概率统计

期末速通版本

作者：WHU-cs 2024 级祝天赐
邮箱 zhutianci@whu.edu.cn

ElegantLATEX Program

目录

第 1 章 核心分布：直觉与本质

内容提要

- ❑ 本章我们不搞题海战术，而是要建立对概率分布的“物理直觉”。为什么是这三个分布？
- ❑ 泊松分布：描述“单位时间内来了多少人/发了多少次故障”。
- ❑ 指数分布：描述“还要等多久下一个人才来”。
- ❑ 正态分布：描述“大量微小误差叠加后的总结果”。

1.1 泊松分布 (Poisson Distribution)

笔记 [核心直觉：稀有事件的计数] 想象你在盯着学校门口，假设学生到达是随机的。泊松分布关心的是：在一个固定的时间段内（比如 **1** 小时），到底来了几个人？

它常用于描述稀有事件的发生次数，比如：一页书上的错别字数、一小时内收到的微信数、服务器一分钟内的访问量。

定义 1.1 (公式与参数)

若随机变量 $X \sim P(\lambda)$，则其分布律为：

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

- 含义：k 是发生的次数。
- 参数 λ：代表该时间段内事件发生的平均速率（强度）。如果平均每小时来 5 个人，那 $\lambda = 5$。

♣

1.2 泊松分布：二项分布的极限 ⋆

定义 1.2 (泊松分布的物理模型)

假设我们观察一个固定的时间段 T，该时段内事件发生的平均次数为 λ。为了研究事件发生的具体概率，我们采用“无限分割法”：

1. 将时间段 T 等分成 n 个极小的时间片，当 $n \to \infty$ 时，每个时间片 $\Delta t \to 0$。
2. 假设 **1**：在极小的时间片内，事件要么发生 1 次，要么不发生（发生 2 次及以上的概率是高阶无穷小，忽略不计）。
3. 假设 **2**：每个时间片内发生事件的概率为 p_n。由于总平均次数为 λ，故 $n \cdot p_n = \lambda$，即每次试验成功的概率 $p_n = \frac{\lambda}{n}$。
4. 各个时间片之间是相互独立的（伯努利试验序列）。

♣

定理 1.1 (泊松极限定理)

在上述假设下，二项分布 $B(n, \frac{\lambda}{n})$ 当 $n \to \infty$ 时的极限即为泊松分布。

♡

证明 设 X 为 n 个时间片中事件发生的总次数，则 X 服从二项分布：

$$P(X=k) = C_n^k p^k (1-p)^{n-k}, \quad 其中 p = \frac{\lambda}{n}$$

展开组合数并代入 p：

$$P(X=k) = \frac{n(n-1)\cdots(n-k+1)}{k!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

为了求极限，我们将式子重新分组：

$$P(X=k)=\frac{\lambda^k}{k!}\cdot\underbrace{\frac{n(n-1)\cdots(n-k+1)}{n^k}}_{\text{第一部分}}\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\text{第二部分}}\cdot\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\text{第三部分}}$$

现在令 $n\to\infty$，我们逐项分析极限：

- 第一部分：分子有 k 项，分母是 n^k。当 n 很大时，$(n-1)\approx n$，故：

$$\lim_{n\to\infty}\frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-k+1}{n}=1\cdot1\cdots1=1$$

- 第二部分：这是重要极限的定义形式 $\lim_{x\to\infty}(1-\frac{1}{x})^x=e^{-1}$，故：

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n=e^{-\lambda}$$

- 第三部分：当 $n\to\infty$ 时，$\frac{\lambda}{n}\to0$，故：

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=(1-0)^{-k}=1$$

将三部分合起来，得证：

$$\lim_{n\to\infty}P(X=k)=\frac{\lambda^k}{k!}\cdot1\cdot e^{-\lambda}\cdot1=\frac{\lambda^k e^{-\lambda}}{k!}$$

1.3 指数分布 (Exponential Distribution)

笔记 [核心直觉：等待时间] 指数分布和泊松分布是“孪生兄弟”。如果“单位时间内来的人数”服从泊松分布，那么“两个人到达的时间间隔”就服从指数分布。

它关心的是：你要等多久，下一件事才会发生？

定义 1.3 (公式与参数)

若 $X\sim E(\lambda)$，其概率密度函数 (PDF) 为：

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

- 参数 λ：和泊松分布的 λ 含义一样，是发生速率。
- 期望：$E(X)=1/\lambda$。直觉：如果你平均每小时接 5 个电话 ($\lambda=5$)，那你平均等待 1/5 小时（12 分钟）就能接到下一个。

♣

性质 [无记忆性 (Memorylessness)] 这是指数分布最牛的性质：

$$P(X>s+t\mid X>s)=P(X>t)$$

人话解释：假设你在等公交车（且公交车服从指数分布），你已经等了 10 分钟没车来。再等 5 分钟车会来的概率，和你刚到站时等 5 分钟车会来的概率是一模一样的！过去的等待通过，不会影响未来。

1.4 指数分布：泊松过程的等待时间 ⋆

笔记 [核心思想] 指数分布并不是凭空出现的。它是泊松过程的“时间视角”。如果事件发生的次数服从泊松分布，那么相邻两次事件的间隔时间必然服从指数分布。

定理 1.2 (从泊松推导指数密度)

设某事件发生的次数服从参数为 λ 的泊松过程。令随机变量 T 为等待第一次事件发生的时间。证明：T 的概率密度函数为 $f(t) = \lambda e^{-\lambda t}$ $(t > 0)$。

证明 我们使用分布函数法 **(CDF Method)** 进行推导。

第一步：建立等价事件事件 $\{T > t\}$ 表示 "等待时间超过了 t"。这等价于：在时间段 $[0, t]$ 内，事件发生的次数 $N(t)$ 为 0。

第二步：利用泊松分布计算概率时间段长为 t，单位时间平均发生 λ 次，故该时间段内的平均发生次数为 λt。根据泊松分布公式 $P(N = k) = \frac{\mu^k e^{-\mu}}{k!}$，其中 $\mu = \lambda t, k = 0$：

$$P(T > t) = P(N(t) = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

第三步：求分布函数 $F(t)$

$$F(t) = P(T \le t) = 1 - P(T > t) = 1 - e^{-\lambda t} \quad (t \ge 0)$$

(当 $t < 0$ 时，时间不可能为负，故 $F(t) = 0$)

第四步：求导得密度函数 $f(t)$ 对分布函数求导即得概率密度函数 (PDF)：

$$f(t) = F'(t) = (1 - e^{-\lambda t})' = 0 - (-\lambda)e^{-\lambda t} = \lambda e^{-\lambda t}$$

综上所述，指数分布的密度函数为：

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

1.5 正态分布 (Normal / Gaussian Distribution)

笔记 [核心直觉：误差的叠加] 为什么正态分布在自然界中无处不在？因为根据 中心极限定理，如果一个量是由许多相互独立的微小因素叠加而成的，它就服从正态分布。

例子：人的身高（受基因、营养、环境等无数微小因素影响）、测量误差、热噪声。

定义 1.4 (高斯分布)

若 $X \sim N(\mu, \sigma^2)$，其密度函数为：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

这个公式看着吓人，其实就是个钟形曲线：

- $(x - \mu)^2$：决定了关于 μ 对称。
- $e^{-\cdots}$：决定了离中心越远，概率衰减越快。
- σ：决定了钟形的胖瘦（波动大小）。

关于正态分布，我们后面还会介绍到关于它的更多的应用，以此来强化我们对其的直观感受，而此处只需记住其公式即可

第 2 章 多维随机变量：从点到面

在本章中，我们不再孤立地看某一个变量，而是研究两个（或多个）变量之间的相互关系。

内容提要

- ❑ 联合分布：X 和 Y 同时发生的概率（全貌）。
- ❑ 边缘分布：不管 Y 取啥，只看 X 的概率（投影/影子）。
- ❑ 条件分布：已知 $Y=y$ 发生了，此时 X 的规律（切片）。
- ❑ 独立性与相关性：它们到底有什么关系？

2.1 联合、边缘与条件分布

笔记 [直观理解：三维蛋糕] 对于连续型二维随机变量 (X,Y)，其概率密度函数 $f(x,y)$ 就像一个起伏不平的 ** 地形图 ** 或 ** 蛋糕表面 **。

- 体积：曲面下方的总体积必须是 1（归一化）。
- 边缘密度 $f_X(x)$：把蛋糕向 x 轴压缩投影，得到的侧面轮廓。
- 条件密度 $f_{X|Y}(x|y)$：沿着 $Y=y$ 切一刀，剩下的那个切面的形状（归一化后）。

定义 2.1 (核心公式速查)

- 边缘密度（积分掉不需要的变量）：

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)dy$$

- 条件密度（联合除以边缘）：

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)} \quad (\text{类比} P(A|B)=\frac{P(AB)}{P(B)})$$

- 独立性判断（充要条件）：

$$f(x,y)=f_X(x)\cdot f_Y(y)$$

如果联合密度能拆成“关于 x 的函数”乘以“关于 y 的函数”，且定义域是矩形区域，则 X,Y 独立。♣

2.2 极值分布公式

定理 2.1 (极值分布公式推导)

设总体分布函数为 $F(x)$，密度为 $f(x)$。

1. 最大值的分布 $F_{max}(z)$

$$P(Z_{max}\le z)=P(X_1\le z, X_2\le z,\ldots,X_n\le z)$$

因为独立，概率相乘：

$$F_{max}(z)=P(X_1\le z)\cdots P(X_n\le z)=[F(z)]^n$$

2. 最小值的分布 $F_{min}(z)$ 先求 $P(Z_{min}>z)$（大家都大于 z）：

$$P(Z_{min}>z)=P(X_1>z,\ldots,X_n>z)=[1-F(z)]^n$$

所以分布函数为：

$$F_{min}(z)=1-[1-F(z)]^n$$

♡

第 3 章 变量代换：怎么求 $Y=g(X)$？

内容提要

- 很多时候我们知道 X 的分布，但需要求 $Y=X^2, Y=e^X$ 或 $Z=X+Y$ 的分布。
- 这一章全是计算题干货，掌握“两大通法”，保你大题不丢分。

3.1 一维变换：万能的分布函数法 (CDF Method)

笔记 [核心心法] 不要去背那个带导数的公式（$f_Y(y)=f_X[h(y)]|h'(y)|$），那个公式只对单调函数有效，遇到 $Y=X^2$ 直接暴毙。

永远使用“分布函数法”三部曲：

1. 写定义：$F_Y(y)=P(Y\le y)=P(g(X)\le y)$。
2. 换元：把不等式 $g(X)\le y$ 转化为 X 的范围（这一步最关键，要在积分区域上画图）。
3. 求导：算出 $F_Y(y)$ 后，求导得到 $f_Y(y)=F'_Y(y)$。

例题 3.1 经典考题：卡方分布的雏形 设 $X\sim N(0,1)$，求 $Y=X^2$ 的概率密度函数。

解 1. 确定范围：因为 $X^2\ge 0$，所以当 $y<0$ 时，$f_Y(y)=0$。下面只讨论 $y\ge 0$。

2. 写分布函数：

$$F_Y(y)=P(Y\le y)=P(X^2\le y)=P(-\sqrt{y}\le X\le \sqrt{y})$$

3. 转化为 X 的积分（或者直接用 **CDF** 表示）：

$$F_Y(y)=F_X(\sqrt{y})-F_X(-\sqrt{y})$$

4. 两边对 y 求导（注意链式法则）：

$$f_Y(y)=f_X(\sqrt{y})\cdot(\sqrt{y})'-f_X(-\sqrt{y})\cdot(-\sqrt{y})'$$

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

$$f_Y(y)=\frac{1}{2\sqrt{y}}[f_X(\sqrt{y})+f_X(-\sqrt{y})]$$

代入标准正态密度 $f_X(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$，且注意到 f_X 是偶函数：

$$f_Y(y)=\frac{1}{2\sqrt{y}}\cdot 2\cdot\frac{1}{\sqrt{2\pi}}e^{-(\sqrt{y})^2/2}=\frac{1}{\sqrt{2\pi y}}e^{-y/2}\quad (y>0)$$

注：这就是自由度为 *1* 的 χ^2 分布。

例题 3.2 真题实战：看似复杂实则简单的线性变换 设随机变量 X 的概率密度为：

$$f(x)=C_n\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\quad -\infty<x<+\infty$$

其中 C_n 为正常数，n 为正整数。求随机变量 $Y=\frac{1}{\sqrt{n}}X$ 的概率密度 $f_Y(y)$。

解 第一步：写出 Y 的分布函数定义

$$F_Y(y)=P(Y\le y)=P\left(\frac{1}{\sqrt{n}}X\le y\right)$$

第二步：转化为 X 的范围因为 $\frac{1}{\sqrt{n}}>0$，不等号不改变方向：

$$F_Y(y)=P(X\le\sqrt{n}y)=F_X(\sqrt{n}y)$$

第三步：利用链式法则求导两边对 y 求导得到密度函数：

$$f_Y(y) = F'_Y(y) = f_X(\sqrt{n}y) \cdot (\sqrt{n}y)' = \sqrt{n} f_X(\sqrt{n}y)$$

第四步：代入并化简（关键步骤）把 $f_X(x)$ 表达式中的 x 全部替换成 $\sqrt{n}y$：

$$\begin{aligned} f_Y(y) &= \sqrt{n} \cdot C_n \left(1 + \frac{(\sqrt{n}y)^2}{n}\right)^{-\frac{n+1}{2}} \\ &= \sqrt{n} C_n \left(1 + \frac{ny^2}{n}\right)^{-\frac{n+1}{2}} \\ &= \sqrt{n} C_n (1+y^2)^{-\frac{n+1}{2}} \end{aligned}$$

结论：

$$f_Y(y) = \sqrt{n} C_n (1+y^2)^{-\frac{n+1}{2}}, \quad -\infty < y < +\infty$$

笔记 [观察] 如果 $n = 1$，这实际上就是柯西分布（Cauchy Distribution）的某种变体。这类题目不用害怕 $f(x)$ 长得复杂，只要代入进去，里面的系数往往会刚好消掉。

3.2 二维变换的本质：重积分通法

定义 3.1 (通用公式)

设 $Z = g(X, Y)$，求 $f_Z(z)$。我们不背特定公式，而是直接从定义出发：

$$F_Z(z) = P(g(X,Y) \le z) = \iint_{D_z} f(x,y)dxdy$$

其中积分区域 $D_z = \{(x,y) \mid g(x,y) \le z\}$。

算出 $F_Z(z)$ 后，求导即得 $f_Z(z)$。

笔记 [核心心法：画图是灵魂] 解决这类问题的关键不在于积分技巧，而在于能不能正确画出不等式 $g(x,y) \le z$ 描述的区域。

3.3 三大经典模型的几何视角

3.3.1 1. 和的分布 $Z = X + Y$（线性扫描）

- **不等式**：$x + y \le z \implies y \le z - x$。
- **几何意义**：直线 $y = -x + z$ 左下方与定义域的交集。
- **动图想象**：一条斜率为 -1 的直线从左下方向右上方扫描。

3.3.2 2. 积的分布 $Z = XY$（双曲线扫描）

- **不等式**：$xy \le z$。
- **几何意义**：
  - 若 $x > 0$，则 $y \le z/x$（双曲线下方）。
  - 若 $x < 0$，则 $y \ge z/x$（双曲线上方，注意不等号变向!）。
- **难点**：积分区域往往是由双曲线和坐标轴围成的“曲边梯形”。

3.3.3 3. 商的分布 $Z = X/Y$（旋转扫描）

- **不等式**：$X/Y \le z$。

- **几何意义**：这是一束过原点的射线。
  - 若 $y>0$，则 $x\le zy$（直线左侧）。
  - 若 $y<0$，则 $x\ge zy$（直线右侧）。
- **动图想象**：一条过原点的直线像雷达一样旋转扫描，扫过的角度即为概率。

例题 3.3 实战：商分布的重积分法 设 $X,Y\sim E(1)$ 独立（指数分布），求 $Z=X/Y$ 的密度。

解 1. 确定区域：$x>0,y>0$（第一象限）。$X/Y\le z\implies x\le zy$。由于 x,y 均为正，积分区域 D_z 是第一象限中直线 $x=zy$ 的上方部分（夹在 y 轴和直线之间）。

2. 建立重积分：

$$F_Z(z)=\iint_{D_z}e^{-x}e^{-y}dxdy$$

最好先积 x（从 0 积到 zy），再积 y（从 0 到 $+\infty$）：

$$F_Z(z)=\int_0^{+\infty}dy\int_0^{zy}e^{-x-y}dx$$

3. 计算内层积分：

$$\int_0^{zy}e^{-x-y}dx=e^{-y}(1-e^{-zy})=e^{-y}-e^{-(1+z)y}$$

4. 计算外层积分：

$$F_Z(z)=\int_0^{+\infty}(e^{-y}-e^{-(1+z)y})dy=\left[-e^{-y}+\frac{e^{-(1+z)y}}{1+z}\right]_0^{+\infty}$$

$$F_Z(z)=(0-0)-(-1+\frac{1}{1+z})=1-\frac{1}{1+z}=\frac{z}{1+z}\quad(z>0)$$

5. 求导：

$$f_Z(z)=F_Z'(z)=\frac{1}{(1+z)^2},\quad z>0$$

注：这就是著名的 F 分布 (变形) 或 *Log-Logistic* 分布的一种。你看，完全不需要背公式！

例题 3.4 实战：均匀分布的乘积（反直觉的结果） 设 $X,Y\sim U(0,1)$ 且相互独立。求 $Z=XY$ 的概率密度函数。

解 1. 几何建模联合密度 $f(x,y)=1$，定义域为单位正方形 $D=\{(x,y)\mid 0\le x\le 1,0\le y\le 1\}$。

我们要计算分布函数 $F_Z(z)=P(XY\le z)$。

- 当 $z<0$ 时，$F_Z(z)=0$；当 $z>1$ 时，$F_Z(z)=1$。
- 关键在于 $0<z<1$ 时。不等式 $xy\le z$ 对应的区域是双曲线 $y=z/x$ 的下方。

2. 积分策略：算补集（抠补法）直接算双曲线下方的面积比较麻烦（要分段积分）。我们不妨算右上方的补集区域 A^c（即 $XY>z$ 的部分），然后用 1 减去它。

补集区域 $D_{>z}$ 满足：$z<x\le 1$ 且 $z/x<y\le 1$。

$$P(XY>z)=\int_z^1dx\int_{z/x}^1 1\cdot dy$$

3. 执行积分先积 y：

$$\int_{z/x}^1dy=1-\frac{z}{x}$$

再积 x：

$$\int_z^1\left(1-\frac{z}{x}\right)dx=[x-z\ln x]_z^1$$

代入上下限：

$$=(1-z\ln 1)-(z-z\ln z)=1-z+z\ln z$$

4. 得到分布函数

$$F_Z(z)=1-P(XY>z)=1-(1-z+z\ln z)=z-z\ln z\quad(0<z<1)$$

5. 求导得密度

$$f_Z(z) = F'_Z(z) = (z)' - (z\ln z)' = 1 - \left(1 \cdot \ln z + z \cdot \frac{1}{z}\right) = 1 - (\ln z + 1) = -\ln z$$

结论：

$$f_Z(z) = \begin{cases} -\ln z, & 0 < z < 1 \\ 0, & \text{其他} \end{cases}$$

第 4 章 条件期望：预测的艺术

内容提要

- ❑ 条件期望 $E(X|Y)$ 到底是一个数还是一个随机变量？
- ❑ 它是现代统计学和机器学习（如回归分析）的理论基石。
- ❑ 本章核心工具：重期望公式（Law of Iterated Expectation），又称“双重期望公式”。

4.1 核心概念：从“切片”到“函数”

定义 4.1 (两个层面的定义)

1. 作为数值 $E(X|Y=y)$：这是一个普通的实数，表示“在 Y 已经取定为 y 的条件下，X 的平均值”。
$$E(X|Y=y)=\int_{-\infty}^{+\infty}x\cdot f_{X|Y}(x|y)dx$$
2. 作为随机变量 $E(X|Y)$：这是一个关于 Y 的函数 $g(Y)$。因为 Y 是随机的，所以 $g(Y)$ 也是随机的。它代表了“在观察到 Y 之后，我们对 X 的最佳估计”。 ♣

笔记 [直观理解] 假设 X 是这种身高的分布，Y 是性别。

- $E(X|Y=\text{男})=175$ (这是一个具体的数)。
- $E(X|Y)$ 是一个随机变量，它有两个取值：如果抽到男，值就是 175；如果抽到女，值就是 163。
- 一句话总结：$E(X|Y)$ 是 Y 的函数，但在求期望的外层 $E[\cdot]$ 看来，它就是 X 的替身。

4.2 四大性质（解题外挂）

性质 [条件期望的运算规则] 设 $g(Y)$ 是关于 Y 的任意函数（即一旦 Y 确定，$g(Y)$ 就已知）。

1. 常数性：$E(c|Y)=c$。如果 X 已经由 Y 决定了，那期望就是它自己。
2. 线性：$E(aX_1+bX_2|Y)=aE(X_1|Y)+bE(X_2|Y)$。
3. 剔除已知因子 **(Taking out what is known)**：
$$E[g(Y)\cdot X\mid Y]=g(Y)\cdot E(X|Y)$$
口诀：既然在这个条件下 Y 是已知的，那 $g(Y)$ 就是个常数，可以直接提到期望外面去。
4. 重期望公式 **(Tower Property / Adam's Law)**：
$$E[E(X|Y)]=E(X)$$
含义：“估计的估计”平均下来就是“真实的平均”。这是解决复杂期望问题的核武器。

例题 4.1 巧用条件期望：秒杀几何分布 设随机变量 $X\sim G(p)$（表示首次成功所需的试验次数）。求 $E(X)$。

传统做法： 计算级数 $\sum_{k=1}^{\infty}k\cdot p(1-p)^{k-1}$，需要错位相减法，计算量大且容易错。**条件期望做法：** 利用第一次试验的结果进行分解。

解 我们根据第一次试验是否成功来使用全期望公式。记 $A=\{\text{第一次试验成功}\}$。
$$E(X)=E(X|A)P(A)+E(X|\bar{A})P(\bar{A})$$
分析两种情况：

1. 如果第一次成功 (A 发生)：概率为 p。此时试验结束，总次数 $X=1$。
$$E(X|A)=1$$

2. 如果第一次失败 ($\bar{A}$ 发生)：概率为 $1-p$。此时浪费了 1 次机会，一切重头再来。由于过程的无记忆性，之后所需的平均次数依然是 $E(X)$。加上刚才浪费的那 1 次，总期望为 $1+E(X)$。

$$E(X|\bar{A}) = 1 + E(X)$$

建立方程：

$$E(X) = 1 \cdot p + [1 + E(X)] \cdot (1-p)$$

求解 $E(X)$：

$$E(X) = p + 1 - p + (1-p)E(X)$$

$$E(X) = 1 + (1-p)E(X)$$

$$pE(X) = 1 \implies E(X) = \frac{1}{p}$$

注：这就是递归法。它避免了复杂的级数求和，直接把概率问题变成了代数方程。

4.3 协方差与相关系数：量化关系

定义 4.2 (计算核心)

- 协方差定义式：

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

- 协方差计算式（做题专用）：

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

- 相关系数：

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

♣

4.4 五大核心恒等式 (运算律)

性质 [线性运算律] 协方差的运算性质非常像乘法，满足交换律、结合律和分配律。设 a,b,c,d 为常数：

1. 对称性：

$$Cov(X,Y) = Cov(Y,X)$$

2. 常数“透明”性（常数与变量无相关）：

$$Cov(X,c) = 0$$

3. 双线性（系数提出的倍数关系）：

$$Cov(aX,bY) = ab\,Cov(X,Y)$$

4. 平移不变性（加减常数不影响波动关系）：

$$Cov(X+a,Y+b) = Cov(X,Y)$$

5. 分配律（最重要！）：

$$Cov(X_1+X_2,Y) = Cov(X_1,Y) + Cov(X_2,Y)$$

综合应用公式：

$$Cov(aX+b,cY+d) = ac\,Cov(X,Y)$$

4.5 方差的展开公式（必考）

定理 4.1 (和差方差公式)

这就像代数中的完全平方公式 $(a+b)^2 = a^2 + b^2 + 2ab$。

$$D(X \pm Y) = D(X) + D(Y) \pm 2Cov(X, Y)$$

特别注意：

- 只有当 X, Y 不相关（$Cov = 0$）时，才有 $D(X \pm Y) = D(X) + D(Y)$。
- 如果 X, Y 独立，则一定不相关，公式成立。

笔记 [推广：项数更多的展开]

$$D\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D(X_i) + 2 \sum_{1 \le i < j \le n} Cov(X_i, X_j)$$

如果是独立同分布 (i.i.d) 的变量求和，后面那一堆 Cov 全是 0，于是 $D(\sum X_i) = nD(X)$。

4.6 “不相关”与“独立”

表 4.1: 独立 vs 不相关

关系	结论
独立 $\Rightarrow$ 不相关	**成立**。独立意味着完全没关系，自然没有线性关系。
不相关 $\Rightarrow$ 独立	**不一定**。$Cov = 0$ 只排除了线性关系，可能存在 $Y = X^2$ 这种非线性关系。
二维正态分布情形	**等价**。对于 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，$\rho = 0 \Leftrightarrow$ 独立。

例题 4.2 经典反例：不相关但不独立 设 $X \sim U(-1, 1)$，令 $Y = X^2$。显然 X, Y 不独立（Y 完全由 X 决定）。但计算协方差：

- $E(X) = 0$
- $E(XY) = E(X \cdot X^2) = E(X^3) = \int_{-1}^{1} \frac{1}{2} x^3 dx = 0$ (奇函数积分)
- $Cov(X, Y) = E(XY) - E(X)E(Y) = 0 - 0 = 0$

结论： X 和 X^2 不相关，但它们紧密相连。

第 5 章 三大核心分布的期望与方差推导

内容提要

- ❑ 本章属于"内功心法"。
- ❑ 掌握这些推导，你不仅记住了公式，更复习了微积分和级数的基本功。
- ❑ 核心技巧：级数下标变换、分部积分、换元积分。

5.1 泊松分布 $X \sim P(\lambda)$

笔记 [推导策略] 离散型求期望，核心难点在于消去分母的 $k!$。技巧：利用 $k/k! = 1/(k-1)!$ 进行"降阶"，然后凑出 e^λ 的泰勒展开式。对于方差，尽量算 $E[X(X-1)]$ 而不是直接算 $E(X^2)$。

证明 [**1.** 期望 $E(X)$ 的推导]

$$E(X) = \sum_{k=0}^{\infty} k \cdot P(X=k) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!}$$

当 $k=0$ 时第一项为 0，故求和从 $k=1$ 开始。消去 k：

$$E(X) = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}$$

令 $j = k-1$（下标平移），则 $\lambda^k = \lambda \cdot \lambda^j$：

$$E(X) = \lambda e^{-\lambda} \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!}}_{\text{此即 } e^\lambda} = \lambda e^{-\lambda} \cdot e^\lambda = \lambda$$

证明 [**2.** 方差 $D(X)$ 的推导] 利用公式 $D(X) = E(X^2) - [E(X)]^2$。我们先求二阶矩 $E(X^2)$。直接求 $\sum k^2 \frac{\lambda^k}{k!}$ 较难，不如利用 ** 二阶阶乘矩 **：

$$X^2 = X(X-1) + X \implies E(X^2) = E[X(X-1)] + E(X)$$

计算 $E[X(X-1)]$：

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!}$$

$k = 0, 1$ 时项为 0，求和从 $k=2$ 开始。消去 $k(k-1)$：

$$= e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!}$$

令 $j = k-2$，则 $\lambda^k = \lambda^2 \cdot \lambda^j$：

$$= \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^2 e^{-\lambda} \cdot e^\lambda = \lambda^2$$

代回方差公式：

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

5.2 指数分布 $X \sim E(\lambda)$

笔记 [推导策略] 连续型求期望，核心是分部积分法 (**Integration By Parts**)。记住公式：$\int u dv = uv - \int v du$。这里需要反复对 $xe^{-\lambda x}$ 这种形式进行积分。

证明 [**1.** 期望 $E(X)$ 的推导]

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^{+\infty} x \cdot \lambda e^{-\lambda x} dx$$

令 $u = x, dv = \lambda e^{-\lambda x}dx$，则 $du = dx, v = -e^{-\lambda x}$。

$$E(X) = \left[-xe^{-\lambda x}\right]_0^{+\infty} - \int_0^{+\infty} (-e^{-\lambda x})dx$$

第一项：利用洛必达法则可知 $\lim_{x\to\infty} xe^{-\lambda x} = 0$，故第一项为 $0 - 0 = 0$。

$$E(X) = \int_0^{+\infty} e^{-\lambda x}dx = \left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^{+\infty} = 0 - (-\frac{1}{\lambda}) = \frac{1}{\lambda}$$

证明 [**2.** 方差 $D(X)$ 的推导] 先求 $E(X^2)$：

$$E(X^2) = \int_0^{+\infty} x^2 \cdot \lambda e^{-\lambda x} dx$$

分部积分：令 $u = x^2, dv = \lambda e^{-\lambda x}dx$，则 $du = 2xdx, v = -e^{-\lambda x}$。

$$E(X^2) = \underbrace{\left[-x^2 e^{-\lambda x}\right]_0^{+\infty}}_{0} + \int_0^{+\infty} 2xe^{-\lambda x}dx$$

提取常数 2 和 $1/\lambda$（构造期望的形式）：

$$= \frac{2}{\lambda}\int_0^{+\infty} x \cdot \lambda e^{-\lambda x}dx = \frac{2}{\lambda} \cdot E(X) = \frac{2}{\lambda} \cdot \frac{1}{\lambda} = \frac{2}{\lambda^2}$$

计算方差：

$$D(X) = E(X^2) - [E(X)]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

5.3 正态分布 $X \sim N(\mu, \sigma^2)$

笔记 [推导策略] 核心技巧是换元法（令 $t = \frac{x-\mu}{\sigma}$）和利用奇偶性。不要硬算 e^{-x^2} 的原函数（算不出来的），要利用 $\int_{-\infty}^{+\infty} t\phi(t)dt = 0$ 这种性质。

证明 [**1.** 期望 $E(X)$ 的推导]

$$E(X) = \int_{-\infty}^{+\infty} x\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

令 $t = \frac{x-\mu}{\sigma}$，则 $x = \sigma t + \mu, dx = \sigma dt$。

$$E(X) = \int_{-\infty}^{+\infty} (\sigma t + \mu)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

拆分为两项：

$$= \sigma \underbrace{\int_{-\infty}^{+\infty} t \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt}_{I_1} + \mu \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt}_{I_2}$$

- 对于 I_1：被积函数是奇函数，且积分区间对称，故积分为 0。
- 对于 I_2：被积函数是标准正态分布 $N(0,1)$ 的密度，全积分必为 1。

$$\therefore E(X) = \sigma \cdot 0 + \mu \cdot 1 = \mu$$

证明 [**2.** 方差 $D(X)$ 的推导] 利用定义 $D(X) = E[(X-\mu)^2]$ 计算更简便。

$$D(X) = \int_{-\infty}^{+\infty} (x-\mu)^2\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

同样令 $t = \frac{x-\mu}{\sigma}$，则 $(x-\mu)^2 = \sigma^2 t^2$：

$$D(X) = \int_{-\infty}^{+\infty} \sigma^2 t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2}} dt$$

对积分部分 $\int_{-\infty}^{+\infty} t^2 e^{-t^2/2} dt$ 使用分部积分：拆分 $t^2 e^{-t^2/2} = t \cdot (te^{-t^2/2})$。令 $u = t, dv = te^{-t^2/2}dt$，则 $du = dt, v = -e^{-t^2/2}$。

$$\int_{-\infty}^{+\infty} t^2 e^{-t^2/2} dt = \underbrace{\left[-te^{-t^2/2}\right]_{-\infty}^{+\infty}}_{0} - \int_{-\infty}^{+\infty} -e^{-t^2/2} dt$$

$$= \int_{-\infty}^{+\infty} e^{-t^2/2} dt = \sqrt{2\pi} \quad \text{(高斯积分)}$$

代回原式：

$$D(X) = \frac{\sigma^2}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = \sigma^2$$

第 6 章 极限定理的基石：两大不等式

内容提要

- 极限定理（大数定律、CLT）是概率论的皇冠。
- 要摘取皇冠，必须先掌握两个核心工具：**Markov** 不等式（控制尾部）和 **Chebyshev** 不等式（控制波动）。

6.1 Markov 不等式：非负变量的边界

定理 6.1 (Markov Inequality)

设 X 为非负随机变量（即 $P(X \geq 0) = 1$），且期望 $E(X)$ 存在。则对于任意常数 $a > 0$：

$$P(X \geq a) \leq \frac{E(X)}{a}$$

笔记 [直观理解]

- 这是一个很宽松的上界，但它非常通用（只要求期望存在，且非负）。
- 含义：一个平均值很小的变量，取极大值的概率不可能很大。
- 比如：全班平均分 60 分（且不能为负分），那么 120 分以上的人数比例最多是 $60/120 = 1/2$？(哦不对，分数不能超 100，但逻辑是这个逻辑：平均值限制了极端值的比例)。

证明 [利用示性函数证明] 定义示性函数 $I_{\{X \geq a\}}$：当 $X \geq a$ 时取 1，否则取 0。由于 $X \geq 0$，显见以下不等式恒成立：

$$a \cdot I_{\{X \geq a\}} \leq X$$

(当 $X < a$ 时，左边为 0，右边 ≥ 0；当 $X \geq a$ 时，左边为 a，右边 $\geq a$。)

两边同时求期望：

$$E[a \cdot I_{\{X \geq a\}}] \leq E(X)$$

$$a \cdot E[I_{\{X \geq a\}}] \leq E(X)$$

因为示性函数的期望即为概率 $P(X \geq a)$：

$$a \cdot P(X \geq a) \leq E(X) \implies P(X \geq a) \leq \frac{E(X)}{a}$$

6.2 Chebyshev 不等式：波动的控制

笔记 [核心思想] Markov 不等式只用了“期望”信息。如果我们还知道“方差”，能得到更精确的界吗？Chebyshev 不等式本质上就是把 **Markov** 不等式应用在 $(X-\mu)^2$ 上。

定理 6.2 (Chebyshev Inequality)

设随机变量 X 的期望为 μ，方差为 σ^2。则对于任意 $\varepsilon > 0$：

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

或者写成 3σ 准则的形式（令 $\varepsilon = k\sigma$）：

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

证明 [一句话推导] 我们要估计 $P(|X-\mu| \geq \varepsilon)$。这等价于估计 $P((X-\mu)^2 \geq \varepsilon^2)$。

令 $Y = (X-\mu)^2$。显见 Y 是非负随机变量，且 $E(Y) = D(X) = \sigma^2$。直接对 Y 使用 Markov 不等式（取 $a = \varepsilon^2$）：

$$P(Y \geq \varepsilon^2) \leq \frac{E(Y)}{\varepsilon^2}$$

代回即得：

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

注[Chebyshev 的威力]

- 它不依赖于分布的具体形式！不管你是正态、均匀还是什么奇形怪状的分布，只要方差存在，这个不等式就成立。
- **例：** 取 $k=3$，则 $P(|X-\mu| \geq 3\sigma) \leq 1/9 \approx 11\%$。
- 虽然正态分布中 3σ 外的概率只有 0.3%，Chebyshev 给出的 11% 显得很粗糙，但它是**万能**的。

第 7 章 极限定理：从混沌到秩序

内容提要

- ❑ 为什么我们可以用“频率”来估计“概率”？(大数定律)
- ❑ 为什么自然界中正态分布无处不在？(中心极限定理)
- ❑ 本章是概率论（理论）向数理统计（应用）跨越的桥梁。

7.1 大数定律 (LLN)：稳定性的保证

笔记 [直观含义] 大数定律说的是：“样本均值” $\bar{X}$ 会依概率收敛于“总体均值” μ。

只要样本量 n 足够大，偶然的波动会被互相抵消，最终显现出必然的规律。这是蒙特卡洛模拟 (Monte Carlo) 和所有统计估算的基石。

定理 7.1 (切比雪夫弱大数定律)

设 $X_1, X_2, \dots$ 是独立同分布 (i.i.d) 的随机变量序列，期望 $E(X_i) = \mu$，方差 $D(X_i) = \sigma^2 < \infty$。则对于任意 $\varepsilon > 0$：

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

即样本均值 $\bar{X}$ 依概率收敛于 μ，记为 $\bar{X} \xrightarrow{P} \mu$。

证明 [切比雪夫一秒证法（必背）] 令样本均值 $\bar{X} = \frac{1}{n}\sum X_i$。

- 期望：$E(\bar{X}) = \frac{1}{n}\cdot n\mu = \mu$。
- 方差：$D(\bar{X}) = D(\frac{1}{n}\sum X_i) = \frac{1}{n^2}\sum D(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$。

直接对 $\bar{X}$ 使用切比雪夫不等式：

$$P(|\bar{X} - \mu| \ge \varepsilon) \le \frac{D(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

当 $n \to \infty$ 时，右边 $\to 0$。证毕。

注[辛钦大数定律 (Khintchine’s Law)] 切比雪夫大数定律要求“方差存在”。其实只要 X 独立同分布且 ** 期望存在 **，大数定律就成立。这就是辛钦大数定律（证明需要特征函数，此处略）。

例题 7.1 伯努利大数定律 这是我们在初中就学过的：抛硬币次数越多，正面朝上的频率 $f_n = n_A/n$ 越接近概率 p。

$$\lim_{n\to\infty} P(|f_n - p| < \varepsilon) = 1$$

7.2 中心极限定理 (CLT)：万物归宗

笔记 [核心思想] 大数定律告诉我们 $\bar{X}$ 会收敛到一个常数 μ。中心极限定理告诉我们：在 $\bar{X}$ 收敛到 μ 的过程中，它的误差分布长什么样？

结论：无论 X_i 原来服从什么分布（只要独立同分布且方差有限），它们的和（或均值）在标准化后，都逼近标准正态分布。

定理 7.2 (林德伯格-列维 (Lindeberg-Lévy) 定理)

设 $X_1, X_2, \ldots$ 独立同分布，期望 μ，方差 σ^2。令 $Y_n = \sum_{i=1}^n X_i$。当 n 充分大时：

$$Z_n = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{\sum X_i - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$

♡

注[解题通法] 考试中遇到“求 100 个部件总寿命大于 1000 小时的概率”或者“求 50 次实验平均误差小于 0.1 的概率”，一律按 CLT 处理：

1. 算出总和的期望 $E(\sum) = n\mu$。
2. 算出总和的方差 $D(\sum) = n\sigma^2$。
3. 标准化：$P(\sum X_i > a) \approx 1 - \Phi\left(\frac{a-n\mu}{\sqrt{n}\sigma}\right)$。

7.3 CLT 的终极魅力：从布朗运动到高斯白噪声

内容提要

- ❑ 为什么我们在信号处理、金融模型（如 Black-Scholes）中，总是假设噪声或收益率服从正态分布？
- ❑ 难道上帝真的只掷高斯骰子吗？
- ❑ 答案就在中心极限定理中：正态分布是大量微小随机扰动叠加的必然归宿。

例题 7.2 物理直觉：布朗运动的微观起源 考虑一个悬浮在液体中的花粉粒子。它为什么会做无规则的布朗运动？

1. 微观视角：无数次的“推搡” 在极短的时间 Δt 内，花粉粒子受到了来自四面八方的水分子的成千上万次撞击。

- 设第 i 次撞击带来的微小位移为 X_i。
- 所有的 X_i 都是独立同分布的（水分子运动是混乱且独立的）。
- X_i 的分布可能很复杂（取决于分子速度、角度等），但这不重要！

2. 宏观累加：CLT 的介入 粒子在 Δt 时间内的总位移 ΔW 是所有微小撞击的总和：

$$\Delta W = \sum_{i=1}^{N} X_i$$

由于 N（撞击次数）极大（数量级可能是 10^{20}），根据**中心极限定理 (CLT)**：无论单个分子的撞击 X_i 服从什么分布，它们的总和 ΔW **必然收敛于正态分布**。

3. 数学结论 这解释了为什么布朗运动的增量服从正态分布：

$$W(t+\Delta t) - W(t) \sim N(0, \sigma^2 \Delta t)$$

笔记 [深度：为什么“对时间的微分”是高斯分布？] 你可能会问：“位移是正态的，那速度（微分）呢？”

数学上，标准的布朗运动路径 $W(t)$ 处处连续但处处不可微（分形性质）。但在工程和物理上，我们引入广义函数的概念。

考虑“平均速度”：

$$\frac{W(t+\Delta t) - W(t)}{\Delta t} \sim \frac{N(0, \Delta t)}{\Delta t} = N\left(0, \frac{1}{\Delta t}\right)$$

当 $\Delta t \to 0$ 时，我们称其极限为高斯白噪声 **(Gaussian White Noise)**，记为 $\xi(t) = \frac{dW}{dt}$。

魅力所在：正是因为底层无数次撞击满足 CLT，导致了宏观上的“位移增量”是正态的，进而导致了其形式上的“导数”（白噪声）也是高斯的。这就是为什么我们在做卡尔曼滤波、信号分析时，默认噪声是高斯分布的物理依据。

第 8 章 参数估计：侦探的游戏

内容提要

- 之前我们学的是概率论：已知参数（如 p, λ），预测数据。
- 现在我们进入数理统计：已知数据（手里的样本），反推参数。
- 两大流派：矩估计（模仿派）与极大似然估计（推理派）。

8.1 矩估计 (MoM)：简单粗暴的模仿

笔记 [核心思想] “样本就是总体的缩影。”矩估计的逻辑非常直白：既然我不知道总体的均值 $E(X)$ 是多少，那我就用样本的均值 $\bar{X}$ 去代替它。

口诀：用样本矩替换总体矩。

定义 8.1 (操作步骤)

设总体有 k 个待估参数 $\theta_1, \dots, \theta_k$。

1. 算总体矩：计算前 k 阶原点矩 $\mu_1 = E(X), \mu_2 = E(X^2), \dots$（这些是含 θ 的式子）。
2. 算样本矩：计算样本的原点矩 $A_1 = \bar{X}, A_2 = \frac{1}{n}\sum X_i^2$（这些是具体的数）。
3. 令其相等：建立方程组

$$\begin{cases} E(X) = \bar{X} \\ E(X^2) = \frac{1}{n}\sum X_i^2 \end{cases}$$

4. 反解参数：解出 $\hat{\theta}$ 即为估计量。

例题 8.1 实战：均匀分布 $U(a,b)$ 设 $X \sim U(a,b)$，求 a, b 的矩估计量。

解： 两个未知数，需要建立两个方程。

- 总体均值 $E(X) = \frac{a+b}{2}$，总体二阶矩 $E(X^2) = D(X) + [E(X)]^2 = \frac{(b-a)^2}{12} + (\frac{a+b}{2})^2$。
- 令 $E(X) = \bar{X}$ 和 $D(X) = S_n^2$（利用方差公式简化计算，σ^2 对应样本方差）。

$$\begin{cases} \frac{a+b}{2} = \bar{X} \\ \frac{(b-a)^2}{12} = S_n^2 \end{cases}$$

- 解得：

$$\hat{a} = \bar{X} - \sqrt{3}S_n, \quad \hat{b} = \bar{X} + \sqrt{3}S_n$$

8.2 极大似然估计 (MLE)：谁的嫌疑最大？

笔记 [核心思想] “存在即合理。”事情已经发生了（样本观测值 $x_1, \dots, x_n$ 已经出现了），那么参数 θ 应该是多少，才能让“这件事发生的概率”最大？

我们把概率函数 $P(x|\theta)$ 反过来看，看作是关于 θ 的函数 $L(\theta|x)$，求它的最大值点。

定理 8.1 (似然函数 Likelihood Function)

- 离散型：$L(\theta) = \prod_{i=1}^n P(X_i = x_i; \theta)$
- 连续型：$L(\theta) = \prod_{i=1}^n f(x_i; \theta)$

笔记 [MLE 标准三部曲]

1. 写似然函数：$L(\theta)=\prod f(x_i)$。
2. 取对数（关键）：$\ln L(\theta)=\sum \ln f(x_i)$。（为什么要取对数？因为连乘求导太难，变成连加求导就简单了。且 ln 单调递增，不改变极值点位置。）
3. 求导置零：计算 $\frac{d}{d\theta}\ln L(\theta)=0$，解出的 $\hat{\theta}$ 即为 MLE。

8.3 高危考点：均匀分布的 MLE

笔记 [陷阱] 如果概率密度函数的定义域包含参数 θ（例如 $U(0,\theta)$），绝对不能用求导法！因为此时 $L(\theta)$ 在边界处不可导，或者最大值在边界上。

例题 8.2 经典反例 设总体 $X\sim U(0,\theta)$，求 θ 的极大似然估计。

$$f(x)=\begin{cases}\frac{1}{\theta}, & 0\le x\le\theta\\ 0, & \text{其他}\end{cases}$$

解 1. 写似然函数

$$L(\theta)=\prod_{i=1}^{n} f(x_i)=\begin{cases}\frac{1}{\theta^n}, & \text{当所有}0\le x_i\le\theta\text{时}\\ 0, & \text{只要有一个}x_i>\theta\end{cases}$$

2. 关键逻辑要使 $L(\theta)\ne 0$，必须满足所有观测值都落在 $[0,\theta]$ 内，即 $\max(x_1,\ldots,x_n)\le\theta$。也就是 θ 必须大于等于所有的样本值，即 $\theta\ge x_{(n)}$。

3. 寻找最大值在区间 $[\max(x_i),+\infty)$ 上，函数 $L(\theta)=\frac{1}{\theta^n}$ 是单调递减的。所以，θ 取最小值的时候，$L(\theta)$ 最大。

结论：$\hat{\theta}_{MLE}=\max(X_1,\ldots,X_n)=X_{(n)}$。

对比：此题若用矩估计，$\bar{X}=\theta/2 \implies \hat{\theta}_{MoM}=2\bar{X}$。显然 $X_{(n)}$ 比 $2\bar{X}$ 更合理（因为它保证了估计值一定大于所有样本）。

8.4 MLE 的巅峰应用：线性回归

内容提要

- 以前学线性回归，老师说“我们要让误差平方和最小”，这叫最小二乘法 (Least Squares)。
- 今天我们用 MLE 的视角重新审视它：为什么偏偏是“平方和”？而不是绝对值之和？
- 结论：因为高斯分布的指数部分，刚好就是平方项。

例题 8.3MLE 推导线性回归 设线性模型为 $y_i=\beta_0+\beta_1 x_i+\epsilon_i$ $(i=1,\ldots,n)$。其中误差项 ϵ_i 独立同分布，服从正态分布 $\epsilon_i\sim N(0,\sigma^2)$。试利用极大似然估计法估计参数 β_0,β_1。

笔记 [看图说话] 请注意图中的蓝色高斯曲线。它垂直于 x 轴，意味着给定 x_0 后，y 的取值是一个概率分布。**MLE** 的本质，就是调整红色曲线的位置，使得所有蓝色数据点出现的联合概率（似然值）最大化。

解 1. 确定 y_i 的分布由于 y_i 是常数 $\beta_0+\beta_1 x_i$ 加上随机变量 ϵ_i，故 y_i 也服从正态分布：

$$y_i\sim N(\beta_0+\beta_1 x_i,\sigma^2)$$

其概率密度函数为：

$$f(y_i)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{[y_i-(\beta_0+\beta_1 x_i)]^2}{2\sigma^2}\right\}$$

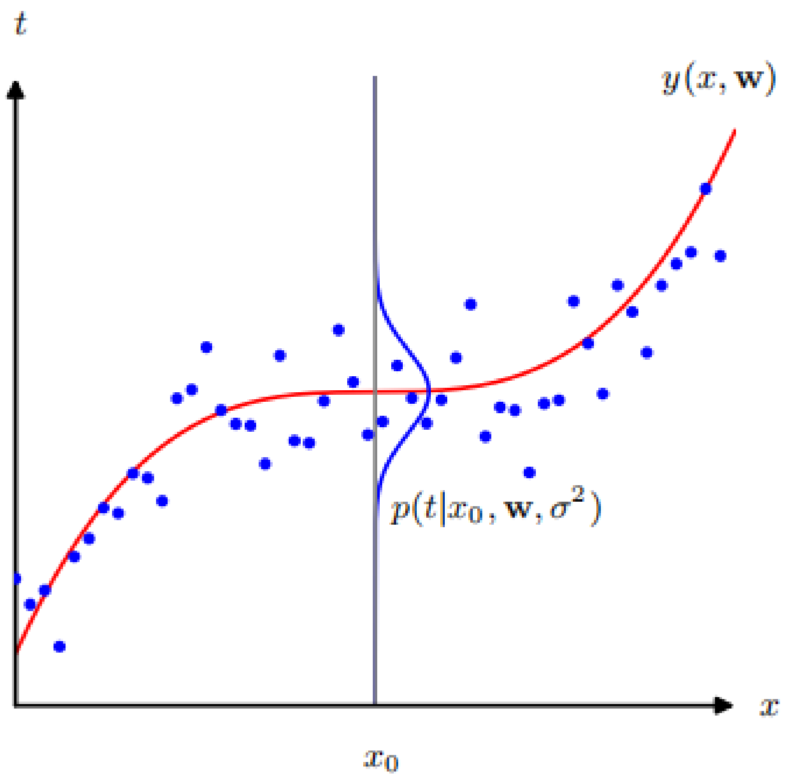

图 8.1: 线性回归的概率视角：数据点围绕回归线做正态波动

2. 写似然函数

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} f(y_i) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i)]^2\right\}$$

3. 取对数似然

$$\ln L = \ln\left[(2\pi\sigma^2)^{-n/2}\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - (\beta_0 + \beta_1 x_i)]^2$$

$$\ln L = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

4. 极大化求解 (见证奇迹的时刻) 我们要寻找 β_0, β_1 使得 $\ln L$ 最大。注意上式中，前两项与 β 无关。第三项前面有个负号。

$$\max_{\beta}(\ln L) \iff \max_{\beta}\left(-\frac{1}{2\sigma^2}\sum(y_i - \hat{y}_i)^2\right) \iff \min_{\beta}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

结论：要让“似然概率最大”(MLE)，就必须让“残差平方和最小”(Least Squares)。这就解释了为什么最小二乘法是线性回归的最佳估计——隐含前提是误差服从正态分布。

8.5 深度思考：为什么偏偏是“平方”？

笔记 [灵魂拷问] 我们在做回归时，习惯性地计算残差平方和 $\sum(y_i - \hat{y}_i)^2$。但你有没有想过：** 为什么不是绝对值之和 $\sum|y_i - \hat{y}_i|$？或者 4 次方之和？**

答案：这完全取决于我们对噪声分布的假设。

定理 8.2 (损失函数与噪声分布的对应关系)

利用极大似然估计 (MLE) 的视角，不同的噪声假设对应着不同的损失函数 (Loss Function)：

1. 高斯噪声 $\to$ 均方误差 **(MSE, L2 Loss)**

$$\epsilon \sim N(0, \sigma^2) \implies f(\epsilon) \propto e^{-\epsilon^2}$$

取对数后，$-\ln L \propto \sum \epsilon^2 = \sum (y_i - \hat{y}_i)^2$。特点：对异常值敏感（平方放大了大误差），计算方便（处处可导）。

2. 拉普拉斯噪声 $\to$ 平均绝对误差 **(MAE, L1 Loss)** 假设噪声服从拉普拉斯分布 (Laplace Distribution)：

$$f(\epsilon) = \frac{1}{2b} \exp\left(-\frac{|\epsilon|}{b}\right)$$

取对数后，指数上的绝对值符号落了下来：

$$-\ln L \propto \sum |\epsilon| = \sum |y_i - \hat{y}_i|$$

特点：对异常值鲁棒（*Robust*），但在 *0* 点不可导。这就是为什么在处理含有很多离群点的数据时，我们要用最小绝对偏差回归 *(LAD)*。

♡

注[一句话总结] 你选择什么样的方法（最小二乘还是最小绝对值），本质上是因为你**相信**数据背后的噪声服从什么样的分布。

- 相信正态分布 $\implies$ 用平方 (L_2)。
- 相信拉普拉斯分布 $\implies$ 用绝对值 (L_1)。

第 9 章 区间估计的前奏：三大抽样分布与枢轴量

内容提要

- ❑ 在统计学中，正态分布 $N(0,1)$ 是“上帝”。
- ❑ 但上帝有三个“亲儿子”：χ^2 分布、t 分布、F 分布。
- ❑ 本章将揭示它们的身世，并利用它们构造枢轴量，为区间估计打下地基。

9.1 正态分布的“拼积木”游戏

笔记 [核心逻辑] 所有的抽样分布，本质上都是标准正态变量 $Z \sim N(0,1)$ 的平方或商。

9.1.1 1. 卡方分布 $\chi^2(n)$：平方和

定义 9.1

设 $X_1, \ldots, X_n$ 独立且都服从标准正态分布 $N(0,1)$。则它们的平方和服从自由度为 n 的 χ^2 分布：

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$$

♣

- **物理意义**：它衡量了 n 个标准正态误差的**总能量**（方差来源）。
- **应用**：专门用来估计**方差** σ^2。
- **性质**：$E(\chi^2) = n, D(\chi^2) = 2n$（具有可加性）。

9.1.2 2. t 分布 $t(n)$：修正的商

定义 9.2

设 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，且 X, Y 独立。则下面的比值服从自由度为 n 的 t 分布：

$$T = \frac{X}{\sqrt{Y/n}} \sim t(n)$$

♣

- **直观理解**：本来 $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ 是标准正态。但如果我们**不知道** σ，只能用样本标准差 S 去代替 σ。因为 S 自己也是随机波动的（不靠谱），所以引入了额外的误差，导致分布的尾巴变“厚”了（Fat Tail）。
- **极限**：当 $n \to \infty$ 时，$S \to \sigma$，$t(n)$ 分布收敛于 $N(0,1)$。

9.1.3 3. F 分布 $F(n_1, n_2)$：方差之比

定义 9.3

设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$，且 U, V 独立。则它们的平均值之比服从 F 分布：

$$F = \frac{U/n_1}{V/n_2} \sim F(n_1, n_2)$$

♣

- **应用**：专门用来比较两个正态总体的**方差**是否相等（ANOVA 的基础）。
- **性质**：$1/F \sim F(n_2, n_1)$（倒数性质）。

9.2 通法：如何构造枢轴量？

内容提要

- 目标：我们要估计参数 θ。
- 枢轴量 $G(X, \theta)$：这是一个混合了样本 X 和参数 θ 的函数，但它最神奇的地方在于——它的分布与 θ 无关（是一个已知的标准分布）。
- 就像一把刻度固定的"尺子"，虽然测量的物体 (θ) 不同，但尺子本身的刻度（分布）是不变的。

表 9.1: 正态总体下的四大枢轴量（背诵版）

待估参数	已知条件	枢轴量构造	服从分布	用途
均值 μ	σ^2 已知	$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	$N(0,1)$	最简单的区间
	σ^2 **未知**	$\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$	$t(n-1)$	**必考**：用 S 替 σ
方差 σ^2	μ 未知	$\dfrac{(n-1)S^2}{\sigma^2}$	$\chi^2(n-1)$	估计波动大小
	(μ 已知)	$\sum \dfrac{(X_i-\mu)^2}{\sigma^2}$	$\chi^2(n)$	(考得少)
方差比 σ_1^2/σ_2^2	μ 未知	$\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$	$F(n_1-1, n_2-1)$	比较精度

笔记 [易错点：自由度]

- 凡是用到样本方差 S^2 的地方，自由度通常都要减 **1**（变为 $n-1$）。
- 原因是计算 S^2 时用到了 $\bar{X}$，消耗了一个自由度（$\sum(X_i-\bar{X})=0$ 这一约束）。

9.3 区间估计的三步走

笔记 [解题通法]

1. 选尺子：根据已知条件（是否知道 σ），从上表中选择合适的枢轴量。
2. 查分位点：根据置信水平 $1-\alpha$，查表找到分布的临界值（如 $z_{\alpha/2}, t_{\alpha/2}(n-1)$）。
3. 反解不等式：

$$P(-c < \text{枢轴量} < c) = 1-\alpha$$

从中反解出 μ 或 σ^2 的范围。

9.4 进阶：双剑合璧——两个正态总体的推断

内容提要

- 现实中我们更常比较两组数据：实验组 vs 对照组、男 vs 女、旧工艺 vs 新工艺。
- 核心问题：两组数据的平均值有差异吗？($\mu_1-\mu_2$)
- 核心问题：两组数据的波动（精度）一样吗？(σ_1^2/σ_2^2)

9.4.1 0. 基础情形：方差 σ_1^2, σ_2^2 已知

笔记 [判别标准] 只要题目里明确给出了总体的标准差 σ_1, σ_2（而不是样本标准差 s_1, s_2），不管样本量 n 是大是小，** 一律用 Z 分布 **。这是最简单、最精确的情况，不需要用 t 分布去模拟，也不需要把方差合起来算。

定理 9.1 (枢轴量)

由于 $\bar{X} \sim N(\mu_1, \sigma_1^2/n_1)$ 且 $\bar{Y} \sim N(\mu_2, \sigma_2^2/n_2)$，且两者独立。根据正态分布的可加性，$\bar{X} - \bar{Y}$ 依然服从正态分布。

枢轴量为标准正态分布：

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

♡

例题 9.1 置信区间公式 $\mu_1 - \mu_2$ 的 $1 - \alpha$ 置信区间为：

$$\left[(\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \quad (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

笔记 [大样本近似] 如果 σ_1, σ_2 未知，但样本量非常大（$n_1, n_2 > 30$），根据大数定律，$s \approx \sigma$。此时也可以套用这个公式（用 s 替换 σ），并查 Z 表而不是 t 表。

9.4.2 1. 均值之差 $\mu_1 - \mu_2$ 的推断

笔记 [前提条件] 这里只讨论考试最常见的情形：

1. 两样本相互独立。
2. 均值 μ_1, μ_2 未知。
3. 方差 σ_1^2, σ_2^2 未知，但假设相等 ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)。

注：如果方差不等（*Behrens-Fisher* 问题），那是研究生课程的内容，本科大概率不考。

定义 9.4 (合并方差 Pooled Variance)

既然假设方差相等，那不如把两组样本合起来，利用所有的自由度来估计这个共同的方差 σ^2。

$$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

理解：这是 S_1^2 和 S_2^2 的加权平均，权重是各自的自由度。

♣

定理 9.2 (枢轴量)

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

♡

例题 9.2 置信区间公式 $\mu_1 - \mu_2$ 的 $1 - \alpha$ 置信区间为：

$$(\bar{x} - \bar{y}) \pm t_{\alpha/2}(n_1 + n_2 - 2) \cdot S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

9.4.3 2. 方差之比 σ_1^2/σ_2^2 的推断

笔记 [应用场景] 比较两台机器的精度。如果置信区间包含 1，说明两台机器精度相当。

定理 9.3 (枢轴量与区间)

枢轴量：$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$。

置信区间（注意左右不对称，且要交叉相除）：

$$\left[\frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\alpha/2}(n_1-1, n_2-1)}, \quad \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\alpha/2}(n_1-1, n_2-1)}\right]$$

技巧：查表时，$F_{1-\alpha/2}(n_1, n_2) = 1/F_{\alpha/2}(n_2, n_1)$。

♡

9.4.4 3. 陷阱：成对数据 (Paired Data)

笔记 [千万别用错公式] 如果题目说是“同一组人减肥前后的体重”或者“同一块地施肥左右两边的产量”，这是成对数据！

- 它们不独立（同一个人的体重肯定相关）。
- 解法：做差 $D_i = X_i - Y_i$。
- 然后把 D_i 看作一个新的单样本，对 D 的均值 μ_D 做 t 检验。
- 此时自由度是 $n-1$（n 为对数），而不是 $2n-2$。

第 10 章 抽样分布：统计量的幕后规律

内容提要

- ❑ 统计量的核心：统计量本身也是随机变量。
- ❑ 核心指标：我们希望统计量既“准”（无偏性），又“稳”（方差小）。
- ❑ 终极谜题：为什么计算样本方差时要除以 $n-1$？

10.1 统计量的评判标准：无偏性

定义 10.1 (无偏性 Unbiasedness)

设 $\hat{\theta}$ 是参数 θ 的估计量。如果 $\hat{\theta}$ 的数学期望恰好等于真值 θ，即：

$$E(\hat{\theta}) = \theta$$

则称 $\hat{\theta}$ 是 θ 的无偏估计量。

直观理解：虽然单次射击可能偏左或偏右，但平均来看，我们的准星是正对靶心的，没有系统性误差。

10.2 样本均值 $\bar{X}$：天选之子

定理 10.1 ($\bar{X}$ 的性质)

设总体 X 的均值为 μ，方差为 σ^2。无论总体服从什么分布，样本均值 $\bar{X} = \frac{1}{n}\sum X_i$ 满足：

1. 无偏性：$E(\bar{X}) = \mu$。
2. 收敛性：$D(\bar{X}) = \frac{\sigma^2}{n}$。

证明 [推导（口算级）] **1.** 期望：利用期望的线性性质：

$$E(\bar{X}) = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum E(X_i) = \frac{1}{n}\cdot n\mu = \mu$$

2. 方差：利用独立性（方差可加）：

$$D(\bar{X}) = D\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n^2}\sum D(X_i) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

笔记 [标准差 vs 标准误]

- σ (Standard Deviation)：描述单个个体的波动。
- $\sigma/\sqrt{n}$ (Standard Error)：描述均值的波动。
- n 越大，$\bar{X}$ 越稳定，这也正是大数定律的基础。

10.3 样本方差 S^2：为什么是 $n-1$？

问题 10.1 如果我们要估计总体的波动 σ^2，直觉告诉我们要算“平均平方误差”。但为什么定义的公式是除以 $n-1$ 而不是 n？

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

证明 [硬核推导：揭秘贝塞尔校正 (Bessel's Correction)] 我们需要证明 $E(S^2) = \sigma^2$。关键在于计算 $\sum(X_i - \bar{X})^2$ 的期望。

第一步：利用恒等式展开

$$\sum(X_i-\bar{X})^2=\sum X_i^2-n\bar{X}^2$$

(这个恒等式类似方差公式 $D(X)=E(X^2)-(EX)^2$，非常重要)

第二步：利用 $E(Y^2)=D(Y)+[E(Y)]^2$ 转换期望

- 对于单个样本 X_i：

$$E(X_i^2)=D(X_i)+[E(X_i)]^2=\sigma^2+\mu^2$$

- 对于样本均值 $\bar{X}$（注意 $D(\bar{X})=\sigma^2/n$）：

$$E(\bar{X}^2)=D(\bar{X})+[E(\bar{X})]^2=\frac{\sigma^2}{n}+\mu^2$$

第三步：求期望并代入

$$\begin{aligned}E\left[\sum(X_i-\bar{X})^2\right]&=\sum E(X_i^2)-nE(\bar{X}^2)\\&=n(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)\\&=n\sigma^2+n\mu^2-\sigma^2-n\mu^2\\&=(n-1)\sigma^2\end{aligned}$$

结论：如果我们除以 n，期望就是 $\frac{n-1}{n}\sigma^2$，这比真实的 σ^2 偏小了（有偏）。为了把这个系数 $(n-1)$ 消掉，我们必须除以 $n-1$：

$$E\left(\frac{1}{n-1}\sum(X_i-\bar{X})^2\right)=\frac{1}{n-1}\cdot(n-1)\sigma^2=\sigma^2$$

10.4 直观解释：自由度的代价

笔记 [为什么会“偏小”？] **1.** 几何解释：$\sum(X_i-a)^2$ 在 $a=\bar{X}$ 时取最小值。真正的方差是数据到真值 μ 的距离，而我们算的是数据到均值 $\bar{X}$ 的距离。因为 $\bar{X}$ 是根据这组数据“凑”出来的，它天然地比 μ 更靠近数据中心。所以 $\sum(X_i-\bar{X})^2<\sum(X_i-\mu)^2$。我们低估了波动，所以要除以一个更小的数 $(n-1)$ 来把结果“拉”回来。

2. 自由度解释：我们有 n 个独立的数据 $X_1,\ldots,X_n$。但在计算 S^2 时，我们用到了 $\bar{X}$。这就引入了一个约束条件：$\sum(X_i-\bar{X})=0$。这导致只有 $n-1$ 个数据是“自由”变动的（最后一个数据被前 $n-1$ 个和均值锁死了）。消耗了一个自由度用来估计 μ，所以剩下的自由度只剩 $n-1$。

第 11 章 假设检验的内核：决策与风险

内容提要

- ❑ 假设检验不是为了证明什么是对的，而是为了 ** 拒绝什么是错的 **。
- ❑ 核心哲学：** 疑罪从无 **（Innocent until proven guilty）。
- ❑ 本章重点：两类错误的定义与计算（必考大题）。

11.1 设立假设：立靶子的艺术

笔记 [黄金法则] 原假设 H_0 **(Null Hypothesis)**：代表"维持现状"、"无显著差异"、"没有效果"。备择假设 H_1 **(Alternative Hypothesis)**：代表"我们想证明的"、"有显著变化"、"新发现"。

判别口诀：

- 相等号 $(=, \leq, \geq)$ 永远在 H_0。
- 我们想极力证明的结论放在 H_1。

例题 11.1 实战判别 **场景 A：**某药厂宣称新药比旧药（疗效 μ_0）更有效。

- 潜台词：我想证明 $\mu > \mu_0$。
- 设定：$H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$（单侧）。

场景 B：检验车床切割的零件直径是否等于标准值 μ_0。

- 潜台词：不论太大还是太小都是次品，我想抓出次品。
- 设定：$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$（双侧）。

11.2 两类错误：弃真与取伪

内容提要

- ❑ 我们做出的决策（拒绝/接受 H_0）与事实真相（H_0 真/假）可能不一致。
- ❑ 这就会导致两种不同性质的错误。

表 11.1: 两类错误的混淆矩阵（以法庭判案为例）

事实真相	我们的裁决 接受 H_0 (判无罪)	 拒绝 H_0 (判有罪)
H_0 **为真** (嫌疑人无辜)	正确 (无罪释放)	**第一类错误 (Type I)** (**弃真**：冤枉好人)
H_0 **为假** (嫌疑人有罪)	**第二类错误 (Type II)** (**取伪**：放过坏人)	正确 (罪有应得)

笔记 [概率定义]

- 第一类错误概率 (α)：

$$\alpha = P(\text{拒绝} H_0 \mid H_0 \text{ 为真})$$

这就是我们也称之为显著性水平的东西。我们通常控制它很小（0.05），意味着"即使冤枉好人，概率也不能超过 5%"。

- 第二类错误概率 (β)：

$$\beta = P(\text{接受}H_0 \mid H_0 \text{ 为假})$$

这是“漏网之鱼”的概率。
- 功效 (**Power**)：$1-\beta$。即“抓坏人一抓一个准”的能力。

问题 11.1 设总体 $X \sim N(\mu, 100)$。我们要检验：$H_0 : \mu = 50 \quad \text{vs} \quad H_1 : \mu < 50$。采用样本量 $n = 25$，拒绝域定为：$\bar{x} \leq 47$。

1. 求犯第一类错误的概率 α。
2. 若真实的 $\mu = 45$，求犯第二类错误的概率 β。

解 已知 $\sigma = 10, n = 25$，故样本均值 $\bar{X}$ 的分布标准差（标准误）为 $\sigma/\sqrt{n} = 10/5 = 2$。

1. 计算 α（弃真）假设 H_0 为真，即 $\mu = 50$。此时 $\bar{X} \sim N(50, 2^2)$。我们却落入了拒绝域 ($\bar{x} \leq 47$) 的概率：

$$\alpha = P(\bar{X} \leq 47 \mid \mu = 50)$$

标准化：

$$= P\left(\frac{\bar{X}-50}{2} \leq \frac{47-50}{2}\right) = P(Z \leq -1.5) = \Phi(-1.5)$$

查表得 $\alpha = 0.0668$。

2. 计算 β（取伪）假设 H_0 为假，且真实的 $\mu = 45$。此时 $\bar{X} \sim N(45, 2^2)$。我们要计算没有拒绝 H_0（即没有落入拒绝域，$\bar{x} > 47$）的概率：

$$\beta = P(\bar{X} > 47 \mid \mu = 45)$$

标准化：

$$= P\left(\frac{\bar{X}-45}{2} > \frac{47-45}{2}\right) = P(Z > 1)$$
$$= 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

结论：在此规则下，我们有 6.68% 的概率冤枉好人（α），有 15.87% 的概率放过坏人（β）。

注[两难困境] 观察上图（或想象一下）：
- 想要减小 α（少冤枉好人），就要把拒绝域的门槛定得更严（比如 $\bar{x} \leq 45$）。
- 这样一来，拒绝域缩小，接受域变大，导致 β（漏网之鱼）必然变大。
- **唯一能同时减小 α 和 β 的方法**：**增加样本量** n（让分布变瘦，重叠面积变小）。

第 12 章 期末实战刷题（真题改编）

内容提要

❑ 这里的题目源自真实考题，难度适中，覆盖面广。 ❑ 建议先遮住答案自己做一遍，再对照解析。

12.1 实战演练（请独立完成）

1. **[多维分布]** 设随机变量 X 与 Y 相互独立，且均服从 $[0,a]$ 上的均匀分布（常数 $a>0$）。定义 $Z=|X-Y|$。试求：
   (a). (X,Y) 的联合概率密度 $f(x,y)$；
   (b). Z 的分布函数 $F_Z(z)$ 和概率密度 $f_Z(z)$。
2. **[假设检验]**
   (a). 在假设检验中，原假设 H_0 不真但被接受，这种判断错误称为第____ 类错误；原假设 H_0 正确但被拒绝，称为第____ 类错误。
   (b). 某零件长度服从正态分布 $N(\mu,\sigma^2)$，标准要求均值为 10.5。今测得 36 个长度数据，算得样本均值 $\bar{x}=11.08$，样本标准差 $s=0.516$。问在显著性水平 $\alpha=0.05$ 下，该零件长度是否符合要求? (参考临界值：$t_{0.025}(35)\approx 2.03$)
3. **[参数估计]** 总体 X 的概率密度函数为

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

   其中 $\theta>1$ 为未知参数，$x_1,\ldots,x_n$ 为样本观测值。
   (a). 求参数 θ 的矩估计量 $\hat{\theta}_1$；
   (b). 求参数 θ 的极大似然估计量 $\hat{\theta}_2$；
   (c). 求 $R=P(X>1)$ 的极大似然估计。
4. **[正态性质]** 设二维随机变量 $(X,Y)\sim N(1,2^2;2,3^2;\frac{1}{3})$（即 $\mu_X=1,\sigma_X=2,\mu_Y=2,\sigma_Y=3,\rho=\frac{1}{3}$），则 $D(X-2Y+5)=$______。
5. **[无偏估计]** 设 $X_1,\ldots,X_n$ 为来自总体 X 的样本，均值 μ，方差 σ^2。若使得统计量 $C\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2$ 为 σ^2 的无偏估计，则常数 $C=$______。
6. **[抽样分布]** 设总体 $X\sim N(\mu,\sigma^2)$，样本均值为 $\bar{X}$。则 $X_1-\bar{X}$ 服从的分布是：

   a) $N(0,\frac{n+1}{n}\sigma^2)$　　b) $N(0,\frac{n-1}{n}\sigma^2)$

   c) $\chi^2(n)$　　d) $\chi^2(n-1)$
7. **[概率性质]** 设事件 A,B 同时发生时，必导致事件 C 发生（即 $AB\subset C$），则有：

   a) $P(C)\geq P(A)+P(B)-1$　　b) $C\subset AB$

   c) $P(AB)\geq P(C)$　　d) $P(C)\leq P(A)+P(B)-P(AB)$
8. **[期望计算]** 对某目标进行射击，每次相互独立，命中率为 $p(0<p<1)$。直到击中目标 n 次为止。试求子弹消耗量 X 的数学期望。
9. **[切比雪夫]** 设 $E(X)=-2,E(Y)=2,D(X)=1,D(Y)=4,\rho_{XY}=-0.5$。根据切比雪夫不等式，$P\{|X+Y|\geq 6\}\leq$______。
10. **高斯分布**设随机变量 $X\sim N(\mu,\sigma^2)$。试求 $E[|X-\mu|^3]$。

12.2 答案与解析

解 [1. 二维均匀分布] **(1)** 联合密度：因为 X,Y 独立且服从 $U[0,a]$，故联合密度为常数：

$$f(x,y)=f_X(x)f_Y(y)=\begin{cases}\frac{1}{a^2}, & 0\le x\le a, 0\le y\le a\\ 0, & \text{其他}\end{cases}$$

(2) Z 的分布：几何法：样本空间是边长为 a 的正方形，面积 $S=a^2$。事件 $\{Z\le z\}\Leftrightarrow\{|X-Y|\le z\}$ 对应对角线附近的条形区域。计算补集 $\{|X-Y|>z\}$（即左上角和右下角两个小三角形）更简单。当 $0\le z\le a$ 时，小三角形边长为 $a-z$，总面积为 $(a-z)^2$。

$$F_Z(z)=P(Z\le z)=1-\frac{(a-z)^2}{a^2}$$

求导得密度：

$$f_Z(z)=F_Z'(z)=\frac{2(a-z)}{a^2},\quad (0<z<a)$$

解 [2. 假设检验] (1) 二（取伪）；一（弃真）。

(2) 双侧 **t** 检验。

- 建立假设：$H_0:\mu=10.5$ vs $H_1:\mu\ne 10.5$
- 统计量：$T=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}=\frac{11.08-10.5}{0.516/6}=\frac{0.58}{0.086}\approx 6.74$
- 拒绝域：查表 $t_{0.025}(35)\approx 2.03$。
- 结论：因为 $|T|=6.74>2.03$，故拒绝 H_0。
- 答：该零件长度不符合要求。

解 [3. 参数估计] 该分布实际上是指数分布的变体（平移）。

(1) 矩估计：计算期望：令 $t=x(\theta-1)$，或直接利用指数分布 $E(Y)=1/\lambda$ 的性质。

$$E(X)=\int_0^{\infty}x(\theta-1)e^{-(\theta-1)x}dx=\frac{1}{\theta-1}$$

令 $E(X)=\bar{X}$，解得：$\bar{X}=\frac{1}{\theta-1}\implies\hat{\theta}_1=\frac{1}{\bar{X}}+1$。

(2) 极大似然估计：似然函数 $L(\theta)=(\theta-1)^n e^{-(\theta-1)\sum x_i}$。取对数：$\ln L=n\ln(\theta-1)-(\theta-1)\sum x_i$。求导：$\frac{n}{\theta-1}-\sum x_i=0\implies\hat{\theta}_2=\frac{1}{\bar{X}}+1$。

(3) R 的估计：利用 MLE 的不变性。先求 R 的表达式：

$$R=P(X>1)=\int_1^{\infty}(\theta-1)e^{-(\theta-1)x}dx=e^{-(\theta-1)}$$

代入 $\hat{\theta}$：

$$\hat{R}=e^{-(\frac{1}{\bar{X}}+1-1)}=e^{-1/\bar{X}}$$

解 [4. 方差运算] **32**。解析：$D(X-2Y+5)=D(X)+(-2)^2D(Y)+2\cdot1\cdot(-2)Cov(X,Y)$。常数 5 的方差为 0。$Cov(X,Y)=\rho\sigma_X\sigma_Y=\frac{1}{3}\cdot2\cdot3=2$。原式 $=2^2+4\cdot3^2-4\cdot2=4+36-8=32$。

解 [5. 无偏估计常数] $\frac{1}{2(n-1)}$。解析：先看单项 $E[(X_{i+1}-X_i)^2]$。X_{i+1},X_i 独立同分布，方差均为 σ^2。$D(X_{i+1}-X_i)=D(X_{i+1})+D(X_i)=2\sigma^2$。又因为 $E(X_{i+1}-X_i)=\mu-\mu=0$。所以 $E[(X_{i+1}-X_i)^2]=D+E^2=2\sigma^2$。求和共有 $n-1$ 项，总期望为 $(n-1)2\sigma^2$。要等于 σ^2，必须 $C\cdot2(n-1)\sigma^2=\sigma^2\implies C=\frac{1}{2(n-1)}$。

解 [6. 抽样分布] **B**。解析：$X_1-\bar{X}$ 是正态变量的线性组合，必为正态，均值为 0。方差计算：$X_1-\bar{X}=\frac{n-1}{n}X_1-\frac{1}{n}X_2-\cdots-\frac{1}{n}X_n$。利用独立性：$D=(\frac{n-1}{n})^2\sigma^2+(n-1)(\frac{1}{n})^2\sigma^2=\frac{(n-1)^2+n-1}{n^2}\sigma^2=\frac{n-1}{n}\sigma^2$。

解 [7. 概率不等式] **A**。解析：$AB\subset C\implies P(AB)\le P(C)$。由加法公式 $P(A\cup B)=P(A)+P(B)-P(AB)\le1$。可得 $P(AB)\ge P(A)+P(B)-1$。故 $P(C)\ge P(AB)\ge P(A)+P(B)-1$。

解 [8. 期望计算] n/p。解析：这是 ** 负二项分布 **（Negative Binomial）。可以理解为：要打中 1 次，平均需要 $1/p$ 发（几何分布期望）。要打中 n 次，因为每次试验独立，就是 n 个几何分布之和。期望 $E(X)=n\cdot\frac{1}{p}=\frac{n}{p}$。

解 [9. 切比雪夫不等式] 1/12。解析：构造随机变量 $Z=X+Y$。$E(Z)=-2+2=0$。$D(Z)=D(X)+D(Y)+$

$2\rho\sqrt{D(X)D(Y)} = 1+4+2(-0.5)\sqrt{1\cdot 4} = 5-2 = 3$。切比雪夫不等式：$P\{|Z-E(Z)| \geq 6\} = P\{|X+Y| \geq 6\} \leq \frac{D(Z)}{6^2} = \frac{3}{36} = \frac{1}{12}$。

解 [10. 高斯分布] **1.** 标准化（化繁为简）直接算 X 的积分太麻烦，先转化为标准正态分布。令 $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$，则 $X-\mu = \sigma Z$。

$$E[|X-\mu|^3] = E[|\sigma Z|^3] = \sigma^3 E[|Z|^3]$$

2. 利用对称性写出积分 Z 的密度函数 $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ 是偶函数，$|z|^3$ 也是偶函数。所以乘积是偶函数，积分区间可以从 $(-\infty,+\infty)$ 变为 $2\times(0,+\infty)$。

$$E[|Z|^3] = \int_{-\infty}^{+\infty} |z|^3 \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz = 2\int_0^{+\infty} z^3 \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$$

提常数出来：

$$= \sqrt{\frac{2}{\pi}}\int_0^{+\infty} z^3 e^{-z^2/2}dz$$

3. 换元积分（凑微分）观察到 $z^3e^{-z^2/2} = z^2\cdot(ze^{-z^2/2})$。令 $t = z^2/2$，则 $dt = zdz$，且 $z^2 = 2t$。积分限：$z\in(0,+\infty) \implies t\in(0,+\infty)$。

$$\int_0^{+\infty} z^2\cdot e^{-z^2/2}\cdot zdz = \int_0^{+\infty}(2t)\cdot e^{-t}\cdot dt = 2\int_0^{+\infty} te^{-t}dt$$

利用分部积分（或者 $\Gamma(2) = 1!$）：

$$\int_0^{+\infty} te^{-t}dt = \left[-te^{-t}\right]_0^{+\infty} + \int_0^{+\infty} e^{-t}dt = 0+1 = 1$$

4. 回代结果

$$E[|Z|^3] = \sqrt{\frac{2}{\pi}}\cdot 2\cdot 1 = 2\sqrt{\frac{2}{\pi}}$$

最后乘上 σ^3：

$$E[|X-\mu|^3] = 2\sqrt{\frac{2}{\pi}}\sigma^3$$