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

从事件到随机变量

$X:\Omega\to\mathbb{R}$ 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。

还引入了分布函数

$F(x)=P(X\le x)$；离散型的 F 是阶梯函数，$P(X=x)=F(x)-F(x^-)$ 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**连续上升的折线**，没有台阶，于是每一点的 $P(X=x)$ 都是 0，概率函数无法描述这个分布。

但折线的**斜率**还在：斜率越大的地方，概率越密集。这个「斜率」该怎么用？

为什么需要这一讲

背景　1998 年 8 月，长期资本管理公司（LTCM）的合伙人包括两位诺贝尔经济学奖得主。其风险模型假定收益率服从**正态分布**：分布曲线呈钟形，取值多集中在平均水平附近；两端的**「尾部」对应大幅上涨或下跌**，概率很小。模型预测：单日亏损超过 4500 万美元的情况，平均每年约一次。

转折　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，**在宇宙的年龄里都不该出现一次**。
模型的计算基于历史数据与分布假设。这个案例提示：**用正态分布描述收益率，可能低估极端亏损的概率。**

问题　正态分布是什么形状，为什么在金融与 AI 模型中广泛使用？对于远离平均水平的大幅亏损，它为什么会给出如此小的概率?

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。

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

许多实际变量可用连续型随机变量建模

- 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

问题

对连续型随机变量，**任何单点的概率都是 0**（上一讲第 27 页）：$P(\text{收益率} = 1\%) = 0$。

所以 $P(X = x)$ 恒为 0，无法描述分布，**需要另一种工具。**

概率密度函数

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

若存在非负函数 f，使随机变量 X 的分布函数对一切 x 都可以写成

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

则称 X 为**连续型随机变量**，f 为 X 的**概率密度函数**（简称密度），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。**在很短的区间 $[x, x+\Delta x]$ 上，概率约为 $f(x)\Delta x$：概率分布在区间上，不集中在单个点上。**

回到上一讲的问题：F 的斜率就是 f。**斜率越大的地方，同样宽的区间里概率越大。**

密度函数的四条性质

- $f(x) \geq 0$

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

  ← 归一化
- $P(x_1 < X \leq x_2) = \int_{x_1}^{x_2} f(t)\,\mathrm{d}t$
- 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。

一个便利的后果

因为 $P(X = x) = 0$，区间端点开闭**无关紧要**：$P(a < X < b) = P(a \leq X \leq b) = \int_a^b f$。**离散型里一般不成立**：端点若是取值点，少算它就少了该点的概率。

连续分布之一：均匀分布

例

若 X 的密度为 $f(x)=c$（$a\le x\le b$），其余为 0，则称 X 服从 $[a,b]$ 上的**均匀分布**，记作 $X\sim U[a,b]$。求 c 与 $F(x)$。

解

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

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

上一讲第 32 页的质点题就是 $a=4$、$b=10$ 的情形。

注意 $c=\frac{1}{b-a}$：区间越短，密度越**高**。$[0,0.1]$ 上的均匀分布密度是 10——**密度可以大于 1**。

均匀分布：神经网络参数初始化

核心思想

若只知道一个量的取值范围，没有理由认为某些位置更可能，就用均匀分布：**长度相同的子区间，概率相同。**

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。太大或太小都会使训练难以进行
- 常用做法之一：在按网络结构算出的对称区间 $[-a, a]$ 内**均匀抽样**

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

- **Xavier 均匀初始化**是一种采用上述抽样方式的参数初始化方法

许多神经网络的训练，第一步就是从这样的均匀分布中抽取参数初值。

用归一化定常数（连续版）

例

X 的密度为 $f(x)=kx+1$（$0\le x\le 2$），其余为 0。求 k、$F(x)$ 及 $P(1.5<X<2.5)$。

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

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

$$P(1.5<X<2.5)=F(2)-F(1.5)$$
$$=1-0.9375=0.0625$$

注意： X 最大只到 2，所以 $F(2.5)=F(2)=1$。超出取值范围的部分密度为 0，积分不再增加。

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**，常用来描述风速、信号幅度这类非负的量。

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：

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

由分布函数求导得到密度；本讲后半求随机变量函数的分布，用的也是这一步。

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

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

解

(1) $\int_{-\pi/2}^{\pi/2} A\cos x\,\mathrm{d}x = 2A = 1 \Rightarrow A=\tfrac12$。

(2) 对 $|x|\le\pi/2$：$F(x)=\displaystyle\int_{-\pi/2}^{x}\tfrac12\cos t\,\mathrm{d}t=\frac{1+\sin x}{2}$，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) $F(\tfrac{\pi}{4})-F(0)=\dfrac{1+\frac{\sqrt2}{2}}{2}-\dfrac12=\dfrac{\sqrt2}{4}\approx 0.354$。

核查 F：$F(-\pi/2)=0$、$F(\pi/2)=1$、处处不减且连续。**求得连续型随机变量的分布函数后，应逐项检查这些性质。**

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

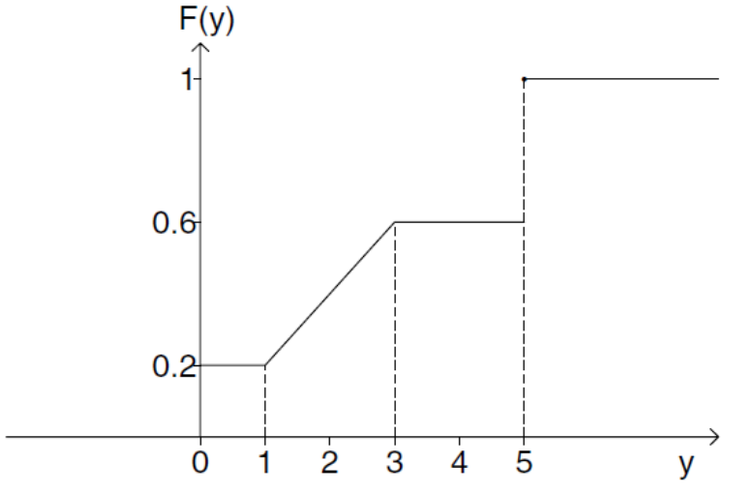
$$Y=\begin{cases}0, & X\le 1\\ 5, & X\ge 3\\ X, & \text{其他}\end{cases}$$

画出 Y 的分布函数。

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**」。



台阶 + 斜坡

它既不离散也不连续，没有概率函数，也没有密度，只能用分布函数描述——分布函数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**：$F(x)=\int_{-\infty}^{x} f$，$f=F'$
- $P(X=x)=0$，于是区间端点开闭不影响概率——这一点只对连续型成立

三个常用的连续分布：**均匀（已讲）、指数、正态**。

连续分布之二：指数分布

定义

若 X 的密度如下（$\lambda > 0$），则称 X 服从参数为 λ 的**指数分布**：

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

两条直接推论

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

特别地，当 $t \ge 0$ 时，$P(X > t) = 1 - F(t) = e^{-\lambda t}$，即 X 大于 t 的概率。

λ 叫**速率**：λ 越大，密度衰减越快，$P(X > t)$ 越小。第 18 页将介绍泊松过程，并说明这一参数与事件发生速率的关系。

指数分布的无记忆性

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**，再工作 t 小时以上的概率。

解

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

(2) 用条件概率：

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

★ 两个答案完全相同

已经工作了 s 小时的元件，剩余寿命超过 t 的概率与新元件相同——元件「不会老化」。

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 个灯泡都还没有失效**。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 服从参数为 $n\lambda$ 的指数分布：参数是单个灯泡的 n 倍。

解 (2)

第一个失效后还剩 $n-1$ 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) 问就不能这样简化。

泊松过程：事件次数与等待时间

同一过程中的两类随机变量

在速率为 λ 的**泊松过程**中，事件以恒定平均速率随机到达，不相交时间区间内的发生次数相互独立。

- 固定时长 t 内的**事件次数** $N(t)$：服从参数为 λt 的**泊松分布**（离散型）
- 到下一次事件的**等待时间** T：服从参数为 λ 的**指数分布**（连续型）

例如，可用泊松过程描述服务器请求的随机到达。

由事件次数推导等待时间的分布

从观察起点到下一次事件的等待时间记为 T。对 $t \geq 0$：

$$\{T > t\} = \{N(t) = 0\}$$

即「等待时间超过 t」等价于「t 内没有事件发生」。由 $P(N(t) = k) = \dfrac{e^{-\lambda t}(\lambda t)^k}{k!}$，取 $k = 0$：

$$P(T > t) = P(N(t) = 0) = e^{-\lambda t} \implies F_T(t) = 1 - e^{-\lambda t}$$

因此，T 服从参数为 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

- 若失业者在很短的时间 Δt 内找到工作的概率约为 $\lambda\Delta t$，且 λ 不随失业时长变化，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 t 个月的人，再过 s 个月仍未找到工作的概率，与刚失业的人相同

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
- 在泊松到达模型下，固定时段内的请求数服从泊松分布，请求间隔服从指数分布

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。**用指数分布描述失业时长，等于假定 λ 不随时间变化；实际研究要先检验这一点。**

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，由样本算出的许多量（如样本平均值）的分布**都有简单的表达式**
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**（第 8 讲）：大量独立随机因素之和近似服从正态分布，即使每个因素本身不服从正态

第 3 条是它被广泛使用的主要原因：**测量误差这类量，往往是许多微小、独立的因素叠加的结果。**

正态分布的定义

定义

若 X 的密度如下（$\sigma > 0$），则称 X 服从参数为 μ、σ^2 的**正态分布**，记作 $X \sim N(\mu, \sigma^2)$：

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

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：

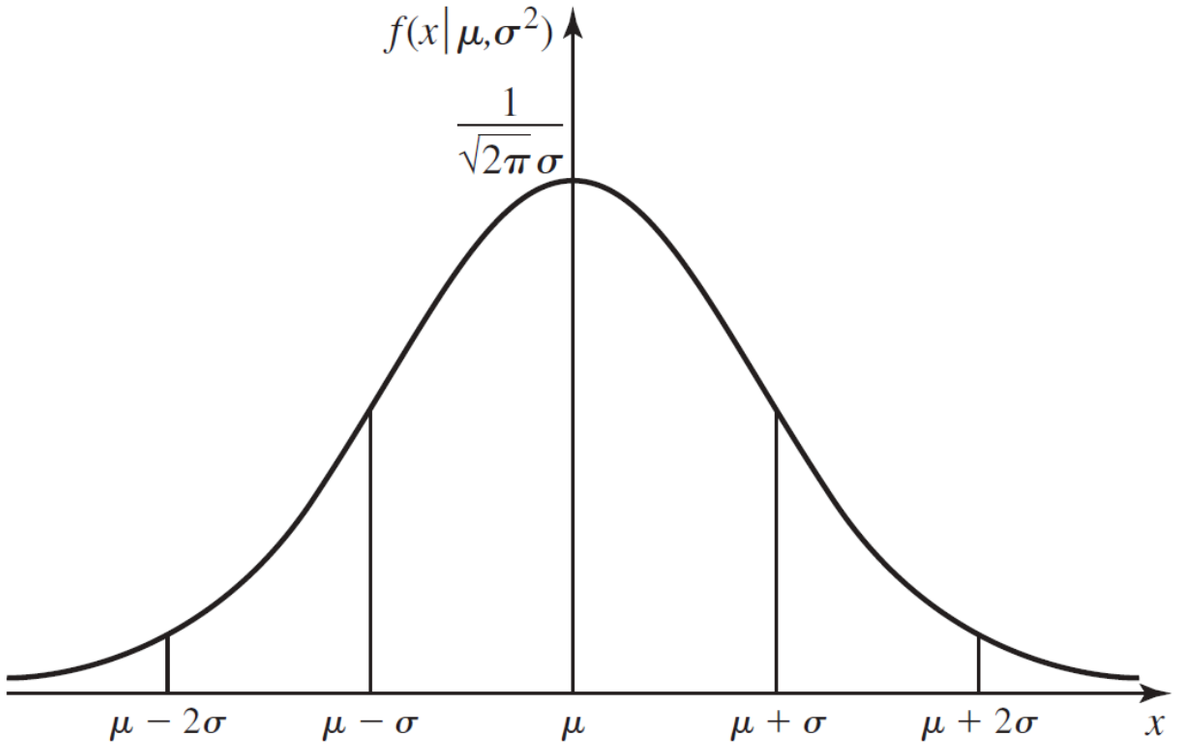
$$I^2 = \iint e^{-(u^2+v^2)/2}\mathrm{d}u\,\mathrm{d}v = \int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r\,\mathrm{d}r\,\mathrm{d}\theta = 2\pi$$

故 $I = \sqrt{2\pi}$，归一化成立。这个积分称为**泊松积分**（与上一讲的泊松分布以同一位数学家命名，但不是一回事）：$e^{-t^2/2}$ 没有初等原函数，一维无法直接积出，化为二重积分后用极坐标即可算出。

参数的含义与归一化系数

指数里是 $-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$：$\mu$ **决定中心位置，**σ **决定曲线的宽窄**。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 $\frac{1}{\sqrt{2\pi}\sigma}$
- 拐点位于 $x=\mu\pm\sigma$：**任一拐点到中心 μ 的水平距离等于 σ**
- 远离中心 μ 时，密度趋于 0

尾部与极端取值的概率

密度曲线两端远离 μ 的区域称为**尾部**。若阈值为 $a=\mu+2\sigma$，则 $P(X>a)$ 等于 a **右侧曲线下的面积**，表示超过该阈值的概率。

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

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

★ 标准化的意义

Φ_0 **没有初等函数表达式**，只能查表或数值计算。有了标准化，**所有正态分布共用一张 Φ_0 表**：任何 $N(\mu, \sigma^2)$ 的概率，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 $\Phi_0(-2)=1-\Phi_0(2)$。

由于密度关于 0 对称，对 $x\geq 0$，$-x$ 左侧与 x 右侧的曲线下面积相等，因此 $\Phi_0(-x)=1-\Phi_0(x)$。

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

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

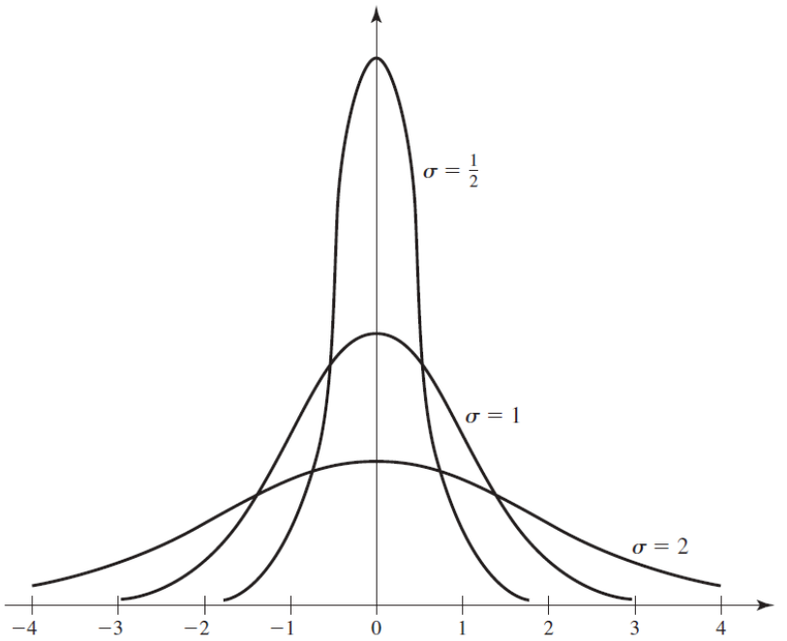
$k\sigma$ 法则：偏离中心不超过 $k\sigma$ 的概率

对任何正态分布

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

对于任意 μ 与 $\sigma > 0$，该概率仅由 k 决定。

k	p_k
1	0.6827
2	0.9545
3	0.9973
5	$1 - 6 \times 10^{-7}$
10	$1 - 1.5 \times 10^{-23}$



不同 μ, σ，同样的 $k\sigma$ 比例

★ 看最后两行

10σ 之外只有 1.5×10^{-23}：每天观测一次，平均约 1.8×10^{20} 年才见到一次，**而宇宙的年龄约 1.4×10^{10} 年。**

回到开场：正态模型下的极端亏损概率

正态模型下的单日亏损概率

若日收益率服从正态分布，单日下跌超过 10σ 的概率约为 7.6×10^{-24}，即第 26 页双侧尾部概率的一半。按每年 252 个交易日估算，对应的平均发生间隔约为 5×10^{20} 年。

实际收益率的「左尾更厚」意味着什么

对相同的亏损阈值，实际收益率低于该值的概率若高于正态模型，就表示实际分布的**左尾概率更大**，也称左尾更厚。比较的是密度曲线左端相应区域的**面积**。

模型假设及其风险含义

以上计算假定**日收益率服从** $N(\mu, \sigma^2)$。如果实际收益率低于同一亏损阈值的概率更大，正态模型就会**低估极端亏损的概率**；使用前应比较实际数据与模型给出的尾部概率。

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」，开场故事里的亏损就落在这 5% 的尾部里。

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，**但剩下 1% 的尾部里损失有多大，VaR 不提供信息**
2. 把 σ 加倍：VaR 也**约加倍**（μ 很小时），风险与所需准备金近似成正比



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

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

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

关键一步：把关于 Y 的事件改写成关于 X 的事件，再用已知的 F_X。

注意 $Y = 100/X$ 是**减函数**（$X > 0$），所以不等号**方向反转**。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（$Z > 0$），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」是连续型最通用的做法。直接把 f 里的 x 换成 $r^{-1}(y)$ 会漏掉一个因子（见下一页）。

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 $0 < y < 1$：

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

求导得 $g(y) = \dfrac{1}{2\sqrt{y}}$ $(0 < y < 1)$。

所得 g 在 $y \to 0^+$ 时趋于无穷，但非负且积分为 1，仍是密度。**密度不等于概率。**

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 $|a|$。

那个 $\frac{1}{|a|}$ 称为**雅可比因子**：取值范围拉伸为 $|a|$ 倍，密度相应缩小为 $\frac{1}{|a|}$，总面积仍为 1。

将此公式用于正态分布，可得到**第 23 页的线性变换结论**：$Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$。

本讲总结

① 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 $P(X=x)=0$，端点开闭不影响概率
- 三个分布：均匀（只知道取值范围时使用）、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）

② 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 $Y=X^2$）

③ 下一讲

到现在为止，我们一次只研究**一个**随机变量。但现实中重要的往往是**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？

开场那个故事还有另一层：**LTCM 的模型并非不知道单个资产的分布，而是低估了它们「一起动」的程度。**

下一讲：多维随机变量及其分布。