

PROBABILITY AND STATISTICS

概率论与数理统计

期中练习册

第 1-3 章；第 4 章至条件矩

Raisetsu41

2026 Autumn

26 October 2026

目录

第 1 章 概率、计数与分布函数	1
第 2 章 变换、期望与矩	23
第 3 章 常见分布族	41
第 4 章 联合分布与条件矩	65

第1章 概率、计数与分布函数

Probability, Counting and Distribution Functions

依据：CB 2e Ch1；本学期 Chapter 1 资料

1.1 从试验到可计算的事件

Experiments, events and probability

PRINCIPLE 先定义样本点和事件，再选择计数或概率规则。

随机试验 (random experiment) 的全部可能结果构成样本空间 (sample space, S)。事件 (event) 是可被赋予概率的结果集合。事件族必须形成 sigma 代数 (sigma-algebra, $\mathcal{B}$)：包含 S 、对补集 (complement) 与可数并 (countable union) 封闭，因此也对可数交封闭。有限空间上可用原子 (atoms, 即不能被事件族进一步区分的非空基本块) 的所有并来构造事件族。

概率测度 (probability measure, P) 满足非负性、 $P(S) = 1$ 和可数可加性 (countable additivity)：互不相交事件的并概率等于概率之和。互斥 (disjoint/-mutually exclusive) 描述集合交为空；独立 (independent) 描述概率乘法关系，两者不是同义词。

$$P(A^c) = 1 - P(A), \quad P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

$$P\left(\bigcup_i A_i\right) \leq \sum_i P(A_i), \quad P(A \cap B) \geq P(A) + P(B) - 1.$$

并集上界 (union bound) 不要求独立；把“全部成功”取补为“至少一个失败”后尤其有用。等可能模型 (equally likely model) 才可写 $P(A) = \#A/\#S$ ，不能把任意不同的结果类别当作等可能样本点。

1.2 排列组合与抽样机制

Counting and sampling mechanisms

PRINCIPLE 先问是否区分顺序、是否放回、对象是否相同，再选公式。

排列 (permutation) 记录顺序, 组合 (combination) 不记录顺序。\$n\$ 个不同对象的有序全排列为 \$n!\$; 选 \$k\$ 个不同对象的无序子集为 \$\binom{n}{k} = n!/[k!(n-k)!]\$。重复字母不可区分时, 除以每种字母重复数的阶乘; 要求连续出现的字块可先视作单个对象, 再检查是否存在重叠。

$$\text{重复字母排列数} = \frac{n!}{n_1! \cdots n_j!}, \quad \text{无标签等大小分组} = \frac{(kn)!}{(k!)^n n!}.$$

无标签 (unlabeled) 分组还需除掉组与组交换造成的重复; 有固定组名时不能除。无放回抽样 (sampling without replacement) 改变后续成功概率, 产生超几何模型; 独立放回且成功概率固定时才能使用二项模型。计数题的分母和分子必须采用同一种顺序口径。

1.3 条件概率、全概率与 Bayes

Conditioning, total probability and Bayes' theorem

PRINCIPLE 先写所求条件方向, 再展开分母中所有可能来源。

条件概率 (conditional probability) 是在已知 \$B\$ 发生后对样本空间重新归一化: \$P(A | B) = P(A \cap B)/P(B)\$, 要求 \$P(B) > 0\$。划分 (partition) 是一组互斥且覆盖全空间的事件 \$C_i\$。

$$P(A) = \sum_i P(A | C_i)P(C_i), \quad P(C_j | A) = \frac{P(A | C_j)P(C_j)}{\sum_i P(A | C_i)P(C_i)}.$$

先验 (prior) \$P(C_j)\$ 表示观察前类型概率; 似然 (likelihood) \$P(A | C_j)\$ 表示给定类型产生观察的概率; 后验 (posterior) \$P(C_j | A)\$ 是观察后更新的类型概率。Bayes 定理通过联合概率连接相反方向的条件, 不能直接把 \$P(A | B)\$ 当 \$P(B | A)\$。

筛查中患病率 (prevalence) 是先验; 灵敏度 (sensitivity) 是患者阳性率; 特异度 (specificity) 是健康人阴性率。假阳性率是 1 减特异度。阳性预测值 (positive predictive value) 才是阳性之后患病的后验, 强烈依赖基准率。

独立性 \$P(A \cap B) = P(A)P(B)\$ 在正概率条件下等价于 \$P(A | B) = P(A)\$。条件独立 (conditional independence) 为 \$P(A \cap B | C) = P(A | C)P(B | C)\$。共同未知类型可能使给定类型独立的观测在边际上依赖, 不能先混合再无依据相

乘。多个事件两两独立 (pairwise independence) 也不自动推出相互独立 (mutual independence)。

1.4 随机变量如何形成分布

Random variables, PMFs, PDFs and CDFs

PRINCIPLE CDF 统一离散与连续；密度是积分的权重，不是单点概率。

随机变量 (random variable, X) 把样本结果映射为实数；其分布 (distribution/law) 说明概率如何分配到数轴。支持集 (support) 记录可出现的取值范围，连续题的开闭端点通常不影响积分，但离散题端点不可忽略。

离散变量的概率质量函数 (probability mass function, PMF) $p(x) = P(X = x)$ 非负、总和为 1。绝对连续变量的概率密度函数 (probability density function, PDF) $f(x)$ 非负、积分为 1，区间概率等于积分，单点概率为 0；密度可大于 1。分布函数 (cumulative distribution function, CDF) 统一定义为 $F(x) = P(X \leq x)$ ，非减、右连续、两端极限为 0 和 1。

$$P(a < X \leq b) = F(b) - F(a), \quad P(X = x) = F(x) - F(x^-),$$

$$P(X < x) = F(x^-), \quad f(x) = F'(x) \quad (\text{绝对连续时几乎处处}).$$

左极限 (left limit) $F(x^-)$ 不包括 x 点；离散 CDF 的跳跃恰好是该点质量。生存函数 (survival function) $S(x) = P(X > x) = 1 - F(x)$ 。从题设候选函数出发，先检验非负、归一化或 CDF 性质，再计算概率；不是所有看似光滑的函数都是分布函数。

1.5 英文概念与原理

English Definitions and Principles

1. Experiments, events and probability

A sample space contains the outcomes of an experiment. Events belong to a sigma-algebra, which contains the full space and is closed under complements and countable unions. In a finite space, all unions of its atoms form such a family.

A probability measure is nonnegative, assigns one to the full space and is countably additive on disjoint events.

Disjointness is a set relation; independence is a probability relation. Complementation, inclusion–exclusion and the union bound do not require independence. Counting favorable outcomes divided by all outcomes is valid only when the chosen elementary outcomes are equally likely.

2. Counting and sampling mechanisms

Permutations retain order and combinations ignore it. Repeated indistinguishable objects require division by their multiplicity factorials. Required consecutive strings can be treated as blocks, subject to overlap checks.

Unlabeled equal-size groups require division both for internal order and for exchange of entire groups; labeled groups do not require the latter factor.

Sampling without replacement changes subsequent success chances and leads to a hypergeometric count. Independent fixed-probability draws lead to a binomial count. Numerator and denominator in a counting probability must use consistent ordering conventions.

3. Conditioning, total probability and Bayes' theorem

Conditional probability restricts and renormalizes the sample space. A partition consists of disjoint events covering it. Total probability sums the contributions of all possible types; Bayes' theorem reverses a conditional direction through joint probabilities.

The prior is a type probability before observation, the likelihood is the observation probability under that type, and the posterior is the updated type probability.

In screening, prevalence is the prior, sensitivity is the true-positive rate, and specificity the true-negative rate. The positive predictive value is a posterior, not sensitivity. Conditional independence does not imply marginal independence when observations share an unknown type.

Pairwise independence also need not imply mutual independence.

4. Random variables, PMFs, PDFs and CDFs

A random variable maps outcomes to numbers. Its support specifies feasible values. A discrete PMF assigns nonnegative point masses summing to one. An absolutely continuous PDF is nonnegative and integrates to one; interval probabilities are integrals and point probabilities vanish. A density may exceed one.

The CDF is $P(X \leq x)$ for every distribution. It is nondecreasing, right-continuous, and has limits zero and one. Its jump is point mass, while the left limit excludes that point. Survival is $1 - F(x)$. Always validate a proposed distribution before using it; differentiating an arbitrary function does not make it a PDF.

P1-01 两枚硬币能否得到均匀的正面数

2026 PS1 Q1; CB 1.6

independence 独立性; sample space 样本空间

Two independent coins have head probabilities u and w . Let X be the number of heads. Can $P(X = 0) = P(X = 1) = P(X = 2)$ hold? Prove your answer, allowing biased coins.

P1-02 事件相交的界与导弹命中

2026 PS1 Q2、Q7; CB 1.13

union bound 并集上界; complement 补事件

1. If $P(A) = 1/3$ and $P(B^c) = 1/4$, can A and B be disjoint? Give a numerical bound.
2. Three missiles each hit with probability 0.7.
Without assuming independence, give a guaranteed lower bound on the probability that all hit.
3. If the three hits are independent, find the probability of at least one hit.

P1-03 掷到首次出现 6

2026 PS1 Q3; CB 1.26

first success 首次成功; geometric distribution 几何分布

A fair die is rolled independently until the first 6 appears. Let T count all rolls, including the successful roll. Find $P(T > 5)$, and explain why $P(T = 5)$ is different.

P1-04 电报误码与后验序列

2026 PS1 Q4; CB 1.41

prior 先验; likelihood 似然; posterior 后验; total probability 全概率

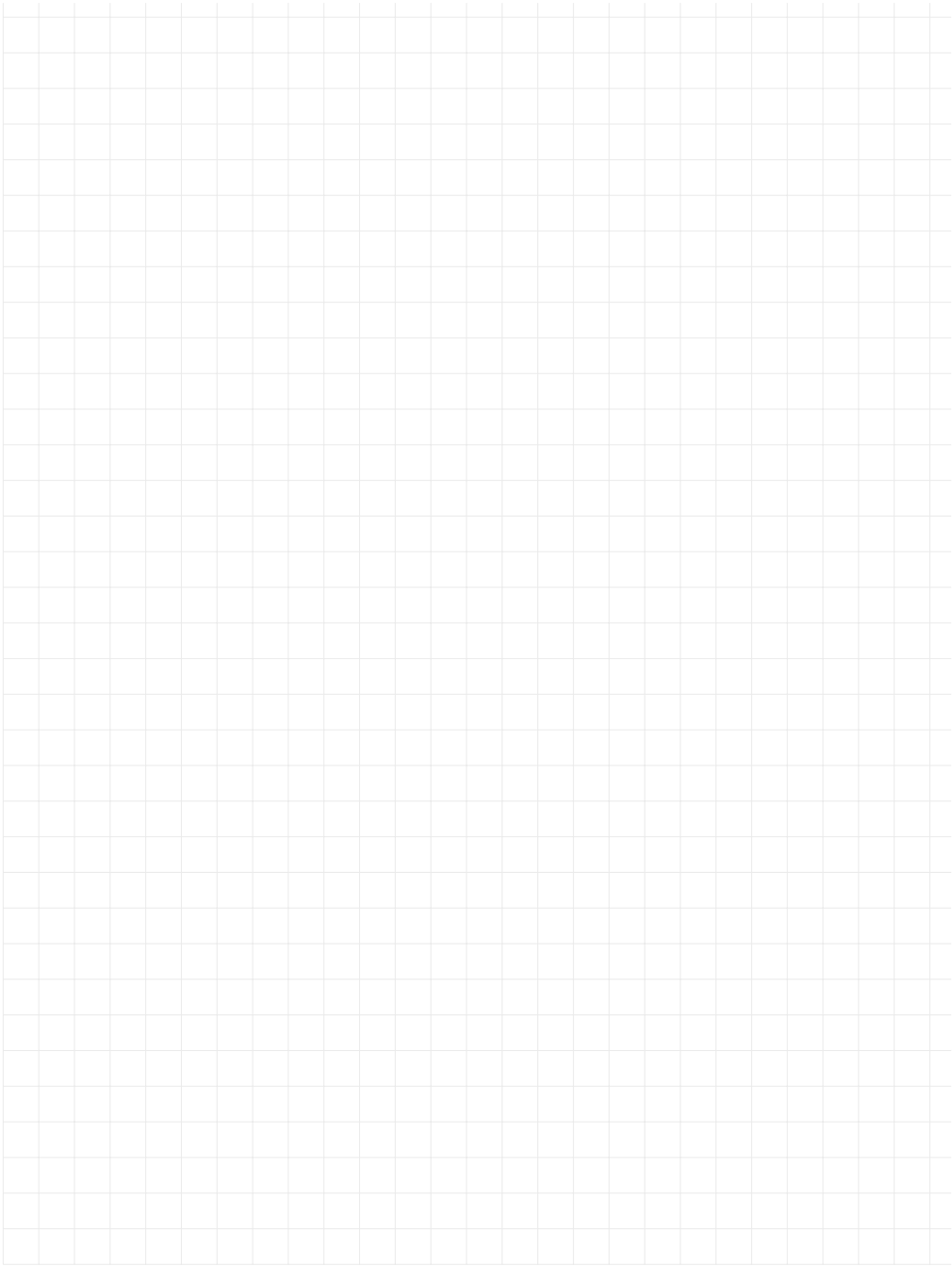
Dots and dashes are sent in the ratio 3 : 4.

A sent dot is received as a dash with probability $1/4$, and a sent dash is received as a dot with probability $1/3$.

1. Given a received dash, find the probability that a dash was sent.
2. Given two received dots, find the probabilities of all four possible sent sequences.

For this part assume independent source symbols and conditionally independent transmission errors across uses.

P1-04 Working space



P1-05 罕见疾病筛查与基准率

2026 PS1 Q10

sensitivity 灵敏度; specificity 特异度; prevalence 患病率

A disease has prevalence 0.01. A test has sensitivity 0.95 and specificity 0.95.

1. Find the probability of disease given a positive test.
2. Replace prevalence by γ .

Derive the posterior as a function of γ , prove that it increases, and find the threshold at which it exceeds $1/2$.

P1-06 尾概率重叠在单点

2026 PS1 Q8

point mass 点概率; inclusion-exclusion 容斥

For an integer-valued random variable, $P(X \geq 2) = a$ and $P(X \leq 2) = b$. Find $P(X = 2)$ and the constraint implied for $a + b$.

P1-07 有限样本空间上的 sigma 代数

2026 PS1 Q9; 生成代数方法拓展

sigma-algebra sigma 代数; closure 封闭性; atom 原子

Let $S = \{1, 2, 3, 4, 5, 6\}$ and

$$\mathcal{B} = \{S, \emptyset, \{1\}, \{2\}, \{1, 2\}, \{3, 4, 5, 6\}\}.$$

Is $\mathcal{B}$ a sigma-algebra? If not, identify a failed condition and write the smallest sigma-algebra containing it.

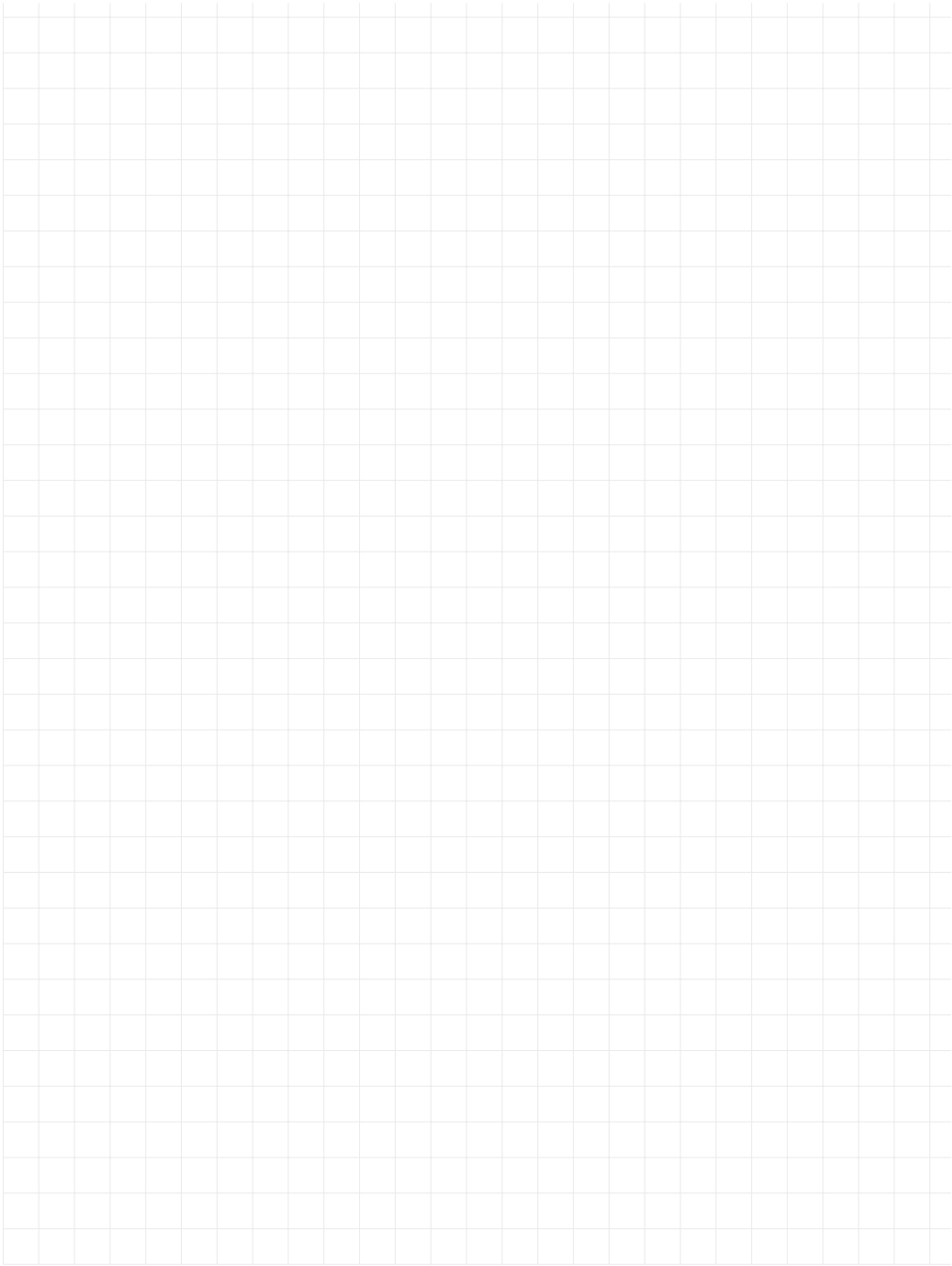
P1-08 不放回抽取与阶梯 CDF

2026 PS1 Q5; CB 1.51; 矩为方法拓展

hypergeometric 超几何分布; without replacement 不放回; CDF 分布函数

Among 30 microwave ovens, 5 are defective. Select 4 uniformly without replacement. Let X count defective ovens selected. Find the support, PMF and complete CDF, and sketch the CDF. Also find its mean and variance using the hypergeometric formulas.

P1-08 Working space



P1-09 重复字母排列与指定字块

TA1 E6

permutation 排列; indistinguishable 相同不可区分; block 字块

Using exactly the letters a,e,e,m,r,r,y, count distinct arrangements (a) without restrictions; (b) containing the consecutive block “eye”; (c) containing both consecutive blocks “ram” and “eye”.

P1-10 无标签分组与过度计数

TA1 E8

unlabeled groups 无标签组; overcounting 重复计数

Count the partitions of 6 distinct people into (a) two unlabeled groups of 3; (b) three unlabeled pairs. Generalize to kn distinct people divided into n unlabeled groups of k people.

P1-11 未知硬币造成的条件依赖

TA1 E12; 第二次投掷为方法拓展

conditional independence 条件独立; mixture 混合

One of three coins is chosen uniformly: two are fair and one has two heads. The chosen coin is tossed twice. Given that the first toss is a head, find the probability that the coin is two-headed and the probability of a head on the second toss. Are the tosses independent before observing the coin type?

P1-12 从条件独立到无条件独立的证明

TA1 E14

partition 划分; conditional independence 条件独立

Let $\{C_i\}$ be a finite or countable partition with $P(C_i) > 0$. Suppose A and B are conditionally independent given every C_i , and B is independent of each C_i . Prove that A and B are independent.

P1-13 从 CDF 跳跃读出概率

TA2 E8

right continuity 右连续; jump 跳跃; left limit 左极限

The CDF is

$$F(x) = \begin{cases} 0 & x < -2, \\ 0.4 & -2 \leq x < 0, \\ 0.5 & 0 \leq x < 1, \\ 0.8 & 1 \leq x < 4, \\ 1 & x \geq 4. \end{cases}$$

Find the PMF, $P(-1 < X < 4)$, and $P(-1 < X < 1 \mid X \geq 0)$.

P1-14 概率函数的合法性与归一化

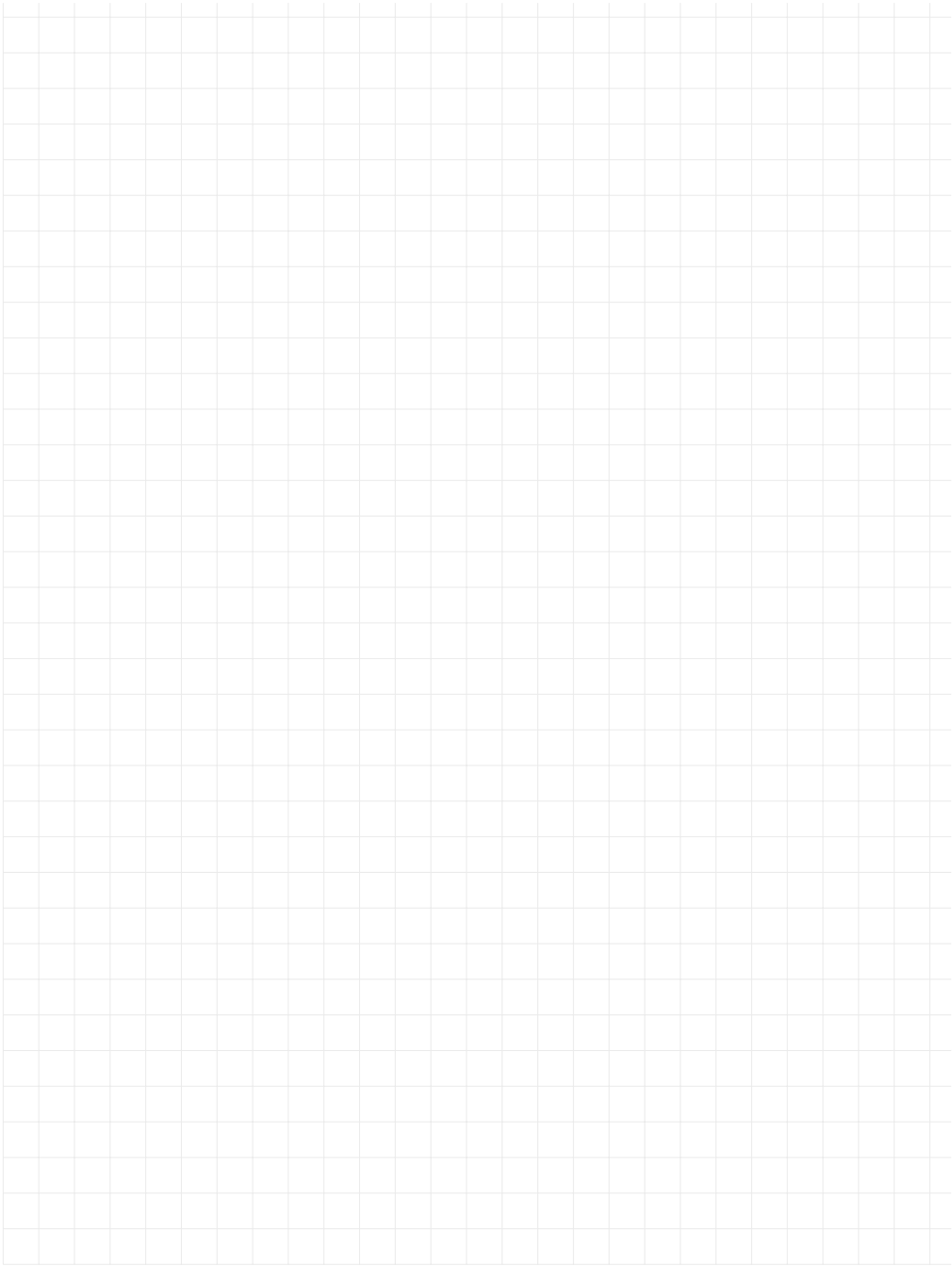
TA2 E9–E11; 同一合法性方法合并

normalization 归一化; nonnegativity 非负性; monotonicity 单调性

Determine which functions can define probability laws, giving reasons.

1. For integers $k \geq 1$: $1/k$, $(-1)^{k+1}/(k \ln 2)$, and $1/[k(k+1)]$.
2. Find c so that $f(x) = c(x+1)^2$ on $-1 < x < 1$ is a PDF.
3. Can $F(x) = (1-x^2)/(1+x^2)$ on the real line be a CDF?

P1-14 Working space



第2章 变换、期望与矩

Transformations, Expectations and Moments

依据: CB 2e Ch2; 本学期 Chapter 2 资料

2.1 函数变换先处理支持与事件

Transformations: support before algebra

PRINCIPLE 每条逆支都必须落在原支持内, 雅可比取绝对值。

变换 (transformation) $Y = g(X)$ 为同一个随机结果重新赋值。最通用方法是先找 g 的值域, 再写 $F_Y(y) = P(g(X) \leq y)$, 把事件还原为 X 的区间, 最后求导。递减变换会改变不等号方向; 平方、绝对值等多对一 (many-to-one) 函数可能产生多个逆像。

$$f_Y(y) = \sum_{j: x_j(y) \text{可行}} f_X(x_j(y)) \left| \frac{dx_j(y)}{dy} \right|.$$

这里 $x_j(y)$ 是逆分支 (inverse branches)。雅可比因子 (Jacobian factor) 将数轴长度变化转换为密度变化, 保证概率不因改单位而改变。不同 y 区间可行逆支数目可能变化, 必须分段。例如原支持 $(-1, 2)$ 上的平方变换, $0 < y < 1$ 有两支, $1 < y < 4$ 只剩正支。

对 $Y = aX + b, a \neq 0$, 密度为 $f_X((y - b)/a)/|a|$ 。结果应检查支持、非负、积分为 1、CDF 衔接连续。连续变量在有限个端点处的密度如何定义不改变分布。

2.2 分位数与概率积分变换

Quantiles and the probability integral transform

PRINCIPLE 分位数反解累计概率, 不是反解密度。

分位数 (quantile) 可定义为 $F^{-1}(u) = \inf\{x : F(x) \geq u\}$ 。中位数 (median) 将概率分成两半, 连续严格递增时解 $F(m) = 1/2$; 均值一般不等于中位数。

$$F \text{ 连续且在支持上严格递增} \Rightarrow F(X) \sim U(0, 1),$$

$$U \sim U(0, 1) \Rightarrow F^{-1}(U) \text{ 具有 CDF } F.$$

第一条为概率积分变换 (probability integral transform)，用事件反解证明： $P(F(X) \leq u) = P(X \leq F^{-1}(u)) = u$ 。逆变换抽样 (inverse-transform sampling) 把均匀数变成目标分布。离散时 $F(X)$ 通常仍是离散，不能无条件声称均匀；广义逆生成法却仍可使用。

2.3 期望、矩与方差

Expectation, moments and variance

PRINCIPLE 先判断存在性；能直接对原变量积分时不必先求新分布。

期望 (expectation) 是按概率加权的平均。无意识统计学家定律 (law of the unconscious statistician, LOTUS) 允许直接用 X 的分布计算 $g(X)$ 的期望：

$$E[g(X)] = \sum_x g(x)p_X(x) \quad \text{或} \quad \int g(x)f_X(x)dx,$$

$$\text{Var } X = E[(X - EX)^2] = EX^2 - (EX)^2.$$

第 k 阶原点矩 (raw moment) 是 EX^k ，中心矩 (central moment) 是 $E[(X - EX)^k]$ 。方差 (variance) 是二阶中心矩，标准差 (standard deviation) 是其平方根。期望线性性 (linearity) 不需要独立， $E(aX + bY) = aEX + bEY$ ；但方差相加需要不相关，独立是充分条件。

$$E(aX + b) = aEX + b, \quad \text{Var}(aX + b) = a^2 \text{Var } X.$$

有限期望要求 $E|X| < \infty$ ；正负部分同时发散时不能相互抵消。对称分布只有在期望存在时才可据对称中心确定均值；Cauchy 是重要反例。混合分布 (mixture) 先随机选类型，原点矩按类型概率加权，再减平方均值求方差；直接加权分量方差会漏掉组间差异。

2.4 矩母函数：积分、导数与识别

MGFs: existence, differentiation and identification

PRINCIPLE 先写收敛区间，再用 0 附近的导数或唯一性。

矩母函数 (moment generating function, MGF) 定义为 $M_X(t) = E(e^{tX})$ 。若其在 0 的某个邻域内有限，则在适当条件下 $M_X^{(k)}(0) = EX^k$ ，且这一邻域的 MGF 唯一决定分布。只写一个代数式却忽略积分发散的 t ，不是完整结果。

$$M_X(0) = 1, \quad EX = M'_X(0), \quad \text{Var } X = M''_X(0) - [M'_X(0)]^2,$$

$$M_{aX+b}(t) = e^{bt} M_X(at), \quad M_{X+Y}(t) = M_X(t) M_Y(t) \quad (\text{独立}).$$

有界支持保证所有实数 t 下有限；指数尾可能只允许半轴；双侧 Laplace 尾要求 $|t| < 1/\beta$ 。正态 MGF 对所有实数存在；Cauchy 没有 0 邻域 MGF。对数正态所有幂矩存在，却没有正 t 的有限 MGF，因此“有全部矩”不自动意味着有通常的 MGF。

推导时优先识别已知归一化积分，例如 Gamma 积分或高斯积分。和的 MGF 相乘必须有独立性；相同边际或期望可加都不能替代这个条件。

2.5 英文概念与原理

English Definitions and Principles

1. Transformations: support before algebra

For $Y = g(X)$, first find the image of the original support. The CDF method rewrites $g(X) \leq y$ as events in X and works even for nonmonotone maps. Inverse-density formulas must sum every feasible branch and include the absolute inverse derivative. The feasible branch count can change across intervals.

The Jacobian compensates for stretching of the coordinate axis. An affine transformation has density $f_X((y - b)/a)/|a|$. Check the new support, nonnegativity, normalization and CDF continuity. Density values at finitely many continuous endpoints do not affect probabilities.

2. Quantiles and the probability integral transform

A quantile inverts cumulative probability, using the generalized inverse if necessary. A continuous median solves $F(m) = 1/2$ and need not equal the mean. For continuous strictly increasing F , the probability integral transform makes $F(X)$ uniform.

Conversely, applying the generalized inverse to a uniform variable generates the target CDF. Discrete $F(X)$ need not be uniform, so conditions matter.

3. Expectation, moments and variance

Expectation is a probability-weighted average. LOTUS computes $E[g(X)]$ directly using the original PMF or PDF, without deriving the transformed law. Raw moments are EX^k ; central moments center at the mean. Variance is $EX^2 - (EX)^2$ and standard deviation is its square root.

Expectation is linear without independence. Variance scales by the square of a multiplier and is invariant to shifts. A finite expectation requires absolute integrability; divergent positive and negative parts cannot cancel. Symmetry alone does not guarantee a mean.

Mixture raw moments are weighted component moments, but mixture variance also reflects differences between component means.

4. MGFs: existence, differentiation and identification

The MGF is $E(e^{tX})$. When finite on a neighborhood of zero, derivatives yield raw moments and the MGF uniquely identifies the distribution. Its convergence domain is part of the answer. Affine changes transform the argument and add an exponential factor; independent sums multiply MGFs.

Bounded support gives finiteness for all real arguments. Exponential and Laplace tails restrict the domain. A normal MGF exists everywhere, while a Cauchy MGF does not exist near zero. A lognormal has all power moments but no finite MGF for positive arguments. Recognizing Gamma or Gaussian integrals often simplifies derivations.

P2-01 河流水位的分布变换

2026 PS1 Q6; CB 1.53

support 支持集; density 密度; affine transformation 仿射变换

Annual maximum river level Y has $F_Y(y) = 0$ for $y < 1$ and $F_Y(y) = 1 - y^{-2}$ for $y \geq 1$. Verify that it is a valid CDF, find its density, and derive the CDF and density of $Z = 10(Y - 1)$.

P2-02 指数变量的线性变换

TA2 E12; TA3 E5; CB 2.1(b)

rate 率参数; Jacobian 雅可比因子

X has density $7e^{-7x}$ for $x > 0$. For $Y = 4X + 3$, find the support, CDF and density, and check normalization. Clearly state whether 7 is a rate or a scale.

P2-03 双侧指数取绝对值

TA3 E1

Laplace distribution Laplace 分布; many-to-one 多对一

X has density $f_X(x) = \alpha e^{-\alpha|x|}/2$ on $\mathbb{R}$, where $\alpha > 0$. Find the CDF and density of $Y = |X|$, and identify its distribution using both rate and scale.

P2-04 抛物线变换的两条逆支

TA3 E2

inverse branch 逆分支; uniform distribution 均匀分布

$X \sim U(0, 2)$ and $Y = X(2 - X)$. Find the support, CDF and density of Y using both a CDF argument and inverse branches.

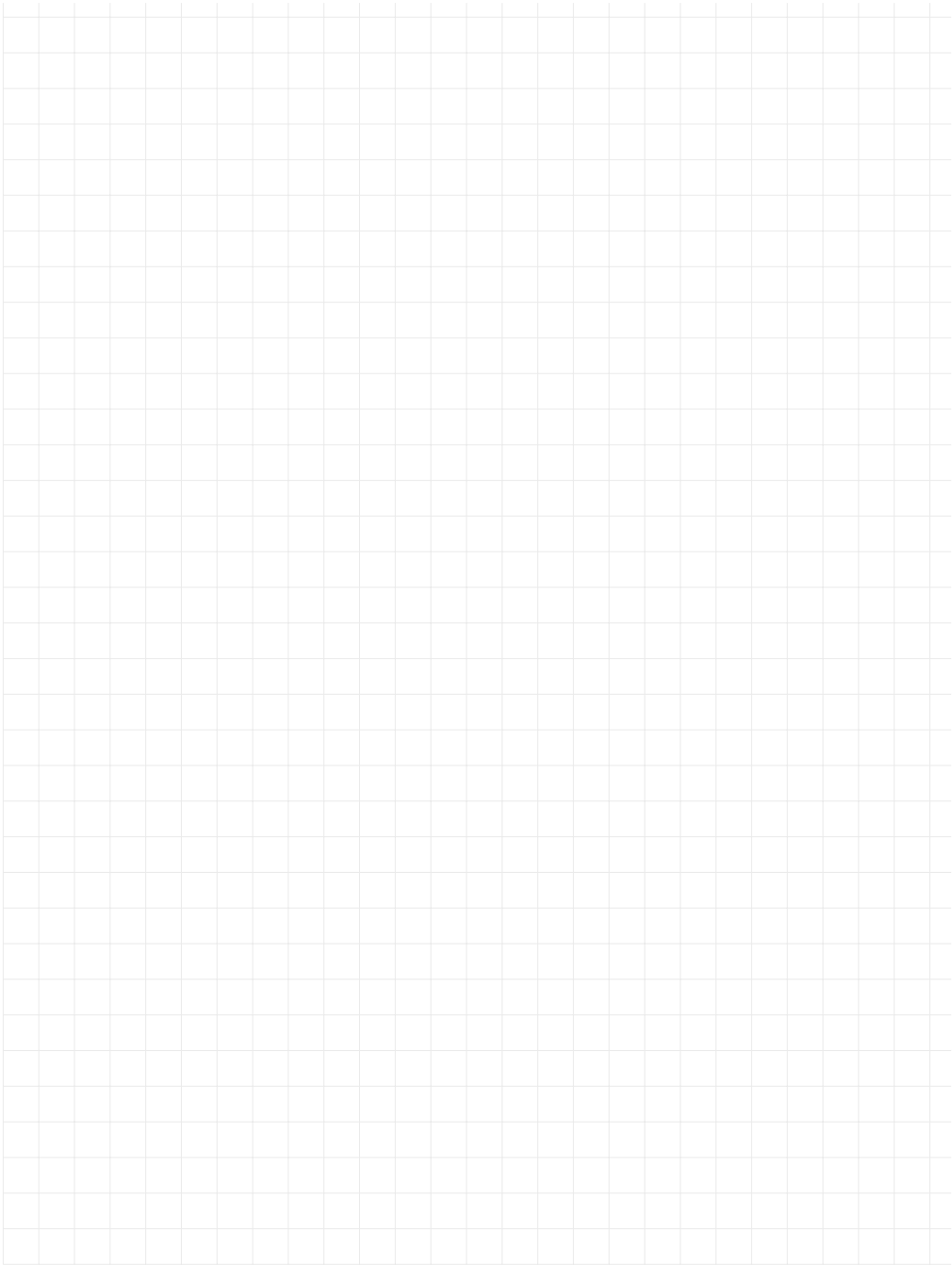
P2-05 非对称支持上的平方变换

TA3 E3

piecewise density 分段密度; LOTUS 无意识统计学家定律

X has density $x^2/3$ for $-1 < x < 2$. Let $Y = X^2$. Find the density and complete CDF of Y , and calculate EY directly from X as a check.

P2-05 Working space



P2-06 概率积分变换与逆变换

TA3 E4

probability integral transform 概率积分变换; quantile 分位数

Suppose X has a continuous strictly increasing CDF F on its support. Prove $F(X) \sim U(0, 1)$ and explain how to generate X from a uniform random variable. Give a discrete counterexample to the unqualified claim that $F(X)$ is always uniform.

P2-07 三种矩的计算与中位数

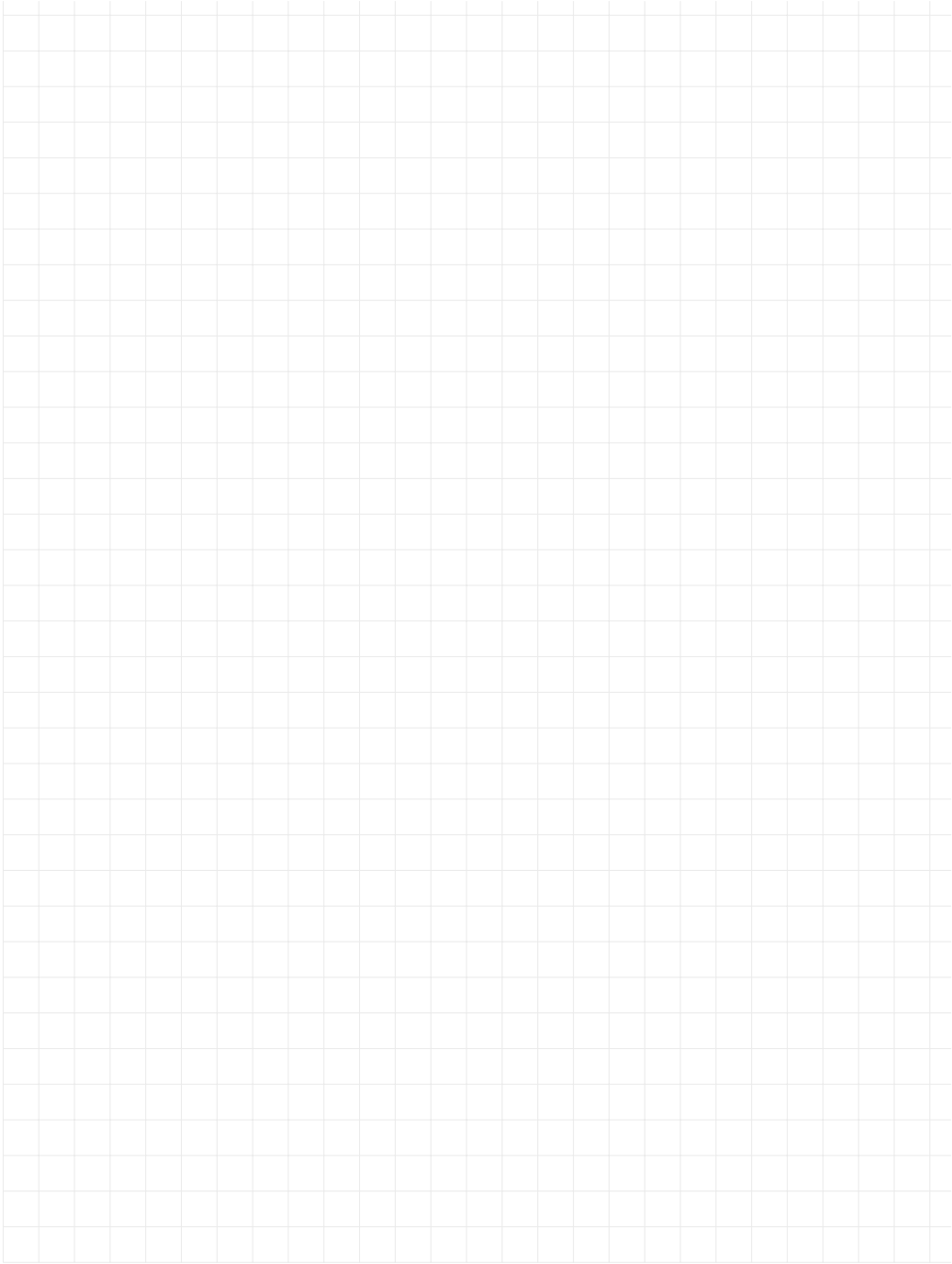
TA3 E6、E7; CB 2.24、2.17(a)

raw moment 原点矩; variance 方差; median 中位数

Find means and variances for:

1. $f(x) = ax^{a-1}$ on $(0, 1)$, $a > 0$. Also find the median and specialize it to $a = 3$.
2. X uniform on $\{1, \dots, n\}$.
3. $f(x) = \frac{3}{2}(x-1)^2$ on $(0, 2)$.

P2-07 Working space



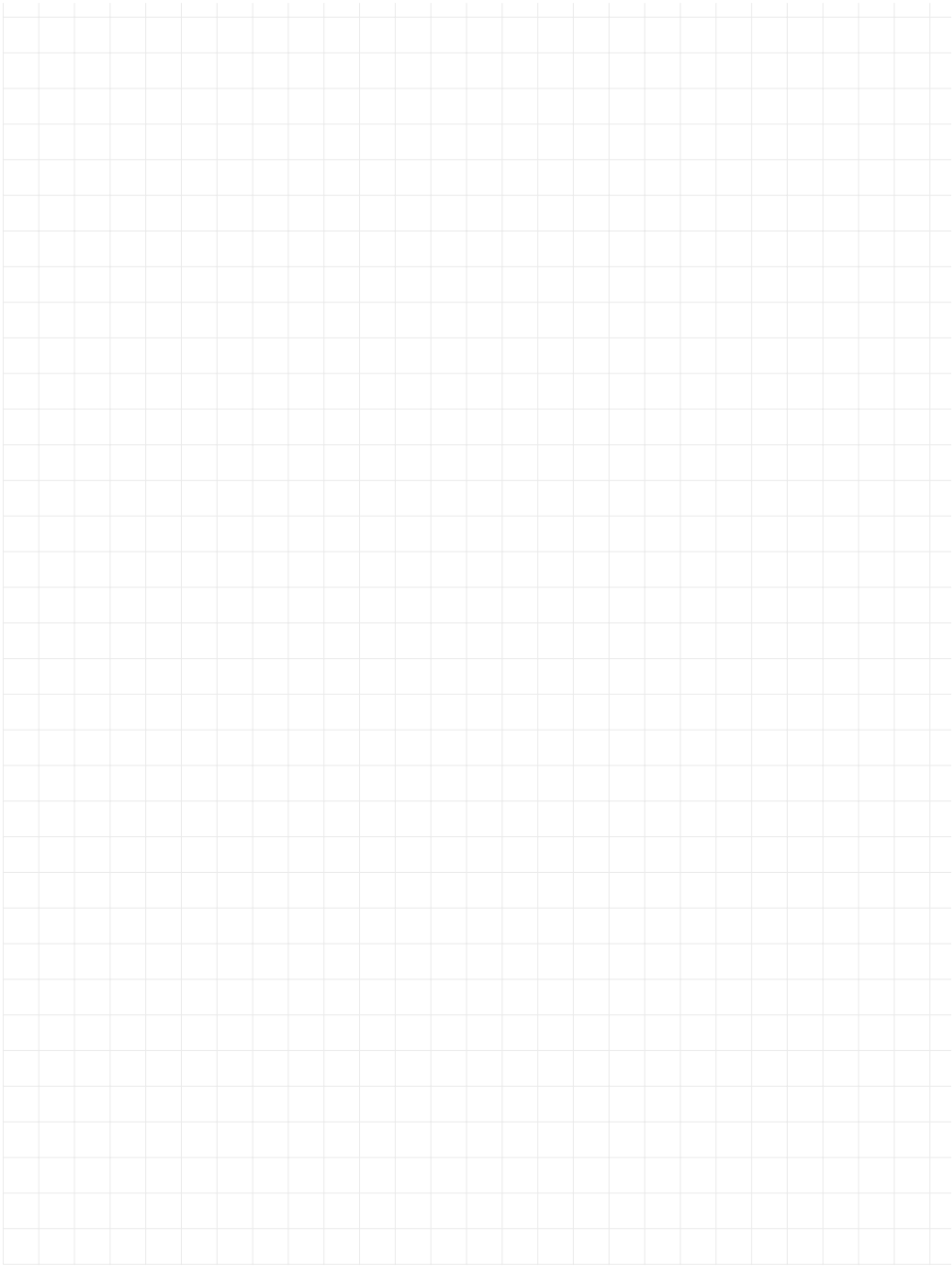
P2-08 混合寿命的生存函数

TA3 E8; CB 2.16

survival function 生存函数; memorylessness 无记忆性

For $t \geq 0$, $P(X > t) = ae^{-\lambda t} + (1 - a)e^{-\mu t}$, where $0 < a < 1$ and $\lambda, \mu > 0$. Find the CDF, density, mean and variance. Is X memoryless when $\lambda \neq \mu$?

P2-08 Working space



P2-09 矩母函数的直接积分

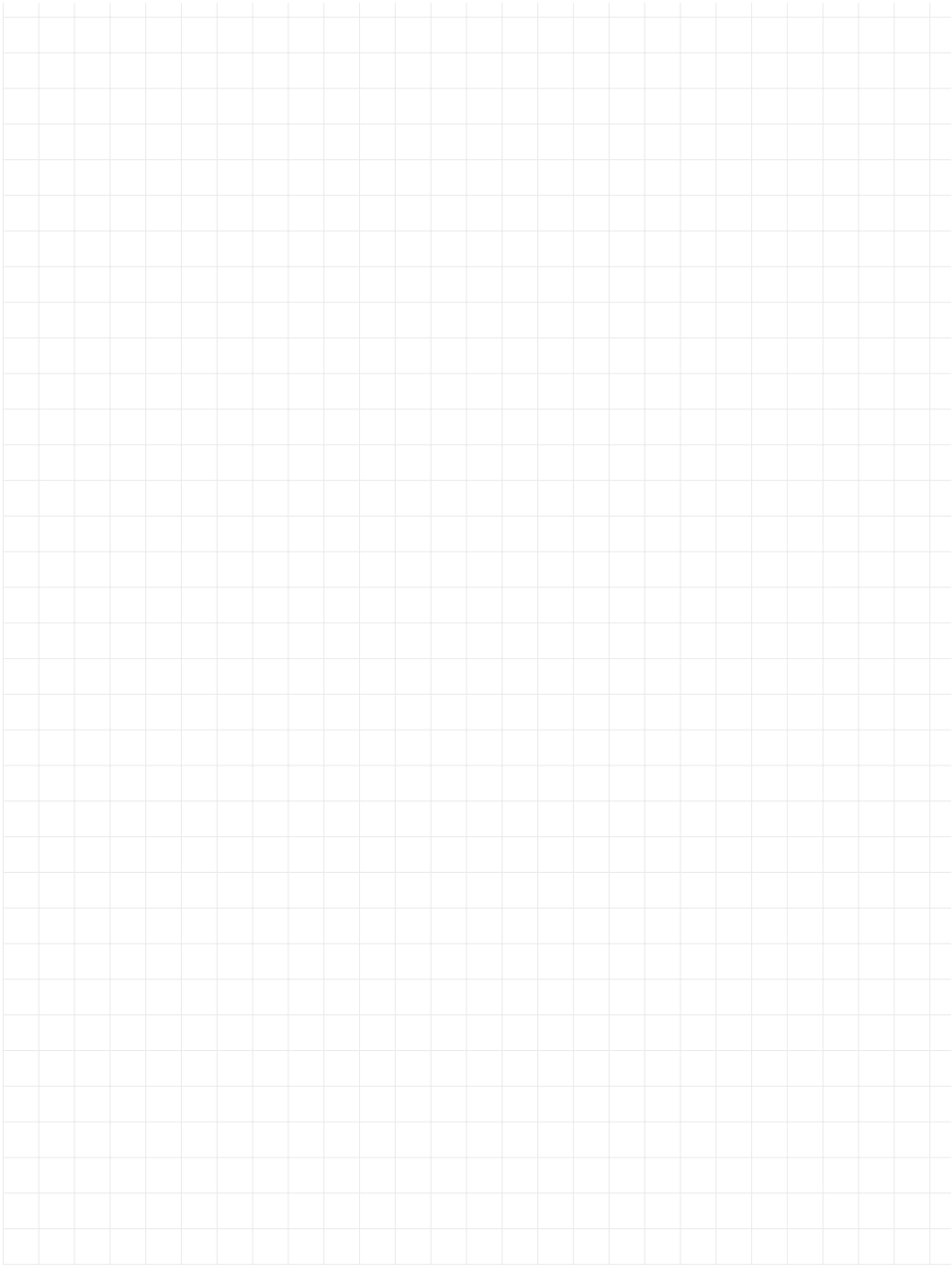
TA3 E11; CB 2.30(b,c)

MGF 矩母函数; domain of convergence 收敛域; location-scale 位置尺度

Find the MGF, its domain, mean and variance for:

1. $f(x) = 2x/c^2, 0 < x < c, c > 0.$
2. $f(x) = e^{-|x-\alpha|/\beta}/(2\beta), x \in \mathbb{R}, \beta > 0.$

P2-09 Working space



第3章 常见分布族

Common Families of Distributions

依据: CB 2e Ch3; 本学期 Chapter 3 资料

3.1 由试验机制选择离散分布

Discrete families from experiment mechanisms

PRINCIPLE 固定次数、停止时刻、不放回抽样和随机到达对应不同模型。

Bernoulli 分布 (Bernoulli distribution) 是一次成功或失败的指示变量 (indicator), 取 1 和 0, 成功率 p 。以下停止规则取 $0 < p < 1$, n, r 为正整数。二项分布 (binomial distribution) 固定独立试验总数 n , 数成功次数。几何分布 (geometric distribution) 停止在第一次成功; 负二项分布 (negative binomial distribution) 停止在第 r 次成功。必须先说数全部试验还是失败次数。

$$P(B_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n.$$

$$P(W = k) = p(1-p)^k, \quad k = 0, 1, \dots$$

$$P(N_r = k) = \binom{k+r-1}{r-1} p^r (1-p)^k, \quad k = 0, 1, \dots$$

试验计数 $T = W + 1, T_r = N_r + r$; 停止在成功时, 最后一次必须成功, 组合数只安排此前的试验。尾事件 $\{T_r > n\} = \{B_n \leq r-1\}$ 将连续等候次数与固定次数计数连接。

超几何分布 (hypergeometric distribution) 从 N 个对象、其中 M 个成功对象中不放回抽 n 个:

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad \max(0, n - N + M) \leq k \leq \min(n, M).$$

抽取不独立, 不能用固定成功率的二项模型精确代替。

Poisson 分布 (Poisson distribution) 用 $\lambda > 0$ 描述计数强度, $P(X = k) = e^{-\lambda} \lambda^k / k!$, $k \geq 0$ 。它可用于适当独立稀疏事件的计数或二项的稀有事件近似, 但不是任何“很多次”都自动适用。

3.2 离散分布的矩与矩母函数

Discrete moments and MGFs

PRINCIPLE 记住参数含义，才能使用下列相互一致的公式。

Distribution	Mean	Variance	MGF
Bernoulli(p)	p	$p(1 - p)$	$1 - p + pe^t$
Binomial(n, p)	np	$np(1 - p)$	$(1 - p + pe^t)^n$
Poisson(λ)	λ	λ	$e^{\lambda(e^t - 1)}$
Geometric failures	$(1 - p)/p$	$(1 - p)/p^2$	$p/[1 - (1 - p)e^t]$
NB failures	$r(1 - p)/p$	$r(1 - p)/p^2$	$\{p/[1 - (1 - p)e^t]\}^r$
Hypergeometric	nM/N	$n\frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}$	finite sum

前三项的 MGF 对所有实数存在；几何与负二项要求 $t < -\ln(1 - p)$ ($0 < p < 1$)。超几何支持有限，MGF 为 $\sum e^{tk}p(k)$ ，对所有实数存在，不必硬记特殊函数形式。不放回产生的负相关使其方差乘以有限总体修正（finite-population correction） $(N - n)/(N - 1)$ ，此式取 $N > 1$ 。总试验计数的 MGF 需额外乘 e^t 或 e^{rt} 。

Poisson 稀释（thinning）指 $N \sim \text{Pois}(\lambda)$ ，每个事件独立以概率 p 被保留，则条件于 N 的计数为二项，边际为 $\text{Pois}(\lambda p)$ 。零截断（zero truncation）是只观察 $X > 0$ ，必须把正整数概率除以 $1 - e^{-\lambda}$ ；截断后不再是原 Poisson，均值方差也随之改变。

3.3 连续分布从区间、等待到正态

Continuous families: intervals, waiting and normality

PRINCIPLE 所有参数化先以密度为准，不凭符号位置猜尺度或方差。

均匀分布（uniform distribution） $U(a, b)$ 在区间等密度 $1/(b - a)$ ，均值 $(a + b)/2$ 、方差 $(b - a)^2/12$ 。离散均匀整数 $1, \dots, n$ 的方差为 $(n^2 - 1)/12$ ，不是把连续公式直接换端点。

指数分布（exponential distribution）以尺度 $\beta > 0$ 写为 $f(x) = \beta^{-1}e^{-x/\beta}$ 、 $x > 0$ ，均值 β 、方差 β^2 ；率（rate） $\lambda = 1/\beta$ 。无记忆性（memorylessness）为 $P(X > s + t \mid X > s) = P(X > t)$ ，表示已等待时间不改变剩余等待分布；一般指数混合不满足它。

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx, \quad \Gamma(a+1) = a\Gamma(a), \quad \Gamma(1/2) = \sqrt{\pi},$$

$$X \sim \Gamma(a, \beta): \quad f(x) = \frac{x^{a-1} e^{-x/\beta}}{\Gamma(a)\beta^a}, \quad EX = a\beta, \quad \text{Var } X = a\beta^2.$$

Gamma 的 a 为形状 (shape), β 为尺度 (scale); 率参数化把 $e^{-x/\beta}$ 改为 $e^{-\lambda x}$, 系数相应变为 $\lambda^a/\Gamma(a)$ 。 $a=1$ 是指数; 整数形状对应多次独立指数等待的和。 $\Gamma(n, 1)$ 的尾概率等于 $\text{Pois}(x)$ 至多 $n-1$ 次的概率, 连接等待时间与计数。

正态分布 (normal distribution) $N(\mu, \sigma^2)$ 的第二参数是方差, 密度为 $e^{-(x-\mu)^2/(2\sigma^2)}/(\sigma\sqrt{2\pi})$ 。标准化 (standardization) $Z = (X - \mu)/\sigma$ 给出标准正态; $\Phi(z) = P(Z \leq z)$, $\Phi(-z) = 1 - \Phi(z)$ 。区间概率为 $\Phi((b - \mu)/\sigma) - \Phi((a - \mu)/\sigma)$, 不能除以 σ^2 。

独立标准正态平方和为自由度 (degrees of freedom) ν 的卡方 (chi-square), $\chi_\nu^2 = \Gamma(\nu/2, 2)$, 均值 ν 、方差 2ν 。

3.4 有限区间、幂变换与厚尾

Beta, transformed families and heavy tails

PRINCIPLE 分布族之间的变换联系, 比孤立背诵密度更可靠。

Beta 分布 (Beta distribution) 在 $(0, 1)$ 上建模比例, 形状参数 $\alpha, \beta > 0$; 归一化常数 $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ 。

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad EX = \frac{\alpha}{\alpha + \beta}, \quad \text{Var } X = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Beta(1, 1) 为均匀。Weibull 分布 (Weibull distribution) 可由尺度 β 的指数取 $1/\gamma$ 次幂构成: $F(y) = 1 - e^{-y^\gamma/\beta}$, 常见 Weibull 形状为 γ 、尺度为 $\beta^{1/\gamma}$ 。矩为 $EY^k = \beta^{k/\gamma}\Gamma(1 + k/\gamma)$ 。

对数正态 (lognormal) 由 $Y = e^Z, Z \sim N(\mu, \sigma^2)$ 构成, 其均值 $e^{\mu+\sigma^2/2}$, 方差 $e^{2\mu+2\sigma^2}(e^{\sigma^2} - 1)$ 。 μ, σ^2 是对数的均值和方差, 不是 Y 的。

Laplace 分布 (Laplace distribution) 密度为 $e^{-|x-\alpha|/\beta}/(2\beta)$, 位置 (location) α 、尺度 β , 均值 α 、方差 $2\beta^2$ 。标准 Cauchy 密度为 $1/[\pi(1+x^2)]$, 尾部衰减慢, $E|X| = \infty$; 对称主值 (principal value) 0 不是均值。厚尾 (heavy tail) 在此提示远端大值可能使矩不存在, 不能只依据图形中心猜测期望。

3.5 连续族的矩母函数与适用区间

Continuous MGFs and their domains

PRINCIPLE 识别和分布时比较 0 邻域的完整函数，不只比较前两个矩。

正态 MGF 的来源是配平方：

$$tx - \frac{(x - \mu)^2}{2\sigma^2} = -\frac{(x - \mu - \sigma^2t)^2}{2\sigma^2} + \mu t + \frac{\sigma^2t^2}{2}.$$

前一项仍是一个平移正态密度的指数，归一化积分为 1，提出与 x 无关的指数因子后即得到下表结果。**Gamma** 则通过将指数衰减率改为 $1/\beta - t$ 来识别归一化积分。

Distribution	MGF	Domain
$U(a, b)$	$(e^{bt} - e^{at})/[(b - a)t]$	$t \in \mathbb{R}$, value 1 at 0
Exp scale β	$(1 - \beta t)^{-1}$	$t < 1/\beta$
Gamma (a, β)	$(1 - \beta t)^{-a}$	$t < 1/\beta$
$N(\mu, \sigma^2)$	$e^{\mu t + \sigma^2t^2/2}$	$t \in \mathbb{R}$
χ^2_ν	$(1 - 2t)^{-\nu/2}$	$t < 1/2$
Laplace (α, β)	$e^{\alpha t}/(1 - \beta^2t^2)$	$ t < 1/\beta$

Beta 支持有界，MGF 对全部实数存在，但本题不要求特殊函数表达。**Cauchy** 与对数正态不能硬套 0 邻域 MGF 识别法。不同分布可能有相同均值方差；只有比较这些数值不能证明同分布。独立性是将和的 MGF 写成乘积的前提。

3.6 英文概念与原理

English Definitions and Principles

1. Discrete families from experiment mechanisms

A Bernoulli indicator records one success. A binomial variable counts successes in a fixed number of independent constant-probability trials. Geometric and negative-binomial variables stop at success one or success r .

State whether the count includes all trials or only failures; the final success is fixed and only earlier positions are arranged.

Hypergeometric counts sample without replacement, with support limited by available successes and failures. Draws are not independent. Poisson counts have nonnegative integer support and intensity λ ; their event assumptions must be checked rather than inferred from a large trial number alone.

2. Discrete moments and MGFs

The table uses geometric and negative-binomial failure counts. Their MGFs require $t < -\ln(1 - p)$; shifting to total trials adds e^t or e^{rt} . Bernoulli, binomial and Poisson MGFs exist for all real t , as does any finite-support MGF.

Negative dependence from sampling without replacement gives the hypergeometric finite-population variance correction $(N - n)/(N - 1)$ for $N > 1$.

Poisson thinning conditions on the original random count, applies a binomial retention model, and sums over the count to obtain $\text{Poisson}(\lambda p)$. Zero truncation instead conditions on a positive value and renormalizes by $1 - e^{-\lambda}$; it changes both moments and the distribution family.

3. Continuous families: intervals, waiting and normality

A continuous uniform law has constant density on an interval. An exponential law with scale β has rate $1/\beta$, mean β and variance β^2 . Its survival function gives memorylessness; a nontrivial mixture of unequal rates loses that property.

The Gamma density defines shape a and scale β , with mean $a\beta$ and variance $a\beta^2$. Rate notation replaces scale by its reciprocal in both the exponent and the normalizing constant. Shape one is exponential, and integer shape corresponds to accumulated exponential waits.

A Gamma waiting-time tail can be expressed as a Poisson count probability.

In $N(\mu, \sigma^2)$, the second parameter is variance and standardization divides by the standard deviation. Standard-normal CDF differences give interval probabilities. The sum of ν independent squared standard normals is chi-square with ν degrees of freedom, equivalently $\text{Gamma}(\nu/2, 2)$.

4. Beta, transformed families and heavy tails

Beta distributions model values in $(0, 1)$ using two shape parameters and a Beta-function normalizer; $\text{Beta}(1, 1)$ is uniform. A Weibull law arises by taking a positive power of an exponential variable, and its usual scale is the corresponding power of the exponential scale.

A lognormal is the exponential of a normal variable. Its parameters describe the logarithm, so its mean is not the exponential of the log mean. Laplace distributions have a location and scale with variance twice scale squared.

A standard Cauchy has such heavy tails that its mean does not exist; a symmetric principal value is not a valid expectation.

5. Continuous MGFs and their domains

Completing the square in the normal MGF integral shifts the normal mean to $\mu + \sigma^2 t$ and leaves the factor $e^{\mu t + \sigma^2 t^2 / 2}$. A Gamma integral instead changes its exponential decay coefficient to $1/\beta - t$.

The table lists both formulas and convergence domains. The uniform formula has a removable singularity at zero. Beta MGFs exist everywhere because the support is bounded, even without a simple elementary formula. Cauchy and lognormal laws do not have the required finite MGF neighborhood.

Equal means and variances alone do not identify a distribution; MGF uniqueness requires agreement on a neighborhood of zero.

P3-01 成功次数与停止规则

课堂离散分布、Quiz 1 方法改编；CB Ch3 分布族

Bernoulli trial Bernoulli 试验；binomial 二项；negative binomial 负二项

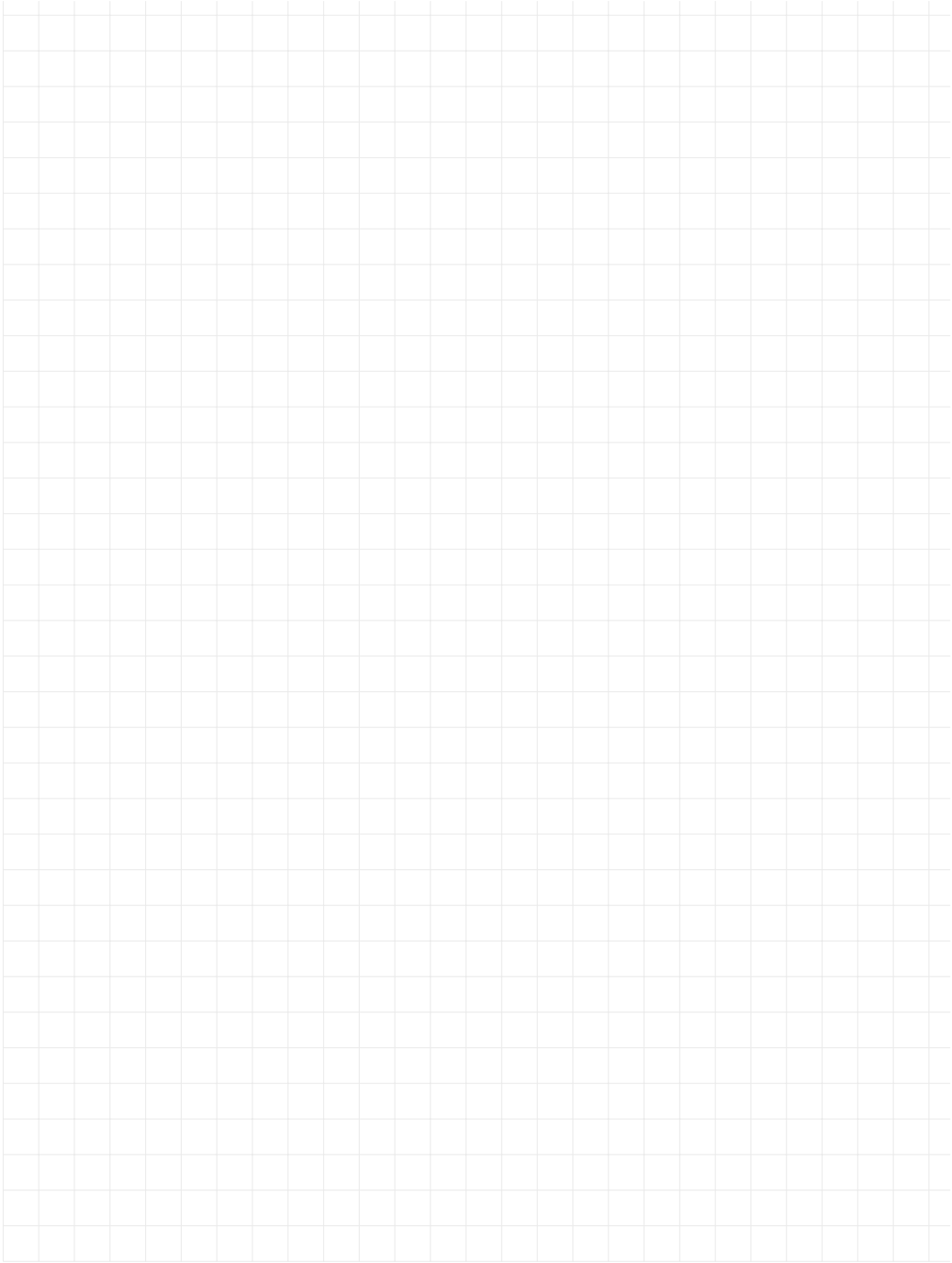
Independent trials have success probability p .

1. Let B_n count successes in a fixed n trials. Give its PMF, support, mean and variance, deriving the moments with indicators.
2. Let T be the trial of the first success and $W = T - 1$ the number of prior failures.

Give both PMFs and moments, and prove the memoryless property for T in the form $P(T > m + n \mid T > m)$.

3. Let N_r count failures before the r th success. Give its PMF, support and moments, explaining the terminal-success factor.

P3-01 Working space



P3-02 Poisson 饼干设计

TA4 E1; CB 3.7

Poisson parameter Poisson 参数; numerical root 数值求根

The number of chips in a cookie is Poisson with mean λ . How large should λ be to ensure that more than 99% of cookies have at least two chips? Set up the inequality, justify uniqueness of the threshold and give a numerical answer.

P3-03 零截断 Poisson

TA4 E2; CB 3.13(a)

zero truncation 零截断; conditional PMF 条件质量函数

$X \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Let Y have the conditional distribution of X given $X > 0$. Find its PMF, CDF, mean and variance. Explain why the mean rises after truncation.

P3-04 二项与负二项的尾事件对应

TA4 E3; CB 3.12

stopping time 停止时刻; tail probability 尾概率

Let T_r be the trial at which the r th success occurs in independent Bernoulli trials with probability p . Let B_n count successes in the first n trials.

For $n \geq r$, prove $P(T_r > n) = P(B_n \leq r - 1)$, give the PMF of T_r , and write the corresponding identity for the failure count $N_r = T_r - r$.

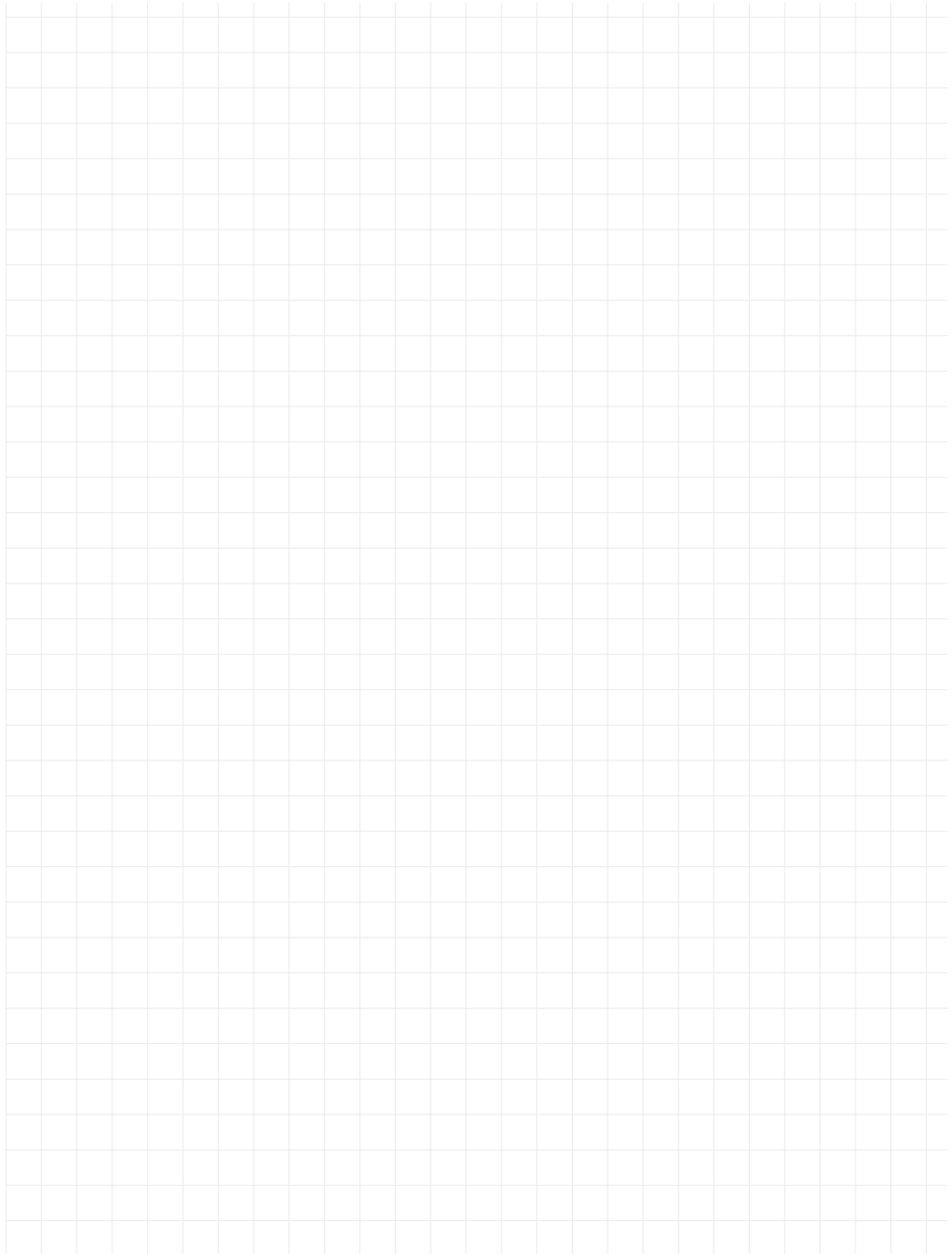
P3-05 Poisson 稀释与被选事件

TA4 E4

thinning 稀释; conditional binomial 条件二项

Let $N \sim \text{Poisson}(\lambda)$ and $Y \mid N = n \sim \text{Binomial}(n, p)$. Derive the marginal distribution of Y directly by summation. Then use $\lambda = 10, p = 1/50$ to find $P(Y = 1)$, $P(Y \geq 1)$ and $P(Y \leq 1)$.

P3-05 Working space



P3-06 高斯积分与 Gamma 半整数

TA4 E5

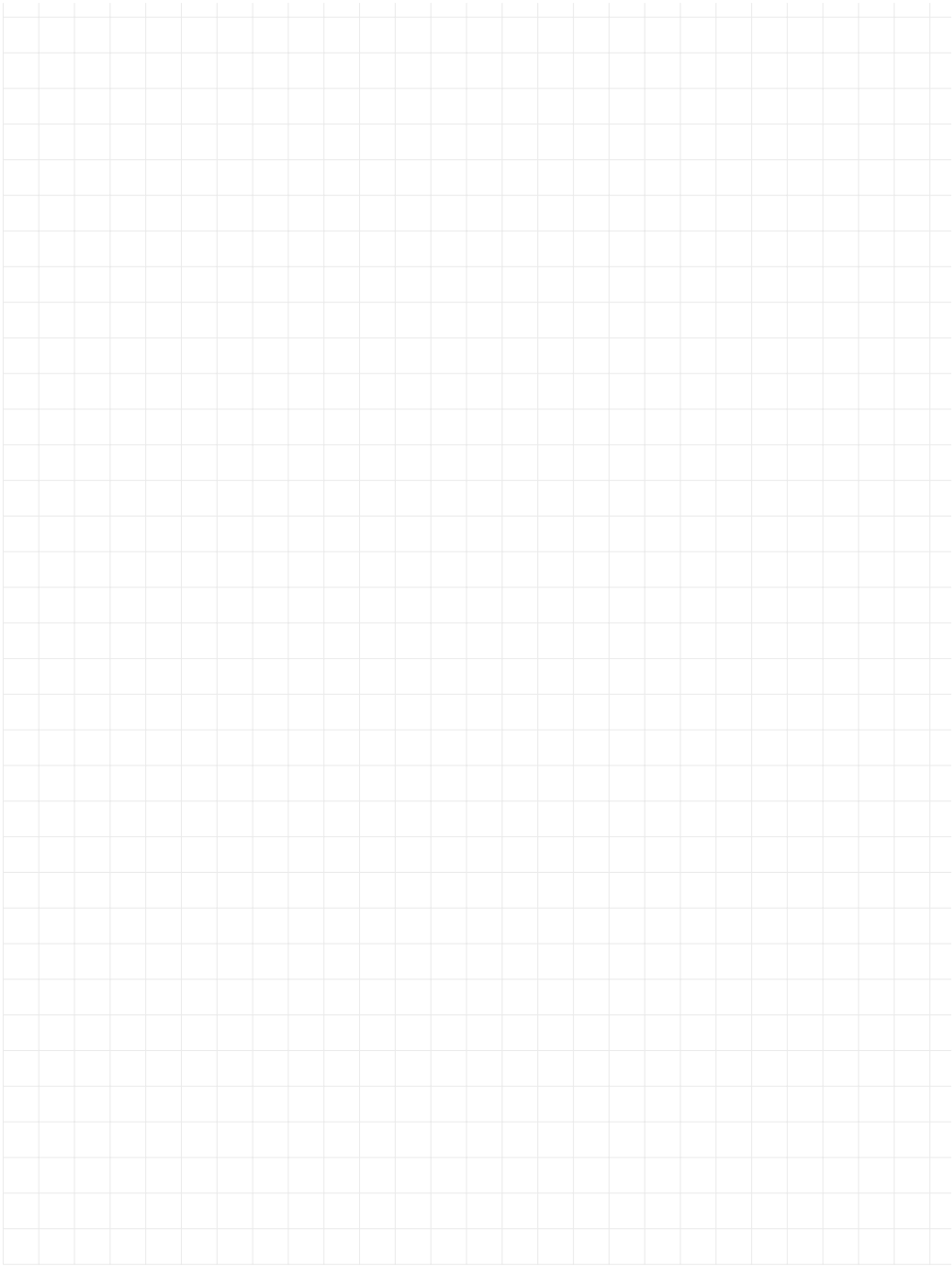
Gaussian integral 高斯积分; polar coordinates 极坐标; Gamma function Gamma 函数

Prove $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ by squaring and using polar coordinates. Deduce

$$\int e^{-x^2/2} dx, \quad \int e^{-(x-\mu)^2/(2\sigma^2)} dx, \quad \int x^2 e^{-x^2/2} dx$$

over the real line, with $\sigma > 0$, and show $\Gamma(1/2) = \sqrt{\pi}$.

P3-06 Working space



P3-07 Gamma、指数与卡方的参数体系

TA4 E6

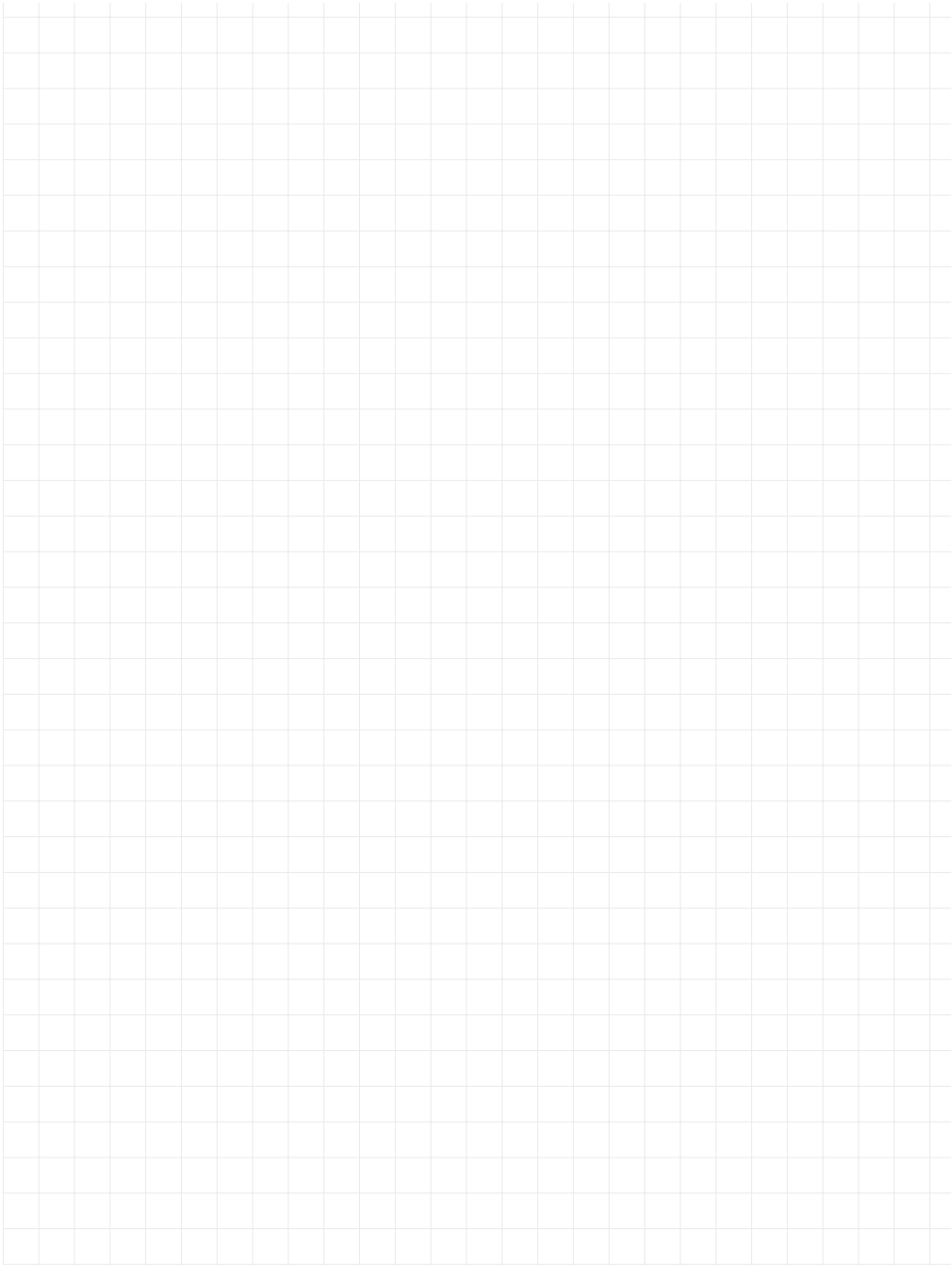
scale 尺度; shape 形状; chi-square 卡方; degrees of freedom 自由度

Use the shape–scale Gamma density

$$f(x) = \frac{x^{a-1}e^{-x/\beta}}{\Gamma(a)\beta^a}, \quad x > 0.$$

1. Derive its MGF and domain, and identify the case $a = 1$.
2. For independent $Z_j \sim N(0, 1)$, derive the MGF of Z_j^2 and identify the law of $\sum_{j=1}^{\nu} Z_j^2$.
3. Give all relevant means and variances and translate Gamma to shape–rate notation.

P3-07 Working space



P3-08 Gamma 等待时间与 Poisson 次数

TA4 E7; CB 3.19

waiting time 等待时间; Poisson count Poisson 计数

Let $G_n \sim \text{Gamma}(n, 1)$ in shape–scale notation, with positive integer n . For each $x \geq 0$, let $N_x \sim \text{Poisson}(x)$. Prove

$$P(G_n > x) = P(N_x \leq n - 1).$$

Use integration by parts, and explain the waiting-time interpretation.

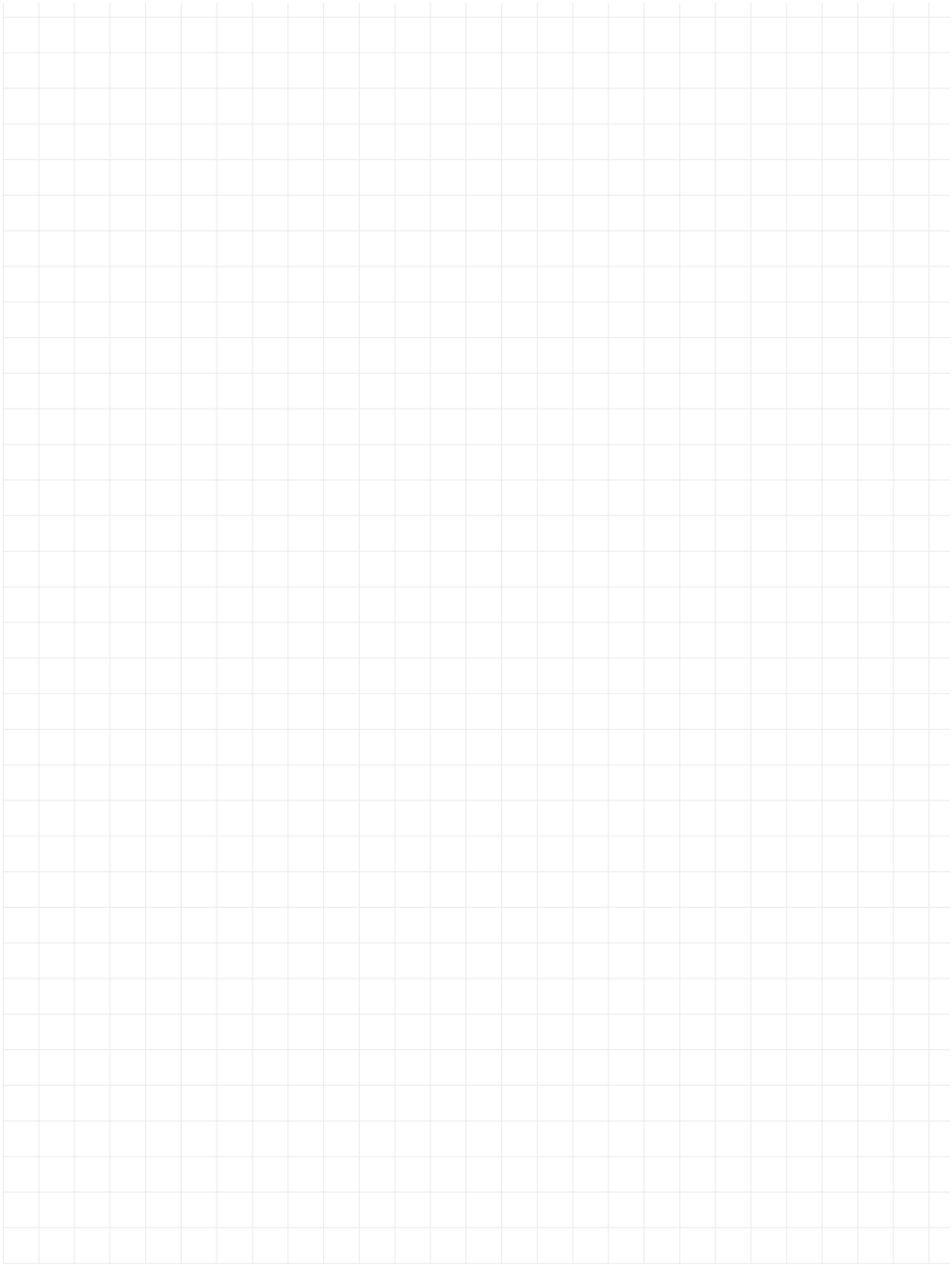
P3-09 指数变量的幂变换与 Weibull

TA4 E8; CB 3.24(a)

Weibull distribution Weibull 分布; raw moment 原点矩

X is exponential with scale $\beta > 0$. For $\gamma > 0$, set $Y = X^{1/\gamma}$. Find the CDF, PDF, EY^k for $k > 0$, the mean and variance. State the equivalent usual Weibull shape and scale.

P3-09 Working space



P3-10 正态标准化、区间与分位数

连续分布课堂笔记；标准化方法改编

standardization 标准化；standard deviation 标准差；percentile 百分位数

$X \sim N(100, 15^2)$. Calculate $P(85 < X \leq 130)$, $P(X \geq 115)$, and the 90th percentile of X . Use $\Phi(1) = 0.84134475$, $\Phi(2) = 0.97724987$ and $\Phi^{-1}(0.9) = 1.28155157$. Explain why the denominator in standardization is 15 rather than 225.

P3-11 对数正态的密度与矩

TA4 E9

lognormal 对数正态; normal MGF 正态矩母函数

$Z \sim N(\mu, \sigma^2)$, $\sigma > 0$, and $Y = e^Z$. Derive the PDF and all real raw moments EY^k . Find the mean and variance and explain why the mean is not e^μ .

P3-12 Beta 归一化与矩

TA4 E10; CB 3.22(d)

Beta function Beta 函数; normalizing constant 归一化常数

$X \sim \text{Beta}(\alpha, \beta)$, with positive parameters and density

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1.$$

Derive EX^k for $k \geq 0$, the mean and variance, using $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

Identify the case $\alpha = \beta = 1$.

P3-13 Cauchy 的对称为何不产生均值

连续分布课堂笔记: Cauchy; CB Ch3

Cauchy distribution Cauchy 分布; principal value 主值; absolute integrability 绝对可积

X has standard Cauchy density $1/[\pi(1 + x^2)]$ on $\mathbb{R}$. Verify normalization and explain rigorously why EX does not exist, even though the density is symmetric. Does its variance exist?

第4章 联合分布与条件矩

Joint Distributions and Conditional Moments

依据: CB 2e Ch4; 本学期 Chapter 4 资料 pp.1-31

4.1 联合分布与边缘分布

Joint distributions and marginalization

PRINCIPLE 一个变量的边际不能恢复两个变量的相互关系。

随机向量 (random vector) (X, Y) 同时记录两个量。联合分布 (joint distribution) 说明它们如何共同取值; 联合 PMF 为 $P(X = x, Y = y)$ 。普通联合 PDF 对二维面积积分给出概率, 它并非所有具有连续边际的随机向量都存在。

联合分布函数 (joint CDF) $F(x, y) = P(X \leq x, Y \leq y)$ 可用于一般分布。它的矩形增量必须非负:

$$P(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c).$$

绝对连续且可微时, 联合密度为混合偏导 $\partial^2 F / \partial x \partial y$ 。边缘化 (marginalization) 将另一变量所有可能取值加总或积分:

$$p_X(x) = \sum_y p(x, y), \quad f_X(x) = \int f(x, y) dy,$$

另一个方向同理。**积分上下限由联合支持集决定**: $0 < x < y < 1$ 中固定 x 时 $y \in (x, 1)$, 固定 y 时 $x \in (0, y)$ 。先画支持再积分, 不应机械对两变量都从 0 到 1。

4.2 独立性要同时检查数值与支持

Independence and the support

PRINCIPLE 密度表达式看似没有另一变量, 不等于联合分布可分离。

两个随机变量独立 (independent random variables) 意味着任意合适集合 A, B 均有 $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ 。等价地, 联合 CDF 在所有点分解为边际 CDF 乘积; 离散时联合 PMF 分解; 有联合密度时几乎处处分解。

$$F_{X,Y}(x, y) = F_X(x)F_Y(y), \quad f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad (\text{相应存在时}).$$

验证不能只比较某一点；否定只需找到一个违例点或事件。非直积支持通常直接揭示依赖：若两边际均允许一个组合而联合不允许，独立不可能。

边际相同不表示联合相同。 (U, U) 与 $(U, 1-U)$ 都有均匀边际，却分别集中在不同直线上，甚至没有普通二维密度。函数期望可直接用联合分布： $E[g(X, Y)] = \sum g(x, y)p(x, y)$ 或 $\iint g(x, y)f(x, y)dxdy$ 。独立且期望存在时 $E(XY) = EXEY$ ；仅这一等式本身并不足以证明独立。

4.3 条件分布是截线后的归一化

Conditional distributions as normalized slices

PRINCIPLE 固定一个变量，只在剩下的可行支持上重新分配概率。

离散条件分布将联合质量除以条件变量边际。连续条件分布用密度比定义，适用于边缘密度正的取值；不能直接用概率比除以 $P(X = x) = 0$ 。

$$p_{Y|X}(y | x) = \frac{p(x, y)}{p_X(x)}, \quad f_{Y|X}(y | x) = \frac{f(x, y)}{f_X(x)}.$$

$$f(x, y) = f_{Y|X}(y | x)f_X(x).$$

条件独立性不是此处自动拥有的性质；只有独立时，条件分布才不随所给变量改变。用最后一式可从一个边际和一个条件分布重建联合，再边缘化和反向条件化。这也解释为什么 Bayes 可以改变条件方向。

给定 $X = x$ 的条件密度应对 y 积分为 1；若密度常数，支持长度必须与常数互为倒数。三角支持中 $X | Y = y$ 是 $U(0, y)$ ， $Y | X = x$ 是 $U(x, 1)$ ；不能把给定值仍当成自由积分变量。

4.4 条件均值与条件方差的两层含义

Conditional means and conditional variances

PRINCIPLE 先给定具体数值求普通矩，再把这个数值换回随机变量。

条件期望 (conditional expectation) $E(Y | X = x)$ 是在给定 x 的条件分布下求平均，得到关于 x 的函数 $m(x)$ 。 $E(Y | X) = m(X)$ 则是一个随机变量，不

是已经平均掉 X 的单个数。条件方差 (conditional variance) 同样是关于给定值的函数。

$$m(x) = \int y f_{Y|X}(y | x) dy, \quad v(x) = \int y^2 f_{Y|X}(y | x) dy - m(x)^2.$$

离散情况将积分换求和。对均匀 $U(a, b)$, 可以直接使用均值 $(a + b)/2$ 、方差 $(b - a)^2/12$; 对条件 Bernoulli 使用成功率 $p(x)$ 得均值 $p(x)$ 、方差 $p(x)[1 - p(x)]$ 。

当前笔记在离散例中通过加权平均核验 $\sum_x E(Y | X = x) p_X(x) = EY$ 。本册相应题目用具体求和或积分核对这一事实, 核心范围停在条件分布与条件矩; 后续估计、充分性、UMVU 与检验理论未作为本次核心题。2024 年卷的最小值题仅作为由 CDF 与独立性推出新分布的衔接练习。

4.5 英文概念与原理

English Definitions and Principles

1. Joint distributions and marginalization

A random vector has a joint law describing simultaneous values, not merely separate marginals. A joint CDF is defined for every law. Its rectangle increments give rectangle probabilities and must be nonnegative.

An ordinary joint PDF exists only for laws absolutely continuous with respect to area; continuous marginals alone do not guarantee it.

Marginalization sums or integrates out the other coordinate. Integration bounds come from the joint support. In the triangle $0 < x < y < 1$, fixing x leaves y between x and one, whereas fixing y leaves x between zero and y .

2. Independence and the support

Independence requires factorization for every pair of measurable events, equivalently for all joint CDF values, all joint masses, or almost everywhere for densities. A single counterexample disproves independence, but one successful point cannot prove it. The support must also be consistent with factorization.

Identical marginals do not determine a joint law. The pairs (U, U) and $(U, 1 - U)$ share uniform marginals but place probability on different lines. Joint expectations sum or integrate the function against the joint law. Independence implies factorization of product moments when defined, but the converse is false.

3. Conditional distributions as normalized slices

A conditional PMF divides joint mass by the conditioning marginal mass. A conditional PDF uses a density ratio where the marginal density is positive; it does not divide by a zero-probability point event. Holding one coordinate fixed gives a slice whose feasible support must be normalized.

Multiplying a marginal by a conditional reconstructs the joint density. One can then integrate for the other marginal and reverse the conditioning. Independence is an additional property under which conditioning leaves the other distribution unchanged. Conditional densities must integrate to one over their slice.

4. Conditional means and conditional variances

The conditional mean given $X = x$ is an ordinary expectation under a conditional law and is a function $m(x)$. Replacing x by X makes $E(Y | X)$ a random variable.

Conditional variance likewise depends on the conditioning value and equals the conditional second moment minus the squared conditional mean.

Recognized conditional families allow their usual moment formulas. The selected exercises check by explicit sums or integrals that averaging the conditional means recovers the marginal mean. Core coverage ends with the joint and conditional distributions and moments in the available notes.

Later inference topics are excluded; the older minimum question is only a CDF-and-independence bridge.

P4-01 联合概率表、条件矩与独立性

Ch4 Part1 课堂材料及本地笔记的离散例

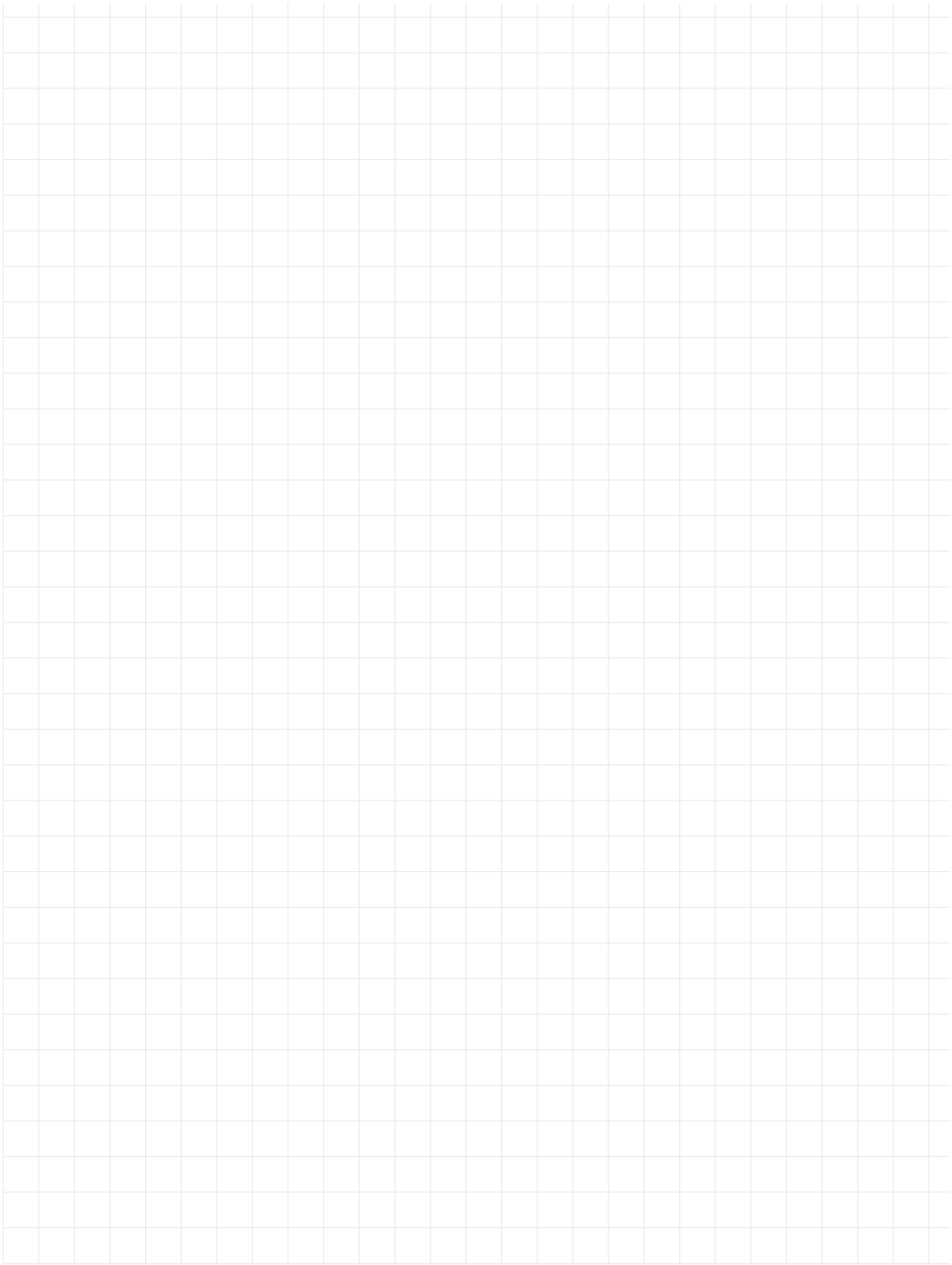
joint 联合; marginal 边际; conditional variance 条件方差

The joint PMF is

$X \backslash Y$	0	1
0	0.2	0.3
1	0.1	0.4

Find both marginals, determine independence, calculate $E(XY)$ and $P(X = 1 | Y = 1)$, and obtain $E(Y | X)$ and $\text{Var}(Y | X)$ as functions of X . Verify numerically that averaging the conditional means gives EY .

P4-01 Working space



P4-02 三角形支持的联合与条件密度

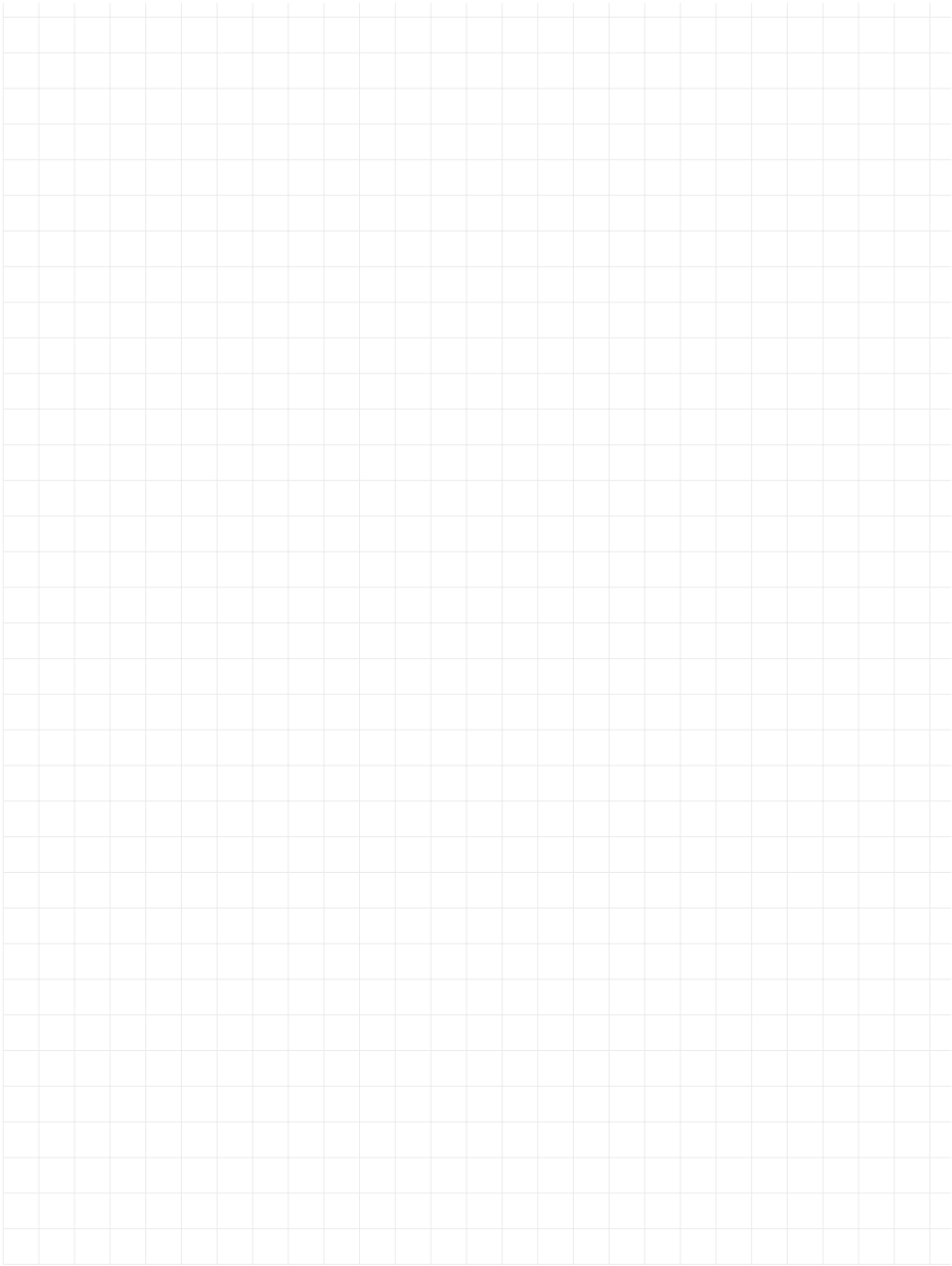
Ch4 Part1 课堂材料: $0 < x < y < 1$ 的三角例

joint support 联合支持; conditional density 条件密度

Let $f_{X,Y}(x,y) = 2$ on $0 < x < y < 1$ and zero elsewhere.

1. Draw the support and find both marginal densities. Are X and Y independent?
2. Find $P(X < 1/2)$, $P(X + Y < 1)$ and $E(XY)$.
3. Find both conditional densities, means and variances.

P4-02 Working space



P4-03 联合 CDF 与矩形概率

Ch4 Part1 的联合 CDF 方法改编

joint CDF 联合分布函数; mixed derivative 混合偏导

Let $F(x, y) = (1 - e^{-x})(1 - e^{-y})$ for $x, y \geq 0$, and zero if either coordinate is negative. Derive the rectangle-probability formula for $P(a < X \leq b, c < Y \leq d)$. Find the joint PDF, determine independence and calculate $P(0 < X \leq 1, 1 < Y \leq 2)$.

P4-04 相同边际为何不能确定联合分布

Ch4 Part1 课堂反例: (U, U) 与 $(U, 1 - U)$

singular joint distribution 奇异联合分布; marginal information 边际信息

Let $U \sim U(0, 1)$. Compare $(X, Y) = (U, U)$ and $(X, Y) = (U, 1 - U)$. Give their marginal distributions and their joint CDFs for $0 \leq x, y \leq 1$. Are they independent? Do they have ordinary joint PDFs with respect to two-dimensional area?

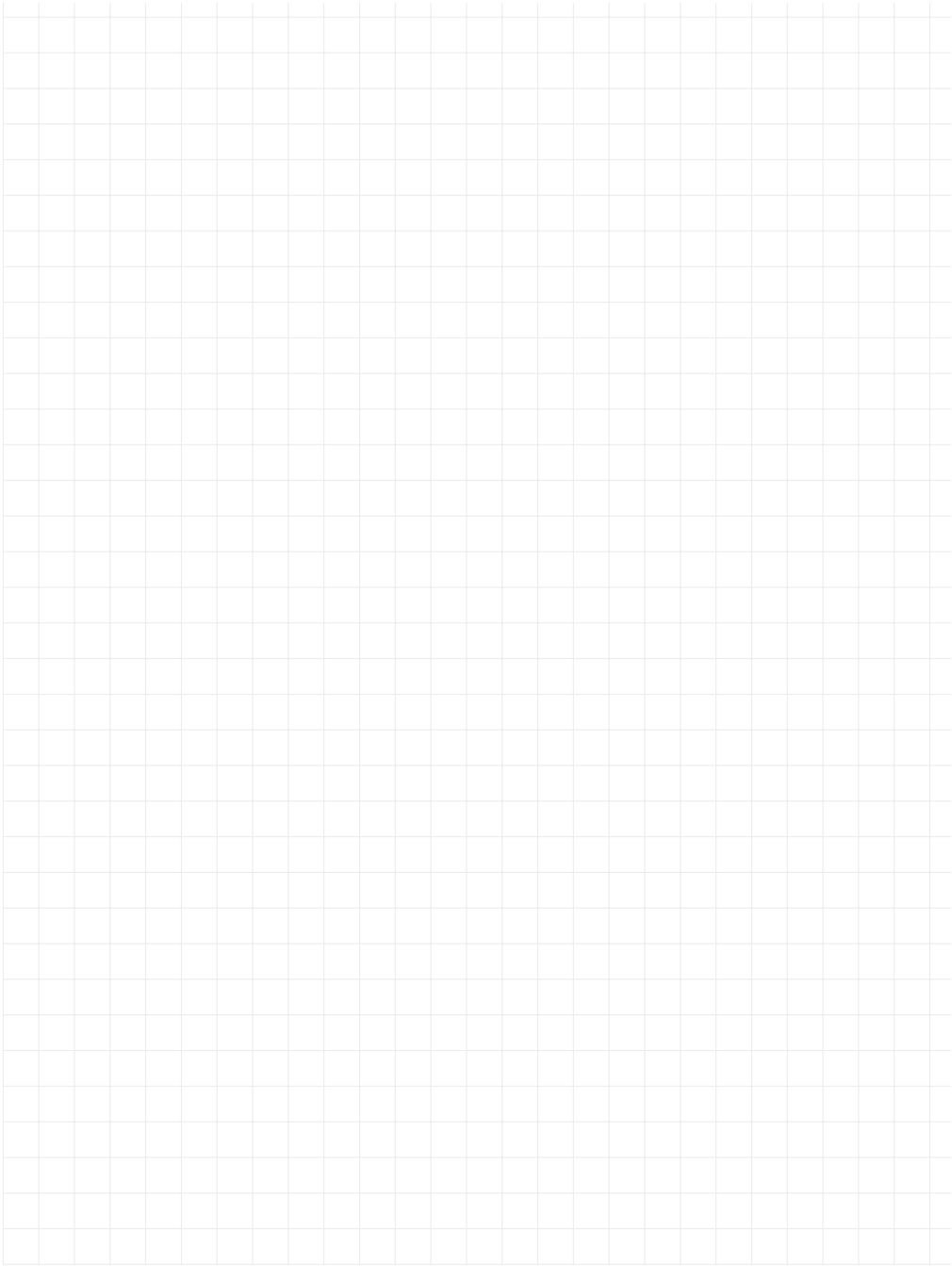
P4-05 由条件密度重建联合密度

Ch4 Part1 条件密度方法改编

reconstruction 重建; shifted exponential 平移指数

$X \sim \text{Gamma}(2, 1)$ in shape–scale notation and, given $X = x$, Y is uniform on $(0, x)$. Find the joint PDF, the marginal distribution of Y , the conditional distribution of X given $Y = y$, and both conditional means and variances.

P4-05 Working space



P4-06 最小值的分布：往年卷衔接题

2024 数理统计期中卷 Q3(a) 的基础分布部分

minimum 最小值；independent sample 独立样本；order statistic 次序统计量

Let $X_1, \dots, X_n$ be independent with common continuous CDF F and density f . Set $M = \min(X_1, \dots, X_n)$. Derive its CDF and density. Then specialize to independent exponential variables with rate λ . This is a bridging exercise, not a statement of this semester's midterm scope.