

复习与总结

第 12 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

把判断变成了一套有明确出错概率的程序

- H_0 受保护：「不拒绝」不是「接受」
- 两类错误： α 由我们选， β 取决于真值距离与样本量
- 四种检验： z 、 t 、 χ^2 、 F ；比均值之前先用 F 查方差

★ 留下的问题

十三讲讲完了。可如果现在问一句：「**这门课到底在讲什么？**」——把随机事件、随机变量、期望方差、大数定律、抽样分布、估计、检验一路列下来，那只是目录，不是答案。

它们之间是怎么接上的？这是最后一次课要回答的事。

为什么需要这一讲

场景 期末复习，一位很用功的同学把十三讲的公式抄成了一张 A4 纸：加法公式、贝叶斯、期望方差的性质、六七个分布的 EX 与 DX 、切比雪夫、中心极限、四个抽样分布、两种估计、四种检验。**密密麻麻，一个不漏。**

冲突 考场上翻到一道题：「某厂抽了 16 件产品，样本均值 2.125，样本标准差 0.017，问长度的波动是否符合规定。」这位同学在纸上找了三分钟——**公式全在，却不知道该用哪一个。**问题不在记性。**一张按讲次排列的公式表，回答不了「该用哪个」——因为它是按我们讲的顺序排的，不是按你要解决的问题排的。**

悬念 那该按什么顺序排？这十三讲其实只做了两件事，方向正好相反。找到那个转折点，整张表就会自己立起来。

本讲安排

- ① **先立骨架**：两个方向、一个接口，把十三讲接成一张图
- ② **再过内容**：八个章节的要点与易错处，按考点密度取舍
- ③ **最后给工具**：一张「该用哪个统计量」的决策树，和答疑

关于考试

- 题型：**选择题、计算题、证明题**
- **可以使用计算器**

三种题型对应三种准备：选择题考概念边界（「不相关能否推出独立」这类），计算题考流程熟练度，证明题考的是**定义会不会用**——绝大多数证明题的第一步就是把定义写出来。

骨架：两个方向，一个接口

整门课只做了两件事

由模型推数据（绪论 – 第 8 讲）

给定分布与参数，算出事件的概率、随机变量的期望方差、以及样本量趋于无穷时的极限行为。

终点是**大数定律与中心极限定理**——它们回答了「样本能有多接近总体」。

由数据推模型（第 9 – 11 讲）

手上只有样本，参数才是要求的東西。

起点是**抽样分布**——它把上半场那些定理，变成了四个可以查表的统计量。

★ 两个方向在哪一点上对接

就在这一个式子上：

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

上半场用它预测样本会落在哪；下半场把它反过来，从样本反推 μ 在哪。

χ^2 、 t 、 F 都是它的衍生品： σ 不知道就用 S 顶替（ $\rightarrow t$ ），要推断 σ^2 本身就看残差平方和（ $\rightarrow \chi^2$ ），要比较两个方差就相除（ $\rightarrow F$ ）。

第一章 随机事件及其概率

要点

- 事件的并、交、差；互斥、对立；**对偶律**
- 广义加法： $P(A + B) = P(A) + P(B) - P(AB)$
- 古典概型与几何概型：**排列数** A_n^k 、**组合数** C_n^k
- 条件概率 $P(B | A) = \frac{P(AB)}{P(A)}$ 与乘法公式
- 全概率： $P(B) = \sum_i P(B | A_i)P(A_i)$
- 贝叶斯： $P(A_m | B) = \frac{P(A_m)P(B | A_m)}{\sum_i P(A_i)P(B | A_i)}$
- 独立性： $P(AB) = P(A)P(B)$

三个高频失分点

1. **互斥与独立混为一谈。** 互斥是「不能同时发生」 ($AB = \emptyset$)，独立是「一个发生不改变另一个的概率」。两个概率为正的事件**互斥就一定不独立**。
2. **古典概型里有序无序不统一。** 分子分母必须用同一种数法——一个用排列一个用组合，结果必错。
3. **贝叶斯题不先算分母。** 分母是全概率公式，把所有可能的原因加一遍；漏掉一种原因，后面全错。

第二章 一维随机变量及其分布

要点

- **分布函数** $F(x) = P(X \leq x)$: 非降;
 $\lim_{x \rightarrow -\infty} F = 0$ 、 $\lim_{x \rightarrow +\infty} F = 1$
- **离散型**: 概率函数 $f(x) = P(X = x)$
- **连续型**: 密度 $f(x) = F'(x)$
- 都要满足**归一化条件**
- **随机变量函数的分布**
 - 离散型: 枚举所有可能取值, 把概率并起来
 - 连续型: **先求分布函数, 再求导**

难点: 定义域

求 $Y = g(X)$ 的分布, **八成的错都出在定义域上**。

做法固定: 先写 $F_Y(y) = P(g(X) \leq y)$, 把 $g(X) \leq y$ **解成关于 X 的区间**, 再与 X 本身的取值范围**取交集**, 最后求导。

g 不单调时 (比如 $Y = X^2$) 要**分段讨论**, 并注意 $y < 0$ 时 $F_Y(y) = 0$ 。

密度不是概率: $f(x)$ 可以大于 1, $P(X = x) = 0$ 对连续型恒成立。 **面积才是概率**。

第三章 多维随机变量及其分布

要点

- 联合分布 $f(x, y)$
- 边缘分布 $f_1(x), f_2(y)$ ——**对另一个变量求和 / 积分**
- 条件分布 $g_1(x | y), g_2(y | x)$ ——**除以边缘, 重新归一化**
- 独立性: $f(x, y) = f_1(x)f_2(y)$ **对一切 (x, y) 成立**
- 两个随机变量函数的分布 (难点): $Z = X + Y$ 、 $Z = XY$ 、 $M = \max\{X, Y\}$ 、 $N = \min\{X, Y\}$

两个坑

1. **定义域 (又是它)**。求边缘时积分限依赖于另一个变量; 把 $\int$ 的上下限写成常数是最常见的错。
2. **边缘分布决定不了联合分布**。边缘一模一样的两个联合分布, 描述的世界可以完全不同 (独立 / 完全相依 / 完全负相依)。**反过来才成立**。

max 与 min 的固定套路

$$P(M \leq z) = P(X \leq z)P(Y \leq z) \text{ (都不超过)}$$

$$P(N > z) = P(X > z)P(Y > z) \text{ (都超过)}$$

第 10 讲那道「 $n \min(X_i)$ 是无偏估计」的题, 用的就是第二条。

第四章 随机变量的数字特征

期望

离散 $EX = \sum_k x_k f(x_k)$; 连续 $EX = \int_{-\infty}^{\infty} x f(x) dx$ 。
存在性要求级数或积分绝对收敛 (第 7 讲那个圣彼得堡悖论)。

函数的期望不必先求分布: $E[r(X)] = \sum r(x) f(x)$ 或 $\int r(x) f(x) dx$ 。

- $E(aX + c) = aE(X) + c$
- $E(X + Y) = E(X) + E(Y)$ ——**不需要独立**
- X, Y 不相关 $\Rightarrow E(XY) = E(X)E(Y)$

条件期望: $E[Y | x] = \sum_y y g_2(y | x)$ 或 $\int y g_2(y | x) dy$ 。

方差与协方差

- $D(aX + c) = a^2 D(X)$ ——**常数不影响方差**
- $D(X) = EX^2 - (EX)^2$ (算方差的常规工具)
- $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$
- $\text{Cov}(X, aY + b) = a \text{Cov}(X, Y)$
- $D(X + Y) = D(X) + D(Y) + 2\text{Cov}(X, Y)$
- $\rho = \frac{\text{Cov}(X, Y)}{\sqrt{DX}\sqrt{DY}} \in [-1, 1]$; $\rho = 0$ 不相关, $|\rho| = 1$ 完全线性相关

★ 选择题最爱考的一条

不相关 ($\text{Cov} = 0$) 是独立的必要非充分条件。 独立一定不相关, 不相关不一定独立——**二元正态是唯一的例外**, 它那里不相关就等价于独立。

几种重要的分布

离散型

分布	f	EX	DX
几何	$(1 - p)^{x-1}p$	$1/p$	$(1 - p)/p^2$
二项	$C_n^k p^k (1 - p)^{n-k}$	np	$np(1 - p)$
泊松	$\frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ

可加性： p 相同的独立二项之和仍是二项（ n 累加）； 独立泊松之和仍是泊松（ λ 累加）。

连续型

分布	EX	DX
均匀 $U(a, b)$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
指数 λ	$1/\lambda$	$1/\lambda^2$
$\Gamma(\alpha, \beta)$	α/β	α/β^2
正态 $N(\mu, \sigma^2)$	μ	σ^2

可加性： k 个独立指数之和是 $\Gamma(k, \lambda)$ ； β 相同的独立 Γ 之和仍是 Γ （ α 累加）。 $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ 。

正态分布：把它单独拎出来

五条性质，后半门课全靠它们

1. $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $EX = \mu$, $DX = \sigma^2$
2. **线性变换还是正态**: $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$
3. **独立正态的线性组合还是正态**: $Z = aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2)$
4. **标准化**: $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$; 查表用 $\Phi_0(-x) = 1 - \Phi_0(x)$
5. $N(0, 1)$ 的平方是 $\Gamma(\frac{1}{2}, \frac{1}{2})$, 也就是 $\chi^2(1)$

另有**二元正态分布**：它是「不相关 $\iff$ 独立」的唯一常见例外。

★ 第 5 条就是那个转折点

第 3 条给出 $\bar{X} \sim N(\mu, \sigma^2/n)$ ——上半场的终点。

第 5 条把正态平方成 χ^2 , k 个一加就是 $\chi^2(k)$ ——下半场的起点。

这两条之间那一步，就是整门课的接缝。 后面的 t （正态除以 χ^2 的根）与 F （两个 χ^2 相比），都由它长出来。

第五章 大数定律与中心极限定理

切比雪夫不等式

$$P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}$$
$$P(|X - EX| < \varepsilon) \geq 1 - \frac{DX}{\varepsilon^2}$$

只需要期望与方差，不需要知道分布——这是它的全部价值，代价是界很松。

大数定律

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n EX_i\right| < \varepsilon\right) = 1$$

中心极限定理

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n EX_i}{\sqrt{\sum_{i=1}^n DX_i}} \leq x\right) = \Phi_0(x)$$

不管原分布是什么，标准化之后都趋于标准正态。

证明题会怎么出

「如何证明大数定律？」——标准答案是用**切比雪夫不等式**：对 $\bar{X}_n$ 用一次， $D(\bar{X}_n) = \sigma^2/n \rightarrow 0$ ，右边的界就压到 0。

这是本课最典型的证明题：**把定义写出来，套一个已知不等式。**

两者的分工：大数定律说 $\bar{X}_n$ 会收敛到 μ ；中心极限定理说它**以多快的速度、按什么形状**收敛。

第六至八章 样本分布 · 参数估计 · 假设检验

四个抽样分布

- $\bar{X} \sim N(\mu, \sigma^2/n)$; $\chi^2(k)$ 是 k 个独立标准正态平方相加 ($EX = k$ 、 $DX = 2k$)
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$; $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$
- $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

参数估计

- 标准：一致、无偏、有效；点估计：矩估计、最大似然估计
- μ 的区间： σ 已知或大样本用 $u_{\alpha/2}$ ， σ 未知小样本用 $t_{\alpha/2}(n-1)$
- σ^2 : $\left(\frac{(n-1)S^2}{\chi_{\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2} \right)$

假设检验：一个正态总体

- σ^2 已知，检验 $H_0: \mu = \mu_0 / \mu \leq \mu_0 / \mu \geq \mu_0$ ——用 z
- σ^2 未知，同上三种 ——用 t
- μ 未知，检验 $H_0: \sigma^2 = \sigma_0^2 / \leq / \geq$ ——用 χ^2

假设检验：两个正态总体

- μ_1, μ_2 未知，检验 σ_1^2 与 σ_2^2 ——用 F
- 方差已知，检验 μ_1 与 μ_2 ——用 z ；方差未知但相等 ——用 $t(n_1 + n_2 - 2)$

三条线：区间估计与检验用**同一批枢轴量**；单侧把 α 全押一边；**比均值前先用 F 查方差**。

交互：知识地图与「该用哪个统计量」

第一页：知识地图

十三讲排成两列，左边「由模型推数据」，右边「由数据推模型」，中间那根弯箭头就是**方向反转**。点任意一格，看这一讲的要点，以及**它从哪里来、往哪里去**。

第二页：该用哪个统计量

回答三四个问题（估计还是检验？期望还是方差？ σ 已知吗？），它给出**该用的那个式子，以及查表时的注意事项**。

这就是开场那位同学缺的东西—— 一张按「你要解决什么问题」排列的表，而不是按讲次排列的表。

① 知识地图

点任意一格，右边显示这一讲的要点与它前后的接口。

概率论 • 由模型推数据

L00 绪论

L01 随机事件与概率定义

L02 古典概率与几何概率

L03 条件概率与独立性

L04 离散型随机变量

L05 连续型随机变量

L06 多维随机变量

L07 数字特征

L08 大数定律与中心极限定理

数理统计 • 由数据推模型

L09 样本及抽样分布

L10 参数估计

L11 假设检验

L12 复习与总结

方向反转

② 该用哪个统计量

L09 • 样本及抽样分布

从哪里来：方向反转：由数据推模型

- 简单随机样本 = 独立同分布（一切的前提）
- S^2 分母是 $n - 1$ ：自由度
- $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1); \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n - 1)$
- $\frac{(n - 1)S^2}{\sigma^2} \sim \chi^2(n - 1); \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F$

往哪里去：有了统计量，但还没真正「估计」过任何东西

整门课其实只有两个方向

- 由模型推数据（绪论 - 第 8 讲）：给定分布与参数，算出各种事件的概率、期望、极限行为。终点是大数定律与中心极限定理——它们说明了「样本能有多接近总体」。
- 由数据推模型（第 9 - 11 讲）：手上只有样本，反推参数。起点是抽样分布——它把上半场那些定理，变成了四个可以查表的统计量。

交互演示：十三讲的知识地图，以及「该用哪个统计量」决策树

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

课程总结

① 回到那张 A4 纸

公式一个不漏，却不知道用哪一个——问题不在记性，在于那张表是**按我们讲的顺序**排的。

换个排法：先问「你要估计，还是要判断」，再问「关于期望，还是关于方差」，再问「 σ 知不知道」。**三个问题就能定位到唯一的那个式子。** 十三讲的内容没有变少，只是换了一个能用的次序。

② 这门课真正教了什么

从 1654 年帕斯卡与费马为「分赌本」通信，到 1935 年费希尔用八杯茶 定义什么叫「有证据」——中间三百年，人类学会的是同一件事：

在没法确定的地方，仍然可以说出有分量的话； 并且能讲清楚，这句话有多大把握、会以多大概率出错。

这不只是考试内容。之后无论做研究、做投资、做产品，面对数据时最该问的永远是那三句：**这批数据是怎么来的？这个结论的不确定性有多大？我一共试了多少次？**

③ 接下来

本课程的自然延伸是**计量经济学**：把第 10 讲的最大似然与最小二乘，从「估一个 μ 」推广到「估一条 $C = a + bY$ 」。

祝考试顺利。