f(x, y) dx \right] dy = \text{LHS} \end{aligned}$$

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。

在离散随机变量的情形下：

$$E(Y) = \sum_{x_i} P(X = x_i) E(Y | x_i)$$

推而广之，对于任意函数 $g(\cdot)$ 都有：

$$E[g(Y)] = E_x E[g(Y) | x]$$

有时省去期望算子 E_x 的下标，需注意对什么变量求期望。

2.5 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。

X 与 Y 相互独立，是随机变量“无关”的最强概念。

线性不相关的概念更弱，仅要求 $\text{Cov}(X, Y) = 0$ 。“相互独立”意味着“线性不相关”，但反之不然。

二者之间还有一个中间层次的无关概念，即“均值独立”(mean-independence)。

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)。

均值独立不是对称的关系，“ Y 均值独立于 X ”不意味着“ X 均值独立于 Y ”。

命题 “ Y 均值独立于 X ” 当且仅当 $E(Y|x) = E(Y)$ 。

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 。

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(均值独立意味着不相关) 如果 Y 均值独立于 X , 或 X 均值独立于 Y , 则 $\text{Cov}(X, Y) = 0$ 。

证明:

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] && \text{(协方差的定义)} \\ &= E_X E_Y [(X - E(X))(Y - E(Y)) | x] && \text{(迭代期望定律)} \\ &= E_X [(X - E(X)) E_Y (Y - E(Y) | x)] && (X - E(X) \text{ 视为常数}) \\ &= E_X [(X - E(X))(E(Y | x) - E(Y))] && \text{(期望算子的线性性)} \\ &= E_X [(X - E(X)) \cdot 0] = 0 && \text{(均值独立的定义)}\end{aligned}$$

总之，“相互独立” \Rightarrow “均值独立” \Rightarrow “线性不相关”。

2.6 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，记为 $X \sim N(\mu, \sigma^2)$ ， μ 为期望， σ^2 为方差。

定义 $Z \equiv \frac{X - \mu}{\sigma}$ ，则 Z 服从标准正态分布，记为 $Z \sim N(0,1)$ 。

标准正态的概率密度记为 $\phi(x)$ ；累积分布函数记为 $\Phi(x)$ 。正态分布也称“高斯分布” (Gaussian distribution)。

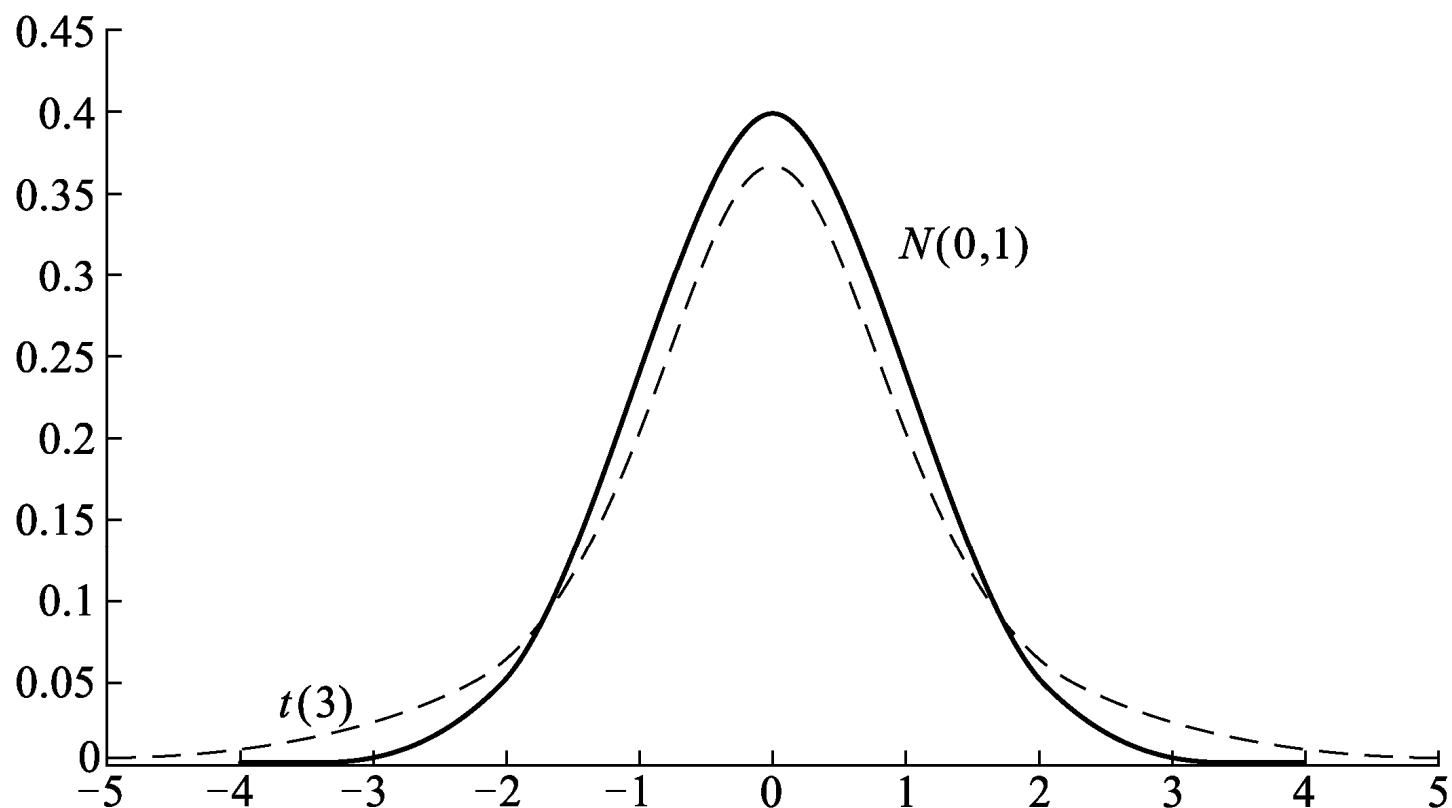


图 2.3 $N(0,1)$ 与 $t(3)$ 的概率密度

如果 n 维随机向量 $\mathbf{X} = (X_1 \ X_2 \ \cdots \ X_n)'$ 的联合密度函数为

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

则称 \mathbf{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 $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 。

多维正态的每个分量都是正态，分量之线性组合仍是正态。

但每个分量均为一维正态不足以保证联合分布为多维正态。

命题 对于 n 个随机变量 X_1, X_2, \dots, X_n , 如果任意线性组合 $k_1 X_1 + \dots + k_n X_n$ (其中 k_1, \dots, k_n 不全为 0) 都服从一维正态分布, 则 $(X_1 \ X_2 \ \dots \ X_n)$ 服从 n 维正态分布。

命题 如果 $(X_1 \ X_2 \ \dots \ X_n)$ 服从 n 维正态分布, 设 Y_1, \dots, Y_m 分别是 $(X_1 \ X_2 \ \dots \ X_n)$ 的线性函数, 则 $(Y_1 \ \dots \ Y_m)$ 也服从多维正态分布。

证明： 考虑 Y_1, \dots, Y_m 的任意线性组合 $k_1 Y_1 + \dots + k_m Y_m$, 其中 k_1, \dots, k_m 不全为 0。由于 Y_1, \dots, Y_m 分别是 $(X_1 \ X_2 \ \dots \ X_n)$ 的线性函数, 故 $k_1 Y_1 + \dots + k_m Y_m$ 也是 $(X_1 \ X_2 \ \dots \ X_n)$ 的线性函数。由于多维正态的线性组合仍为正态, 故 $k_1 Y_1 + \dots + k_m Y_m$ 服从一维正态分布。根据 k_1, \dots, k_m 的任意性可知, $(Y_1 \ \dots \ Y_m)$ 服从多维正态分布。

如果 (X_1, X_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, 有时很容易证明正态变量的独立性。

2. 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 $\{Z_1, \dots, Z_k\}$ 为独立同分布的标准正态, 则其平方和

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

参数 k 为 “自由度” (degree of freedom), 即由 k 个相互独立(自由)的随机变量构成。

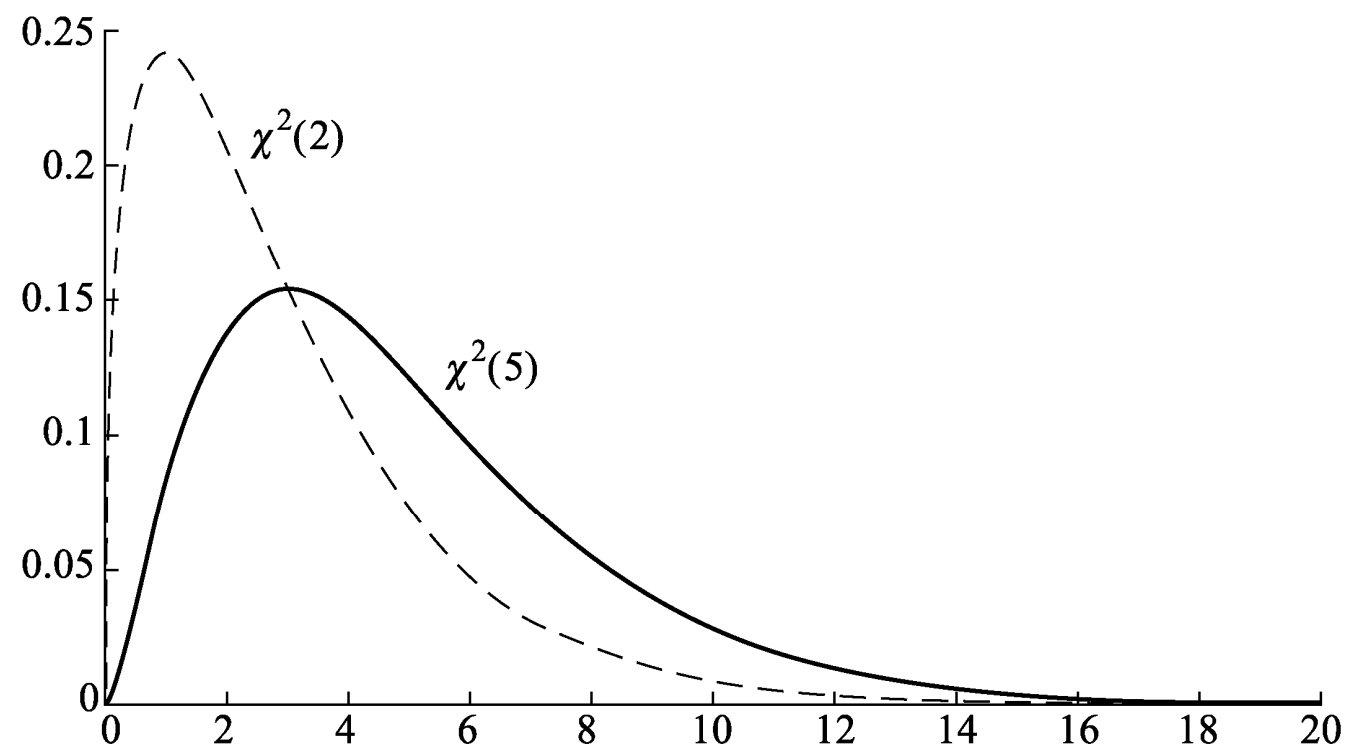


图 2.4 χ^2 分布的概率密度

3. t 分布

假设 $Z \sim N(0, 1)$, $Y \sim \chi^2(k)$, 且 Z 与 Y 相互独立, 则

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

其中, k 为自由度。

t 分布以原点对称, 与标准正态相比, 中间“山峰”更低(但更尖), 两侧有“厚尾”(fat tails)。

当自由度 $k \rightarrow \infty$ 时, t 分布收敛于标准正态分布。

4. 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

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

其中, k_1, k_2 为自由度。

F 分布取值为正, 概率密度形状与 χ^2 分布相似。

如果 $X \sim t(k)$, 可证明 $X^2 \sim F(1, k)$ (参见习题)。

2.7 统计推断的思想

计量经济学的主要方法是数理统计的“统计推断”(statistical inference)。

称我们感兴趣的研究对象全体为“总体”(population)，其中的每个研究对象称为“个体”(individual)。

由于总体包含的个体可能很多，普查成本较高，故常从总体抽取部分个体，称为“样本”(sample)。样本中包含个体的数目称为“样本容量”(sample size)。

通常希望样本为“随机样本”(random sample)，即总体中的每个个体都有相同的概率被抽中，且被抽中的概率相互独立，称为“独立同分布”(independently identically distributed, 简记 iid)。

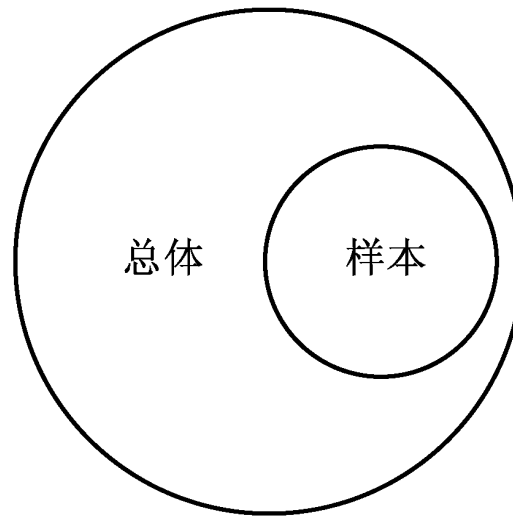


图 2.5 总体与样本

由于样本来自总体，必然带有总体的信息。而统计推断就是根据样本数据对总体性质进行推断的科学。统计推断的主要形式有参数估计(点估计、区间估计)、假设检验及预测等。