随机变量的数字特征

第 7 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

二维随机变量及其函数的分布

先研究联合、边缘、条件分布及独立性，再求由 X, Y 构成的新变量的分布，例如**和、积、最大值与最小值**。

求和的分布时，将 $X + Y \leq z$ 写成平面区域，再用联合分布计算该区域的概率；独立的连续型变量可用卷积求和的密度。

三个常见分布的独立和性质：以下均要求 X, Y 相互独立

分布	两个变量各自的分布	和的分布
二项	$X \sim B(n_1, p)$，$Y \sim B(n_2, p)$	$X + Y \sim B(n_1 + n_2, p)$
泊松	$X \sim \text{Poisson}(\lambda_1)$，$Y \sim \text{Poisson}(\lambda_2)$	$X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$
正态	$X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$	$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

二项分布还要求成功概率相同；此时合并的是两组具有相同成功概率的独立试验。

本讲接着讨论

知道分布后，怎样用几个数描述平均水平、波动大小，以及两个变量的共同变化？

平均延误相同，稳定程度也相同吗？

两个航班的教学示例

设航班甲每次延误10分钟；航班乙有一半概率准点，另一半概率延误20分钟。

按概率计算平均延误

$$\text{甲：}\quad 10 \times 1 = 10$$

$$\text{乙：}\quad 0 \times \frac{1}{2} + 20 \times \frac{1}{2} = 10$$

相同的平均数没有反映全部差别

甲的延误时间固定，乙的延误时间会变化。怎样分别描述平均水平和波动大小？

本讲学习目标

1. 理解**期望**是分布的加权平均，会算常见分布及随机变量函数的期望。
2. 理解**方差与标准差**刻画偏离均值的程度，会算常见分布的方差并运用其性质。
3. 理解**协方差与相关系数**只刻画线性关联，会计算和的方差，并区分独立与不相关。

投资者 A、B 的收益分布不同，谁更好？AI 模型 A、B 的误差分布不同，哪个更优？答案还取决于评价标准。

本讲用几个数概括分布的部分特征，并说明它们的适用边界。

为什么要「数字特征」

用几个数字概括分布的主要特征

- **平均水平：**乘坐这个航班，平均延误多久？
- **偏离平均的程度：**同规格产品的长度在平均值附近如何波动
- **两个量的线性关联：**睡眠时长与学习成绩是否呈线性关联？

先从「平均」开始

某航班过去四天分别延误 $0, 10, 20, 30$ 分钟，平均延误多久？——15 分钟。

换个问法：过去四天里**两天**延误 10 分钟、一天 20 分钟、一天没延误呢？

$$\frac{10+10+20+0}{4} = 10 \times \tfrac{1}{2} + 20 \times \tfrac{1}{4} + 0 \times \tfrac{1}{4} = 10$$

右式是以**观察频率为权重的平均数**。用分布中的概率代替这些频率，就得到期望的计算形式。

期望的定义

定义

离散型：若级数 $\sum_k x_k f(x_k)$ **绝对收敛**，则

$$EX = \sum_{k=1}^{\infty} x_k f(x_k)$$

连续型：若积分 $\int_{-\infty}^{\infty} x f(x)\mathrm{d}x$ **绝对收敛**，则

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

什么叫绝对收敛

把定义中的取值 x 换成 $|x|$，再按概率求和或积分，结果仍须是有限数。

只有有限个可能取值时，这一条件自动满足；有无限多个取值或取值无界时，需要检查。本讲讨论满足这一条件的有限期望。

两个最简单的例子

例（离散）

X 服从 0-1 分布，$P(X=1)=p$。

$$E(X)=0\times(1-p)+1\times p=p$$

伯努利分布的期望就是 p 本身——「平均成功次数」等于「成功概率」。它也可用于计算多个试验中成功总次数的期望。

例（连续）

X 有密度 $f(x)=2x$（$0<x<1$）。

$$E(X)=\int_0^1 x\cdot 2x\,\mathrm{d}x=\frac{2}{3}$$

直觉核对：密度随 x 增大而增大，取值更偏向区间右端，因此期望应**大于**中点 0.5；$\frac{2}{3}$ 与此一致。

二项分布的期望：由定义计算

例

盒中红球比例 p，**有放回**取 n 个，其中 X 个红球。求 EX。

解：利用组合恒等式

$$EX = \sum_{k=0}^{n} k\, C_n^k p^k (1-p)^{n-k}$$

关键一步：$k\, C_n^k = k \cdot \dfrac{n!}{k!(n-k)!} = n \cdot \dfrac{(n-1)!}{(k-1)!(n-k)!} = n\, C_{n-1}^{k-1}$，于是

$$EX = np \sum_{k=1}^{n} C_{n-1}^{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = np \cdot 1 = np$$

末尾那个和式是 $B(n-1, p)$ 的全部概率，等于 1。

第15页将利用期望的可加性给出另一种计算方法。

几何分布的期望：一个错位相减的技巧

$X \sim$ **几何分布**(p)，**求** EX

$$EX = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

两边乘 $(1-p)$ 并把指标平移：

$$(1-p)EX = \sum_{k=1}^{\infty} k(1-p)^{k}p = \sum_{k=1}^{\infty} (k-1)(1-p)^{k-1}p$$

上下两式**相减**，$k-(k-1)=1$：

$$pEX = \sum_{k=1}^{\infty} (1-p)^{k-1}p = 1 \implies EX = \frac{1}{p}$$

结果的含义：成功率 p，平均要试 $1/p$ 次。$p=0.1$ 就平均试 10 次。

以上级数在 $0<p<1$ 时绝对收敛，因此可以逐项相减；$p=1$ 时 $X=1$，结论同样成立。

另外三个分布的期望

泊松 (λ)

$$EX = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} = \lambda$$

$x = 0$ 那项为 0，从 1 开始；提出一个 λ 后剩下的又是全部概率。

指数 (λ)

$$EX = \int_0^{\infty} \lambda x e^{-\lambda x} \mathrm{d}x = \frac{1}{\lambda}$$

分部积分。**速率 λ 的倒数就是平均等待时间**。例如，泊松到达模型中速率为每小时4次，平均等待时间为15分钟。

正态 (μ, σ^2)

$$EX = \mu$$

正态分布有有限期望，由密度关于 μ 对称可得。

小结：六个常见分布的期望

分布	EX	怎么记
伯努利 (p)	p	平均成功次数 = 成功概率
二项 $B(n,p)$	np	做 n 次，每次期望 p
几何 (p)	$1/p$	成功率的倒数
泊松 (λ)	λ	由泊松概率函数计算得到
指数 (λ)	$1/\lambda$	速率的倒数
正态 $N(\mu,\sigma^2)$	μ	对称中心

表中结果可直接用于后续的期望计算；使用前先确认分布及参数含义。

随机变量函数的期望：不必先求分布

定理

令 $Y = r(X)$，若 $E|r(X)| < \infty$，则

$$E[r(X)] = \sum_{\text{所有 } x} r(x) f(x) \qquad \text{或} \qquad E[r(X)] = \int_{-\infty}^{\infty} r(x) f(x)\,\mathrm{d}x$$

直接使用原变量的分布

第5讲介绍了求 $Y = r(X)$ 分布的方法。**但如果只要期望，用 X 的密度直接积就行，不必求 Y 的分布。**

例

$f(x) = 2x$（$0 < x < 1$），求 $E(X^{1/2})$。

$$\int_0^1 x^{1/2} \cdot 2x\,\mathrm{d}x = \frac{4}{5}$$

例

$f(x) = 3x^2$（$0 < x < 1$），求 $E(1/X)$。

$$\int_0^1 \frac{1}{x} \cdot 3x^2\,\mathrm{d}x = \frac{3}{2}$$

二元函数的期望

例

$f(x,y)=1$（$0\le x,y\le 1$），求 $E(X^2+Y^2)$。

$$\int_0^1\int_0^1 (x^2+y^2)\,\mathrm{d}x\,\mathrm{d}y=\frac{1}{3}+\frac{1}{3}=\frac{2}{3}$$

例

$f(x,y)=12y^2$（$0\le y\le x\le 1$），求 $E(XY)$。

$$\int_0^1\int_0^x xy\cdot 12y^2\,\mathrm{d}y\,\mathrm{d}x=\int_0^1 3x^5\,\mathrm{d}x=\frac{1}{2}$$

积分限由三角形取值区域确定：y 从 0 到 x。

若 $E|r(X,Y)|<\infty$，则 $E[r(X,Y)]=\iint r(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$——同样**不必先求 $r(X,Y)$ 的分布**。

期望的三条性质

以下假设 $E|X|$、$E|Y|$ 均有限。

- **线性（对常数）：** $E(aX+c)=aE(X)+c$；特别地 $E(c)=c$、$E(aX)=aE(X)$
- **可加（★ 不需要独立）：** $E(X+Y)=E(X)+E(Y)$
- **独立时可乘：** X,Y 独立 $\Rightarrow E(XY)=E(X)E(Y)$

证明思路

可加性：$E(X+Y)=\iint(x+y)f(x,y)=\iint xf+\iint yf$，前者对 y 积得边缘 f_1，即 $E(X)$，后者同理——**全程没用到独立**。
可乘性：独立时 $f=f_1f_2$，二重积分拆成两个一重积分之积。

两条公式的条件不同

只要期望存在，$E(X+Y)=E(X)+E(Y)$ 就成立，不要求独立。

独立可以推出 $E(XY)=E(X)E(Y)$，反过来不一定；后面会给出反例。

用性质重做两道题

二项分布的期望，第二种解法

$X \sim B(n,p)$ 可写成 n 个独立伯努利之和 $X = \sum_{i=1}^{n} X_i$，每个 $E(X_i) = p$。由**可加性**

$$EX = \sum_{i=1}^{n} E(X_i) = np$$

直接利用可加性，可省去第8页的组合恒等式推导。

配对问题

n 封信随机装进 n 个信封，X 为装对的封数。求 EX。

令 $X_i = 1$ 表示第 i 封装对，则 $P(X_i = 1) = \frac{1}{n}$，$E(X_i) = \frac{1}{n}$：

$$EX = \sum_{i=1}^{n} E(X_i) = n \cdot \frac{1}{n} = 1$$

无论 n 多大，平均装对的封数都是 1。 当 $n \geq 2$ 时，这些 X_i **并不独立** ——可加性不需要独立，这里就用上了。

用独立性拆开期望

例

X_1, X_2, X_3 独立，$E(X_i) = 0$、$E(X_i^2) = 1$。求 $E[X_1^2(X_2 - 4X_3)^2]$。

X_1^2 与 $(X_2 - 4X_3)^2$ 独立，故可乘：

$$= E(X_1^2) \cdot E[(X_2 - 4X_3)^2]$$

$$= 1 \cdot E(X_2^2 - 8X_2X_3 + 16X_3^2) = 1 - 0 + 16 = 17$$

中间项 $E(X_2X_3) = E(X_2)E(X_3) = 0$。

例：两级滤水

第一次滤去 X_1 比例杂质，第二次滤去剩余杂质中的 X_2 比例。 X_1, X_2 独立同分布，$f(x) = 4x^3$（$0 < x < 1$）。 求两次后剩余比例 Y 的期望。

$Y = (1 - X_1)(1 - X_2)$，由独立性

$$E(Y) = \left[1 - E(X_1)\right]^2 = \left(1 - \tfrac{4}{5}\right)^2 = 0.04$$

（$E(X_1) = \int_0^1 x \cdot 4x^3 = \tfrac{4}{5}$。）

两级过滤后的剩余杂质比例的期望为4%；单级过滤后的相应期望为20%。

应用：投资组合的预期收益

场景

组合总收益率 $R_p = \sum_i w_i R_i$（w_i 是各资产权重）。

由期望的线性性

$$E[R_p] = \sum_i w_i E[R_i]$$

组合的预期收益就是各资产预期收益的加权平均——不需要知道资产之间怎么联动。

这还不能说明实际收益有多稳定

这个公式回答了组合平均能获得多少收益。即使平均收益相同，实际收益的波动也可能不同；下一部分再学习怎样描述这种波动。

条件期望

定义

给定 $X = x$，且相应条件分布存在，且在该条件下 $|Y|$ 的期望有限时，Y 的**条件期望**，就是条件分布的期望：

$$E[Y \mid x] = \sum_{\text{所有 } y} y\, g_2(y \mid x) \qquad \text{或} \qquad E[Y \mid x] = \int_{-\infty}^{\infty} y\, g_2(y \mid x)\, \mathrm{d}y$$

注意它是 x 的函数

固定 x 后，$E[Y \mid x]$ 是一个数；随 x 变化，这些数构成一个函数。 第6讲比较的是给定评分后的高分概率；这里进一步考察条件平均值：给定用户对甲片打 5 分，他对乙片的平均打分。

平均数之外，还要描述什么？

回到两个航班

甲的延误始终为10分钟；乙以各1/2的概率延误0分钟或20分钟。两者的期望相同，但甲没有波动，乙会偏离平均值。

接下来的问题

能否用一个非负的数，概括随机变量偏离自身期望的程度？

方差与标准差

先说明一个记号

$E(X^2)$ 是先将 X 平方，再求期望，称为**二阶矩**。它通常不同于 $(EX)^2$。以下假设 $E(X^2) < \infty$。

定义

称 $X - EX$ 为**离差**。离差**平方**的期望称为**方差**：

$$D(X) = E\left[(X - EX)^2\right]$$

离散型 $\sum_x (x - EX)^2 f(x)$，连续型 $\int (x - EX)^2 f(x)\mathrm{d}x$。 其平方根 $\sqrt{D(X)}$ 称为**标准差**。

平方偏差与标准差

离差的期望为0，正负偏差会抵消；平方后不再抵消，也便于推导和的方差。绝对偏差也是可用的离散程度指标。

方差的单位是原单位的平方；标准差与原变量单位相同。

例：两只股票，同样的期望

例

股价 $A \sim U[25, 35]$，$B \sim U[15, 45]$。求各自的期望与方差。

解

$E(A) = E(B) = 30$——**期望完全相同。**

$$D(A) = \int_{25}^{35} (a - 30)^2 \cdot \tfrac{1}{10}\, \mathrm{d}a = \frac{25}{3} \approx 8.33$$

$$D(B) = \int_{15}^{45} (b - 30)^2 \cdot \tfrac{1}{30}\, \mathrm{d}b = 75$$

标准差分别约为 2.89 与 8.66，**后者约为前者的3倍。**

一般地，$U[a, b]$ 的方差是 $\dfrac{(b-a)^2}{12}$——区间宽度变为原来的3倍，方差变为原来的9倍。**本例均值相同，方差与标准差反映了取值范围不同所带来的波动差异。**

方差的三条性质

以下假设 X, Y 均有有限二阶矩。

- **线性变换：** $D(aX+c) = a^2 D(X)$；特别地 $D(c) = 0$、$D(aX) = a^2 D(X)$
- **独立时可加：** X, Y 独立 $\Rightarrow D(X+Y) = D(X) + D(Y)$
- **化简公式（最常用）：** $D(X) = E(X^2) - (EX)^2$

证明思路

化简公式：$E[(X-EX)^2] = E[X^2 - 2X\,EX + (EX)^2] = E(X^2) - 2(EX)^2 + (EX)^2$。

线性变换：$D(aX+c) = E[(aX+c-aEX-c)^2] = a^2 E[(X-EX)^2]$——**常数 c 被消掉了**， 平移不改变离散程度。

为什么方差乘以 a^2?

线性变换后，变量与自身期望的偏差为 $aX+c-E(aX+c) = a(X-EX)$。方差是**偏差平方的期望**，因此乘以 a^2；标准差则乘以 $|a|$。

例如，$D(2X+3) = 4D(X)$，标准差变为原来的2倍；加上常数3不改变波动大小。

几何分布与二项分布的方差

几何分布：计算二阶矩

$$(1-p)E(X^2)=\sum_{k\geq 1}(k-1)^2(1-p)^{k-1}p$$

相减，$k^2-(k-1)^2=2k-1$：

$$pE(X^2)=E(2X-1)=\frac{2}{p}-1$$

$$D(X)=E(X^2)-(EX)^2=\frac{2-p}{p^2}-\frac{1}{p^2}=\frac{1-p}{p^2}$$

二项分布：两种解法

先计算 $E[X(X-1)]$：

$$E(X^2)=E[X(X-1)]+EX=n(n-1)p^2+np$$

$$D(X)=n^2p^2-np^2+np-n^2p^2=np(1-p)$$

利用独立性： $X=\sum X_i$ 独立，每个 $D(X_i)=p(1-p)$，由可加性

$$D(X)=np(1-p)$$

另外三个分布的方差

泊松 (λ)

$$E[X(X-1)] = \lambda^2 \sum_{x \geq 2} \frac{e^{-\lambda}\lambda^{x-2}}{(x-2)!} = \lambda^2$$

$$D(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$$

期望与方差均为 λ。

指数 (λ)

$$D(X) = \frac{1}{\lambda^2}$$

标准差与期望均为 $1/\lambda$。

正态 (μ, σ^2)

$$D(X) = \sigma^2$$

$N(\mu, \sigma^2)$ 的第二个参数为方差；标准差为 σ。

应用：测量误差的平均大小

误差、平方误差与均方误差

设物体的真实长度为常数 c，测量结果为随机变量 Y。误差为 $Y-c$，平方误差为 $(Y-c)^2$。平方误差的期望称为**均方误差**。

与方差的关系

若 $E(Y^2)<\infty$，则

$$E[(Y-c)^2]=D(Y)+(EY-c)^2$$

方差描述测量值围绕自身均值的波动；后一项描述平均测量值与真实值的偏离。

例

真实长度 $c=100$ 毫米，$EY=100.1$ 毫米，$D(Y)=0.04$ 平方毫米，则

$$E[(Y-c)^2]=0.04+(100.1-100)^2=0.05 \text{ 平方毫米}$$

小结：期望与方差对照表

分布	EX	$D(X)$
伯努利 (p)	p	$p(1-p)$
二项 $B(n,p)$	np	$np(1-p)$
几何 (p)	$1/p$	$(1-p)/p^2$
泊松 (λ)	λ	λ
均匀 $U[a,b]$	$(a+b)/2$	$(b-a)^2/12$
指数 (λ)	$1/\lambda$	$1/\lambda^2$
正态 $N(\mu,\sigma^2)$	μ	σ^2

比较各分布的参数与数字特征，计算时确认变量的取值范围及参数约定。

协方差：两个离差乘积的期望

动机

怎样描述两个变量偏离各自平均值时的共同变化？ 例如，数学成绩较高时，语文成绩是否也倾向于较高？

定义

设 $E(X^2)$ 与 $E(Y^2)$ 均有限；这一条件也称为具有有限二阶矩，见第20页。

$$\operatorname{Cov}(X,Y) = E\big[(X-EX)(Y-EY)\big]$$

展开后得到常用的计算式：

$$\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

与期望乘法公式的关系

协方差等于 $E(XY)$ 与 $E(X)E(Y)$ 的差。独立时差为0；不独立时，差也可能为0。

同号离差产生正的乘积，异号离差产生负的乘积；协方差的符号取决于这些乘积按概率加权后的结果。

协方差的六条性质

以下随机变量均有有限二阶矩

1. $\mathrm{Cov}(X, X) = D(X)$
2. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
3. $E(XY) = E(X)E(Y) + \mathrm{Cov}(X, Y)$
4. $D(X + Y) = D(X) + D(Y) + 2\mathrm{Cov}(X, Y)$
5. $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$
6. $\mathrm{Cov}(X, aY + b) = a\,\mathrm{Cov}(X, Y)$

不独立时，两个变量之和的方差怎样计算?

第4条给出一般答案：除了两个方差之和，还要计入两倍的协方差。当协方差为0时，两个方差仍可直接相加。

例：计算两个变量的协方差

联合密度

$f(x,y)=2xy+0.5$（$0\le x,y\le 1$），其他位置为0。求 $\mathrm{Cov}(X,Y)$。

先求各自期望，再求乘积期望

$$E(X)=E(Y)=\int_0^1\int_0^1 x(2xy+0.5)\,\mathrm{d}y\,\mathrm{d}x=\frac{7}{12}$$

$$E(XY)=\int_0^1\int_0^1 xy(2xy+0.5)\,\mathrm{d}y\,\mathrm{d}x=\frac{25}{72}$$

代入协方差的计算式

$$\mathrm{Cov}(X,Y)=\frac{25}{72}-\left(\frac{7}{12}\right)^2=\frac{1}{144}$$

协方差为正。其大小受计量单位影响，后面用相关系数比较线性关联的强弱。

应用：协方差如何影响组合波动

双资产组合

设 $R_p = w_A R_A + w_B R_B$，权重固定且 $w_A, w_B \geq 0$、$w_A + w_B = 1$。

$$D(R_p) = w_A^2 D(R_A) + w_B^2 D(R_B) + 2 w_A w_B \mathrm{Cov}(R_A, R_B)$$

保持其他条件不变时

资产的预期收益、方差与权重给定后，协方差降低不会增加组合方差；两项权重均为正时，组合方差严格降低。组合预期收益保持不变。

比较的是哪一种风险

负协方差使交叉项非正，组合方差不超过两个加权方差项之和。这既不保证组合方差小于任一资产的方差，也不保证组合实际收益优于任一单资产。

交互：相关系数如何改变风险—收益曲线

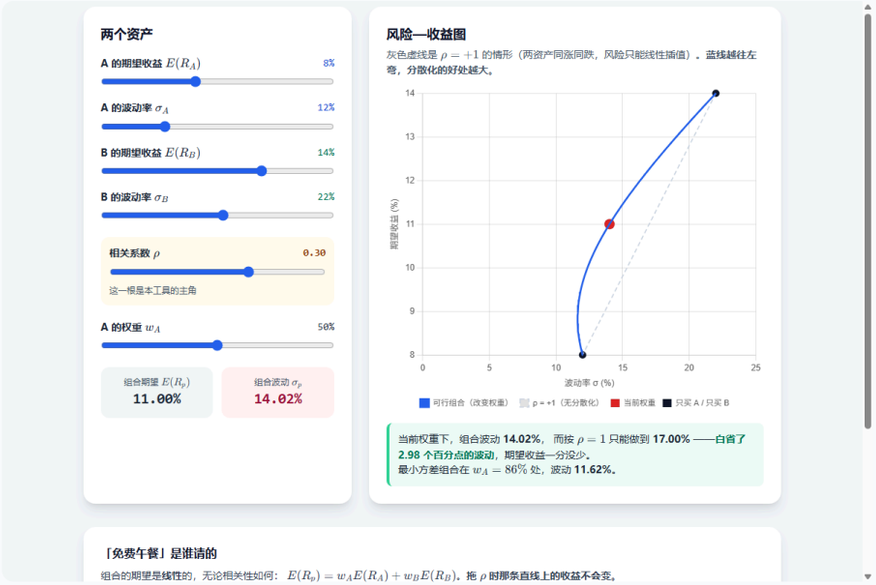
怎么用

两个资产的期望收益与波动率可调，关键是**相关系数**滑块。

固定两项资产的预期收益、波动率和权重范围，比较三种情形

先将相关系数理解为两资产线性关联的标准化指标，定义见下一页；此处权重在0至1之间。

1. 拖 ρ，盯住**纵轴**——期望收益**完全不变**
2. ρ 拖到 **+1**：曲线塌成直线，组合标准差等于两资产标准差的加权平均
3. ρ 拖到 **−1**：曲线碰到纵轴，存在一个权重让波动**恰好为 0**



交互演示：拖动相关系数，观察风险—收益曲线如何弯折

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

相关系数：把协方差标准化

协方差的一个问题：量纲

$\mathrm{Cov}(2X, Y) = 2\mathrm{Cov}(X, Y)$——**换个单位，数就变了。** 用元还是用万元，算出来差一万倍，无法横向比较。

定义相关系数

要求 $0 < D(X), D(Y) < \infty$。

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

可以证明 $|\rho| \le 1$。特别地：

- $|\rho| = 1 \iff X, Y$ **线性**相关（以概率1满足非零斜率的线性关系）
- $\rho = 0$ 称为 X, Y **不相关**
- **独立** $\Rightarrow$ **不相关，反之不成立**（下下页给反例）

$|\rho| \le 1$ 的证明思路

对任意实数 t，$E\left\{\left[(X - EX)t + (Y - EY)\right]^2\right\} \ge 0$。 展开是关于 t 的二次式 $D(X)t^2 + 2\mathrm{Cov}\, t + D(Y) \ge 0$， 故判别式 $4\mathrm{Cov}^2 - 4D(X)D(Y) \le 0$，即 $|\rho| \le 1$。 等号成立当且仅当存在一个 t 使该平方以概率1等于0，即 Y 是 X 的线性函数。

应用：改变单位后，相关系数不变

同一个量，使用不同单位

以万元记录 X，改为以元记录 $X^* = 10000X$；保持 Y 的单位不变。

$$\mathrm{Cov}(X^*, Y) = 10000\mathrm{Cov}(X, Y)$$

标准化抵消单位变化

$$\rho(X^*, Y) = \frac{10000\mathrm{Cov}(X, Y)}{10000\sigma_X\sigma_Y} = \rho(X, Y)$$

正的单位换算系数不会改变相关系数。协方差的数值变大，并不意味着两个变量的关联变强。

不相关不能推出独立

反例

X 等可能取 $-1, 0, 1$，令 $Y = X^2$。

不独立，例如 $P(Y = 0 \mid X = 0) = 1 \neq P(Y = 0) = 1/3$
——知道 X 就完全知道 Y。

但算协方差：$E(X) = 0$，

$$\mathrm{Cov}(X, Y) = E(X^3) - E(X)E(X^2) = 0 - 0 = 0$$

不相关。

非线性关系中的零相关

ρ 只度量**线性**关系。Y 由 X 完全确定，但 $Y = X^2$ 是非线性关系；正、负取值在协方差计算中的贡献恰好抵消。

$\rho = 0$ **只表示没有线性相关，不表示「没有关系」。**

完全线性相关（$|\rho| = 1$）

若 $Y = aX + b$，$a \neq 0$ 且 $0 < D(X) < \infty$，则 $\mathrm{Cov} = aD(X)$、$D(Y) = a^2 D(X)$，

$$\rho = \frac{aD(X)}{|a|D(X)} = \mathrm{sgn}(a)$$

当 $a > 0$ 时，$\rho = 1$。当 $a < 0$ 时，$\rho = -1$。

二元正态分布

定义

这里讨论 $\sigma_1, \sigma_2 > 0$、$|\rho| < 1$ 的非退化二元正态分布，其联合密度为

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

三条性质

- 边缘分布分别是 $X \sim N(\mu_1, \sigma_1^2)$、$Y \sim N(\mu_2, \sigma_2^2)$
- 式子里那个 ρ 就是相关系数 $\rho(X,Y)$
- $\rho = 0$ **是** X, Y **独立的充分必要条件**

关键前提是联合正态

在联合分布为二元正态这一条件下，不相关与独立等价。仅知道两个边缘分布都为正态，不能应用这一结论。

n 维正态分布（选讲）

协方差矩阵正定时的密度

以下密度公式要求协方差矩阵正定；退化正态分布可能没有这种密度。

记 $\boldsymbol{X}, \boldsymbol{\mu}$ 为 n 维列向量，$\boldsymbol{C}$ 为**协方差矩阵** （$c_{ij} = \mathrm{Cov}(X_i, X_j)$），以 $\boldsymbol{x}$ 表示取值，则

$$f(x_1, \ldots, x_n) = (2\pi)^{-n/2} |\boldsymbol{C}|^{-1/2} \exp\left\{-\tfrac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

允许退化情形时的四条性质

- 每个分量都是正态；反之若各分量正态**且相互独立**，则整体是 n 维正态
- 任意**线性组合**是一维正态，反之亦然
- 线性函数 $(Y_1, \ldots, Y_k)$ 仍是多维正态
- **相互独立** $\Longleftrightarrow$ **两两不相关**

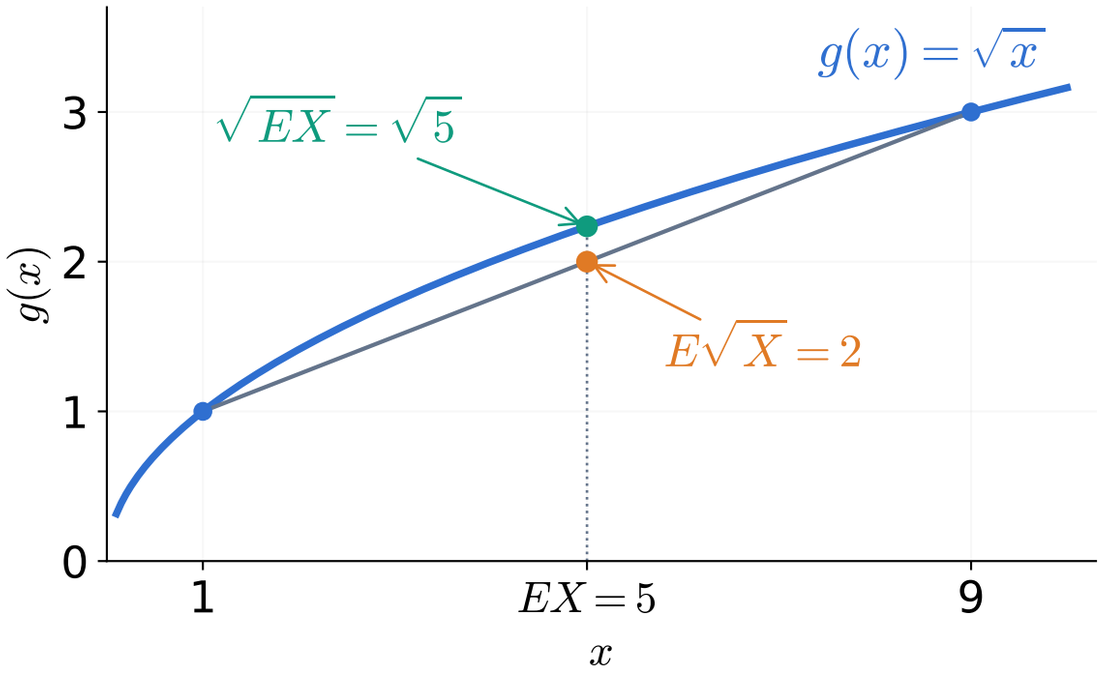
注意第一条的「**且相互独立**」不能去掉： 各分量都正态，不足以保证整体是多维正态。 第四条是二维那条性质的推广，同样依赖联合正态这一前提。

琴生不等式：非线性变换与平均值（选讲）

若 g 在 X 取值所在的区间上为凹函数，且 EX 与 $E[g(X)]$ 均有限，则

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

先变换再求平均，不超过先求平均再变换的结果。若 g 为凸函数，不等号反向。



凹函数的图像不低于任意两点间的连线。

应用：比较稳定投入与波动投入的产出

设某生产过程的投入为 x，产出为 $g(x) = \sqrt{x}$（教学模型）。

投入以相同概率取1或9，平均投入为5，平均产出为 $\frac{1+3}{2} = 2$；稳定投入5时，产出为 $\sqrt{5} \approx 2.24$。

在这个模型中，平均投入相同，稳定投入的平均产出更高。

柯西—施瓦茨不等式：共同变化的限度（选讲）

若 $E(X^2)$、$E(Y^2)$ 均有限，则

$$\big[E(XY)\big]^2 \le E(X^2)E(Y^2).$$

直观含义：共同变化受各自波动大小限制

将 X, Y 换为各自的离差，得到

$$|\mathrm{Cov}(X,Y)| \le \sigma_X \sigma_Y.$$

协方差的绝对值不超过两者标准差的乘积。除以这个乘积后，相关系数满足 $|\rho| \le 1$（两者方差均非零时）。

$|\rho| = 1$ 对应两个变量之间完全确定的线性关系。

应用：估计总测量误差的方差

两段长度的测量误差 X, Y 单位相同，标准差分别为2、3，则

$$-6 \le \mathrm{Cov}(X,Y) \le 6.$$

$$D(X+Y) = 4 + 9 + 2\mathrm{Cov}(X,Y),$$

$$\boxed{1 \le D(X+Y) \le 25}.$$

即使不知道两个误差之间的相关性，仍能给出总误差方差的上下界。

本讲总结：数字特征的含义与边界

平均水平与离散程度

期望概括平均水平，方差概括偏离自身期望的程度。相同的期望或方差可以对应不同的分布。

共同变化与线性关联

协方差参与和的方差计算；相关系数消除了单位影响。独立推出不相关，反向结论需要额外条件。

先检查定义与前提

使用有限期望、方差和相关系数前，分别确认可积性、有限二阶矩及非零方差。数字特征不能替代完整分布，也不能单独确定决策。

本讲总结（续）

② 核心结论

- **期望**：$E(X+Y)=E(X)+E(Y)$ 不需要独立；独立时有 $E(XY)=E(X)E(Y)$，反向不一定成立
- **方差**：$D(X)=E(X^2)-(EX)^2$；$D(aX+c)=a^2D(X)$；$D(X+Y)=D(X)+D(Y)+2\mathrm{Cov}(X,Y)$
- **协方差**就是 $E(XY)-E(X)E(Y)$；独立时该值为0
- **相关系数** $\rho=\mathrm{Cov}/(\sigma_X\sigma_Y)$，$|\rho|\le 1$，只度量线性关系

下一讲：许多变量的平均

当独立同分布变量的个数增加时，样本平均为何趋近总体均值？它在均值附近的随机波动又呈现什么分布？

大数定律讨论前一个问题；中心极限定理讨论适当标准化后的波动。两者是不同的结论。