连续型随机变量与概率密度函数

第 5 讲 • 概率论与数理统计

复旦大学经济学院　ECON130001

上一讲我们做了什么

把事件升级成了随机变量

$X:\Omega\to\mathbb{R}$ 是一个**函数**；离散型用概率函数 $f(x)=P(X=x)$ 描述， 四个基本分布：伯努利 → 几何 → 二项 → 泊松。

也给了一个通用工具

分布函数 $F(x)=P(X\le x)$，离散型是阶梯，且 $P(X=x)=F(x)-F(x^-)$ 就是该点的跳跃高度。

★ 留下的问题

上一讲最后那个例子 $F(x)=\frac{x-4}{6}$ 是一条**光滑上升的曲线**，没有台阶 ——于是 $P(X=x)=0$，概率函数彻底失效。

但曲线的**斜率**还在：斜率大的地方概率就密集。这个「斜率」该怎么用？

为什么需要这一讲

场景　1998 年 8 月，长期资本管理公司（LTCM）。合伙人里有两位诺贝尔经济学奖得主， 风险模型是当时业界最精密的。模型说：**单日亏损超过 4500 万美元的概率， 大约是每年一次。**

冲突　那年 8 月 21 日，基金一天亏了 **5.5 亿美元**——超出模型日常波动尺度十倍以上。 按模型的正态假设，这种规模的单日损失**几十亿年才该出现一次**。
模型的数学没有错，参数也是老老实实用历史数据估的。 错的是**一个看不见的前提：收益率服从正态分布。**

悬念　**正态分布长什么样，为什么它几乎是所有金融与 AI 模型的默认选择， 它的尾巴又薄到什么程度——以至于把「几十亿年一次」和「就在下个月」搞混?**

来源与说明
LTCM 于 1998 年 8–9 月间发生巨额亏损、并由纽约联储组织救助，为金融史公认事实； 合伙人包括 1997 年诺贝尔经济学奖得主 Robert Merton 与 Myron Scholes。 本页引用的具体日损金额与「模型预测频率」为教学化表述， 用于说明正态尾部假设的脆弱性，不作为精确史料。 本讲末尾会用本讲公式给出可自行验算的量级对照。

本讲学习目标

1. 会用**概率密度函数**描述连续型随机变量，并与分布函数互相转换
2. 掌握三个连续分布：**均匀、指数、正态**，尤其是正态的标准化
3. 会求**随机变量函数**的分布——离散型用求和，连续型用「先求 F 再求导」

从「计数」到「测量」

上一讲的收尾例题

在 $[4, 10]$ 上任意抛一个质点，X 为它与原点的距离，落在任一子区间的概率与区间长度成正比。

$$F(x) = \begin{cases} 0, & x < 4 \\ \dfrac{x-4}{6}, & 4 \le x < 10 \\ 1, & x \ge 10 \end{cases}$$

现实中更多的量要「测」而不是「数」

- 股票的日收益率
- 下一季度的 GDP 增长率
- 自动驾驶汽车与前车的距离

挑战

对连续变量，**任何单点的概率都是 0**：$P(\text{收益率} = 1\%) = 0$。

所以概率函数 $f(x) = P(X = x)$ 恒为 0，没有任何信息。**需要一个全新的工具。**

概率密度函数

定义（probability density function, p.d.f.）

对连续型随机变量 X，若其分布函数可以写成

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

则称 f 为 X 的**概率密度函数**，记作 $X \sim f(x)$。

★ 密度不是概率

$f(x)$ 本身**不是** X 取 x 的概率——它甚至可以大于 1。 有意义的是它在一段区间上的**积分**。

类比：密度 $\mathrm{kg/m^3}$ 不是质量，乘上体积才是质量。 **概率是「密度 × 长度」，摊在区间上而不是堆在点上。**

回到上一讲的问题：F 的斜率就是 f。 **斜率大的地方，同样宽的区间里装的概率就多。**

密度函数的四条性质

- $f(x) \geq 0$

  ← 非负性

- $\displaystyle\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1$

  ← 归一化

- $P(x_1 < X \leq x_2) = \displaystyle\int_{x_1}^{x_2} f(t)\,\mathrm{d}t$
- 在 f 的所有连续点上，$F'(x) = f(x)$

证明思路

第三条：由 $F(x_2) - F(x_1) = \int_{-\infty}^{x_2} f - \int_{-\infty}^{x_1} f$ 与上一讲的 $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$ 得到。第四条是微积分基本定理。完整证明见教师笔记。

一个便利的后果

因为 $P(X = x) = 0$，区间端点开闭**无关紧要**：$P(a < X < b) = P(a \leq X \leq b) = \int_a^b f$。**离散型里完全不成立**——那里少算一个端点就少一块概率。

连续分布之一：均匀分布

例

若 ξ 的密度为 $\varphi(x)=\lambda$（$a\le x\le b$），其余为 0，则称 ξ 服从 $[a,b]$ 上的**均匀分布**。求 $F(x)$。

解

由归一化 $\lambda(b-a)=1$，得 $\lambda=\dfrac{1}{b-a}$。积分得

$$F(x)=\begin{cases}0, & x<a\\ \dfrac{x-a}{b-a}, & a\le x<b\\ 1, & x\ge b\end{cases}$$

这正是上一讲那道 $[4,10]$ 质点题的一般形式。

注意 $\lambda=\frac{1}{b-a}$：区间越短，密度越**高**。$[0,0.1]$ 上的均匀分布密度是 10——**密度大于 1 完全正常**。

均匀分布：神经网络参数初始化

核心思想

当我们对一个量**一无所知**，只能确定它的范围时， 均匀分布是「最公平」的假设——不偏袒任何取值。

AI 应用：训练前要给参数一个初值

- 训练深度模型前，成千上万个参数都要赋初值。太大太小都会让模型学不动
- 最常见的做法：在一个精心算出的对称区间 $[-a, a]$ 内**均匀抽样**

$$W \sim U[-a, a]$$

- 这就是 **Xavier 初始化**等方法的核心思想

可以说，**均匀分布是许多复杂 AI 模型训练的起点**。 区间宽度 a 怎么定？要用到方差——第 7 讲。

用归一化定常数（连续版）

例

$\varphi(x)=kx+1$（$0\le x\le 2$），其余为 0。求 k、$F(x)$ 及 $P(1.5<\xi<2.5)$。

归一化：$\int_0^2(kx+1)\mathrm{d}x=2k+2=1\Rightarrow k=-\frac{1}{2}$。

$$F(x)=-\frac{1}{4}x^2+x,\quad 0\le x\le 2$$

$$P(1.5<\xi<2.5)=F(2)-F(1.5)=1-0.9375=0.062$$

注意：ξ 最大只到 2，所以 $F(2.5)=F(2)=1$。积分上限要落在支撑集内。

例

$f(x)=cxe^{-x^2/2}$（$x>0$），其余为 0。求 c。

换元 $u=x^2/2$：

$$\int_0^\infty cxe^{-x^2/2}\mathrm{d}x=c\int_0^\infty e^{-u}\mathrm{d}u=c=1$$

故 $c=1$。

这个分布叫**瑞利分布**，是二维正态的模长——第 6 讲会见到。

反过来：由分布函数求密度

例

$F(x)=e^{x-3}$（$x\le 3$），$F(x)=1$（$x>3$）。求密度函数。

对 F 求导：

$$f(x)=\begin{cases}e^{x-3}, & x\le 3\\ 0, & x>3\end{cases}$$

「先求 F 再求导」是连续型的标准动作，本讲后半还会用到。

例

$F(x)=0$（$x<0$）、$a+b\cos x$（$0\le x<\pi/4$）、1（$x\ge \pi/4$）。求 $a+b$。

连续型的分布函数是连续函数（没有跳跃），所以在 $x=0$ 处

$$\lim_{x\to 0^-}F(x)=0=a+b\cos 0=a+b$$

故 $a+b=0$。

这道题考的不是计算，是**「连续型 $\Rightarrow$ F 连续」这条性质**。

一道综合题

例

$\varphi(x)=A\cos x$（$|x|\le \pi/2$），其余为 0。 求 (1) A；(2) 分布函数；(3) $P(0<\xi<\pi/4)$。

解

(1) $\int_{-\pi/2}^{\pi/2} A\cos x\,\mathrm{d}x = 2A = 1 \Rightarrow A=\frac{1}{2}$。

(2) 对 $|x|\le\pi/2$：$F(x)=\displaystyle\int_{-\pi/2}^{x}\tfrac{1}{2}\cos t\,\mathrm{d}t=\frac{1+\sin x}{2}$，于是

$$F(x)=\begin{cases}0, & x<-\frac{\pi}{2}\\ \dfrac{1+\sin x}{2}, & |x|\le\frac{\pi}{2}\\ 1, & x>\frac{\pi}{2}\end{cases}$$

(3) $F(\frac{\pi}{4})-F(0)=\dfrac{1+\frac{\sqrt{2}}{2}}{2}-\dfrac{1}{2}=\dfrac{\sqrt{2}}{4}\approx 0.354$。

核查一下 F：$F(-\pi/2)=0$、$F(\pi/2)=1$、处处不减且连续。**凡是求出 F，都该这样验一遍。**

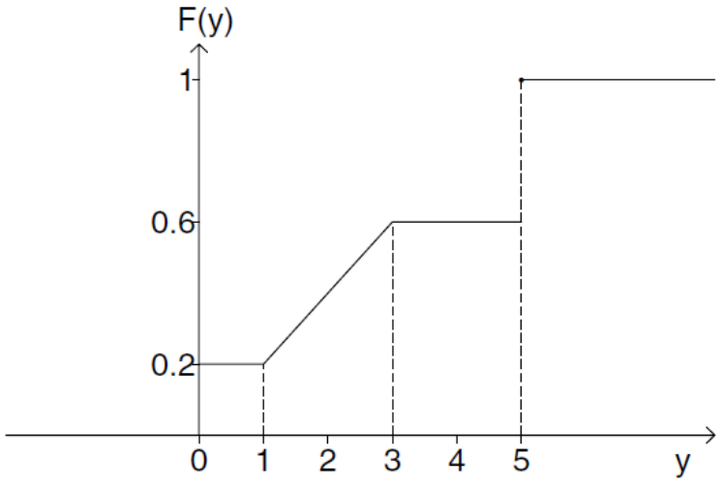
一个既不离散也不连续的例子

例

X 服从 $[0,5]$ 上的均匀分布，定义

$$Y=\begin{cases}0, & X\le 1\\ 5, & X\ge 3\\ X, & \text{其他}\end{cases}$$

画出 Y 的分布函数。



台阶 + 斜坡

解的形状

$P(Y=0)=P(X\le 1)=\frac{1}{5}$，$P(Y=5)=P(X\ge 3)=\frac{2}{5}$—— **这两点上有跳跃**；而在 $(1,3)$ 上 $Y=X$，F 连续上升。

所以 F_Y 是「**两级台阶 + 中间一段斜坡**」。

它既不离散也不连续。分布函数的价值就在这里——它对任何随机变量都有定义。

小结：连续型的三件事

- **密度不是概率**，$f(x)$ 可以大于 1；概率是它在区间上的积分
- F **与** f **互为积分/求导**：$F(x) = \int_{-\infty}^{x} f$，$f = F'$
- $P(X = x) = 0$，于是区间端点开闭无所谓——但这只对连续型成立

接下来三个具体分布：**均匀（已讲）、指数、正态**。

连续分布之二：指数分布

定义

X 服从参数为 λ 的**指数分布**，若

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & \text{其他}\end{cases}$$

两条直接推论

- $F(x)=1-e^{-\lambda x}\ \ (x>0)$
- $P(a<X<b)=e^{-\lambda a}-e^{-\lambda b}$

特别地，$P(X>t)=e^{-\lambda t}$——「活过 t」的概率，本讲后面反复要用。

λ 叫**速率**：λ 越大，密度衰减越快，事件来得越急。 它就是上一讲泊松分布里那个 λ——下面会证明这不是巧合。

指数分布的无后效性

例

电子元件寿命服从参数 λ 的指数分布。求 (1) 寿命大于 t 的概率；(2) **已工作了 s 小时后**，再活过 t 小时的概率。

解

(1) $P(X > t) = \int_t^{\infty} \lambda e^{-\lambda x} \mathrm{d}x = e^{-\lambda t}$

(2) 用条件概率：

$$P(X > t+s \mid X > s) = \frac{P(X > t+s)}{P(X > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$$

★ 两个答案一模一样

已经用了 s 小时这件事，对元件的剩余寿命毫无影响——它「不会变老」。

这是上一讲几何分布无记忆性的**连续版本**。 而且可以证明：指数分布是唯一具有这个性质的连续分布。

指数分布例题：n 个灯泡

例

n 个灯泡同时点亮，寿命相互独立且都服从参数 λ 的指数分布。求 (1) 到**第一个**灯泡失效的时间 Y_1 的分布；(2) 从第一个失效到**第二个**失效的时间 Y_2 的分布。

解 (1)

$Y_1 > t$ 意味着**所有 n 个都还没坏**。由独立性

$$P(Y_1 > t) = \prod_{i=1}^{n} P(X_i > t) = \left(e^{-\lambda t}\right)^n = e^{-n\lambda t}$$

所以 Y_1 服从参数 $n\lambda$ 的指数分布——n 个一起等，来得快 n 倍。

解 (2)

第一个坏掉后还剩 $n-1$ 个，而由**无记忆性**， 这 $n-1$ 个「和新的一样」。于是同理，Y_2 服从参数 $(n-1)\lambda$ 的指数分布。

没有无记忆性，第 (2) 问根本没法这么简单地做。

泊松与指数：同一过程的两个侧面

核心思想

两者描述的是同一个随机过程（**泊松过程**）的两个问法：

- **泊松分布**（离散）：在固定时间内，会发生多少次？
- **指数分布**（连续）：等到下一次发生，需要多长时间？

例：公交车到站平均每小时 λ 辆。「一小时来几辆」服从泊松，「我要等多久」服从指数。

从泊松推出指数

设速率为 λ，T 是到下一次事件的等待时间。关键的一步翻译：

$$\{T > t\} \iff \{\text{在长度 } t \text{ 的区间内事件发生了 } \mathbf{0} \text{ 次}\}$$

由泊松分布 $P(\text{时长 } t \text{ 内发生 } k \text{ 次}) = \dfrac{e^{-\lambda t}(\lambda t)^k}{k!}$，取 $k = 0$：

$$P(T > t) = e^{-\lambda t} \implies F(t) = 1 - e^{-\lambda t}$$

正是参数 λ 的指数分布。

指数分布：失业持续期与服务器排队

失业持续时间的建模

- 若失业者找到工作的「瞬时概率」是常数 λ，则所需时间 T 服从指数分布
- **无记忆性的含义：** $P(T > t+s \mid T > t) = P(T > s)$—— 已失业很久的人，未来找到工作的概率和刚失业的人一样

排队论与服务系统

- 用户请求到达服务器的**间隔时间**常被建模为指数分布
- 这是排队论的基础，用于服务器动态扩容与云端任务调度
- 与上一讲呼应：请求数服从泊松，请求间隔服从指数

但左边这个模型的假设要小心

无记忆性意味着**失业时长不影响再就业难度**。现实中往往相反： 技能折旧、雇主对长期失业者的筛选，都会让 λ 随时间下降。 **模型给出的是一个基准，偏离基准的地方才是研究对象。**

连续分布之三：正态分布

它是概率论与数理统计中最重要的分布，有三个原因

1. 若总体服从正态分布，其**随机抽样的特征有简单的表达式**——第 9 讲整章都依赖这一条
2. 大量实际观测（测量误差、生物性状等）确实**近似**服从正态分布
3. **中心极限定理**：即使原变量不服从正态，样本的某些函数也会趋近正态——第 8 讲

第 3 条是它真正的统治力来源：**正态分布不是因为世界本来正态， 而是因为「大量微小独立因素之和」必然趋近正态。**

但请记住开场那个故事——**「近似」和「精确」之间的差距，全在尾巴上。**

正态分布的定义

定义

X 服从参数为 μ, σ^2 的**正态分布**，若密度函数为

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

记作 $X \sim N(\mu, \sigma^2)$。

怎么证明它的积分是 1

记 $I = \int_{-\infty}^{\infty} e^{-t^2/2}\mathrm{d}t$，考虑 I^2 化为二重积分后换成极坐标：

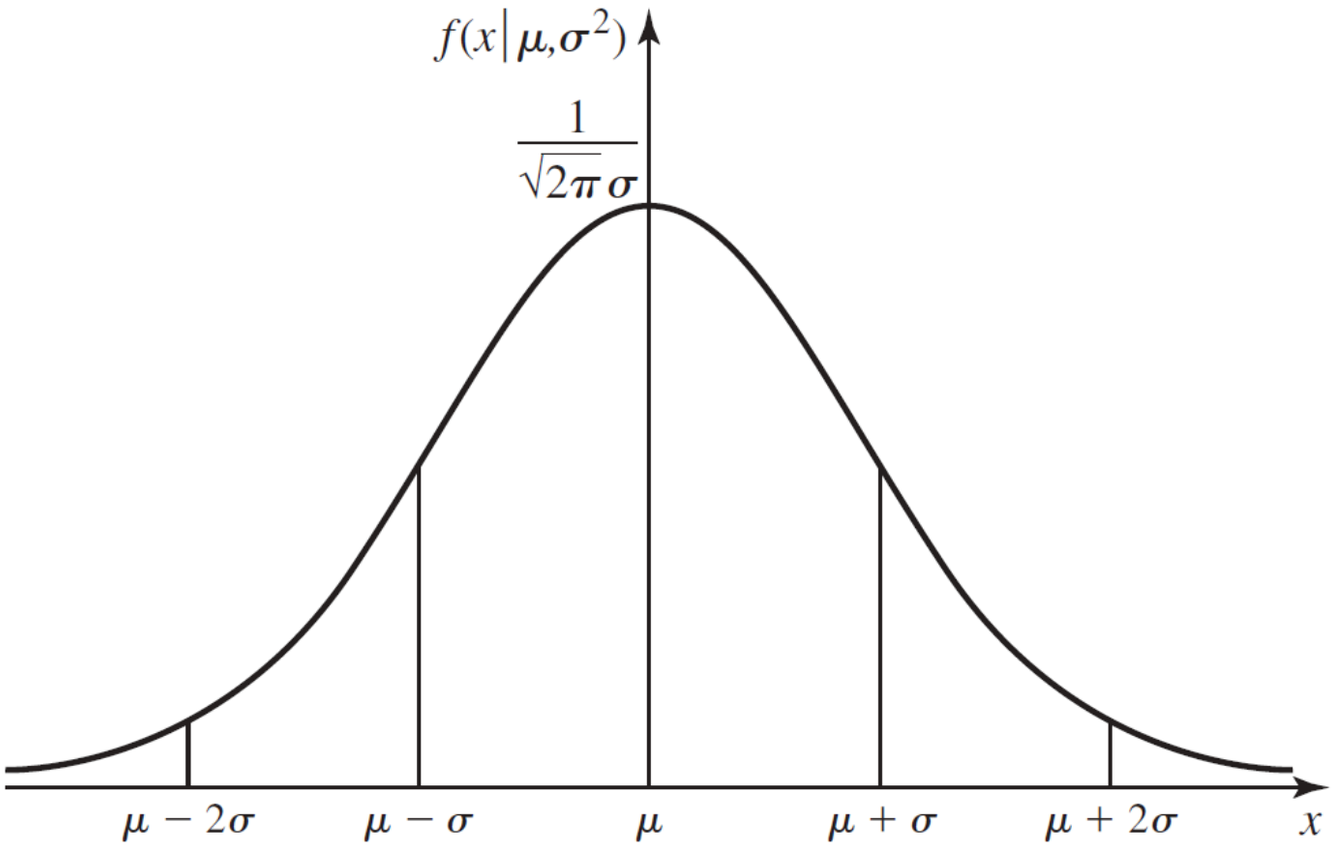
$$I^2 = \iint e^{-(u^2+v^2)/2}\mathrm{d}u\,\mathrm{d}v = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r\,\mathrm{d}r\,\mathrm{d}\theta = 2\pi$$

故 $I = \sqrt{2\pi}$，归一化成立。这是经典的**Poisson 积分**技巧 ——一维算不出来，升到二维反而能算。完整推导见教师笔记。

读这个式子

指数里是 $-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$：$\mu$ **决定中心位置，**σ **决定胖瘦。** 前面的 $\frac{1}{\sqrt{2\pi}\sigma}$ 只是为了让总面积等于 1。

正态分布的形状



钟形曲线

四个特征

- 关于 $x=\mu$ **对称**
- 在 $x=\mu$ 处取最大值 $\frac{1}{\sqrt{2\pi}\sigma}$
- 在 $x=\mu\pm\sigma$ 处有**拐点**——σ 是可以「看见」的
- 两侧以 $e^{-x^2/2}$ 的速度趋于 0，**衰减极快**

最后一条就是开场故事的伏笔：$e^{-x^2/2}$ **的衰减比任何指数都快**，所以正态分布认为极端事件「几乎不可能」。 真实收益率的尾巴要厚得多。

标准化：把所有正态分布变成一个

线性变换

若 $X \sim N(\mu, \sigma^2)$ 且 $Y = aX + b$（$a \neq 0$），则

$$Y \sim N(a\mu + b,\ a^2\sigma^2)$$

特别地，取 $a = \frac{1}{\sigma}$、$b = -\frac{\mu}{\sigma}$：

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

标准正态分布

$N(0, 1)$ 称为**标准正态分布**，其密度与分布函数记为

$$\phi_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad \Phi_0(x) = \int_{-\infty}^{x} \phi_0(u)\,\mathrm{d}u$$

★ 为什么这件事这么重要

Φ_0 **没有初等函数表达式**，只能查表或数值计算。 但有了标准化，**全世界只需要一张表**——任何 $N(\mu, \sigma^2)$ 的问题 都先化成 Z 再查。

查表要用的三条性质

- 一般正态的密度：$\phi(x) = \dfrac{1}{\sigma}\phi_0\left(\dfrac{x-\mu}{\sigma}\right)$
- 一般正态的分布函数：$\Phi(x) = \Phi_0\left(\dfrac{x-\mu}{\sigma}\right)$
- **对称性：**$\Phi_0(-x) = 1 - \Phi_0(x)$

第三条是查表的关键

表上通常**只印 $x \geq 0$ 的值**。要查 $\Phi_0(-2)$， 就用 $\Phi_0(-2) = 1 - \Phi_0(2)$。

它来自密度关于 0 对称这个几何事实——**不用背，画个图就出来了。**

标准化例题

例

$X \sim N(5, 4)$（即 $\mu = 5$、$\sigma = 2$），求 $P(1 < X < 8)$。

解：三步

① 标准化 令 $Z = \dfrac{X - 5}{2} \sim N(0, 1)$：

$$P(1 < X < 8) = P\left(\frac{1 - 5}{2} < Z < \frac{8 - 5}{2}\right) = P(-2 < Z < 1.5)$$

② 拆成分布函数之差

$$= \Phi_0(1.5) - \Phi_0(-2)$$

③ 用对称性处理负值再查表

$$= \Phi_0(1.5) - [1 - \Phi_0(2)] = 0.93319 - (1 - 0.97725) = 0.91044$$

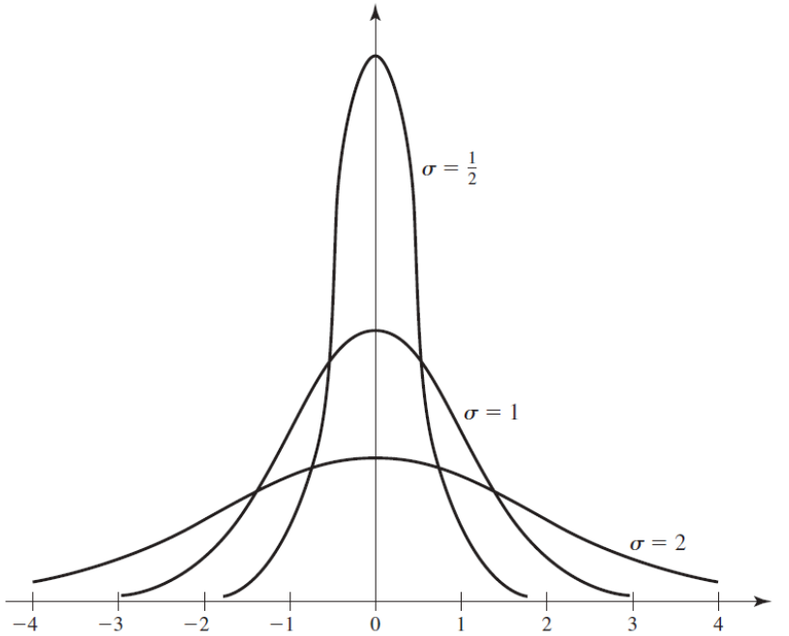
$k\sigma$ 法则：与 μ, σ 无关的那些数

对任何正态分布

$$p_k = P(|X - \mu| \le k\sigma) = P(|Z| \le k)$$

右边只与 k 有关，和 μ, σ 无关。

k	p_k
1	0.6826
2	0.9544
3	0.9974
4	0.99994
5	$1 - 6 \times 10^{-7}$
10	$1 - 2 \times 10^{-23}$



不同 μ, σ，同样的 $k\sigma$ 比例

★ 看最后两行

10σ 之外只有 2×10^{-23}。每天观测一次，平均要等 10^{20} 年 ——

比宇宙年龄还长十亿倍。

回到开场：尾巴有多薄

把 LTCM 的数字放进这张表

若日收益率真的服从正态分布，那么一个 10σ 的单日损失， 平均要 $\dfrac{1}{2\times 10^{-23}}\approx 5\times 10^{22}$ 个交易日才出现一次——约合 2×10^{20} 年。

但这类事件在真实市场上，几十年里出现过不止一次

1987 年 10 月 19 日、1998 年 8–9 月、2008 年秋、2020 年 3 月…… **「几十亿年一次」和「十年一次」的差距，不在计算，在假设。**

正态分布的尾部按 $e^{-x^2/2}$ 衰减，**快得不像话**。 真实收益率的尾部是幂律式的，厚得多。

这不是说正态分布没用——它是**最好的第一近似**， 也是第 8–12 讲全部推断工具的基础。 但用它做**极端风险**的判断时，必须知道自己在假设什么。

应用：风险价值 VaR

例

某股票日收益率 $R \sim N(0.0005,\ 0.02^2)$，持有 100 万元。 问：未来一天，有 95% 的把握损失不超过多少？

解

等价于求临界值 r 使 $P(R < r) = 0.05$。

① 标准化 $\quad P\left(Z < \dfrac{r - 0.0005}{0.02}\right) = 0.05$

② 查表 $\quad$ 下 5% 分位点 $\dfrac{r - 0.0005}{0.02} \approx -1.645$

③ 解出 $\quad r \approx -0.0324$，即 -3.24%

$$\text{VaR} = 100\ \text{万} \times 3.24\% = 32\,400\ \text{元}$$

读法：**「平均每 20 个交易日，会有 1 天亏得比 32 400 元多」**。 注意它没有回答「那一天会亏多少」——而那正是开场故事里出事的地方。

交互：正态分布与风险价值

怎么用

拖 μ、σ 与置信水平，看红色尾部面积与 VaR 数字同步变化。

当堂要做的两个实验

1. 把置信水平从 95% 拖到 99%：VaR 变大，**但尾巴里剩下的 1% 依然可以任意糟**
2. 拖 μ 和 σ：下方三个 $k\sigma$ 概率**纹丝不动**——标准化的意义

交互演示：拖动参数与置信水平，观察尾部面积与 VaR

https://fin-tech.fun/courseware/probability-statistics-undergrad/toolkit/normal_var.html

应用：正态分布是模型可解释性的根基

场景：用线性回归预测房价

$$\text{房价} = \beta_0 + \beta_1 \times \text{面积} + \epsilon$$

核心假设：无法解释的误差项 $\epsilon \sim N(0, \sigma^2)$。

因此才能做参数检验

基于正态假设，才能对系数 β_1 做 t 检验， 判断「面积」是否真有统计意义上的影响。

第 12 讲

因此才能给预测区间

也正是基于它，才能为预测的房价给出一个置信区间， 而不只是一个孤零零的数。

第 11 讲

没有这个假设，模型只能给出一个点预测，无法回答「这个预测有多可信」。 正态分布是连接机器学习与统计推断的桥梁——它把「预测」变成了「可以被质疑和检验的预测」。

随机变量函数的分布

动机

我们常常关心的不是 X 本身，而是它的某个函数。 例：X 是汽车时速，我们想知道跑完一百公里的用时 $Y = 100/X$。

通用做法：先求 Y 的分布函数

$$F_Y(y) = P(Y \le y) = P\left(\frac{100}{X} \le y\right) = P\left(X \ge \frac{100}{y}\right) = 1 - F_X\left(\frac{100}{y}^-\right)$$

核心动作：把关于 Y 的事件翻译成关于 X 的事件， 再用已知的 F_X。

注意 $Y = 100/X$ 是**减函数**，所以不等号**翻了向**。 这是本节最容易错的地方——每次都要先判断单调方向。

离散型：把落到同一点的概率加起来

定理

离散型随机变量 X 有概率函数 f，令 $Y = r(X)$，则 Y 的概率函数为

$$g(y) = P[r(X) = y] = \sum_{x:\, r(x)=y} f(x)$$

例

X 在 1–9 上均匀分布，$Y = |X - 5|$。

Y 取 $0, 1, 2, 3, 4$。$Y = 0$ 只来自 $X = 5$；$Y = k$（$k \geq 1$）来自 $X = 5 \pm k$ **两个**值：

$$g(y) = \begin{cases} \frac{1}{9}, & y = 0 \\ \frac{2}{9}, & y = 1, 2, 3, 4 \end{cases}$$

例

X 等概率取 $-3, \ldots, 3$（七个值），$Y = X^2 - X$。

逐个代：$X = 0, 1 \to Y = 0$；$X = -1, 2 \to Y = 2$；$X = -2, 3 \to Y = 6$；$X = -3 \to Y = 12$。

$$g(y):\ \tfrac{2}{7}, \tfrac{2}{7}, \tfrac{2}{7}, \tfrac{1}{7} \quad (y = 0, 2, 6, 12)$$

连续型：先求 G，再求导

定理

连续型随机变量 X 有密度 f，令 $Y = r(X)$，则

$$G(y) = P[r(X) \le y] = \int_{x:\, r(x) \le y} f(x)\,\mathrm{d}x$$

若 Y 也是连续型，其密度为 $g(y) = \dfrac{\mathrm{d}G(y)}{\mathrm{d}y}$。

例

Z 是队列的服务速率，有连续分布函数 F。平均等候时间 $Y = 1/Z$，求 G。

$$G(y) = P(Y \le y) = P\left(\frac{1}{Z} \le y\right) = P\left(Z \ge \frac{1}{y}\right) = P\left(Z > \frac{1}{y}\right) = 1 - F\left(\frac{1}{y}\right)$$

倒数第二步用到 $P(Z = 1/y) = 0$——连续型才能这样把 $\ge$ 换成 $>$。

「先求分布函数、再求导」是连续型的唯一通法。 直接对密度做代换是错的——密度不是概率，不能像换元那样搬。

连续型函数分布的两个例子

例：$Y = X^2$

$X \sim U[-1,1]$，密度 $f(x) = \frac{1}{2}$。求 $Y = X^2$ 的密度。

对 $0 < y < 1$：

$$G(y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}\,\mathrm{d}x =$$

求导：

$$g(y) = \frac{1}{2\sqrt{y}}, \quad 0 < y < 1$$

注意 g 在 $y \to 0^+$ 时趋于无穷，但它仍是合法密度（积分为 1）——再次说明密度不是概率。

线性变换的一般公式

X 有密度 f，$Y = aX + b$（$a \neq 0$），则

$$g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right)$$

证明思路

设 $a > 0$：$G(y) = P(aX + b \le y) = F\left(\frac{y-b}{a}\right)$，对 y 求导得 $\frac{1}{a} f\left(\frac{y-b}{a}\right)$。$a < 0$ 时不等号翻向，多出一个负号，合起来即 $|a|$。完整证明见教师笔记。

那个 $\frac{1}{|a|}$ 就是**雅可比因子**：变量被拉伸 a 倍，密度必须压扁 a 倍，总面积才守恒。

把它用在正态分布上，正好给出前面那条 $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$。

本讲总结

① 回到开场那个问题

正态分布的尾部按 $e^{-x^2/2}$ 衰减，快到 10σ 之外的概率只有 2×10^{-23} ——按这个模型，那种损失「10^{20} 年才出现一次」。

可它在几十年里出现了好几次。**数学没有错，参数也没估错， 错的是「收益率服从正态分布」这个看不见的前提。** 与第 3 讲那个结构化产品的教训完全一样：出事的从来不是公式，是公式的假设。

② 核心结论

- **密度不是概率**：f 可大于 1、可趋于无穷；概率是 $\int_a^b f$
- F 与 f 互为积分与导数；连续型 $P(X=x)=0$，端点开闭无所谓
- 三个分布：均匀（无知时最公平）、指数（无记忆、泊松过程的等待时间）、正态（$k\sigma$ 与标准化）
- 随机变量函数：离散型**求和**，连续型**先求 G 再求导**

本讲总结（续）

③ 易错提醒

- **密度 $\neq$ 概率。** $f(x)$ 可以大于 1，$f(x)=0.3$ 不表示概率是 0.3
- 查表前**必须标准化**，且负值要用 $\Phi_0(-x)=1-\Phi_0(x)$
- 求 $Y=r(X)$ 的分布时，**不能直接对密度做代换**，会漏掉雅可比因子 $\frac{1}{|a|}$
- r 不单调时要**分段讨论**并把各段概率相加（如 $Y=X^2$）

④ 下一讲

到现在为止，我们一次只研究**一个**随机变量。 可现实里要紧的往往是**几个量之间的关系**： 两只股票的收益会不会一起跌？身高与体重怎么联动？

而这正是开场那个故事真正的病根—— **LTCM 的模型不是不知道单个资产的分布，是低估了它们「一起动」的程度。**

下一讲：多维随机变量及其分布。