中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

第三章 随机变量和一元概率分布

洪永淼
中国科学院数学与系统科学研究院
中国科学院大学经济与管理学院

目　录

为什么需要引入随机变量 (Random variable)?

- 概率空间因随机试验而异使其在实际应用中存在诸多不便。
- 为了发展统一的概率理论，我们考虑从原始**样本空间**（sample space) S 到一个新的样本空间 Ω 的映射 X，其中 Ω 由一组实数构成。
- 这种转换 $X: S \rightarrow \Omega$ 被称为随机变量。在许多应用中，我们可能只对实验结果的某些特定方面感兴趣，而不是对结果本身感兴趣。一个适当定义的随机变量 X 会更好地满足我们的需求。

定义 3.1

[随机变量]： 随机变量 $X(\cdot)$ 是从样本空间 S 到实数集 $\mathbb{R}$ 的 $\mathbb{B}$-可测映射（或点函数），满足对每个基本结果 $s \in S$，都存在唯一的实数 $X(s)$ 与之对应。随机变量 X 可能取的所有实数值的集合，也称为 X 的值域，构成了新的样本空间，记为 Ω。

- 根据函数的定义，函数 $X : S \to \Omega$ **无需为一一映射。**
  - ✓ 有可能两个基本结果 $s_1, s_2 \in S$ 的随机变量 X 值是一样的，即 $X(s_1) = X(s_2)$。

例 3.1：[抛硬币]

- 抛硬币时，样本空间 $S=\{H,T\}$。定义随机变量 $X(\cdot)$，满足
  - ✓ $X(H)=1$，$X(T)=0$，
- 新样本空间：$\Omega=\{1,0\}$。

例 3.2：[选举]

- 选举某个候选人时，样本空间 $S=\{\text{成功},\text{失败}\}$。定义随机变量 $X(\cdot)$，满足
  - ✓ $X(\text{成功})=1$，$X(\text{失败})=0$，
- 新样本空间：$\Omega=\{1,0\}$。

S 和 Ω 的基本结果的数目不必相等

例 3.3：[抛硬币]

- 假设抛三枚质地均匀的硬币，其样本空间

$$S=\{TTT,TTH,THT,HTT,HHT,HTH,THH,HHH\}$$

- 令 $X(\cdot)$ 表示正面朝上的个数，则有
  - ✓ $X(TTT)=0$,
  - ✓ $X(TTH)=1$，$X(THT)=1$，$X(HTT)=1$,
  - ✓ $X(HTH)=2$，$X(THH)=2$,
  - ✓ $X(HHH)=3$。
- 因此，新样本空间 $\Omega=\{0,1,2,3\}$。

例 3.4：[掷骰子]

- 掷骰子试验的样本空间 $S = \{1, 2, 3, 4, 5, 6\}$。定义 $X(s) = s$，则 $\Omega = S$。此为恒等变换。

◆ 问题 3.1

若 S 的基本结果的数目是可数的，那么：

(1) Ω 的基本结果的数目是否可大于 S 的基本结果的数目？

(2) Ω 的基本结果的数目是否可小于 S 中的基本结果的数目？

例 3.5：[二元变量]

- 假设 $S = \{s : -\infty < s < +\infty\}$。
  - ✓ 若 $s > 0$，定义 $X(s) = 1$；
  - ✓ 若 $s \leq 0$，定义 $X(s) = 0$。

- 在例 3.5 中，随机变量 X 仅取两个可能的值，故称为**二元随机变量。**
- 二元变量在经济学有广泛的应用，例如对非对称经济周期转折点的预测（如 Neftci, 1984）。

随机变量举例

- 随机变量的例子还有很多，例如
  - ✓ 主观幸福感
  - ✓ 投资者情绪指数 (sentiment index)
  - ✓ 经济政策不确定性 (economic policy uncertainty, EPU) 指数
- 上述指标通常基于社交媒体平台 (如，微博和脸书) 和新媒体的文本数据构造。
- 文本数据是非结构化大数据，大数据包括结构化、半结构化和非结构化数据。

原样本空间 S 上的 $P(\cdot) \Longrightarrow$ 新样本空间 Ω 上的 $P_X(\cdot)$

- 首先，假设样本空间 S 包含有限个基本结果：

$$S = \{s_1, \cdots, s_n\}$$

  且概率函数 $P: \mathbb{B} \to [0,1]$，其中 $\mathbb{B}$ 是从 S 生成的 σ 域。

- 同时，定义随机变量 $X: S \to \mathbb{R}$，其值域为

$$\Omega = \{x_1, \cdots, x_m\}$$

  其中 m 和 n 未必相等。

原样本空间 S 上的 $P(\cdot) \Longrightarrow$ 新样本空间 Ω 上的 $P_X(\cdot)$ (Cont.)

- 则随机变量 X 的概率函数 $P_X\colon \Omega \to \mathbb{R}$ 可定义如下：

$$\begin{aligned} P_X(x_i) &\equiv P(X = x_i) \\ &= P(C_i) \end{aligned}$$

其中 C_i 是原始样本空间 S 中的一个事件，满足 $C_i = \{s \in S\colon X(s) = x_i\}$。

- 换言之，在新样本空间 Ω 上定义的概率函数 $P_X(\cdot)$ 可由原始概率函数 $P(\cdot)$ 推导而得。

原样本空间 S 上的 $P(\cdot) \Longrightarrow$ 新样本空间 Ω 上的 $P_X(\cdot)$ (Cont.)

- 更正式地，对任意一个实数集合 $A \in \mathbb{B}_\Omega$，其中 $\mathbb{B}_\Omega$ 是一个从新样本空间 Ω 生成的 σ 域，可定义概率函数 $P_X\colon \mathbb{B}_\Omega \to \mathbb{R}$，满足

$$P_X(A) = P(C_A) = P[s \in S: X(s) \in A]$$

其中 $C_A = \{s \in S: X(s) \in A\}$ 表示在原始样本空间 S 中，其随机变量 X 取值落入实数集合 $A \subset \Omega$ 的那些基本结果的集合。

◆ **课后练习**

试证明，当 S 为可数样本空间时，诱导概率函数 $P_X(\cdot)$ 满足概率函数的三个条件。

例 3.6： [抛硬币]

- 若抛掷三枚硬币，则样本空间

$$S = \{HHH, HTH, HHT, THH, THT, TTH, HTT, TTT\}$$

- 令随机变量 $X(\cdot)$ 表示随机试验中正面朝上的个数。则新的样本空间或 X 的值域为 $\Omega = \{0, 1, 2, 3\}$
- 假设需要计算概率 $P(0 \le X \le 1)$。
- 定义

$$\begin{aligned} C &= \{s \in S: 0 \le X(s) \le 1\} \\ &= \{TTT, TTH, THT, HTT\} \end{aligned}$$

- 则

$$\begin{aligned} P(0 \le X \le 1) &= P(C) \\ &= P(TTT) + P(TTH) + P(THT) + P(HTT) \\ &= 1/2 \end{aligned}$$

定义 3.2

[可测函数 (Measurable Function)]：若对任何实数 a，有 $\{s \in S: X(s) \le a\} \in \mathbb{B}$，则称函数 $X: S \to \mathbb{R}$ 是 $\mathbb{B}$-可测的（即关于 σ 域 (σ –field) $\mathbb{B}$ 是可测度的）。

- **可测函数**确保了对 $\mathbb{B}_\Omega$ 中的任何实数子集 A，相对应的等价事件 C 总落在 $\mathbb{B}$ 中，因此概率 $P_X(A)$ 均有定义。
- 若 $X(\cdot)$ 不是可测函数，则对于实数集 $\mathbb{R}$ 上的一些 σ 域，存在某些无法定义概率的子集。
- 本课程假设所有随机变量 X 关于从样本空间 S 生成的 σ 域 $\mathbb{B}$ 都是可测的。

定理 3.1

令 $\mathbb{B}$ 表示从样本空间 S 生成的 σ 代数 (σ-algebra)。若 $f(\cdot)$ 和 $g(\cdot)$ 为 $\mathbb{B}$-可测实值函数，且 c 为实数，则函数 $c\cdot f(\cdot)$，$f(\cdot)+g(\cdot)$，$f(\cdot)\cdot g(\cdot)$ 以及 $|f(\cdot)|$ 均为 $\mathbb{B}$-可测函数。

- 若序列 $X_1(s), X_2(s), \cdots$ 是可测函数，则通过极限运算构造的函数 $Z(s)$ 也是可测函数。如：

✓ $Z(s)=\lim\limits_{i\to\infty} Z_i(s)$

✓ $Z(s)=\lim\limits_{n\to\infty}\sup\limits_{1\le i\le n}|Z_i(s)|$

随机变量 X 的诱导概率 $P_X(\cdot)$ 满足概率函数的定义

1. 由于对任意事件 $C_A=\{s\in S: X(s)\in A\}\in\mathbb{B}$，其中 $\mathbb{B}$ 是从原始样本空间 S 生成的 σ 域，均有 $0\le P(C_A)\le 1$。由此可得 $1\ge P_X(A)=P(C_A)\ge 0$。
2. 由于 $S=\{s\in S\text{：}X(s)\in\Omega\}$ 意味着 $P_X(\Omega)=P(S)=1$。
3. 我们将概率函数定义的条件 (3) 的讨论限制在从 Ω 生成的 σ 域 $\mathbb{B}_\Omega$ 中的两个互不相交事件 A_1 和 A_2。
   - $A_1\cup A_2$ 的诱导概率由下式给出：$P_X(A_1\cup A_2)=P(C)$, 其中 $C=\{s\in S: X(s)\in A_1\cup A_2\}$。
   - $C=\{s\in S: X(s)\in A_1\}\cup\{s\in S: X(s)\in A_2\}=C_1\cup C_2$, 其中 C_1 和 C_2 是 S 中的两个互斥事件。
   - 因此， $P_X(A_1\cup A_2)=P_X(A_1)+P_X(A_2)$。

目　录

◆ **问题 3.3**

如何刻画随机变量 X?

定义 3.3

[累积分布函数 (Cumulative Distribution Function)]：随机变量 X 的 CDF 定义为

$$F_X(x) = P(X \leq x), \qquad \text{对任意 } x \in \mathbb{R}$$

其中函数 F 的下标 X 表示该函数是随机变量 X 的 CDF。

定理 3.2

[$F_X(\cdot)$ 的性质]：假设 $F_X(\cdot)$ 是随机变量 X 的 CDF，则

(1) $\lim\limits_{x\to-\infty} F_X(x) = 0$, $\lim\limits_{x\to\infty} F_X(x) = 1$；

(2) $F_X(x)$ 为单调非递减函数，即对任意的 $x_1 < x_2$，有 $F_X(x_1) \leq F_X(x_2)$；

(3) $F_X(x)$ 为 x 的右连续函数，即对任意 x 以及 $\delta > 0$，有

$$\lim_{\delta\to0^+} [F_X(x+\delta) - F_X(x)] = 0$$

定理 3.3

令 $a < b$，则

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

证明：

- 注意事件 $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$，且事件 $\{X \leq a\}$ 和 $\{a < X \leq b\}$ 互不相交，故有

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

- 由 CDF $F_X(x)$ 的定义即得证。

证毕。

定理 3.4

对任意实数 b,

$$P(X > b) = 1 - F_X(b)$$

例 3.7： [抛硬币]

假设 $F(x)$ 是 CDF。定义 $G(x) = 1 - F(-x)$。$G(x)$ 是 CDF 吗?

解： 依次检验 $G(\cdot)$ 是否满足 CDF 的三个基本性质：

1.
$$G(-\infty) = 1 - F[-(-\infty)] = 1 - F(\infty) = 1 - 1 = 0$$
$$G(\infty) = 1 - F(-\infty) = 1 - 0 = 1$$

2. 对任意 $x_1 < x_2$，有
$$G(x_1) = 1 - F(-x_1)$$
$$G(x_2) = 1 - F(-x_2)$$
$$\begin{aligned} G(x_2) - G(x_1) &= [1 - F(-x_2)] - [1 - F(-x_1)] \\ &= F(-x_1) - F(-x_2) \\ &\geq 0 \qquad (-x_1 > -x_2 \text{ 且 } F(\cdot) \text{ 为 CDF}) \end{aligned}$$

3. 然而，$G(x)$ 是左连续函数，而未必是右连续函数。

例 3.9： [混合分布 (Mixture of Distribution)]

假设 $F_1(x)$ 和 $F_2(x)$ 是 CDF。问线性组合

$$F(x) = pF_1(x) + (1-p)F_2(x)$$

是否也为 CDF？其中 p 为常数。

解：

- 当 $0 \leq p \leq 1$ 时，答案是肯定的。
- 此处，$F(x)$ 通常称为分布 $F_1(x)$ 和 $F_2(x)$ 的混合。
- 两分布的混合可提供很多灵活性，比如刻画偏度与厚尾现象。

混合分布的实践应用

- 在经济观测数据中，部分观测值由某一个分布生成，而其余的观测值则由另一个分布生成。
- 存在两个互斥的状态，即状态 1 和状态 2，分别以概率 p 和 $1-p$ 发生。假设随机变量 X 在状态 1 发生时服从分布 $F_1(x)$，在状态 2 发生时服从分布 $F_2(x)$。因此，X 的分布 $F(x)$ 为 $F_1(x)$ 和 $F_2(x)$ 的混合分布。
- 例如，金融市场上资产收益率 X 在经济扩张和经济萧条时可能服从不同分布。

定义 3.4

[同分布 (Identical Distributions)]：令 $\mathbb{B}_\Omega$ 为从 Ω 生成的包含所有具有 (a,b)，$[a,b)$ 及 $[a,b]$ 形式的实数区间的最小 σ 域，若对 $\mathbb{B}_\Omega$ 中的任意集合 A，有

$$P(X \in A) = P(Y \in A)$$

则称两个随机变量 X 和 Y 服从同分布。

◆ 问题 3.4

同分布是否意味着 $X = Y$?

例 3.10：[抛硬币]

- 假设分别抛 n 次五角和一元的硬币，分别考察以下关于 X 和 Y 的两个定义：

  (1) X 为五角硬币正面朝上的次数，Y 为一元硬币正面朝上的次数；

  (2) X 和 Y 均为五角硬币正面朝上的次数。

- 对上述两个定义，X 和 Y 均为同分布。然而，在第一种情况下 X 和 Y 相互独立，而在第二种情况下 $X = Y$。
- 需要指出，虽然 $X = Y$ 可推出 X 和 Y 同分布，但同分布不意味着 $X = Y$。X 和 Y 不一定要定义在同一样本空间上，只需两者的分布函数一致即可。

◆ **问题 3.5**

为什么 CDF $F_X(x)$ 可刻画随机变量 X 的概率分布?

- CDF 的定义表明：随机变量 X 的概率函数 $P(\cdot)$ 决定了 CDF $F_X(x)$，也可从 CDF $F_X(x)$ 推得随机变量 X 的概率函数 $P(\cdot)$。
- 换言之，$P(\cdot)$ 和 $F_X(x)$ 包含相同的关于随机变量 X 的概率分布信息。

定理 3.5

当且仅当

$$F_X(x) = F_Y(x), \quad \text{对所有 } -\infty < x < \infty$$

两个随机变量 X 和 Y 同分布。

例 3.11：[收入分配与洛仑兹曲线]

- 在经济学，**洛仑兹曲线** (Lorenz Curve) 和**基尼系数** (Gini Coefficient) 常用于刻画**收入不均等程度**。
- **洛伦兹曲线**用图形的方式描述收入的（经验）累积概率分布 CDF。它描述了最贫穷的 $x\%$ 家庭或人口的收入占全社会总收入的 $y\%$。
- **完全均等线（$y = x$）**：完全均等的收入分配是每个家庭收入都相等。该情形下，对所有 $x \in [0, 100]$ 而言，全社会最贫穷的 $x\%$ 人口总是占有全社会总收入的 $x\%$。
- **基尼系数**：可以看作是平等直线与洛伦兹曲线之间的面积与平等直线以下三角形的面积之比。

例 3.11 (Cont.):

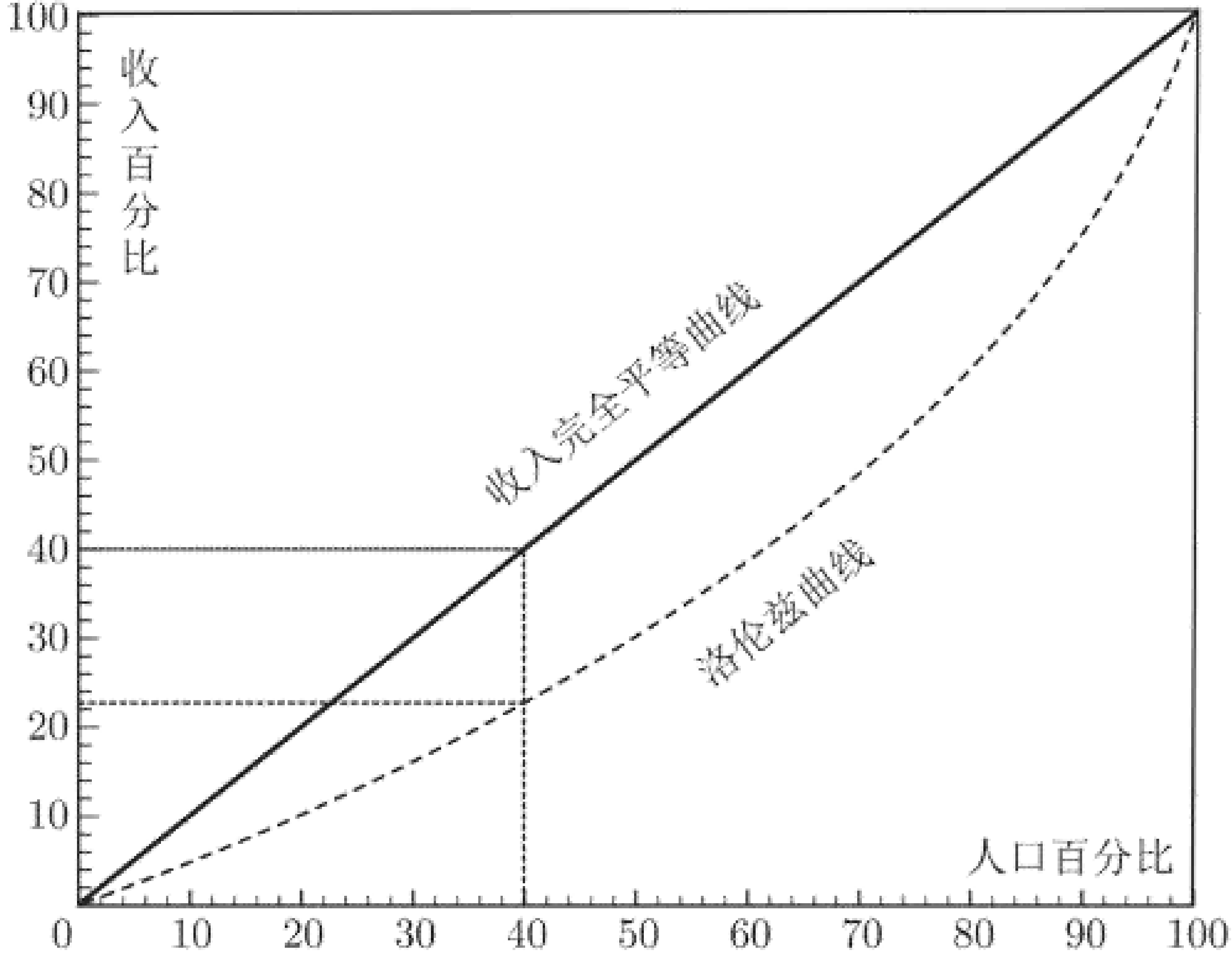


图 3.2：洛仑兹曲线、基尼系数与收入不均等

例 3.12：[一阶随机占优 (First Order Stochastic Dominance)]

若两个分布 $F(\cdot)$ 和 $G(\cdot)$ 对实数轴上所有 x 均满足 $F(x) \leq G(x)$，且存在 x 使得严格不等式成立，则称分布 $F(\cdot)$ 一阶随机占优于 $G(\cdot)$。

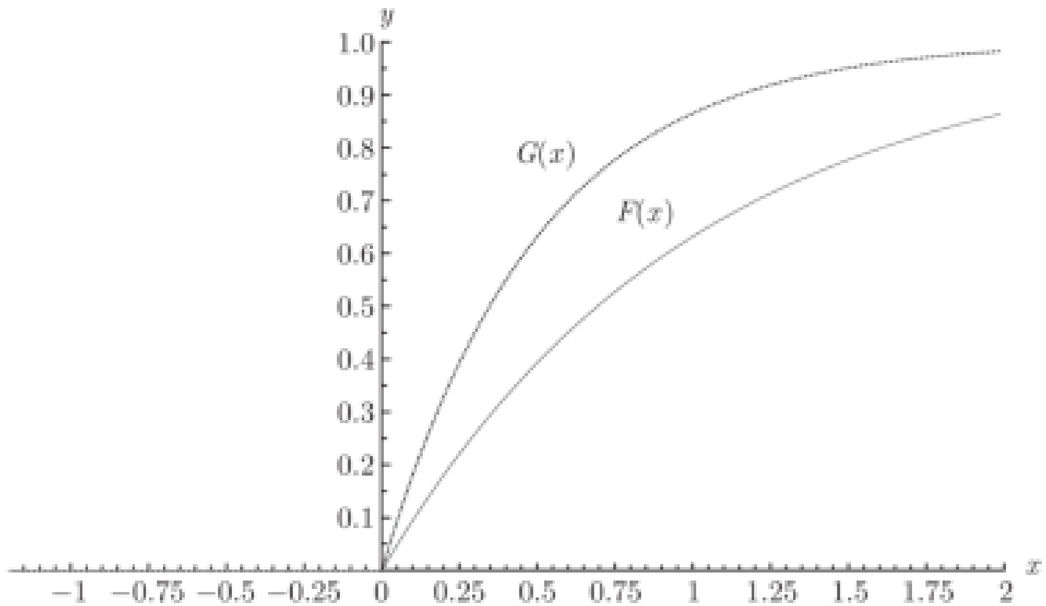


图 3.3：$F(\cdot)$ 一阶随机占优于 $G(\cdot)$

一阶随机占优的含义

- 若令 x 表示收入水平，$F(x) \leq G(x)$ 表示：
  - ✓ 分布 $F(\cdot)$ 中收入不超过 x 的人口比例 $\leq$ 分布 G(·) 中收入不超过 x 的人口比例。
  - ✓ 即分布 $G(\cdot)$ 中的贫困人口比例 $\geq$ 分布 $F(\cdot)$ 中的人口比例。
- 在金融学中，若投资组合分布 $F(\cdot)$ 一阶随机占优于另一投资组合分布 $G(\cdot)$，则对所有 x，有 $P(X_F > x) \geq P(X_G > x)$，即投资组合 $F(\cdot)$ 的收益超过任一水平 x 的概率总比投资组合 $G(\cdot)$ 的收益超过 x 的概率大。因此，最大化预期效用的投资者会更偏好 $F(\cdot)$。

例 3.13：[二阶随机占优 (Second Order Stochastic Dominance)]

若对任意 $x \in (-\infty, \infty)$，有

$$\int_{-\infty}^{x} F(y)dy \leq \int_{-\infty}^{x} G(y)dy$$

且存在 x 使得严格不等式成立，则称概率分布 $F(\cdot)$ 二阶随机占优于概率分布 $G(\cdot)$。

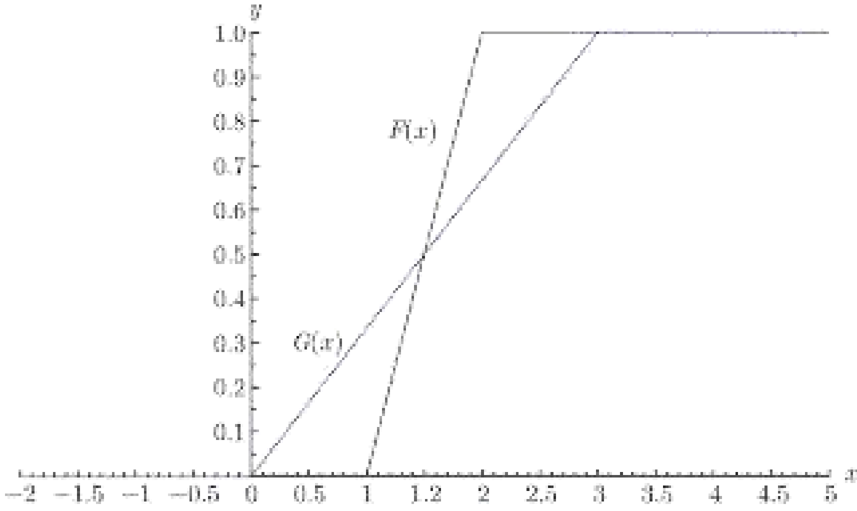


图 3.4：$F(\cdot)$ 二阶随机占优于 $G(\cdot)$

一阶随机占优与二阶随机占优

- 显然，一阶随机占优是二阶随机占优的充分条件。
- 直观上，偏爱更多收益的经济主体会偏好一阶随机占优分布，而偏爱更多收益但厌恶风险的经济主体会偏好二阶随机占优分布。
- 风险规避型的经济主体总是会偏好分布 $F(\cdot)$，因为对于任何递增且凹的效用函数 $u(\cdot)$，我们有：

$$\int_{-\infty}^{\infty} u(x)dF(x) \geq \int_{-\infty}^{\infty} u(x)dG(x)$$

- 当且仅当 $F(\cdot)$ 对 $G(\cdot)$ 具有二阶随机占优时。

目 录

定义 3.5

[离散随机变量 (Discrete Random Variables, DRV)]：若随机变量 X 可能的取值是有限个或可列个，则称其为离散随机变量。

例 3.14:

离散随机变量 X 对应的样本空间 Ω 仅包含可列个基本结果。例如，$\Omega = \{1, 2, 3, 4, 5, 6\}$ 或 $\Omega = \{0, 1, 2, \cdots\}$。

定义 3.6

[概率质量函数 (Probability Mass Function, PMF)]：离散随机变量 X 的概率质量函数定义为

$$f_X(x) = P(X = x),\quad 对所有\ x \in \mathbb{R}$$

定理 3.6

[PMF的性质]：

（1）对所有 $x \in \mathbb{R}$，$0 \le f_X(x) \le 1$；

（2）$\sum_{x\in\Omega} f_X(x) = 1$。

定义 3.7

[离散随机变量的支撑 (Support)]：离散随机变量 X 在实数集 $\mathbb{R}$ 上概率为正的所有点构成的集合称为 X 的支撑集合，记为

$$\mathrm{Support}(X) = \{x \in \mathbb{R}: f_X(x) > 0\}$$

因此有

$$\mathrm{Support}(X) = \Omega$$

- 直观上，X 的支撑是 X 取严格正概率的所有可能点构成的集合。
- 虽然 $f_X(x)$ 定义在整个实数集 $\mathbb{R}$ 上，但是，离散随机变量 X 的支撑以及支撑所包含的所有点的概率值可以充分地刻画该随机变量的概率分布。
- PMF 可用概率直方图描述。

定义 3.8

[概率直方图 (Probability Histogram)]：概率直方图是由一系列矩形构成的描述离散随机变量概率函数的图形。图中相邻的两个矩形之间没有空隙，并且每个矩形分别以具有严格正概率的每个值 x 为中心，其高度对应于 PMF 在点 x 处的概率值。

定理 3.7

假设离散随机变量 X 的 PMF 为 $f_X(x)$，则其 CDF

$$F_X(x) = P(X \leq x) = \sum_{y \leq x} f_X(y), \text{对任意 } x \in \mathbb{R}$$

其中，加和符号是指对 Ω 中所有小于或等于 x 的 y 值进行求和。

例 3.16：

假设随机变量 X 服从如下 PMF

$$f(x) = \begin{cases} \dfrac{1}{N}, & x = 1, \cdots, N \\ 0, & \text{其他} \end{cases}$$

求其 CDF $F_X(x)$。

解：

- 因为 N 个整数 $\{1,\cdot\cdot\cdot,N\}$ 发生的概率均相同，该分布也称为**离散型均匀分布** (discrete uniform distribution)。

解 (Cont.):

- 为计算 CDF $F_X(x)$，其中 $x \in \mathbb{R}$，我们将实数轴划分为 $N+1$ 段:
  - ✓ 情形1：$x < 1$。此时事件 $\{X \le x\}$ 为空集 $\emptyset$，即
  $$F_X(x) = \sum_{x_i \le x} f_X(x_i) = 0$$
  - ✓ 情形2：$1 \le x < 2$。此时事件 $\{X \le x\} = \{1\}$，即
  $$F_X(x) = \sum_{x_i \le x} f_X(x_i) = f_X(1) = \frac{1}{N}$$
  - ✓ 情形3：$2 \le x < 3$。此时事件 $\{X \le x\} = \{1, 2\}$，即
  $$F_X(x) = \frac{2}{N}$$

解 (Cont.):

- ✓ 情形 j：$j-1 \le x < j$，$2 \le j \le N$。此时事件 $\{X \le x\} = \{1,\cdots,j-1\}$，即

$$F_X(x) = \frac{j-1}{N}$$

- ✓ 情形 $N+1$：$x \ge N$。此时事件 $\{X \le x\} = \{1,\cdots,N\}$，即

$$F_X(x) = 1$$

- 汇总可得

$$F_X(x) = \begin{cases} 0, & x < 1 \\ j/N, & j \le x < j+1, 1 \le j < N \\ 1, & x \ge N \end{cases}$$

◆ 问题 3.7

假定 X 是 CDF 为 $F_X(x)$ 离散随机变量，是否可通过 $F_X(x)$ 求得 $f_X(x)$?

- 现用如下定理总结回答该问题。

定理 3.8

假设离散随机变量 X 的 CDF 为 $F_X(x)$，其支撑集由一系列的点 $\{x_1 < x_2 < \cdots\}$ 构成。则其 PMF 为

$$f_X(x_i) = \begin{cases} F_X(x_i), & i = 1 \\ F_X(x_i) - F_X(x_{i-1}), & i > 1 \end{cases}$$

CDF 与 PMF 的关系

- 定理 3.8 表明，可通过对 CDF $F_X(x)$ 求差分获得 PMF。
- $f_X(x)$ 和 $F_X(x)$ 在描述离散随机变量 X 的概率分布时是等价的。
- 给定 $f_X(x)$ 或 $F_X(x)$，均可获得随机变量 X 的完整概率法则。

目　录

定义 3.9

[连续随机变量 (Continuous Random Variables, CRV)]： 若随机变量 X 的累积分布函数$F_X(x)$ 是实数集上的连续函数，则称其为连续随机变量。反之，若 $F_X(x)$ 是阶梯函数，则称 X 为离散随机变量。

◆ 问题 3.8

对连续随机变量 X 可定义 PMF $f_X(x)$ 吗?

- 对任意常数 $\epsilon > 0$，有 $\{X = x\} \subset \left\{x - \frac{\epsilon}{2} < X \leq x + \frac{\epsilon}{2}\right\}$。根据 $F_X(x)$ 的连续性，对任意实数 x，可得

$$
\begin{aligned}
0 &\leq P(X = x) \\
&\leq P\left(x - \frac{\epsilon}{2} < X \leq x + \frac{\epsilon}{2}\right) \\
&= F_X\left(x + \frac{\epsilon}{2}\right) - F_X\left(x - \frac{\epsilon}{2}\right) \\
&\to 0,\ \text{当}\ \epsilon \to 0\ \text{时}
\end{aligned}
$$

- 因此，对所有实数点 x，有 $P(X = x) = 0$
- 即，若 X 为连续随机变量，则 X 取任何单点值的概率为零。

连续随机变量取任何单点值的概率为零

- 为直观理解，类比考虑卫星飞越中国领空的例子。假设某卫星从西到东飞越中国领空需要一个小时，飞过福建省需 2 分钟，飞越厦门市需 1 秒钟，不难理解，卫星飞越厦门大学经济楼的时间接近 0 秒。
- 对连续随机变量 X 而言，对所有 x 点都有 $P(X=x)=0$，该结果具有重要的含义。
  - ✓ 首先，由于 PMF $f_X(\cdot)$ 对连续随机变量是一个奇异函数，故不适于描述连续随机变量。
  - ✓ 其次，对连续随机变量 X，有

$$P(a<X\leq b)=P(a\leq X<b)$$
$$=P(a\leq X\leq b)$$

◆ 问题 3.9

由于 PMF $f_X(\cdot)$ 不适于描述连续随机变量 X，需要寻求其他工具。

在什么条件下，存在一个函数 $f_X(x)$ 使得 $F_X(x) = \int_{-\infty}^{x} f_X(y)dy$

成立？如何解释函数 $f_X(x)$ ？

定义 3.10

[绝对连续（Absolute Continuity）]：如果函数 $F(x)$ 在实数集上连续且几乎处处 (almost everywhere) 可导（即在几乎所有实数点 x 可导），则称函数 $F:\mathbb{R}\to\mathbb{R}$ 为关于勒贝格测度（Lebesgue measure）的绝对连续函数。

“几乎处处”是什么意思呢?

- 直观上，在 $\mathbb{R}$ 的任意有限区间内，$F_X(x)$ 仅在有限个点或无限但可数个点处不可求导。注意连续可导函数（即函数处处可导且导数是连续的）是绝对连续的。

定义 3.11

[概率密度函数（Probability Density Function, PDF)]：假设连续随机变量 X 的分布函数 $F_X\colon \mathbb{R} \to \mathbb{R}$ 绝对连续。则存在函数 $f_X(x)$，使得

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy, \text{对所有 } x \in (-\infty, \infty)$$

其中，函数 $f_X(x)$ 称为 X 的概率密度函数（PDF）。

定义 3.11 的含义

- 当 $F_X(x)$ 不满足绝对连续条件时，X 可能不具有上述关系。
- 对存在导数 $F_X'(x)$ 的点 x，由上述定义可得

$$f_X(x) = \frac{dF_X(x)}{dx} = F_X'(x)$$

- 若 $F_X(x)$ 绝对连续，则概率密度函数 $f_X(x)$ 几乎处处存在。
- 由于 $f_X(x)$ 相当于 CDF $F_X(x)$ 的斜率，故可取大于 1 的值。

◆ **问题 3.11**

如何解释 PDF $f_X(x)$?

- 对事件 $X \in \left(x-\frac{\epsilon}{2}, x+\frac{\epsilon}{2}\right]$，其中 $\epsilon > 0$，其中 $\epsilon > 0$ 为任意小的常数，根据中值定理可得

$$P\left(x-\frac{\epsilon}{2} < X \leq x+\frac{\epsilon}{2}\right) = F_X\left(x+\frac{\epsilon}{2}\right) - F_X\left(x-\frac{\epsilon}{2}\right)$$

$$= \int_{x-\frac{\epsilon}{2}}^{x+\frac{\epsilon}{2}} f_X(y)dy$$

$$= f_X(\bar{x})\epsilon$$

其中 $\bar{x}$ 是位于 $x-\frac{\epsilon}{2}$ 和 $x+\frac{\epsilon}{2}$ 之间的一个点。

◆ **问题 3.11 (Cont.)**

如何解释 PDF $f_X(x)$?

- 尽管 $f_X(x)$ 本身不是概率测度，但它与 X 在以 x 为中心的小邻域内取值的概率成正比例。因此，$f_X(x)$ 刻画了 X 在以 x 为中心的小邻域内取值的概率的相对大小。特别地，$f_X(x)$ 可用于描述连续随机变量 X 的概率分布的形状，包括众数 (mode)、偏度 (skewness) 或对称性 (symmetry)、厚尾 (heavy tails) 等。

◆ 问题 3.12

给定 $F_X(x)$，$f_X(x)$ 是否唯一?

- 给定 CDF $F_X(x)$，可在 $F_X(x)$ 可导处求得 PDF $f_X(x) = F_X'(x)$。
  - ✓ 若 $F_X(x)$ 在整个实数轴上可导，则 $f_X(x) = F'(x)$ 是唯一的。
  - ✓ 然而，若 $F_X(x)$ 在某些点不可导，$f_X(x)$ 在这些点上没有定义。

◆ 问题 3.13

在导数 $F'_X(x)$ 不存在的点上，如何定义 PDF $f_X(x)$ 的值?

- 可在这些点上任意定义 $f_X(x)$。连续随机变量 X 的性质 $P(X = x) = 0$ 意味着，在一个单点或无限可数点序列上改变连续随机变量 X 的 PDF $f_X(x)$ 的取值，将不会影响 X 的概率分布函数。

CONTINUE

- 例如，考虑如下两个 PDF：

$$f_X(x)=\begin{cases}e^{-x}, & 0<x<\infty\\ 0, & \text{其他}\end{cases}$$

和

$$f_X(x)=\begin{cases}e^{-x}, & 0\le x<\infty\\ 0, & \text{其他}\end{cases}$$

- 注意这两个 PDF 仅在 $x=0$ 处取值不同，并且 $P(X=x)=0$。因此，这两个 PDF 对应的概率分布在实数轴的任何子集 A 上具有相同的概率 $P_X(A)$。

CONTINUE

- 更一般地，若两个 PDF 仅在有限个点或无限可数点序列上取值不同，其对应的 CDF 或概率函数将完全相同。
- 相反地，离散随机变量的 PMF $f_X(x)$ 在其支撑中的任意点的取值都不能改变，因为任何这种变化将改变离散随机变量的概率分布。
- 如上所述，一般情况下，PDF 并不唯一。然而，在许多应用中，连续的 PDF $f_X(x)$ 比不连续的 PDF 更方便。由于这个原因，一般要求 PDF $f_X(x)$ 在实数轴 $\mathbb{R}$ 上连续。

定理 3.9

[PDF 的性质]：当且仅当

(1) 对所有 $x \in \mathbb{R}$，$f_X(x) \geq 0$;

(2) $\int_{-\infty}^{\infty} f_X(x)dx = 1$

函数 $f_X(x)$ 是连续随机变量 X 的 PDF。

证明：[必要性]

- 若 $f_X(x)$ 为 PDF，则根据定义可知 $F_X(x) = \int_{-\infty}^{x} f_X(y)dy$ 为 CDF。由中值定理可得，$F_X(x+\delta) - F_X(x) = f_X(\bar{x})\delta$，其中 $\bar{x}$ 是在 x 与 $x+\delta$ 之间的一个取值。
- 因为 $F_X(x)$ 是非递减函数，故 $f_X(\bar{x}) \geq 0$，又由 $F_X(\infty) = 1$，可得条件 (2)。

证明 (Cont.):

[充分性]

- 构造函数 $F_X(x) = \int_{-\infty}^{x} f_X(y)dy$ 并证明 $F_X(x)$ 为 X 的 CDF。
- 首先，根据中值定理及 $f_X(x) \geq 0$ 可知，$F_X(x)$ 为非减函数。
- 其次，由 $\int_{-\infty}^{\infty} f_X(x)dx = 1$，有 $F_X(-\infty) = 0$ 且 $F_X(\infty) = 1$。
- 同时，由可知，$F_X(x)$ 为连续函数，当 $\delta \to 0$ 时，

$$F_X(x+\delta) - F_X(x) = \int_{x}^{x+\delta} f_X(y)dy \to 0$$

- 故 $F_X(x)$ 为右连续函数。因此，$F_X(x)$ 为连续随机变量 X 的 CDF。

证毕。

定义 3.12

[**连续随机变量的支撑**]：连续随机变量 X 的支撑定义为

$$\text{Support}\,(X) = \{x \in \mathbb{R}: f_X(x) > 0\}$$

其中 $f_X(x)$ 是 X 的 PDF。

- 连续随机变量 X 的支撑是实数轴 $\mathbb{R}$ 上具有严格正 PDF $f_X(x)$ 的所有点的集合。这意味着，
  - ✓ X 在其**支撑中**任意一点的极小邻域内取值的**概率均为正**。
  - ✓ X 在其**支撑之外**任意一点的极小邻域内取值的**概率均为零**。
- 因此，在计算连续随机变量的概率时只需关注 X 的支撑。

引理 3.1

[位置-尺度函数类 (Location-Scale family)]：令 $f_X(x)$ 表示随机变量 X 的 PDF，并且 μ 和 $\sigma > 0$ 为任意两个常数。则函数

$$g_X(x) = \frac{1}{\sigma} f_X\left(\frac{x-\mu}{\sigma}\right)$$

仍然是 PDF。

引理 3.1 含义

- 任何一对参数值 (μ,σ) 对应一个 PDF。
  - ✓ 由参数 μ 刻画的函数类 $f_X(x-\mu)$ 称为标准 PDF $f_X(x)$ 的**位置函数类**，其中参数 μ 称为这类函数的位置参数；
  - ✓ 由参数 σ 刻画的函数类 $\frac{1}{\sigma}f_X\left(\frac{x}{\sigma}\right)$ 称为标准 PDF $f_X(x)$ 的**尺度函数类**，其中参数 σ 称为这类函数的尺度参数；
  - ✓ 由参数 (μ,σ) 刻画的函数类 $\frac{1}{\sigma}f_X\left(\frac{x-\mu}{\sigma}\right)$ 称为标准 PDF $f_X(x)$ 的**位置-尺度函数类**，其中 μ 和 σ 分别是位置参数和尺度参数。

连续随机变量 vs. 离散随机变量

X 类型	离散型	连续型
累积分布函数	$F_X(x) = P(X \le x), \forall x \in \mathbb{R}.$	
概率函数	PMF $f_X(x) = P(X = x)$	PDF $f_X(x) = \frac{dF_X(x)}{dx} = F_X'(x)$
概率函数性质	(1) $0 \le f_X(x) \le 1$ (2) $\sum_{x\in\Omega} f_X(x) = 1$	(1) $f_X(x) \ge 0$ (2) $\int_{-\infty}^{\infty} f_X(x)dx = 1$
概率函数含义	对应取值 x 的概率	与 X 落在以 x 为中心的小邻域内的概率成正比： $f_X(x) \propto P\left(x - \frac{\epsilon}{2} < X \le x + \frac{\epsilon}{2}\right)$

定义 3.13

[离散-连续型混合分布 (Mixed Distribution of Discrete and Continuous Components)]：如果随机变量 X 的 CDF 在每个非零概率的实数点都不连续，而在其他实数点都连续，且并不是一个纯粹阶跃函数，则称其服从连续离散混合分布。分布函数在非连续点处与离散随机变量情形类似，其跳跃的高度决定了 X 在该点的概率值；而在其他点上，X 的行为与连续随机变量相同。对包括离散和连续元素的混合随机变量 X，其 CDF 是两个 CDF 的加权平均，其中一个为离散随机变量的 CDF，另一个为连续随机变量的 CDF，权重非负且其和为 1。

离散-连续型混合分布

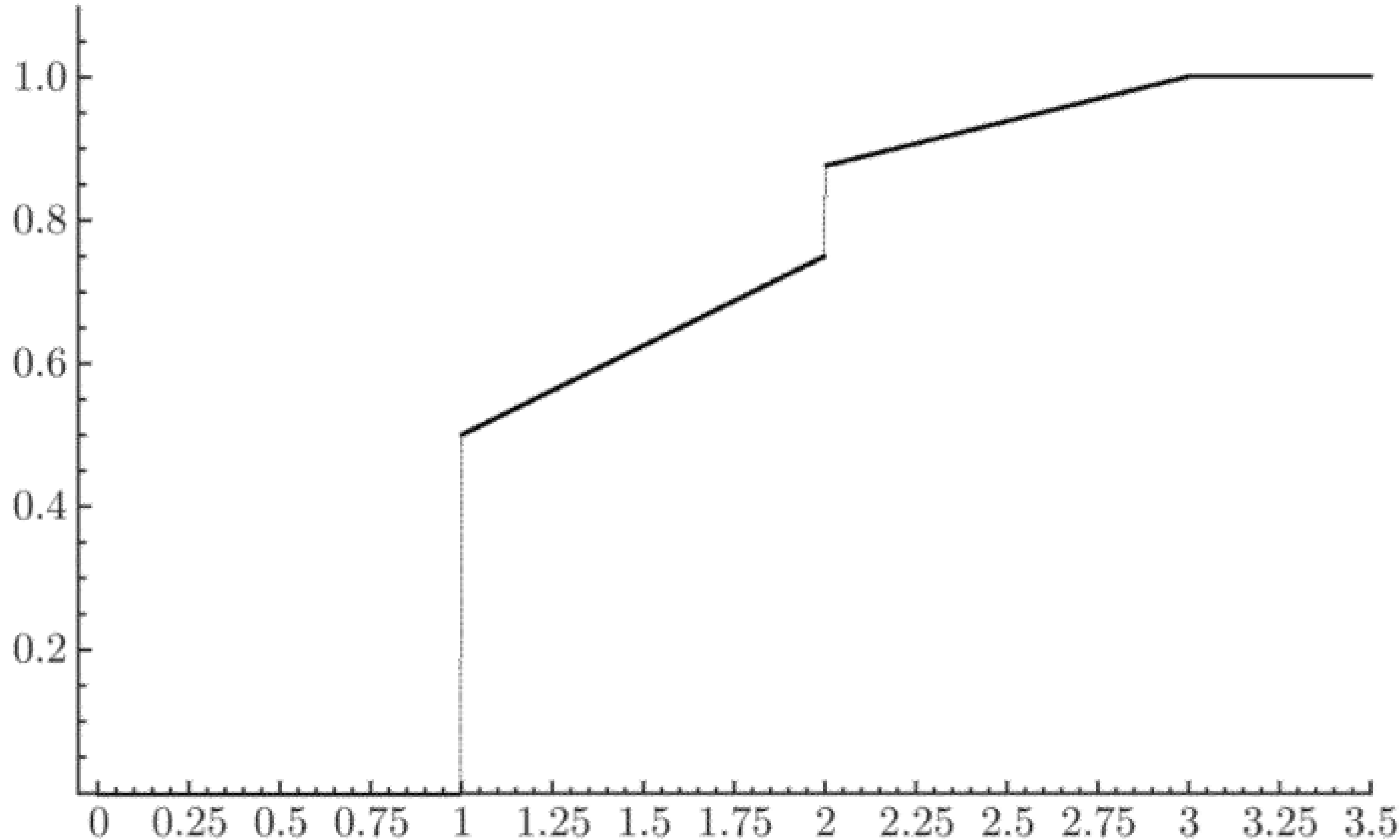


图 **3.8**：具有离散和连续元素的混合随机变量的 CDF

现实中的离散-连续型混合随机变量

- 例如，某随机经济变量 X^* 服从连续分布，
  - ✓ 当 $X^* > c$ 时，可观察到 X^* 的值；
  - ✓ 当 $X^* \leq c$ 时，则只能观察到 c 值。此时，观测变量

$$X = \begin{cases} X^*, & X^* > c \\ c, & X^* \leq c \end{cases}$$

- 这里 $P(X = c) > 0$ 且对所有 $x > c$，$P(X = x) = 0$，故 X 服从离散连续型混合分布。该分布称为左侧受限分布。
  - ✓ 当 X^* 为被辞退雇员失业久期（duration）、工人的保留工资 (reservation wage) 等时，会出现此类分布。
- 归并回归模型（censored regression model）就是用于对此类随机变量建模的计量经济学模型（参见洪永淼，2011，第九章）。

定理 3.10

[勒贝格分解 (Lebesgue's Decomposition)]：任意 CDF $F_X(x)$ 可写成如下形式

$$F_X(x) = \sum_{i=1}^{3} a_i F_i(x)$$

其中

- 权重 $a_i \geq 0$，$i = 1,2,3$，$a_1 + a_2 + a_3 = 1$，
- $F_1(x)$ 为“绝对连续”，
- $F_2(x)$ 为具有有限或无限但可数个“跳跃”的“阶梯函数”，
- $F_3(x)$ 为“奇异”函数，也就是连续并且几乎处处导数为零的函数。

目 录

◆ 问题 3.14

假设 $g:\mathbb{R}\to\mathbb{R}$ 为实值可测函数，则 $Y=g(X)$ 是一个新的随机变量。如何求 Y 的概率分布?

例 3.19：[消费函数 (Consumption Function)]

- $Y=g(X)$，其中
  - ✓ X 为收入，
  - ✓ Y 为消费。
- 通常假设消费为收入的线性函数，即

$$Y=\alpha+\beta X$$

例 3.20：[期权价格 (Options Prices)]

根据著名的 Black & Scholes (1973) 期权定价公式，欧式看涨期权 (call option) 价格 Y 为波动率 X 的非线性函数：

$$Y = P_0\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$

其中，

- $d_1 = \frac{\ln(P_0/K)+(r+X^2/2)T}{X\sqrt{T}}$，$d_2 = \frac{\ln(P_0/K)+(r-X^2/2)T}{X\sqrt{T}}$
- 函数 $\Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-u^2/2}du$ 是标准正态分布的 CDF（参见第四章），
- P_0 是初始时刻的股票价格，K 是执行价格，
- r 是连续复合无风险利率，
- T 是期权到期时间。

例 3.21：[资产收益率 (Asset Returns)]

假设 P_t 为第 t 期的资产价格。则其对数收益率

$$X_t = \ln(P_t) - \ln(P_{t-1}) \approx \frac{P_t - P_{t-1}}{P_{t-1}}$$

近似等于从 $t-1$ 期到 t 期的相对价格变化率。

3.5.1 离散情形

◆ **问题 3.15**

给定离散随机变量 X 的 PMF $f_X(x)$，如何求得 PMF $f_Y(y)$?

- 对离散随机变量 X，通用公式如下：

$$f_Y(y) = \sum_{x\in\Omega_X(y)} f_X(x)$$

其中，对任意给定实数值 y，

$$\Omega_X(y) = \{x \in \Omega_X : g(x) = y\}$$

- 直观上，$Y = g(X)$ 是从样本空间 Ω_X 到新样本空间 Ω_Y 的映射。Y 的 PMF $f_Y(y)$ 可通过 PMF $f_X(x)$ 定义为诱导概率函数，即对任意实数 y，

$$f_Y(y) = P(Y = y) = P[X \in \Omega_X(y)]$$

例 3.22:

假定 X 服从概率分布:

X	-2	-1	0	1	2
$f_X(x)$	0.2	0.1	0.1	0.3	0.3

求 $Y = X^2 + X$ 的 PMF。

解: 当 $X = -2, -1, 0, 1, 2$, 有 $Y = X^2 + X = 2, 0, 0, 2, 6$, 故 Y 的支撑为 $\Omega_Y = \{0, 2, 6\}$.

- 同时,

$$P(Y = 0) = P_X(X = -1) + P_X(X = 0) = 0.2$$

$$P(Y = 2) = P_X(X = -2) + P_X(X = 1) = 0.5$$

$$P(Y = 6) = P_X(X = 2) = 0.3$$

- 因此,Y 的概率分布为

Y	0	2	6
$f_Y(y)$	0.2	0.5	0.3

3.5.2 连续情形

◆ 问题

假设 $g(\cdot)$ 为连续函数，则当 X 为连续随机变量时，Y 也是连续随机变量。现在探讨给定 X 的 PDF $f_X(x)$，如何求解 Y 的PDF $f_Y(y)$。

(1) CDF 方法

- **基本思想：** 先求得 Y 的 CDF $F_Y(y)$，后对其求导，得 PDF $f_Y(y)=F'_Y(y)$。
- **步骤一：** 用 $F_X(x)$ 表示 $F_Y(y)$， 即：

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P[g(X) \le y] \\ &= P\left[X \in \Omega_{g^{-1}}(y)\right] \end{aligned}$$

其中 $\Omega_{g^{-1}}(y)=\{x\in\Omega_X\colon g(x)\le y\}$ 为 Ω_X 的一个子集。

- **步骤二：** 对 CDF $F_Y(y)$ 关于 y 求导，得

$$f_Y(y)=F'_Y(y)$$

- **步骤三：** 检验 $f_Y(y)$ 是否为 PDF（即对任意实数 y，检验 $f_Y(y)\ge 0$ 和 $\int_{-\infty}^{\infty} f_Y(y)dy=1$。

例 3.23：

假设连续随机变量 X 的 PDF 为

$$f_X(x) = \begin{cases} 1, & -\dfrac{1}{2} < x < \dfrac{1}{2} \\ 0, & \text{其他} \end{cases}$$

求如下随机变量的 PDF $f_Y(y)$。

（1） $Y = a + bX$，$b \neq 0$;

（2） $Y = X^2$;

（3） $Y = |X|$。

解：

- 求解此类问题的关键在于识别 Y 的所有可能取值，即 Y 的支撑。为此有必要画出 $Y=g(X)$ 的曲线图。

(1) 当 $Y=a+bX$ 时，有

$$\begin{aligned} F_Y(y) &= P(Y\le y)\\ &= P(a+bX\le y)\\ &= P(bX\le y-a) \end{aligned}$$

- 当 $b<0$ 时，除以 b 将改变不等式的方向。

- 因此需要分别讨论 $b>0$ 和 $b<0$ 的两种情形。

解 (Cont.):

情形 (1)： 当 $b > 0$ 时，有

$$F_Y(y) = P(Y \le y) = P(bX \le y - a) = P[X \le (y-a)/b] = F_X\left(\frac{y-a}{b}\right)$$

- ✓ 故有

$$F_Y(y) = F_X\left(\frac{y-a}{b}\right) = F_X(z)$$

- ✓ 其中 $z = (y-a)/b$。

解 (Cont.):

- ✓ 采用链式微分法则 (chain rule of differentiation)，得

$$
\begin{aligned}
f_Y(y) &= F_Y'(y) \\
&= F_X'(z)\frac{dz}{dy} \\
&= f_X(z)\frac{1}{b} \\
&= f_X\left(\frac{y-a}{b}\right)\frac{1}{b} \\
&= 1 \cdot \frac{1}{b}, -\frac{1}{2} < \frac{y-a}{b} < \frac{1}{2}
\end{aligned}
$$

- ✓ 因此

$$
f_Y(y) = \begin{cases} \frac{1}{b}, & a - \frac{b}{2} < y < a + \frac{b}{2} \\ 0, & \text{其他} \end{cases}
$$

解 (Cont.):

- ✓ 由于随机变量 X 的 PDF $f_X(\cdot)$ 在区间 $\left(-\frac{1}{2},\frac{1}{2}\right)$ 上各点的取值为常数，其分布称为区间 $\left(-\frac{1}{2},\frac{1}{2}\right)$ 上的均匀分布。因为 Y 是 X 的线性变换，故 Y 在区间 $\left(a-\frac{b}{2},\ a+\frac{b}{2}\right)$ 上也服从均匀分布。
- ✓ 参数 a 和 b 分别改变均匀分布的均值（mean）和尺度(scale)，其中 a 称为位置参数，b 称为尺度参数。

情形 (2)： 当 $b<0$ 时，有

$$
\begin{aligned}
P(Y\le y) &= P(bX\le y-a)\\
&= P\left(X\ge \frac{y-a}{b}\right)\\
&= 1-F_X\left(\frac{y-a}{b}\right)
\end{aligned}
$$

解 (Cont.):

✓ 因此，$F_Y(y) = 1 - F_X(z)$，其中 $z = (y-a)/b$。对其求导可得

$$f_Y(y) = 0 - F_X'(z)\frac{dz}{dy}$$
$$= -f_X(z)\frac{1}{b}$$
$$= -f_X\left(\frac{y-a}{b}\right)\frac{1}{b}$$
$$= -\frac{1}{b}, -\frac{1}{2} < \frac{y-a}{b} < \frac{1}{2}$$

✓ 故 Y 的 PDF 为

$$f_Y(y) = \begin{cases} -\frac{1}{b}, & a + \frac{b}{2} < y < a - \frac{b}{2} \\ 0, & \text{其他} \end{cases}$$

解 (Cont.):

(2) 由于 $Y = X^2$ 且 $\text{Support}(X) = \left(-\frac{1}{2}, \frac{1}{2}\right)$，$Y$ 在 $\left[0, \frac{1}{4}\right)$ 上取非负值。令 $y \geq 0$，有

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P\left(X^2 \leq y\right) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

- 其中，用到了事件 $\{X^2 \leq y\}$ 和事件 $\{-\sqrt{y} \leq X \leq \sqrt{y}\}$ 两者等价这一事实。

解 (Cont.):

- 对其求导可得

$$f_Y(y) = \frac{d}{dy}[F_X(\sqrt{y}) - F_X(-\sqrt{y})]$$

$$= F_X'(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) - F_X'(-\sqrt{y})\left(-\frac{1}{2\sqrt{y}}\right)$$

$$= f_X(\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right) + f_X(-\sqrt{y})\left(\frac{1}{2\sqrt{y}}\right)$$

- 由此可得

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{y}}, & 0 < y < \frac{1}{4} \\ 0, & \text{其他} \end{cases}$$

解 (Cont.):

(3) 因为 $Y=|X|$，Y 为非负随机变量。令 $y\geq 0$，有

$$\begin{aligned}F_Y(y) &= P(Y\leq y)\\ &= P(|X|\leq y)\\ &= P(-y\leq X\leq y)\\ &= F_X(y)-F_X(-y)\end{aligned}$$

- 对其求导可得

$$\begin{aligned}f_Y(y) &= F_X'(y)-F_X'(-y)(-1)\\ &= f_X(y)+f_X(-y)\end{aligned}$$

解 (Cont.):

- 因此，

$$f_Y(y) = \begin{cases} 2, & 0 \le y < \dfrac{1}{2} \\ 0, & 其他 \end{cases}$$

- 该函数是区间 $\left[0, \frac{1}{2}\right)$ 上的均匀密度函数，即均匀分布随机变量的绝对值仍然是均匀分布 (但区间更小)。

例 3.27：

- 假设连续随机变量 X 的 PDF 为

$$f_X(x)=\begin{cases}\frac{3}{8}(x+1)^2, & -1<x<1\\ 0, & \text{其他}\end{cases}$$

- 定义

$$Y=\begin{cases}1-X^2, & X\le 0\\ 1-X, & X>0\end{cases}$$

- 求随机变量 Y 的 PDF $f_Y(y)$。

解：

- **步骤一：** Y 的支撑集为 $0 < y \leq 1$.

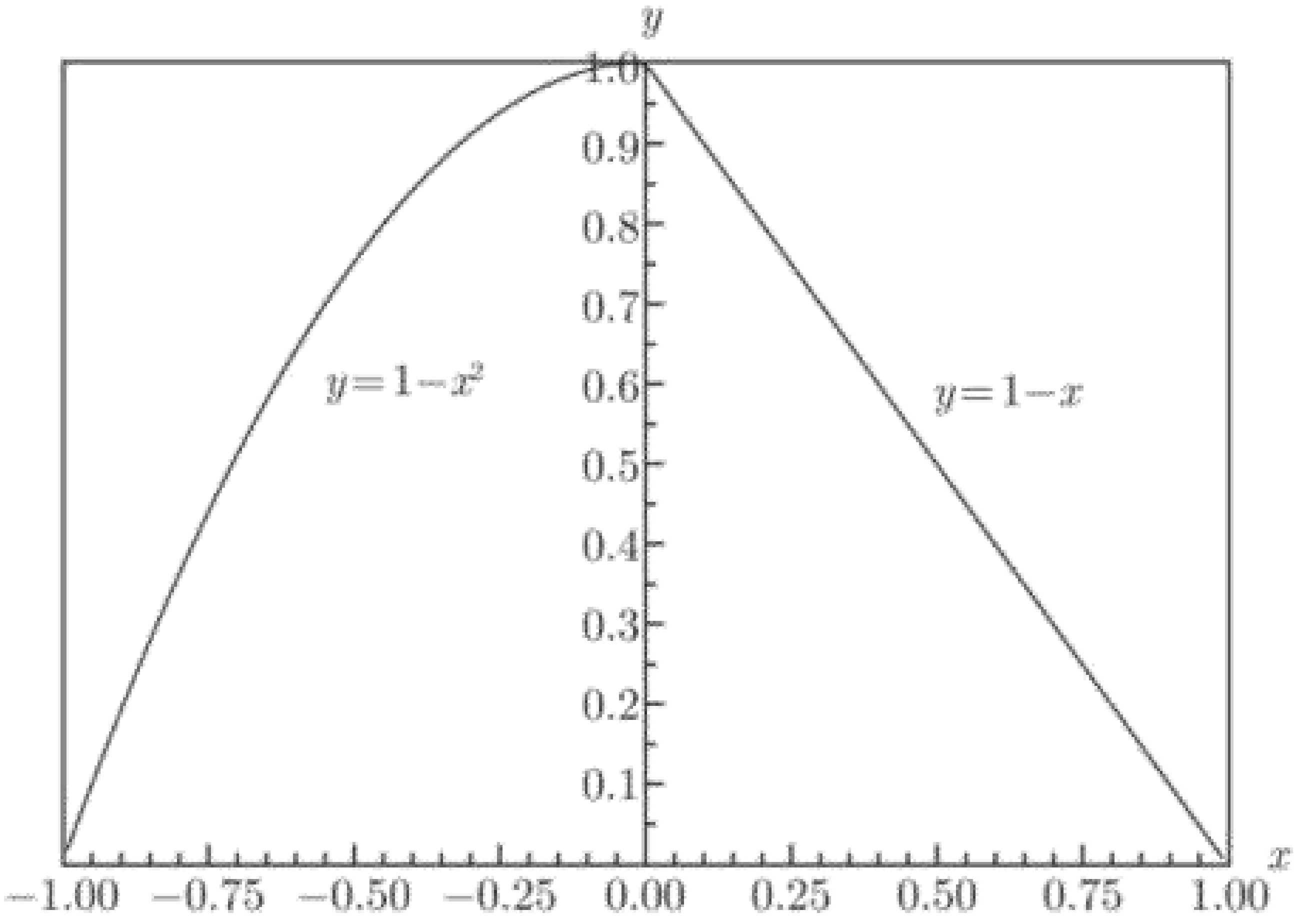


图 3.9：例 3.27 中 Y 的支撑

解 (Cont.):

- **步骤二：**对 $0 < y \leq 1$，有

$$F_Y(y) = P(Y \leq y) = P(-1 < X \leq -\sqrt{1-y} \text{ 或 } 1-y \leq X < 1)$$
$$= P(-1 < X \leq -\sqrt{1-y}) + P(1-y \leq X < 1)$$
$$= F_X(-\sqrt{1-y}) - F_X(-1) + F_X(1) - F_X(1-y)$$

- 其中，使用了集合 $\{-1 < X \leq -\sqrt{1-y}\}$ 和集合 $\{1-y \leq X < 1\}$ 为互斥事件这一性质。

解 (Cont.):

- 因此

$$
\begin{aligned}
f_Y(y) = F_Y'(y) &= F_X'(-\sqrt{1-y})\frac{1}{2\sqrt{1-y}} + F_X'(1-y) \\
&= f_X(-\sqrt{1-y})\frac{1}{2\sqrt{1-y}} + f_X(1-y) \\
&= \frac{3}{8}(1-\sqrt{1-y})^2\frac{1}{2\sqrt{1-y}} + \frac{3}{8}(2-y)^2, 0 < y < 1
\end{aligned}
$$

- 即，Y 的 PDF 为

$$
f_Y(y) = \begin{cases} \frac{3}{8}(1-\sqrt{1-y})^2\frac{1}{2\sqrt{1-y}} + \frac{3}{8}(2-y)^2, & 0 < y < 1 \\ 0, & \text{其他} \end{cases}
$$

定理 3.11

[概率积分变换 (Probability Integral Transform)]：假设随机变量 X 有连续且严格单调递增的 CDF $F_X(x)$。定义 $Y = F_X(X)$，即

$$Y = \int_{-\infty}^{X} f_X(x)dx$$

则 Y 服从区间 $[0,1]$ 上的均匀分布，即 Y 的 PDF 对于 $0 \leq y \leq 1$ 为 $f_Y(y) = 1$，对于 y 的其他点则取值为 0。

证明：

- $Y=F_X(X)$ 的支撑单位区间 $[0,1]$。令 $y\in[0,1]$，有

$$F_Y(y)=P(Y\le y)$$
$$=P[F_X(X)\le y]$$

- 因 $F_X(x)$ 严格递增，其反函数，记作 $F_X^{-1}(y)$，存在且严格递增。对任意实数 x，有

$$F_X^{-1}[F_X(x)]=x$$

- 根据反函数运算法则，可得

$$\begin{aligned}F_Y(y)&=P(Y\le y)\\&=P[F_X(X)\le y]\\&=P\{F_X^{-1}[F_X(X)]\le F_X^{-1}(y)\}\\&=P\left[X\le F_X^{-1}(y)\right]\\&=F_X\left[F_X^{-1}(y)\right]\\&=y,\qquad y\in[0,1]\end{aligned}$$

证明 (Cont.):

- 因此，Y 的 PDF 为

$$f_Y(y) = \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{其他} \end{cases}$$

- 该分布为单位区间 $[0,1]$ 上的均匀分布，称为标准均匀分布，记作 $U[0,1]$。

证毕。

定理 3.11的意义

- 定理 3.11 的结果不仅具有重要理论价值，而且可用于模拟生成任意概率分布的随机数。
  - ✓ 计算机科学家在利用计算机生成服从均匀分布的随机数方面已取得较为成熟的研究成果。
  - ✓ 许多算法生成的仿随机数通过了几乎所有关于均匀分布的检验。
  - ✓ 目前，绝大多数成熟的统计软件包均有可靠的均匀分布随机数生成器。
  - ✓ 更多关于仿随机数生成机制的探讨，可参考 Devroye (1985) 或 Ripley (1987)。

均匀分布的重要意义

- 基于均匀分布的随机数，可通过概率积分变换生成任意连续分布的随机数。
- 具体而言，为了生成服从某个特定分布 $F_X(\cdot)$ 的观测值，首先借助计算机生成一个服从标准均匀分布 $U[0,1]$ 的实现值，记为 y，然后根据等式 $F_X(x)=y$ 求解 x。则 $x=F_X^{-1}(y)$ 即为服从特定分布 $F_X(\cdot)$ 的随机变量 X 的一个实现值。
- 例如，考虑生成服从所谓指数分布的随机数。
  - ✓ 若随机变量 X 的 PDF 满足当 $x\geq 0$ 时为 $f_X(x)=e^{-x}$，而当 $x<0$ 时取值为 0，则称其服从标准指数分布，记作 $EXP(1)$。

均匀分布的重要意义 (Cont.)

✓ X 的 CDF 为

$$F_X(x)=\begin{cases}1-e^{-x}, & x\geq 0\\ 0, & \text{其他}\end{cases}$$

定义 $Y=F_X(X)=1-e^{-X}$，则 Y 服从 $U[0,1]$ 分布。

✓ 现在使用计算机从标准均匀分布 $U[0,1]$ 生成一个数 y，则

$$x=-\ln(1-y)$$

即为标准指数分布的一个实现值。

检验分布模型拟合优度

- 为了检验随机变量 X 是否符合假设分布 $F_0(\cdot)$，可以首先计算概率积分变换 $Y=F_0(X)$，然后通过随机样本 $\{Y_i=F_0(X_i)\}_{i=1}^n$ 检查 Y 是否服从均匀分布 $U[0,1]$。

- 当且仅当随机样本 $\{X_i\}_{i=1}^n$ 符合假设分布 $F_0(\cdot)$ 时，Y_i 将服从 $U[0,1]$ 分布。

- **QQ 图 (QQ-plot)**：可用于检验 $X_1,\cdots,X_n$ 是否服从预设分布 $F_0(\cdot)$。定义如下函数

$$\hat{F}_Y(y)=\frac{1}{n}\sum_{i=1}^n \mathbf{1}(Y_i\le y)$$

- 若 X_i 服从假设分布 $F_0(\cdot)$，Y_i 将服从 $U[0,1]$ 分布，从而 $\hat{F}_Y(y)$ 在 $0\le y\le 1$ 上近似为 45 度线。反之，则不近似在 45 度线。

CDF 方法 vs. 变换法

- 以上介绍的 CDF 法是求 $Y = g(X)$ 的 PDF 的通用方法。
  - ✓ 先求得 $Y = g(X)$ 的 CDF $F_Y(y)$，
  - ✓ 再对 $F_Y(y)$ 求导得 PDF $f_Y(y)$。
- 在某些情形下，CDF 方法可能十分繁琐。
- 若 $g(\cdot)$ 为严格单调函数，可采用变换法获得 $Y = g(X)$ 的 PDF $f_Y(y)$ 的计算公式。

(2) 变换法 (Transformation Method)

定理 3.12

[单变量变换 (Univariate Transformation)]：假设连续随机变量 X 的 PDF 为 $f_X(x)$，且函数 $g:\mathbb{R}\to\mathbb{R}$ 为严格单调且在 X 的支撑上可导。

- 则对随机变量 $Y=g(X)$ 在其支撑上的任意取值 y，有

$$f_Y(y)=f_X(x)\frac{1}{|g'(x)|}$$

其中 x 是 X 的支撑上满足 $g(x)=y$ 的唯一数值；

- 对不在 Y 的支撑上的点 y，则 $f_Y(y)=0$。

证明:

- 首先考虑 $g(x)$为严格递增函数的情形。对严格递增函数 $g(x)$，存在唯一的严格递增反函数 $g^{-1}(y)$，满足 $g^{-1}[g(x)]=x$。
- 对 Y 支撑上的任意 y，有

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P[g(X) \le y] \\ &= P[X \le g^{-1}(y)] \\ &= F_X[g^{-1}(y)] \end{aligned}$$

- 根据链式求导法则，可得

$$\begin{aligned} f_Y(y) &= F_Y'(y) \\ &= F_X'[g^{-1}(y)]\frac{d}{dy}g^{-1}(y) \\ &= f_X(x)\frac{1}{g'(x)} \end{aligned}$$

证明 (Cont.):

- 上述推导用到了 $x = g^{-1}(y)$，以及
$$\frac{d}{dy} g^{-1}(y) = \frac{1}{g'(x)}$$
- 此式可通过对下式求导得出 $g^{-1}(y) = x$, 其中 $y = g(x)$。
- 接下来，考虑 $g(x)$ 为单调递减函数的情形。采用类似的推导过程可得
$$f_Y(y) = -f_X(x) \frac{1}{g'(x)}$$
- 其中 $x = g^{-1}(y)$。因此，综合单调递增和单调递减两种情形，得
$$f_Y(y) = f_X(x) \frac{1}{|g'(x)|}$$
- 其中，对 Y 支撑上的任意 y，有 $x = g^{-1}(y)$。**证毕。**

定理 3.12 的含义

- 为更好地理解定理 3.12，考察事件 $y-\epsilon<Y<y+\epsilon$ 的概率，其中常数 $\epsilon>0$。当 ϵ 很小时，

$$\mathrm{P}(y<Y<y+\epsilon)\sim f_Y(y)\epsilon$$

- 简单起见，假设 $g(\cdot)$ 为严格单调递增函数。则 X 的等价事件为 $x<X<x+\delta$，其中 $x=g^{-1}(y),\delta=\frac{d}{dy}g^{-1}(y)\epsilon=\frac{1}{g'(x)}\epsilon$。
- 对于小 $\delta>o$，此等价事件的概率为

$$P(x<X<x+\delta)\sim f_X(x)\delta$$

- 由于关于 X 和 Y 的两个等价事件具有相同概率，因此 $f_Y(y)\epsilon=f_X(x)\delta$，故 $f_Y(y)=f_X(x)\frac{1}{g'(x)}$。如图 3.13 所示。

定理 3.12 的含义 (Cont.)

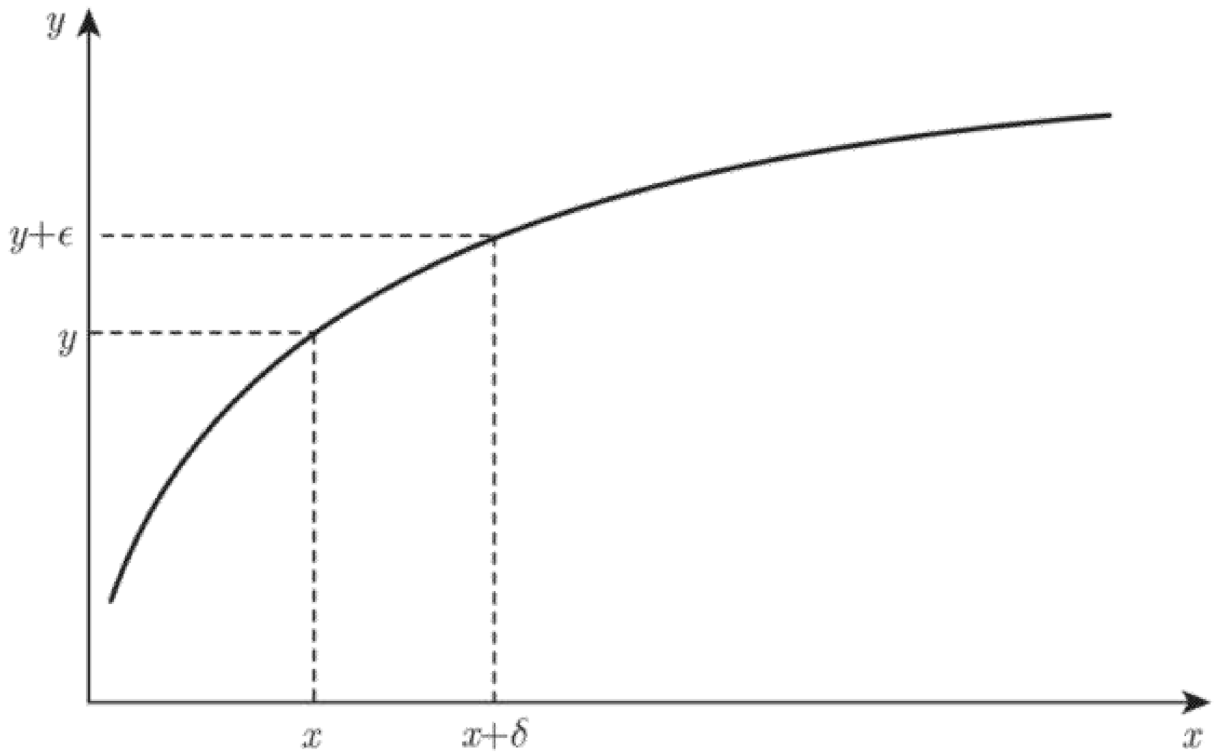


图 3.13： 定理 3.12 单变量变换

- 若 $Y = g(X)$ 在实数轴上的多个区间都分别严格单调，可将上述单变量变换定理（定理 3.12）扩展到更一般的情形。

定理 3.13

假设当 $x \in A_i$ 时，$g(x) = g_i(x)$，其中 $i = 1,\cdots,k$，对任意 i，$g_i(x)$ 均是区间 A_i 上的严格单调（单调递增或单调递减）且可导的函数。又假设 k 个区间 $\{A_i\}$ 互不相交且 $\cup_{i=1}^{k} A_i = \mathbb{R}$。则对 $Y = g(X)$ 的支撑 Ω_Y 上任何 y 值，有

$$f_Y(y) = \sum_{i=1}^{k} f_X[g_i^{-1}(y)] \frac{1}{|g_i'[g_i^{-1}(y)]|}$$

如何理解定理 3.13

- 为了直观解释定理 3.13，考虑共有 $k=3$ 个子区间的一个简单例子，如图 3.14 所示。

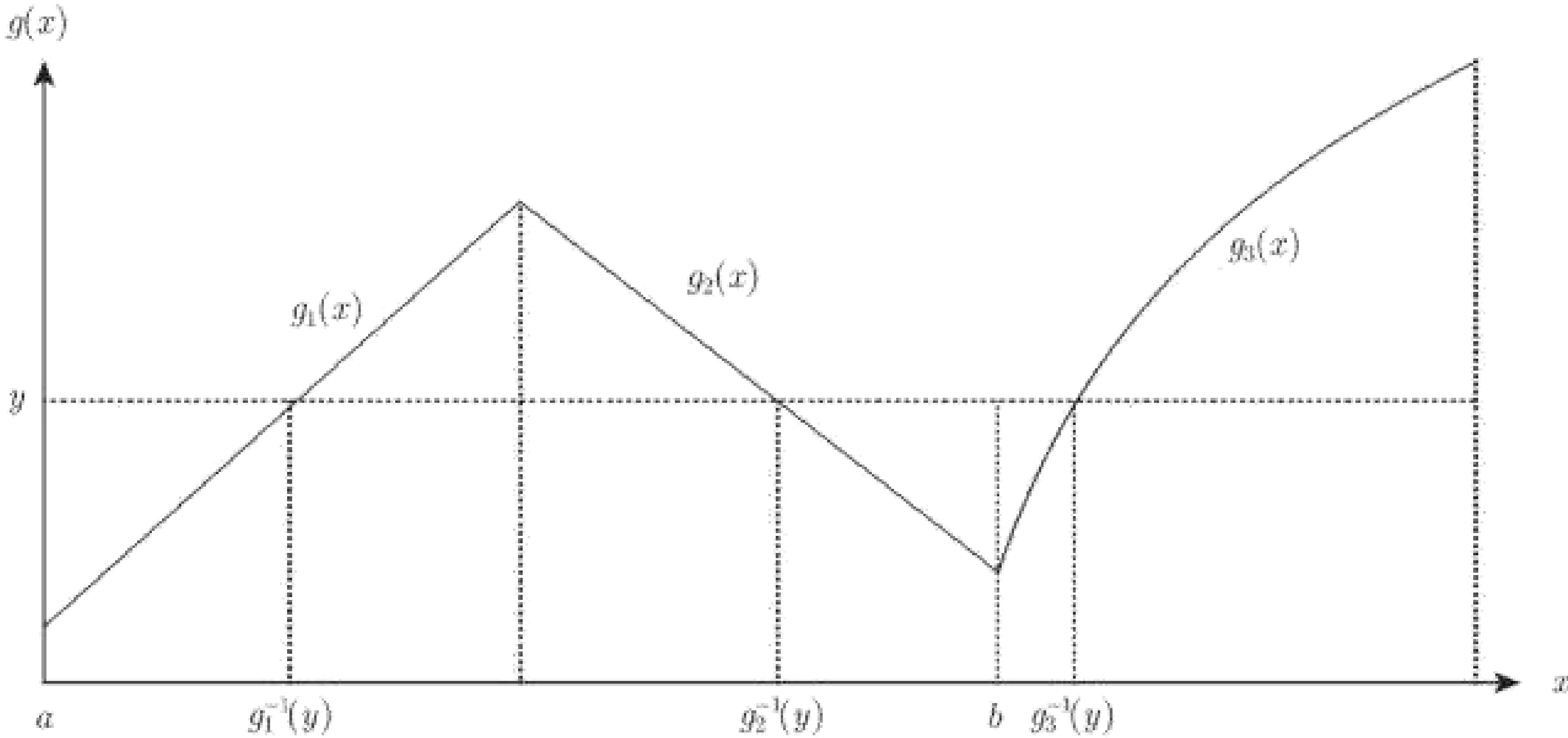


图 3.14：定理 3.13 在 $k=3$ 个子区间的情形

如何理解定理 3.13 (Cont.)

- 对于给定的 y 值，有

$$
\begin{aligned}
P(Y \le y) &= P[g(X) \le y] \\
&= P\left[a \le X \le g_1^{-1}(y)\right] + P\left[g_2^{-1}(y) \le X \le b\right] + P\left[b \le X \le g_3^{-1}(y)\right] \\
&= F_X\left[g_1^{-1}(y)\right] - F_X(a) + F_X(b) - F_X\left[g_2^{-1}(y)\right] + F_X\left[g_3^{-1}(y)\right] - F_X(b)
\end{aligned}
$$

- 对上式求导，可得

$$
\begin{aligned}
F_Y(y) &= f_X\left[g_1^{-1}(y)\right]\frac{1}{g_1'(y)} + f_X\left[g_2^{-1}(y)\right]\frac{1}{-g_2'(y)} + f_X\left[g_3^{-1}(y)\right]\frac{1}{g_3'(y)} \\
&= \sum_{i=1}^{3} f_X\left[g_i^{-1}(y)\right]\frac{1}{|g_i'(y)|}
\end{aligned}
$$

目　录

定理 3.14

[期望 (Expectation)]：假设随机变量 X 的 PMF/PDF 为 $f_X(x)$，则可测函数 $g(X)$ 的期望

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dF_X(x)$$

$$= \begin{cases} \sum_{x \in \Omega_X} g(x) f_X(x), & X \text{ 为离散随机变量} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx, & X \text{ 为连续随机变量} \end{cases}$$

其中假设式中的求和或积分存在。

如何解释期望 $E[g(X)]$？

- 假设我们通过大量独立重复实验生成了 X 的大量实现值，并计算这些重复实验中 $g(X)$ 的所有实现值的平均值：

$$E[g(X)] = \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} g(x_j)。$$

- 或者，假设 X 取 m 个不同的值 $\{x_1, x_2, \ldots, x_m\}$。那么在大量 N 次重复实验中，每个具体值 x_j 被重复 N_j 次，因此：

$$E[g(X)] = \lim_{N\to\infty} \sum_{j=1}^{m} \frac{N_j}{N} g(x_j)。$$

如何解释期望 $E[g(X)]$？ (Cont.)

- 积分 $\int_{-\infty}^{\infty} g(x)dF_X(x)$ 是黎曼-斯蒂尔杰斯积分 (Riemann-Stieltjes integral)。对于离散情形，求和是对 X 的 Ω_X 支撑上所有可能取值加总。
- 若 $E|g(X)| = \infty$，则称 $E[g(X)]$ 不存在。换言之，为避免收敛问题，定义 3.14 要求:
  - ✓ 离散随机变量 X 满足 $\sum_{x\in\Omega_X} |g(x)|f_X(x) < \infty$，
  - ✓ 连续随机变量满足 $\int_{-\infty}^{\infty} |g(x)|f_X(x)dx < \infty$。

如何解释期望 $E[g(X)]$？ (Cont.)

- 存在另一种计算期望 $E[g(X)]$ 的方法。由于 $Y=g(X)$ 是随机变量并且具有 PMF/PDF $f_Y(y)$，则

$$E[g(X)] = E(Y) = \begin{cases} \sum_{y\in\Omega_X} y f_Y(y), & Y \text{ 为离散随机变量} \\ \int_{-\infty}^{\infty} y f_Y(y)dy, & Y \text{ 为连续随机变量} \end{cases}$$

如何解释期望 $E[g(X)]$？ (Cont.)

- 由于期望 $E(\cdot)$ 是一个积分或求和运算，故为线性算符。因此，对任意两个可测函数 $g_1(\cdot)$ 和 $g_2(\cdot)$，以及 a 和 b 两个常数，有

$$E[ag_1(X)+bg_2(X)]=aE[g_1(X)]+bE[g_2(X)]$$

- 假设 $g(\cdot)$ 为凹函数。凹函数的一个例子是 $g(x)=\ln(x)$。
- **詹森不等式 (Jensen's inequality)**：对期望 $E[g(X)]$ 存在的任意凹函数与概率分布，有

$$E[g(X)] \le g[E(X)]$$

詹森不等式在行为经济学中的应用

- 假设 $g(\cdot)$ 为经济主体的效用函数，则风险厌恶型主体的效用函数 $g(\cdot)$ 是凹函数。此时，詹森不等式有很好的经济含义：不确定条件下的期望效用小于或等于确定收入为均值 $E(X)$ 时所带来的效用。换言之，在不确定条件下，经济主体偏好确定收入 $E(X)$。

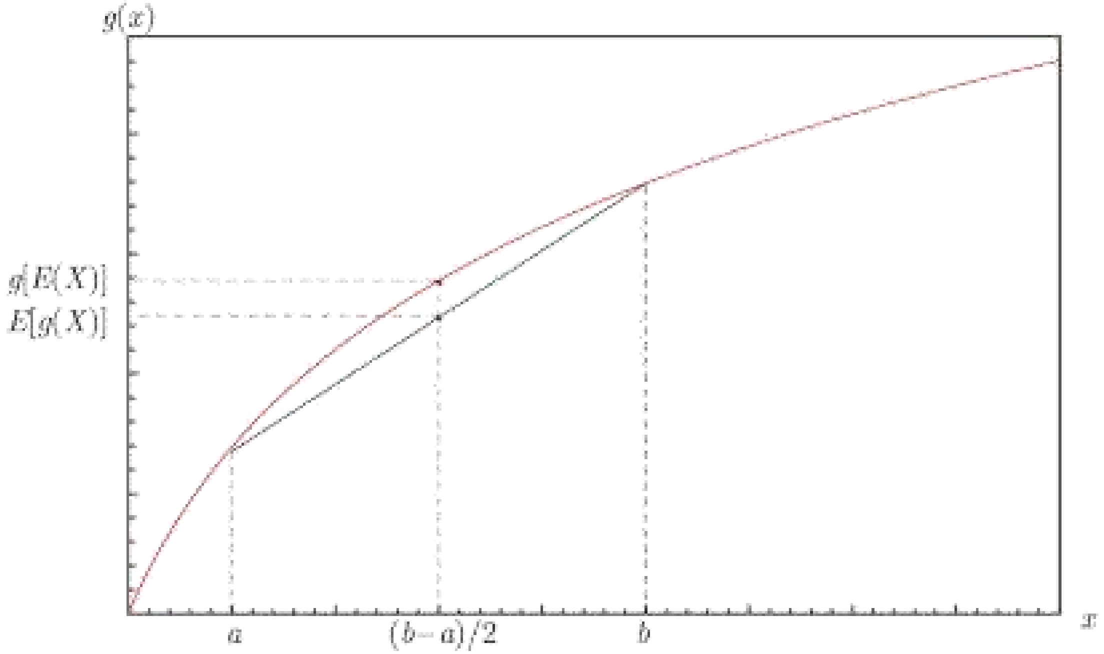


图 3.15：詹森不等式的经济学解释

詹森不等式在金融学中的应用

- 假设 $g(\cdot)$ 为金融衍生产品的支付函数，则 $g(\cdot)$ 一般为凸函数。
- 比如，Black & Scholes (1973) 的欧式看涨期权是一种在到期日购买标的资产（如，股票）的权利。
  - ✓ 购买价 K，即执行价，是预设的。投资者可以在到期日以价格 K 购买资产或放弃购买资产的权利。
  - ✓ 欧式看涨期权的支付函数为

$$g(X) = \max(X - K, 0)$$

  - ✓ 其中，X 表示到期日的资产价格。该支付函数为凸函数，在 $X = K$ 处存在转折点。负凸函数是凹函数。

目录

定义 3.15

[均值 (Mean)]：随机变量 X 的均值定义为

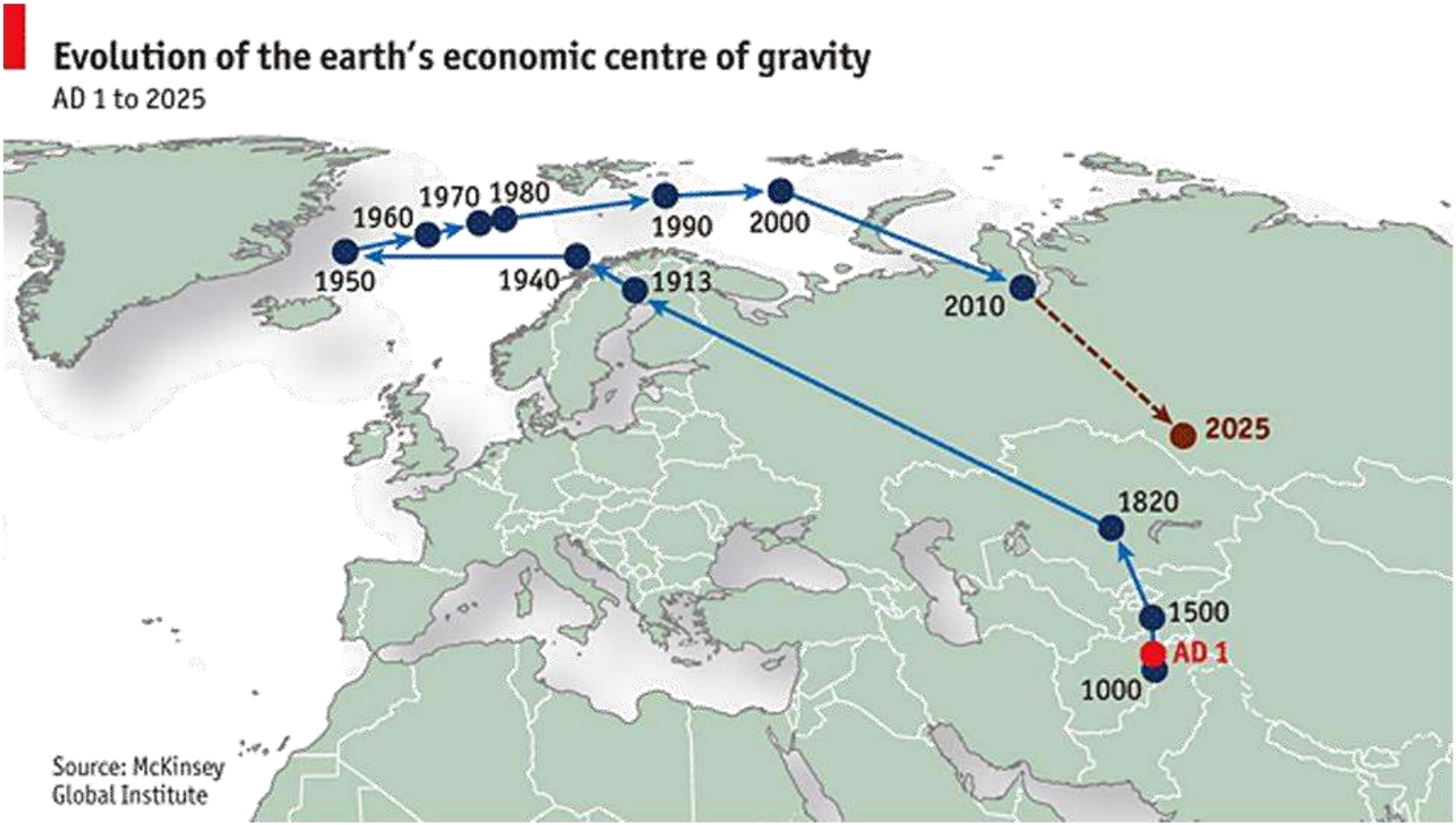
$$\mu_X = E(X) = \begin{cases} \sum_{x\in\Omega_X} x f_X(x), & X \text{ 为离散随机变量} \\ \int_{-\infty}^{\infty} x f_X(x)dx, & X \text{ 为连续随机变量} \end{cases}$$

其中，求和符号表示对离散随机变量 X 的支撑 Ω_X 上的所有可能取值求和。

均值

- 均值 μ_X 是 X 的期望，又称为 X 的一阶矩，可视为一个**“位置”**参数。
- 在**大量独立重复试验**中，$\mu_X = E(X)$ 可视为随机变量 X 实现值的平均的极限。从这个角度来看，均值度量了 X 分布的中心位置。
- 假设 X 代表一定时期内某资产收益率并且该资产收益率的分布不随时间改变，则 μ_X 可解释为该资产的长期平均收益率。

均值概念的应用：世界经济引力重心



Source: The Economist (2012)

均值概念的应用：世界经济引力重心 (Cont.)

- 世界经济引力重心与世界经济活动集聚的概念不能混为一谈。假设世界上只有中国和美国两个地方，两国拥有以 GDP 为衡量方式的等量经济活动，则全球经济引力重心恰好落在中美两国的中间，而且重心上的经济活动正好为零。
- 在这个例子中，有两个集聚：一个在中国，一个在美国，但只有一个重心，恰好落在两国中间。
- 世界经济引力重心并非人们寻求经济财富的地点，但它反映了一个重要的趋势——随着中国比美国发展得越来越快，经济活动大幅东移。关于人们寻求财富的地点，适用的概念是经济集聚而非经济引力重心。

定理 3.14

假设 $E(X^2)$ 存在，则

$$\mu_X = \underset{a}{\operatorname{argmin}} E(X-a)^2$$

证明： 上述最小化问题的一阶条件为

$$\left.\frac{dE(X-a)^2}{da}\right|_{a=a^*} = \left.\frac{d}{da}\int_{-\infty}^{\infty}(x-a)^2 dF_X(x)\right|_{a=a^*} = 0$$

交换求导和积分的次序可得

$$\left.-2\int_{-\infty}^{\infty}(x-a)dF_X(x)\right|_{a=a^*} = -2\int_{-\infty}^{\infty}(x-a^*)dF_X(x) = 0$$

则有

$$a^* = \frac{\int_{-\infty}^{\infty} x dF_X(x)}{\int_{-\infty}^{\infty} dF_X(x)} = \mu_X$$

证毕。

◆ **问题 3.16**

$X = \mu_X$ 发生的概率是否最大？换言之，概率 $P(X = \mu_X)$ 是否最大?

- 答案是否定的。
  - ✓ 例如，假设 X 以相同概率取值 1 和 0，即 $P(X = 1) = P(X = 0) = 1/2$，则 $\mu_X = E(X) = 1/2$。
  - ✓ 因此，$P(X = \mu_X) = 0$，即 $X = \mu_X$ 发生的概率最小。

定义 3.16

[方差 (Variance) 与标准差 (Standard Deviation)]：随机变量 X 的**方差**定义为

$$\sigma_X^2 = E(X-\mu_X)^2$$

$$= \begin{cases} \sum\limits_{x\in\Omega_X} (x-\mu_X)^2 f_X(x), & X\text{为离散随机变量} \\ \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx, & X\ \text{为连续随机变量} \end{cases}$$

其中 Ω_X 是 X 的支撑。$\sigma_X = \sqrt{\sigma_X^2}$ 称为 X 的**标准差。**

方差

- 方差 σ_X^2 作为 X 分布的**尺度参数**，度量了概率分布在其均值周围的分散程度。在经济学中，σ_X^2 通常解释为对**不确定性的测度。**
- 在 σ_X^2 的极端情况下，$X = \mu_X$ 的概率为 1，此时 X 不存在任何变动。简单起见，考察 X 为离散随机变量的情形。因为

$$\sigma_X^2 = \sum_{x \in \Omega_X} (X - \mu_X)^2 f_X(x) = 0$$

当且仅当

$$(x - \mu_X)^2 f_X(x) = 0, \text{对所有 } x \in \Omega_X$$

- 故有 $x = \mu_X$ 且 $f_X(\mu_X) = 1$。
- 即当 $\sigma_X^2 = 0$ 时，X 只有一个可能取值，即 μ_X。这是所谓退化分布 (degenerate distribution) 的一个例子。

变异系数：$V = \frac{\sigma_X}{\mu_X}$

- 在概率论与统计学中，**变异系数** (coefficient of variation, CV) 又称**相对标准差** (relative standard deviation, RSD)，是概率分布或频率分布离散程度的一种标准测度。
- 变异系数体现了总体均值的变异程度。因为数据标准差必须建立在数据均值的基础上，所以变异系数非常有用。变异系数的实际值不依赖于度量单位，故变异系数是一个无量纲量。对单位不同或均值差异很大的数据集进行比较时，应使用变异系数而非标准差。
- **变异系数测度的主要缺陷：**当均值接近于零时，变异系数趋于无穷大，因此该系数对均值的微小变化很敏感。

定理 3.15

$$\sigma_X^2 = E(X^2) - \mu_X^2$$

证明：

- 根据公式$(a-b)^2 = a^2 - 2ab + b^2$，可得

$$\begin{aligned}\sigma_X^2 &= E[(X-\mu_X)^2] \\ &= E(X^2) - E(2\mu_X X) + E(\mu_X^2) \\ &= E(X^2) - 2\mu_X E(X) + \mu_X^2 \\ &= E(X^2) - \mu_X^2\end{aligned}$$

证毕。

- **注：**一般将 σ_X^2 称为**二阶中心矩** (second central moment)，而 $E(X^2)$ 则称为 X 的**二阶矩** (second moment)。

定理 3.16

若 $Y=a+bX$，则:

(1) $\mu_Y=a+b\mu_X$;

(2) $\sigma_Y^2=b^2\sigma_X^2$。

- 参数 a 和 b 可分别解释为**位置**和**尺度**参数。
  - ✓ 位置和尺度参数都将影响 Y 的均值，
  - ✓ Y 的方差仅受尺度参数 b 的影响，而不受位置参数 a 的影响。

线性变换 $Y = a + bX$ 的均值和方差在经济学中的应用

- 假设有两种资产：一种无风险资产的总收益等于 1，另一种风险资产的随机收益为 X。
- 假设 a 和 b 分别表示投资者对无风险资产和风险资产的投资比重（称为投资组合权重），则 Y 是该投资组合的总收益。
- 计算该投资组合的预期收益 μ_Y 和风险（由 σ_Y^2 测度）十分重要。
- 这里，预期收益 μ_Y 包括风险资产的预期收益和无风险资产的预期收益，然而，投资组合 Y 的风险，由 $\sigma_Y^2 = b^2\sigma_X^2$ 度量，仅依赖于风险资产 X，与无风险资产无关。

定义 3.17

[k 阶矩（k-th Moment）和 k 阶中心矩（k-th Central Moment）]：

- 随机变量 X 的 k 阶矩定义为

$$E(X^k)=\begin{cases}\sum_{x\in\Omega_X} x^k f_X(x), & X\text{ 为离散随机变量}\\ \int_{-\infty}^{\infty} x^k f_X(x)dx, & X\text{ 为连续随机变量}\end{cases}$$

其中 Ω_X 为 X 的支撑。

- 类似地，随机变量 X 的 k 阶中心矩定义为

$$E(X-\mu_X)^k=\begin{cases}\sum_{x\in\Omega_X} (x-\mu_X)^k f_X(x), & X\text{ 为离散随机变量}\\ \int_{-\infty}^{\infty} (x-\mu_X)^k f_X(x)dx, & X\text{ 为连续随机变量}\end{cases}$$

例 3.29：[圣彼得堡悖论 (St. Petersburg Paradox)]

- 在抛掷一枚质地均匀硬币的一系列独立重复试验中，如果一名参与者抛掷第 x 次才首次获得正面朝上，则其将获得 2^x 元奖金。显然，首次正面朝上出现在第 x 次抛掷的概率为 $f_X(x)=\left(\frac{1}{2}\right)^x, x=1,2,\cdots$ 。
- 该游戏的期望收益为

$$E\left(2^X\right)=\sum_{x=1}^{\infty} 2^x\left(\frac{1}{2}\right)^x$$
$$=\infty$$

- 然而，若要求一个厌恶风险的参与者支付 1000 元，他通常不会选择玩这个游戏。为什么呢？因为风险厌恶型的参与者不仅考虑游戏的预期收益，而且也考虑游戏的风险。

例 3.30：[投资组合的选择]

投资者如何选择最优的资产投资组合?

- 令 μ 和 σ^2 分别表示某投资组合在持有期内的收益率的均值和方差。
- 假设某投资者为风险厌恶型，即偏好高收益低风险。具体而言，其效用函数 $U(\mu,\sigma^2)$ 是 μ 和 σ^2 的函数，满足 $\frac{\partial}{\partial\mu}U(\mu,\sigma^2)>0$ 和 $\frac{\partial}{\partial\sigma^2}U(\mu,\sigma^2)<0$。
  - ✓ $\frac{\partial}{\partial\mu}U(\mu,\sigma^2)>0$ 意味着预期收益越高，效用越高；
  - ✓ $\frac{\partial}{\partial\sigma^2}U(\mu,\sigma^2)<0$ 意味着风险越大，效用越低。

例 3.30 (Cont.)：

- 一个例子是

$$U(\mu, \sigma^2) = a\mu - \frac{b}{2}\sigma^2$$

  - ✓ 其中，$a > 0$ 和 $b > 0$ 为偏好参数。
  - ✓ 投资者将在预算约束条件下最大化效用函数 $U(\mu, \sigma^2)$。

例 3.30 (Cont.)：

- 现在假设金融市场上有一种随机收益率 X 的风险资产，以及一种收益率为常数 r 的无风险资产。
- 进一步假设投资者共有 I 元，其中 z 元投资于风险资产，其余 $I-z$ 元投资于无风险资产。
- 则该投资组合的
  - ✓ **总收益率**：$Y=zX+(I-z)r$
  - ✓ **预期收益率**：$\mu_Y=z\mu_X+(I-z)r$
  - ✓ **方差**：$\sigma_Y^2=z^2\sigma_X^2$
  - ✓ **效用函数**：$U\left(\mu_Y,\sigma_Y^2\right)=a[z\mu_X+(I-z)r]-\frac{b}{2}z^2\sigma_X^2$

例 3.30 (Cont.)：

- 通过最大化该投资者的效用函数，可得最优组合权重 z^*，即，投资者将选择最大化 $U(\mu_Y,\sigma_Y^2)$ 的组合权重 z 值。相应的一阶条件为

$$a(\mu_X - r) - bz^*\sigma_X^2 = 0$$

- 由此可得，投资者在风险资产上的最优投资为

$$z^* = \frac{a(\mu_X - r)}{b\sigma_X^2}$$

例 3.30 (Cont.)：

- 现在，考察以下两类特殊情形。
  - ✓ **情形 (1)：$\boldsymbol{b}=0$。该投资者为风险中性 (risk-neutral)**。若 $\mu_X>r$，投资者将只投资风险资产，即 $z^*=I$ 或 $z^*=\infty$ (当允许借贷时)。
  - ✓ **情形 (2)：$\boldsymbol{b}=\infty$。投资者极度厌恶风险 (risk-averse)**。即使 $\mu_X>r$，投资者也会选择 $z^*=0$，即仅投资无风险资产。
- 一般情况下，z^* 是一个内点解，这意味着投资者在风险资产和无风险资产之间对投资额进行分配。
- 两种资产的投资相对比例取决于投资者风险偏好参数 a 和 b 的相对值以及风险资产的预期收益和无风险资产收益之间的差异。

一、二阶矩在经济学中的应用：夏普比

- 实际应用中，投资者常采用的一种称为**夏普比（Sharpe ratio)**投资组合决策准则，其定义为

$$\text{Sharpe ratio} = \frac{\mu_Y}{\sigma_Y}$$

其中，μ_Y 是投资组合的预期超额收益率，σ_Y 为超额收益率的标准差。

- 这是变异系数倒数在金融市场上的应用。直观上，夏普比可解释为单位风险的预期收益。厌恶风险的投资者偏好高夏普比的资产组合。这与上述例题中最优投资组合的思想一致。

三、四阶矩在经济学中的应用

- 除了前两阶矩外，高阶矩尤其三、四阶矩在经济学也越来越受到重视。
- 例如，金融学中，证券的系统风险通常是由充分分散化的投资组合的方差贡献度来测度的。然而，大量证据表明，仅采用均值和方差常常无法充分描述证券收益率的分布特征。这促使经济学家关注三阶矩即偏度。

三阶中心矩 & 偏度

- 对随机变量 X，三阶中心矩 $E(X-\mu_X)^3$ 度量其概率分布的“**偏度**”或“**非对称性**”。正偏度指的是分布函数的右尾比左尾长，负偏度则情况相反。现有文献将偏度（skewness）定义为**标准化的三阶中心矩**：

$$S_X=\frac{E(X-\mu_X)^3}{\sigma_X^3}$$

- 通过除以 σ_X^3 进行标准化处理，使偏度不受分布尺度的影响。
  - ✓ S_X 为正意味着概率密度函数的右侧尾部比左侧更长；
  - ✓ S_X 为负则情况相反。
- 在金融学，偏度可用以测度金融危机，因为当大的损失比大的收益发生的概率更大时，$E(X-\mu_X)^3$ 为负且绝对值大。

例 3.31 [存在偏度的资本资产定价模型 (Capital Asset Pricing Model, CAPM)]：

Kraus & Liezenkerger (1976) 使用了定义在均值、标准差以及偏度的三次方根基础上的投资者效用函数。

四阶中心矩 & 峰度

- 四阶中心矩 $E(X-\mu_X)^4$ 测度**分布的厚尾程度**。厚尾意味着 X 更容易出现极端数值。在金融市场中，这可能导致金融破产或违约的发生。
- 峰度 (kurtosis) 定义为**标准化的四阶中心矩**：

$$K_X=\frac{E(X-\mu_X)^4}{\sigma_X^4}$$

- 标准化使得峰度不受分布尺度的影响。分布的峰度度量了分布的高突或平坦程度。换言之，它揭示了概率分布在中心的集中程度，尤其是分布的厚尾程度。

四阶中心矩 & 峰度 (Cont.)

- 峰态类型
  - ✓ $K_X < 3$ 的概率分布称为**低峰态**（platykurtic）（平坦或细尾），
  - ✓ $K_X > 3$ 的概率分布称为**尖峰**（leptokurtic）（细长或厚尾）。
  - ✓ 若概率分布有 $K_X = 3$，则称为**常峰态**（mesokurtic）。
- 为何将 $K_X = 3$ 作为划分基准?
  - ✓ 这主要是因为著名的正态分布 $N(\mu, \sigma^2)$ 的峰度 $K_X = 3$。

例 3.32：

假设正态随机变量 X 的均值和方差分别为 μ 和 σ^2，则其 PDF 为

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

可以计算出 $\mu_X = \mu, \sigma_X^2 = \sigma^2, S_X = 0$ 以及 $K_X = 3$。

目　录

定义 3.18

[α-分位数 (α-quantile)]：假设 X 的 CDF 为 $F_X(x)$。给定 $\alpha \in (0,1)$，则分布 $F_X(x)$ 的 α-分位数，记为 $Q(\alpha)$，满足方程

$$P[X \leq Q(\alpha)] = \alpha$$

或者等价地

$$F_X[Q(\alpha)] = \alpha$$

当 $F_X(x)$ 严格递增时，有

$$Q(\alpha) = F_X^{-1}(\alpha)$$

其中，$F_X^{-1}(\alpha)$ 为 $F_X(x)$ 的反函数。换言之，分位数 $Q(\alpha)$ 是 CDF $F_X(x)$ 的反函数。若 $F_X(x)$ 非严格增，其 α-分位数定义为

$$Q(\alpha) = \inf\{x \in \mathbb{R}: \alpha \leq F_X(x)\}$$

α-分位数的图形表示

- 假设 $f_X(x) = F'_X(x)$ 几乎处处存在，则

$$\int_{-\infty}^{Q(\alpha)} f_X(x)dx = \alpha$$

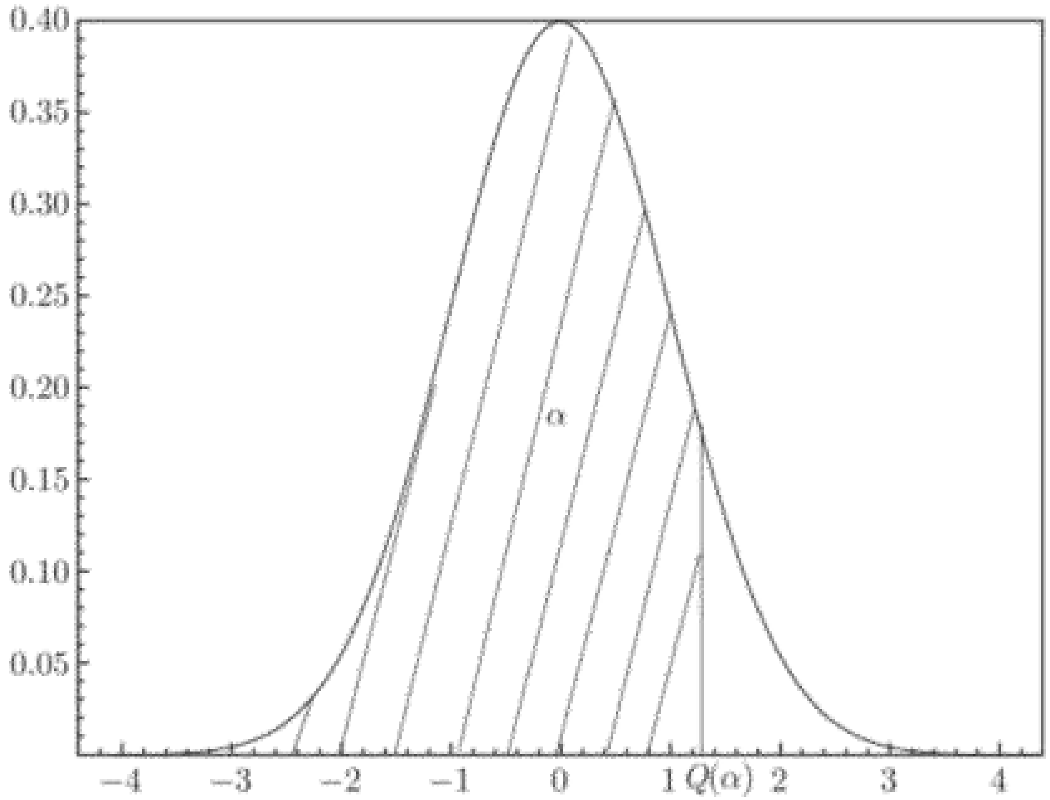


图 3.20：α-分位数的另一种图形表述

分位数的特殊情况：中位数

- 当 $\alpha = 1/2, m = Q\left(\frac{1}{2}\right)$ 称为 X 的分布 $F_X(x)$ 的中位数。当 PDF $f_X(x)$ 存在时，有

$$\int_{-\infty}^{m} f_X(x)dx = \frac{1}{2}$$

分位数的应用：中产阶级的度量

- 实际上，收入中位数常用于刻画一个社会的中产阶级。
- 美国皮尤研究中心 (Pew Research Center) 将美国中产阶级定义为三口之家，其家庭收入在美国家庭收入中位数的 67% 到两倍之间。

◆ 问题 3.18

如何解释中位数呢？均值和中位数有何关系呢？

- 若 $F_X(m)=1/2$ 有解，则中位数 $m=Q\left(\frac{1}{2}\right)$ 是将分布 $F_X(x)$ 划分为两等份的断点或临界值。
- 然而，尽管均值 μ_X 可很好地度量对称分布或者近似对称分布的中心位置，但它在度量高偏度分布的中心位置时，会产生一定的误导。
- 相反，中位数 $m=Q\left(\frac{1}{2}\right)$ 受奇异值的影响较小，从而能够更稳健地度量概率分布的中心位置。
- 比如，若一个社会的收入分布是高右偏度，则收入中位数可能比收入均值更适合度量该社会的“平均”收入。

定理 3.17

中位数 m 是最小化平均绝对误差的最优解，即

$$m = \arg\min_{a} E|X - a|$$

分位数的应用：

风险价值（VaR）与量化金融风险管理（Financial Risk Management）

- 在金融文献中，当 X 为某投资组合在给定持有期的收益率时，对应于水平 α（如 $\alpha = 0.01$）的 $-Q(\alpha)$ 通常称为风险价值。
- 直观上，$-Q(\alpha)$ 是实际损失以概率 α 超过的临界值。
- 风险价值被国际清算银行和大多数商业银行等金融机构用以设定银行资本水平。

例 3.33：[风险价值与 J. P. 摩根风险度量 (J. P. Morgan Risk Metrics)]

- 投资组合在一定时期内，在水平 α 上的风险价值 $V_t(\alpha)$ 定义为

$$P[X_t < -V_t(\alpha) \mid I_{t-1}] = \alpha$$

  - ✓ 其中 X_t 是投资组合在持有期间 t 的收益率，
  - ✓ 而 I_{t-1} 表示第 $t-1$ 期可获取的信息。

例 3.33 (Cont.)：

- 风险价值 $V_t(\alpha)$ 是实际损失以概率 α 超越的临界值。
- 它是决定资本充足水平的基础，以将极端损失事件出现的概率控制在 α 以内。当资本准备未达到该临界值时，则引起破产或违约的恶性事件出现的概率将超过 α。显然，$V_t(\alpha)$ 是水平 α 上的负的条件分位数。
- 风险价值概念被 Adrian & Brunnermeier (2016) 拓展为 CoVaR 概念，用于测度金融市场的系统风险 (systemic risk)。

目　录

定义 3.19

[矩生成函数 (Moment Generating Function，MGF)]：随机变量 X 的 MGF 定义为

$$M_X(t) = E\left(e^{tX}\right)$$
$$= \begin{cases} \sum_{x\in\Omega_X} e^{tx} f_X(x), & X \text{ 为离散随机变量} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, & X \text{ 为连续随机变量} \end{cases}$$

如果上述期望对于 t 在 0 的某个邻域内存在，则称 $M_X(t)$ 对于 t 在 0 的某个小邻域内是存在的（即存在某个常数 $\epsilon > 0$，使得对任意 $t \in (-\epsilon, \epsilon)$，$E(e^{tX})$ 总是存在的)。若在 0 的任意小邻域内，上述期望都不存在，则称 $M_X(t)$ 对 X 的分布不存在。

◆ 问题 3.19

MGF $M_X(t)$ 存在是否意味着 X 的各阶矩均存在?

答案是肯定的。

- 若 $M_X(t)$ 对于所有的 $t \in (-\epsilon, \epsilon)$ 都存在，这意味着 $M_X(t)$ 在 $t = 0$ 处的所有阶导数都存在。为什么呢？
- 根据麦克劳林级数展开，对所有实数 $t \in (-\epsilon, \epsilon)$，有

$$e^{tx} = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}$$

- 则 $M_X(t)$ 的麦克劳林级数展开为

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E\left(X^k\right)$$

CONTINUE

- 若存在某个 k 有 $E(X^k)=\infty$，则对所有 $t>0$，$M_X(t)$ 均不存在。
- 因此，若 $M_X(t)$ 对所有 $t\in(-\epsilon,\epsilon)$ 都存在，那么 X 的各阶矩都必须存在。
- 上述级数展开说明，$M_X(t)$ 在 0 的邻域内包含所有矩信息，并且提供了一种通过 MGF 获得各阶矩的方法。

定理 3.18

若 MGF $M_X(t)$ 对于 t 在 0 的某个邻域内存在，则 $M_X(0)=1$。

证明：

根据定义 $M_X(t)=\int_{-\infty}^{\infty} e^{tx} dF_X(x)$，有

$$M_X(0)=\int_{-\infty}^{\infty} dF_X(x)=1$$

证毕。

定理 3.19

若 MGF $M_X(t)$ 对于 t 在 0 的某个邻域内存在，则对所有正整数 $k=1,2,\cdots$，有 $M_X^k(0)=E(X^k)$

证明： 对任意给定整数 $k>0$ 以及所有 $t\in(-\epsilon,\epsilon)$，有

$$
\begin{aligned}
M_X^{(k)}(t) &= \frac{d^k}{dt^k}\int_{-\infty}^{\infty} e^{tx}\,dF_X(x) \\
&= \int_{-\infty}^{\infty} \frac{d^k}{dt^k}(e^{tx})\,dF_X(x) \\
&= \int_{-\infty}^{\infty} x^k e^{tx}\,dF_X(x)
\end{aligned}
$$

令 $t=0$，得

$$
M_X^{(k)}(0)=\int_{-\infty}^{\infty} x^k\,dF_X(x)=E\left(X^k\right)
$$

证毕。

- 定理 3.19 表明，对任意 $t \in (-\epsilon, \epsilon)$，若 $M_X(t)$ 存在，则可对 $M_X(t)$ 在原点处求导获得 X 的各阶矩。由于 $M_X(t)$ 可用以生成各阶矩，该函数因此被称为**矩生成函数。**
- 对 $k = 1, 2,$ 有

$$M_X^{(1)}(0) = \mu_X$$

$$M_X^{(2)}(0) = E\left(X^2\right) = \sigma_X^2 + \mu_X^2$$

定理 3.20

若 $Y=a+bX$，其中 a 和 b 为两个常数，并且对在 0 的某个小邻域内的所有 t，X 的 MGF $M_X(t)$ 存在。则对在 0 的某个小邻域内的所有 t，Y 的 MGF 存在且为

$$M_Y(t)=e^{at}M_X(bt)$$

例 3.35：

假设随机变量 X 的均值为 μ，方差为 σ^2，并且 MGF $M_X(t)$ 对在 0 的某个小邻域内的所有 t 都存在。定义标准化随机变量

$$Y=\frac{X-\mu}{\sigma}$$

这对应于变换 $Y=a+bX$ 中 $a=-\frac{\mu}{\sigma},b=\frac{1}{\sigma}$ 的情形。则 MGF

$$M_Y(t)=e^{-\frac{\mu}{\sigma}t}M_X\left(\frac{t}{\sigma}\right)$$

◆ **问题 3.20**

如何用 MGF 刻画 X 的概率分布?

定理 3.21

[MGF的唯一性]：假设两个随机变量 X 和 Y 的 MGF $M_X(t)$ 和 $M_Y(t)$ 在 0 的某个小邻域 $N_\epsilon(0)=\{t\in\mathbb{R}: -\epsilon<t<\epsilon\}$ 存在。当且仅当对任意 $z\in\mathbb{R}$，$F_X(z)=F_Y(z)$（即同分布），则对所有 $t\in N_\epsilon(0)$，X 和 Y 都有相同的 MGF $M_X(t)$ 和 $M_Y(t)$。

启发式证明：

- 该定理的证明依赖于拉普拉斯变换（Laplace transform）（参见 Feller, 1971）。根据 MGF $M_X(t)$ 的定义，对连续随机变量，有

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

此为拉普拉斯变换。

- 拉普拉斯变换的关键特征在于它具有唯一性。若存在正数 ϵ，使得对任意 $|t| < \epsilon$，$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$ 均存在，则给定 $M_X(t)$，仅有一个函数 $f_X(x)$ 满足上述拉普拉斯变换。
- 基于这一事实，可知该定理的合理性。然而，该定理的正式证明具有相当的技术性，且并不能提供更多的启示。

例 3.36：

假设离散随机变量 X 有 MGF

$$M_X(t) = \frac{1}{2} + \frac{1}{4}e^{-t} + \frac{1}{4}e^{t}, -\infty < t < \infty$$

求 X 的概率分布。

解：

- 根据 MGF 的定义，有

$$M_X(t) = E\left(e^{tX}\right)$$

$$= \sum_x e^{tx} f_X(x)$$

$$= \frac{1}{2}e^{0\cdot t} + \frac{1}{4}e^{(-1)\cdot t} + \frac{1}{4}e^{1\cdot t}$$

CONTINUE

解 (Cont.):

- 因此猜测 X 的 PMF 为

$$f_X(x)=\begin{cases}\frac{1}{4}, & x=-1\\ \frac{1}{2}, & x=0\\ \frac{1}{4}, & x=1\end{cases}$$

- 容易验证猜测分布 $f_X(x)$ 的 MGF 为

$$M_X(t)=\frac{1}{2}+\frac{1}{4}e^{-t}+\frac{1}{4}e^{t},-\infty<t<\infty$$

- 现在论证猜测分布 $f_X(x)$ 是与给定的 $M_X(t)$ 相对应的唯一分布。若不然，假设存在另一个分布 $h_X(x)$ 可生成同样的 MGF $M_X(t)$。则由 MGF 的唯一性定理，这个分布必然与猜测分布 $f_X(x)$ 相同。

例 3.37：

若随机变量 X 均值为 0，方差为 2，且 MGF 为

$$M_X(t) = a\left(1 + be^{-2t} + e^{-t} + e^t + ce^{2t}\right), -\infty < t < \infty$$

(1) 求常数 a, b, c 的值；

(2) X 的 PMF 是什么？证明其确实是 X 的 PMF。

解：(1)

- 根据定理 3.18 和 3.19，有

$$\begin{aligned} M(0) &= 1 \\ M'(0) &= \mu = 0 \\ M''(0) &= E\left(X^2\right) \\ &= \sigma^2 + \mu^2 \\ &= 2 + 0^2 \\ &= 2 \end{aligned}$$

CONTINUE

解 (Cont.):

- 则

$$M_X(0) = a(3 + b + c) = 1$$
$$M_X'(0) = a(-2b - 1 + 1 + 2c) = 0$$
$$M_X''(0) = a(4b + 1 + 1 + 4c) = 2$$

- 求解上述联立方程，得 $a = 1/5$， $b = c = 1$。

(2)

- 用类似例 3.36 的推理方法，可猜测对应给定 $M_X(t)$ 的概率分布为 PMF $f_X(x) = 1/5$，$x = -2, -1, 0, 1, 2$；对于其他 x 值，$f_X(x) = 0$。
- 根据 MGF 的唯一性定理，这是对应给定 $M_X(t)$ 的唯一的概率分布。

定理 3.22

[MGF 的收敛性]：假设 $\{X_n, n = 1, 2, \cdots\}$ 为随机变量序列，每个随机变量 X_n 有 MGF $M_n(t)$ 和 CDF $F_n(x)$。进一步假设，对 t 在 0 的某个小邻域内的任意取值，有

$$\lim_{n\to\infty} M_n(t) = M_X(t)$$

其中，$M_X(t)$ 是与随机变量 X 的 CDF F_X 相对应的 MGF。则对所有连续点 x（即在该点，$F_X(x)$ 是连续的），有

$$\lim_{n\to\infty} F_n(x) = F_X(x)$$

定理 3.22 的重要性

- MGF 的收敛性意味着依分布收敛。
  - ✓ 实际应用中，$F_n(x)$ 和 $M_n(t)$ 一般都未知，但若能证明二者分别收敛于已知极限函数 $F_X(x)$ 和 $M_X(t)$，则可用这些已知的极限函数分别近似未知的 $F_n(x)$ 和 $M_n(t)$。

◆ 问题

假设对所有正整数 k，有 $E(X^k) = E(Y^k)$，那么随机变量 X 和 Y 是否有相同分布?

例 3.40:

- 假设随机变量 X 和 Y 分别具有如下 PDF

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} e^{-\frac{1}{2}(\ln x)^2}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$f_Y(y) = \begin{cases} f_X(y)[1 + \sin(2\pi \ln y)], & y > 0 \\ 0, & y \leq 0 \end{cases}$$

- 其中 $f_X(x)$ 和 $f_Y(y)$ 分别为 X 和 Y 的 PDF。
- 证明对所有正整数 k，有 $E(X^k) = E(Y^k)$ 。

解： (1)

- 首先，根据例 3.26 可知，X 服从所谓的标准对数正态分布 (lognormal distribution)，即 $Z = \ln X \sim N(0,1)$。因此

$$E\left(X^k\right) = E\left[\left(e^Z\right)^k\right] = E\left(e^{kZ}\right) = M_Z(k) = e^{\frac{1}{2}k^2}$$

- 其中，用到了 $M_Z(t) = e^{\frac{1}{2}t^2}$ 为标准正态随机变量 (standard normal distribution) $Z \sim N(0,1)$ 的 MGF 这一事实。

解 (Cont.):

(2)

- 注意到

$$E\left(Y^k\right)=E\left(X^k\right)+\int_0^\infty x^k f_X(x)\sin(2\pi\ln x)dx$$

- 为证明 $E(X^k)=E(Y^k)$，只需证明

$$\int_0^\infty x^k f_X(x)\sin(2\pi\ln x)dx=0$$

- 定义变换 $v=\ln x-k$，则积分

$$\int_0^\infty x^k f_X(x)\sin(2\pi\ln x)dx=\int_{-\infty}^\infty \left(e^{v+k}\right)^k\frac{1}{\sqrt{2\pi}}e^{-(v+k)}e^{-\frac{(v+k)^2}{2}}\sin[2\pi(v+k)]e^{(v+k)}dv$$

$$=\int_{-\infty}^\infty \left(e^{v+k}\right)^k\frac{1}{\sqrt{2\pi}}e^{-\frac{(v+k)^2}{2}}\sin[2\pi(v+k)]dv$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-\frac{1}{2}(v^2-k^2)}\sin(2\pi v)dv=0$$

注意

- 例 3.39 说明，各阶矩的集合无法唯一确定一个概率分布。
- 但是，有一特殊情形，即当随机变量为有界支撑时（参见 Billingsley，1995），可采用各阶矩构成的集合唯一刻画概率分布。

定理 3.25

假设 $F_X(x)$ 和 $F_Y(y)$ 为两个存在有界支撑的 CDF。则对所有整数 $k=1,2,\cdots$，$E(X^k)=E(Y^k)$，当且仅当对任意 $z\in(-\infty,\infty)$，$F_X(z)=F_Y(z)$。

证明:

- 因为 X 和 Y 均为有界支撑，故存在一个充分大的常数 $M>0$，使得 $P(|X|\le M)=1$ 和 $P(|Y|\le M)=1$。

证明 (Cont.)：

- 又因 $E|X|^k \le M^k$ 和 $E|Y|^k \le M^k$，有

$$
\begin{aligned}
M_X(t) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} E\left(X^k\right) \\
&\le \sum_{k=0}^{\infty} \frac{t^k}{k!} E|X|^k \\
&\le \sum_{k=0}^{\infty} \frac{(tM)^k}{k!} \\
&= e^{tM} < \infty, \text{对任意 } t < \infty
\end{aligned}
$$

- 类似地，有

$$M_Y(t) \le e^{tM} < \infty$$

证明 (Cont.):

- 根据 MGF 的公式 $M_X(t)=\sum_{k=0}^{\infty}\frac{t^k}{k!}E\left(X^k\right)$ 可知，若对所有整数 $k>0$，有 $E(X^k)=E(Y^k)$，则对属于 0 的某邻域任意 t 都有 $M_X(t)=M_Y(t)$。由 MGF 的唯一性定理，对任意 $z\in(-\infty,\infty)$，必有 $F_X(z)=F_Y(z)$。
- 另外，若对所有 $z\in(-\infty,\infty)$ 有 $F_X(z)=F_Y(z)$ 以及 $P(|X|\le M)=1$，$P(|Y|\le M)=1$，则对于任意给定正整数 k，矩 $E(X^k)$ 和 $E(Y^k)$ 必然存在，且对所有 $k>0$ 有 $E(X^k)=E(Y^k)$。
- 因此，两个有界支撑的随机变量 X 和 Y，有如下性质：即对所有整数 $k>0$，$E(X^k)=E(Y^k)$，当且仅当对任意 $z\in(-\infty,\infty)$，$F_X(z)=F_Y(z)$ 成立。**证毕。**

目　录

定义 3.20

[特征函数 (Characteristic Function)]：假设随机变量 X 的CDF为 $F_X(x)$，则其特征函数定义为

$$\varphi_X(t)=E\left(e^{\mathbf{i}tX}\right)$$
$$=\int_{-\infty}^{\infty}e^{\mathbf{i}tx}dF_X(x)$$

其中 $\mathbf{i}=\sqrt{-1}$，并且

$$e^{\mathbf{i}tx}=\cos(tx)+\mathbf{i}\sin(tx)$$

傅里叶变换 (Fourier transform) & 逆傅里叶变换 (inverse Fourier transform)

- 根据上述定义，$\varphi_X(t)$ 是分布函数 $F_X(x)$ 的**傅里叶变换。**

$$\varphi_X(t) = E\left(e^{\mathrm{i}tX}\right) = \int_{-\infty}^{\infty} e^{\mathrm{i}tx} dF_X(x)$$

- 傅里叶变换的基本性质：$\varphi_X(t)$ 和 $F_X(x)$ 包含相同的关于概率分布的所有信息，并且一一对应。
- 特别地，若随机变量 X 的 PMF/PDF $f_X(x)$ 存在，则可通过**逆傅里叶变换**求得 PMF/PDF，即

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\mathrm{i}tx} \varphi_X(t) dt$$

- 对于不存在解析形式 PDF 的概率分布，该反演公式在求概率密度函数时非常有用。

定理 3.26

[特征函数的性质]

(1) 对任意概率分布，其特征函数 $\varphi_X(t)$ 总存在而且有界，即对 $-\infty < t < \infty$，有 $|\varphi_X(t)| \le 1$；

(2) $\varphi_X(0) = 1$；

(3) $\varphi_X(t)$ 在 $(-\infty, \infty)$ 上连续；

(4) $\varphi_X(-t) = \varphi_X(t)^*$，其中 $\varphi_X(t)^*$ 代表 $\varphi_X(t)$ 的复共轭 (complex conjugate)；

(5) 假设 $Y = a + bX$，其中 a 和 b 为任意实常数，则

$$\varphi_Y(t) = e^{\mathrm{i}at}\varphi_X(bt)$$

(6)若 MGF $M_X(t)$ 对在 0 的某邻域内的所有 t 都存在，则对所有 $t \in (-\infty, \infty)$，有 $\varphi_X(t) = M_X(\mathrm{i}t)$。

特征函数相比 MGF 的优势

- 特征函数 $\varphi_X(t)$ 对所有分布以及所有实数集上的 t 总存在，而 MGF $M_X(t)$ 对某些分布并不存在，即使是 t 在 0 的一个很小邻域内。从这一角度而言，特征函数 $\varphi_X(t)$ 比 MGF $M_X(t)$ 更一般化。
- 另外，定理 3.26 (1) 之所以成立，是因为

$$|\varphi_X(t)| \leq \int_{-\infty}^{\infty} \left|e^{\mathrm{i}tx}\right| dF_X(x) = \int_{-\infty}^{\infty} dF_X(x) = 1 < \infty$$

例 3.41：

- 假设随机变量 X 服从柯西分布 (Cauchy distribution) Cauchy(0,1)，其 PDF 为

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, -\infty < x < \infty$$

- 则其特征函数存在且为

$$\varphi_X(t) = e^{-|t|}, -\infty < t < \infty$$

- 注意 $\varphi_X(t)$ 在 $t = 0$ 处存在转折点，故其在原点处不可导。这与 X 当 $k \geq 1$ 时不存在任何矩 $E(X^k)$ 的事实一致（因此柯西分布的 MGF 不存在）。

定理 3.27

假设 X 的 k 阶矩存在，则 $\varphi_X(t)$ 对 $t \in (-\infty, \infty)$ 是 k 阶可导的，且

$$\varphi_X^{(k)}(0) = \mathbf{i}^k E\left(X^k\right)$$

证明:

- 给定 $E|X|^k < \infty$，因对任意 t，

$$\begin{aligned} \left|\varphi_X^{(k)}(t)\right| &= \left|\frac{d^k}{dt^k}\int_{-\infty}^{\infty} e^{\mathbf{i}tx} dF_X(x)\right| \\ &= \left|\int_{-\infty}^{\infty} (\mathbf{i}x)^k e^{\mathbf{i}tx} dF_X(x)\right| \\ &\le \int_{-\infty}^{\infty} |x|^k dF_X(x) \\ &= E|X|^k < \infty \end{aligned}$$

- 故 k 阶导数 $\varphi_X^{(k)}(t)$ 存在。

- 又因 $\varphi_X^{(k)}(t) = \int_{-\infty}^{\infty} (\mathbf{i}x)^k e^{\mathbf{i}tx} dF_X(x)$，令 $t=0$，得

$$\begin{aligned} \varphi_X^{(k)}(0) &= \int_{-\infty}^{\infty} (\mathbf{i}x)^k dF_X(x) \\ &= \mathbf{i}^k E\left(X^k\right) \end{aligned}$$

证毕。

定理 3.28

[特征函数的唯一性]：假设两个随机变量 X 和 Y 的特征函数分别为 $\varphi_X(t)$ 和 $\varphi_Y(t)$。则 X 和 Y 具有同分布，当且仅当对所有 $t \in (-\infty, \infty)$，有 $\varphi_X(t) = \varphi_Y(t)$。

证明:

- 根据定义，特征函数 $\varphi_X(t)$ 为 CDF $F_X(x)$ 的傅里叶变换，即

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{\mathrm{i}tx} dF_X(x)$$

- 由于分布函数与其傅里叶变换一一对应，故对每个给定分布 $F_X(x)$，有唯一的特征函数 $\varphi_X(t)$。
- **证毕。**

注意

- 这里需要检验对实数集上的所有 t，特征函数 $\varphi_X(t)$ 和 $\varphi_Y(t)$ 是否相等。这与 MGF 的情况不同。对 MGF $M_X(t)$ 和 $M_Y(t)$，只需检验它们对在 0 的某个邻域内的所有 t 是否相等即可。

定理 3.29

[特征函数的收敛]：假设随机变量序列 $\{X_n\}$ 的 CDF 和特征函数分别为 $F_n(x)$ 和 $\varphi_n(t)$。又设随机变量 X 的 CDF 和特征函数分别为 $F_X(x)$ 和 $\varphi_X(t)$。令 $n \to \infty$。

(1) 若对 $F_X(x)$ 的所有连续点 x，有 $F_n(x) \to F_X(x)$，则对任意 $t \in (-\infty, \infty)$，有 $\varphi_n(t) \to \varphi_X(t)$；

(2) 若对任意 $t \in (-\infty, \infty)$，有 $\varphi_n(t) \to \varphi_X(t)$，则对 $F_X(x)$ 的所有连续点 x，有 $F_n(x) \to F_X(x)$。

特征函数的应用

- 在某些情况下，使用特征函数 $\varphi_X(t)$ 比分布函数更为方便。例如，所谓的稳态分布是用特征函数定义的（参看第四章第4.3.4节），因为稳态分布的 PDF $f_X(x)$ 没有解析形式。
- 金融学中，仿射跳跃扩散模型 （参考 Duffie et al., 2000）作为一类常见的连续时间模型，其（转移）概率密度函数不存在解析形式。这导致极大似然估计不可行。然而，这类模型具有解析形式的（条件）特征函数，故可用其特征函数进行参数估计和统计推断。
- 在金融学，具有解析形式的特征函数的另一个重要例子是列维（Levy）过程。该随机过程是现代金融建模的基石。最著名的列维过程的一个特例是布朗运动。许多金融模型，包括Black & Scholes（1973）和 Cox et al.（1985）的连续时间模型，均假设布朗运动。

目　录

总结

- 随机变量的概念其重要之处在于量化了随机试验结果。
- 为了描述随机变量的概率分布，引入了**累积分布函数**（CDF），并且将随机变量分为**离散型**和**连续型**两种基本类别，分别引入**概率质量函数**（PMF）和**概率密度函数**（PDF）以刻画这两类随机变量的**概率分布**；还介绍了**随机变量可测变换概率分布**的推导方法。
- 另外，可用 CDF（或等价的 PMF/PDF）定义**各阶矩**以刻画概率分布的总体特征，并提供了关于一些矩的经济解释及其若干应用。
- 此外，还探讨了**矩生成函数** MGF 和**特征函数**，二者都可通过求导运算生成各阶矩（当矩存在时）。更重要地，两者均可唯一刻画随机变量的概率分布。

X 类型	离散型	连续型
累积分布函数	$F_X(x) = P(X \le x), \forall x \in \mathbb{R}.$	
概率函数	PMF: $f_X(x) = P(X = x)$	PDF: $f_X(x) = \frac{dF_X(x)}{dx} = F_X'(x)$
概率函数性质	(1) $0 \le f_X(x) \le 1$ (2) $\sum_{x\in\Omega} f_X(x) = 1$	(1) $f_X(x) \ge 0$ (2) $\int_{-\infty}^{\infty} f_X(x)dx = 1$
概率函数含义	对应取值 x 的概率	与 X 落在以 x 为中心的小邻域内的概率成正比： $f_X(x) \propto P\left(x - \frac{\epsilon}{2} < X \le x + \frac{\epsilon}{2}\right)$
期望 $\boldsymbol{E}[\boldsymbol{g}(\boldsymbol{X})]$	$\sum_{x\in\Omega_X} g(x)\, f_X(x)$	$\int_{-\infty}^{\infty} g(x)\, f_X(x)\mathrm{d}x$
均值 $\boldsymbol{\mu_X} = \boldsymbol{E}(\boldsymbol{X})$	$\sum_{x\in\Omega_X} x\, f_X(x)$	$\int_{-\infty}^{\infty} x\, f_X(x)dx$
方差 $\boldsymbol{\sigma_X}^2 = \boldsymbol{E}(\boldsymbol{X} - \boldsymbol{\mu_X})^2$	$\sum_{x\in\Omega_X}(x - \mu_X)^2\, f_X(x)$	$\int_{-\infty}^{\infty}(x - \mu_X)^2\, f_X(x)dx$
$\boldsymbol{\alpha}$-分位数$\boldsymbol{Q}(\boldsymbol{\alpha})$	$F_X[Q(\alpha)] = P\big(X \le Q(\alpha)\big) = \alpha$	$F_X[Q(\alpha)] = \int_{-\infty}^{Q(\alpha)} f_X(x)dx = \alpha$

中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

Thank You！