

probability
probability

第九章

回 归 分 析

probability
probability

§9.1 相关关系与回归分析

§9.2 一元回归分析

§1 相关关系与回归分析

一. 相关关系与回归函数

在现实世界中存在大量的变量, 它们有相互依存、相互制约的关系, 一般分为两类:
确定性关系与非确定性关系.

Ex.1 球的直径 X 与球的体积 V 之间有确定的函数关系

$$V = \frac{1}{6} \pi X^3$$

Ex.2 随机信号

$$X(t) = A \cos(\omega t + \Theta), \quad t \in R,$$

由振幅 $A \sim N(\mu, \sigma^2)$ 和初相角 $\Theta \sim U(-\pi, \pi)$ 确定.

还有另一类变量关系，如：

Ex.3 农作物产量 Y 与降雨量 X_1 ，氮、磷、钾的施肥量 X_2 、 X_3 、 X_4 之间的关系.

EX. 4 产品的价格 X 与需求量 M 之间存在关系.

特点 变量间的关系无法用确定的函数来明确表述.

问题 如何描述变量间的各类关系?

将作为考察目标的变量称为**因变量**(记为 Y), 而将影响它的各个变量称为**自变量**或**可控变量**, 记为

$$(X_1, X_2, \cdots X_k)$$

1.确定性的函数关系

用第三章方法可求随机变量函数的分布.

若已知随机变量 X 的分布就可以确定函数

$$V = \frac{1}{6} \pi X^3$$

的分布.

由振幅和初相角的分布可以完全确定随机信号

$$X(t) = A \cos(\omega t + \Theta), \quad t \in R,$$

的分布.

2.非确定性的相关关系

EX. 5 已测得非线性系统的输入和输出信号, 如何确定输入和输出之间的关系? (§9.5非线性交调的频率设计)



设想 构造某种函数来描述输入和输出之间的非确定关系.

随机变量的相关关系引进:

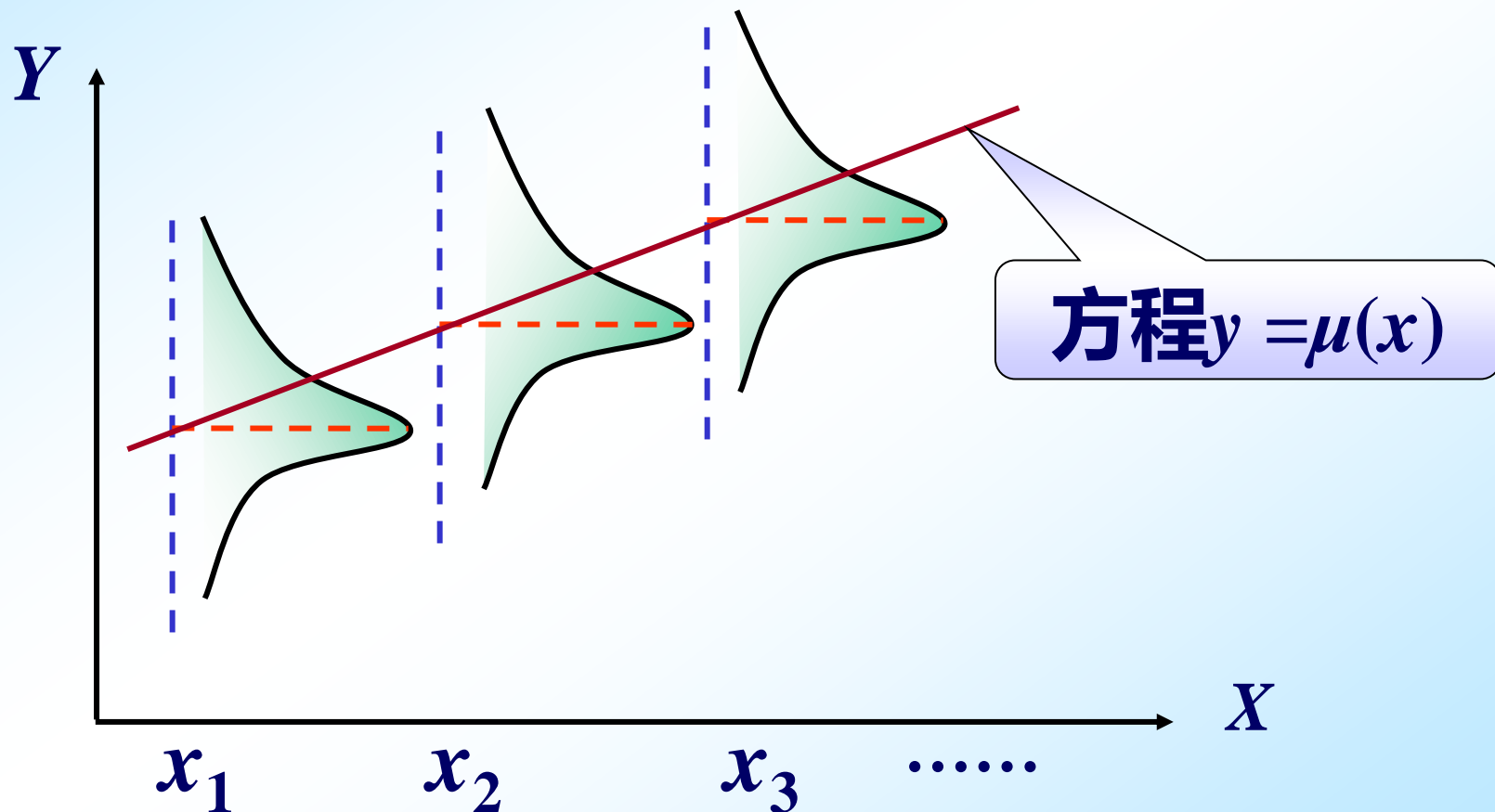
考虑单个因变量 Y 与单个自变量 X 的情形.

在“ $X=x$ ”时, Y (连续型)的条件数学期望为

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

1) $\mu(x)$ 可理解为在“ $X=x$ ”的条件下, 随机变量 Y 取值的集中点;

对于 X 的不同取值 x_1, x_2, \dots, x_n



2) 方程 $y = \mu(x)$ 描述了 Y 与 X 间非确定性的关系.

将可控变量 X_1, X_2, \dots, X_k 的取值记为 x_1, x_2, \dots, x_k , 若条件数学期望:

$$\mu(x_1, x_2, \dots, x_k) = E(Y | X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$$

存在, 称 Y 与 X_1, X_2, \dots, X_k 具有**相关关系**.

相关关系是一种非确定性关系

定义9.1.1 称

$$\mu(x_1, x_2, \cdots, x_k) = E(Y | X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k)$$

为 Y **关于** X_1, X_2, \cdots, X_k **的回归函数, 方程**

$$y = \mu(x_1, x_2, \cdots, x_k)$$

称为 Y **对** X_1, X_2, \cdots, X_k **的回归方程.**

注 回归函数是确定性的函数.

回归分析是从回归函数出发处理相关关系的方法.

3.回归模型的引进

若 Y 关于 $X_1, X_2, \cdots X_k$ 的回归函数为

$$y = \mu(x_1, x_2, \cdots, x_k)$$

设想: $Y = \mu(x_1, x_2, \cdots, x_k) + \text{随机误差}$

得数学模型:

$$Y = \mu(x_1, x_2, \cdots, x_k) + \varepsilon$$

$$\varepsilon = Y - \mu(x_1, x_2, \cdots, x_k)$$

可视为随机误差, 通常要求:

其它未知的、
未考虑的因素
以及随机因素
的影响所产生.

1) $E(\varepsilon)=0$;

2) $D(\varepsilon)=E(\varepsilon^2)=\sigma^2$ 尽可能小.

注意到 $\sigma^2 = E[Y - \mu(x_1, x_2, \dots, x_n)]^2$
 σ^2 是用回归函数近似因变量 Y 产生的**均方误差**.

建立模型涉及**三个问题**:

1) 确定对因变量 Y 影响显著的自变量;

2) 确定回归函数 $\mu(x)$ 的类型;

3) 对参数进行估计.

} 本章内容