ESTR 2018 概率论

①②③④⑤⑥⑦⑧⑨⑩

Sheldon M. Ross, "A First Course in Probability", Pearson, 9th Edition, 2014

Axiom	公理
antenna	天线
consecutive	连续的
Proposition	可翻译为命题，经证明且 interesting，但没有 Theorem 重要，比较常用.
without replacement	不放回（抽签）
die dice	骰子
partition	分类，分割
the event A_1 of interest	感兴趣的事件 A_1

> 即题目考虑的事件，这么写就不用再抄一遍

disjoint events	不相交事件（不会同时发生）
amount to	相当于
feasible	可行的，可实现的
parity	奇偶性
deck	扑克牌堆
playoff NBA	季后赛
bivariate function	双变量函数
stock	股票
urn	瓮
A is congruent to B modulo 3	A 与 B 模 3 同余
congruent	全等
Covariance	协方差
Uncorrelated	不相关
monotonically increasing	单调递增
converge to	收敛于
sprinkle	撒

Lecture 1 基本概念

1.1 样本空间

> Sample Space

The $\color{red}{sample\ Space\ \Omega}$ is the set of all possible outcomes.

1.2 事件

> Event

An $\color{blue}{event\ E}$ is a subset of the sample space Ω.

If Ω has n elements, it has 2^n events (subsets).

> 每个有或没有.

1.3 概率公理（一）

> Probability Axioms (Probability in Finite Spaces)

The probability of an event is the **sum of the probability of its outcomes.**

1. $\color{red}{probabilities}$ are **non-negative**,
2. $\color{red}{probabilities}$ add up to **one**.

1.4 事件概率

> Probability Calculation for Events

The probability of an $\color{blue}{event}$ is $\color{red}{the\ sum\ of\ the\ probabilities\ of\ its\ outcomes}$.

> 如果多个元素触发同一个事件，显然这个事件的概率是各个元素发生的概率之和。（注意，确保各元素独立）

Uniform Probability Law

If the outcomes in Ω are $\color{red}{equally\ likely}$, then the probability of event A will be

$P(A) = \frac{Number\ of\ outcomes\ in\ A}{Number\ of\ outcomes\ in\ \Omega} = \frac{|A|}{|\Omega|}$

i.e., the $\color{red}{probability\ calculation}$ simplifies to $\color{blue}{counting}$.

注意保证各结果是等可能，且独立的. 不要默认这两个成立，要再三确认.

$|A|$ means the size of A.

Lecture 2 计数原理

2.1 乘法原理

> Proposition (Product Rule)

Suppose that

- there are $n\ possible\ outcomes\ for\ Experiment\ 1$,
- **For every outcome of Experiment 1**, there are $m\ possible\ outcomes\ for\ Experiment\ 2$.

Then, there are $m \times n$ possible outcomes for the two experiments.

> 组合计数的基本原理之一. 详见 `ENGG 2440 离散数学`.

2.2 排列数

> Permutations

A **permutation** of n different objects is an arrangement of the objects into an $ordered\ sequence$.

For n different objects, there exists $n!$ different permutations:

$n! = n \times (n-1) \times \cdots \times 2 \times 1$

> 感叹号：阶乘，factorial

例：2个dice，不同值的概率

$\frac{6\times5}{6\times6} = \frac{5}{6}$

例：6个dice，不同值的概率

$\frac{6!}{6^6} = \frac{720}{46656} \approx 0.015$

例：2个dice，和为 7

$\frac{6}{6\times6} = \frac{1}{6}$

> 1+6，2+5，3+4 两倍，共 6 种.

法二：乘法原理

$\frac{6\times1}{6\times6} = \frac{1}{6}$

> 每个第一个骰子，都有一种第二个来配.

例：2个dice，和为 odd

法一：

奇数：奇数+偶数（奇+奇，奇+偶，偶+奇，偶+偶，等可能）

$\frac{1}{2}$

法二：条件概率（第一个摇到某个数）

$\frac{1}{6}\times\frac{1}{2}+\frac{1}{6}\times\frac{1}{2}+\frac{1}{6}\times\frac{1}{2}+\frac{1}{6}\times\frac{1}{2}+\frac{1}{6}\times\frac{1}{2}+\frac{1}{6}\times\frac{1}{2}=\frac{1}{2}$

法三：乘法原理

$\frac{6\times3}{6\times6}=\frac{1}{2}$

例：2个dice，第一个比第二个大

$\frac{1+2+3+4+5}{6\times6}=\frac{15}{36}$

法二：$\frac{36-6}{2}=15$，等大剔除，概率平均

例：三个人，生日的星期相同

一周7天，总可能：7^3

相同：7

$\frac{7}{7^3}=\frac{1}{49}$

例：三个人，生日的星期不同

一周 7 天，总可能：7^3

星期不同组数：$C_3^7\cdot A_3^3=7\times6\times5=210$

$\frac{210}{7^3}=\frac{30}{49}$

2.3 生日悖论

> birthday paradox

x 个学生（$x\in[1,365]$），生日有重合的概率.

解：

365 天，无重合：$C_x^{365}A_x^x=\frac{365!}{x!(365-x)!}\cdot x!=\frac{365!}{(365-x)!}$

所有结果：365^x

有重合概率：$\frac{365^x-C_x^{365}A_x^x}{365^x}=1-\frac{365!}{365^x(365-x)!}$

生日悖论：对于给定 x，上述概率值可能比直观想象的要大很多.

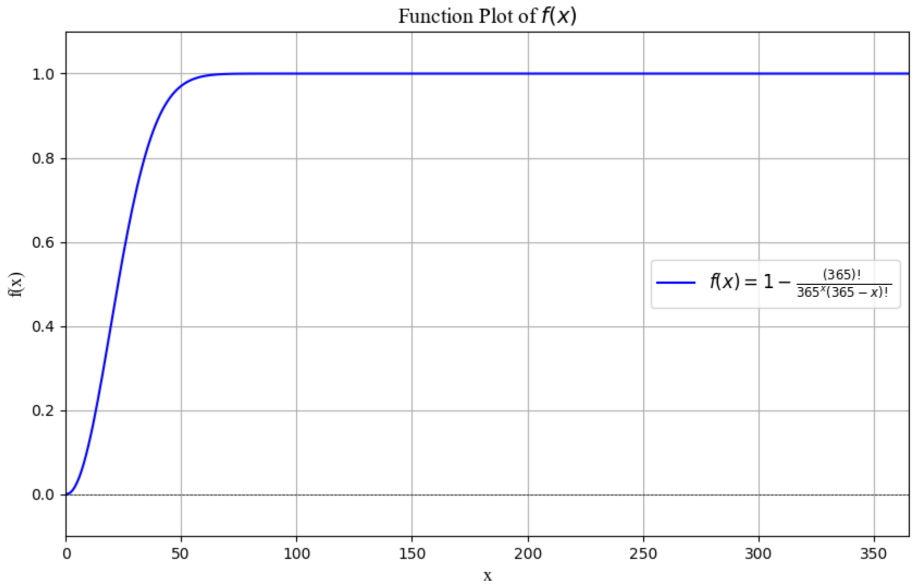
令 $x=110$

$1-\frac{365!}{365^{110}(365-110)!}\approx0.9999999895>0.9$

令 $x=50$

$1-\frac{365!}{365^{50}(365-50)!}\approx0.9704>0.9$

用 `python` 作出函数 $f(x)=1-\frac{365!}{365^x(365-x)!}$（$x\in[0,365]$）的图像.




```python
# pip install numpy matplotlib
# 安装 numpy 和 matplotlib 库

import numpy as np
import matplotlib.pyplot as plt
from math import factorial, log

# 定义函数 f(x)
def f(x):
    if x > 365:
        return 1  # 超出范围时返回 1
    return 1 - np.exp(log(factorial(365)) - x * log(365) - log(factorial(365 - x)))

# 生成 x 值
x_values = np.arange(0, 366)
# 计算对应的 f(x) 值
y_values = [f(x) for x in x_values]

# 绘图
plt.figure(figsize=(10, 6))
plt.plot(x_values, y_values, label=r'$f(x) = 1 - \frac{(365)!}{365^x (365 - x)!}$',
color='blue')
plt.title('Function Plot of $f(x)$', fontsize=14, fontname='Times New Roman')
plt.xlabel('x', fontsize=12, fontname='Times New Roman')
plt.ylabel('f(x)', fontsize=12, fontname='Times New Roman')
plt.ylim(-0.1, 1.1)
plt.xlim(0, 365)
plt.axhline(0, color='black', lw=0.5, ls='--')
plt.axvline(0, color='black', lw=0.5, ls='--')
plt.grid()
plt.legend(prop={'size': 12, 'family': 'Times New Roman'})
plt.show()
```

2.4 组合数

> 又叫 Binomial Coefficient，二项式系数.

Given a set S of size n, the number of subsets of size k will be

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

The number of possible arrangements of $\color{red}{k\ object\ of\ Type\ A}$ and $\color{blue}{n-k\ object\ of\ Type\ B}$ into an $\color{red}{ordered\ sequence}$ will be $\binom{n}{k}$.

> 因为从 n 个里抽 k 个，和抽 $n-k$ 个是一样的.

例：8 红 2 蓝，抽 2 个，不放回，不同色概率

$$\frac{8\times 2 \color{red}{+2\times 8}}{\binom{10}{2}} = \frac{32}{45} > \frac{1}{2}$$

> 第一个红第二个蓝，或第一个蓝第二个红

2.5 多项式系数

> Multinomial Coefficient，二项式系数的推广.

For a set S of size n, the number of partitioning of the set to partitions of size $k_1,\ k_2,\ \cdots,\ k_t$ (note that $n = k_1 + k_2 + \cdots + k_t$) will be

$$\binom{n}{k_1, k_2, \cdots, k_t} = \frac{n!}{k_1!k_2!\cdots k_t!}$$

The number of permutations of $\color{red}{k_1\ objects\ of\ Type\ 1}$, $\color{red}{k_2\ objects\ of\ Type\ 2}$, ..., and $\color{blue}{k_t\ objects\ of\ Type\ t}$ will be

$$\binom{n}{k_1, k_2, \cdots, k_t}$$

不放回抽签，连抽 t 次，第 1 次抽 k_1, ..., 第 t 次抽 k_t，有 $\binom{n}{k_1, k_2, \cdots, k_t}$ 种抽法.

> 例：分班级，分队伍

逻辑理解：

若每次抽完都要排列，则无论班级/队伍存在与否，总 arrangements 都相同.

> $A_n^n = n!$. 班级/队伍就像隐形的隔板，存不存在都是那么多种排列方式.

若只抽签不排列，则需要除以 $A_{k_1}^{k_1},\ A_{k_1}^{k_1},\ \cdots,\ A_{k_t}^{k_t}$. 因此结果为 $\frac{n!}{k_1!k_2!\cdots k_t!}$.

> 班级/队伍内部顺序无关.

数学推导：

$$
\begin{aligned}
\binom{n}{k_1, k_2, \cdots, k_t} &= \binom{n}{k_1} \times \binom{n-k_1}{k_2} \times \cdots \times \binom{n-\cdots-k_{t-1}}{k_t} \\
&= \frac{n!}{k_1!\,\cancel{(n-k_1)!}} \times \frac{\cancel{(n-k_1)!}}{k_2!\,\cancel{(n-k_1-k_2)!}} \times \cdots \times \frac{\cancel{k_t!}}{k_t!0!} \\
&= \frac{n!}{k_1!k_2!\cdots k_t!}
\end{aligned}
$$

2.6* 如何选模型

> How to choose a probability model?

Approach 1: Common Sense

> For example: If there is no reason to favor one outcome over another, we assign the same probability to the outcomes.

Approach 2: Frequency of Occurrence

> We assign the probabilities based on the fraction of times that an outcome occurs when the experiment is performed many times under the same conditions.

频率估计概率.

Note: The more times we repeat the experiment the more accurate the model will be.

Example: A probability model for the gender of newborn babies in Hong Kong

> Common Sense: Equally-likely outcomes in $\Omega = \{boy,\ girl\}$
> Frequency of Occurrence: rely on the published data (Male is more)

Approach 3: Ask the Market

> We can ask the field experts and use their averaged assigned scores as the probabilities.

比赛结果预测.

Lecture 3 进阶概念

> Probability Axioms and Models

概念回顾：`1.1 样本空间`，`1.2 事件`.

3.1 交集

> Intersection of Events

The **intersection of events** happens when $\textit{all the events occur}$. We denote the intersection of events A and B with $A \cap B$.

3.2 并集

> Union of Events

The **union of events** happens when $\textit{at least one of the events occur}$. We denote the union of events A and B with $A \cup B$.

3.3 补集

> Complement of Event

The $\textit{complement of Event } A$ (denotes by A^c) is $A's$ **opposite event**. In other words, A^c happens **if and only** if A does not happen.

3.4 互斥事件

> Disjoint Events

We call events $A_1, A_2, ...$ **disjoint events** (or **mutually exclusive events**) if $\textit{the intersection of every two events } A_i,\ A_j$ $(i \neq j)$ is $\textit{the null event}$:

$\forall i \neq j,\ A_i \cap A_j = \emptyset$

> 两两不相交，为互斥事件

3.5 概率公理（二）

复习：Probability in Finite Spaces and Probability Axioms

> 见 `1.3 概率公理（一）`.

A $\textit{probability assignment } P$ to **sample space** Ω should satisfy the following $\textit{three axioms}$:

1. For every event A, $\mathbb{P}(A) \geq 0$,
2. $\mathbb{P}(\Omega) = 1$,
3. If events $A_1,\ A_2,\ \cdots$ are disjoint, $\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots$

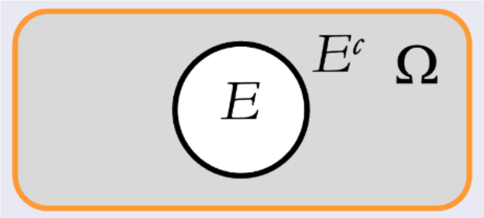
3.6 概率计算法则

> Rules for Probability Calculation

Complement Rule

For every event E and its complement E^c:

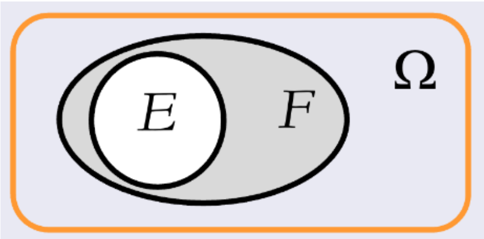
$\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$



Difference Rule

If events $E,\ F\ satisfy\ E \subseteq F$, then:

$\mathbb{P}(F \cap E^c) = \mathbb{P}(F) - \mathbb{P}(E)$



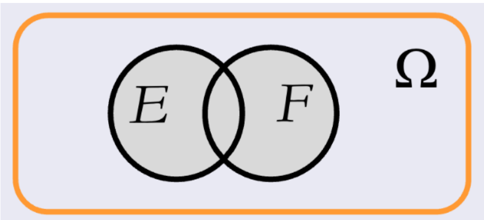
As a result, if $E \subseteq F$ then $\mathbb{P}(E) \leq \mathbb{P}(F)$.

Inclusion-Exclusion Principle

> 容斥原理

For events $E,\ F$:

$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$



可拓展：

$\mathbb{P}(E \cup F \cup G) = \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) - \mathbb{P}(E \cap F) - \mathbb{P}(F \cap G) - \mathbb{P}(G \cap E) + \mathbb{P}(E \cap F \cap G)$

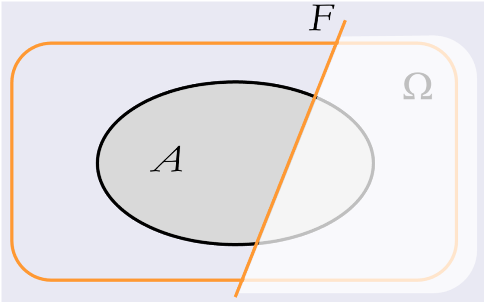
Lecture 4 条件与独立

> Conditional Probability & Independence

4.1 条件概率

> Conditional Probability

The **Conditional Probability** $\mathbb{P}(A|F)$ represents the probability of event A $assuming\ that\ event\ F\ happened$.



The **conditional probability** of A with respect to $\textit{reduced sample space } F$ is given by the formula

$$\mathbb{P}(A|F) = \frac{\mathbb{P}(A \cap F)}{\mathbb{P}(F)}$$

4.1.1 性质

> Properties of Conditional Probability

Conditional probabilities $\mathbb{P}(\cdot|F)$ are probabilities over $\textit{reduced sample space } F$ and satisfy probability axioms:

1. For every A, $\mathbb{P}(A|F) \geq 0$,
2. $\mathbb{P}(F|F) = 1$,
3. For disjoint events $A,\ B$: $\mathbb{P}(A \cup B|F) = \mathbb{P}(A|F) + \mathbb{P}(B|F)$

Uniform Probability Law for Conditional Probabilities

Under $\textit{equally-likely outcomes in } F$:

$$\mathbb{P}(A|F) = \frac{\textit{Number of outcomes in } A \cap F}{\textit{Number of outcomes in } F} = \frac{|A \cap F|}{|F|}.$$

> 注意条件，等可能结果.
>
> 关于 Uniform Probability Law，可查看 `1.4 事件概率`.

4.2 交事件概率

> The Multiplication Rule

For events $E_1,\ E_2$ we can write the probability of $\textit{their intersection } E_1 \cap E_2$ as

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)$$

In general, for every $n \textit{ events } E_1,\ E_2,\ \cdots,\ E_n$, the multiplication rule says

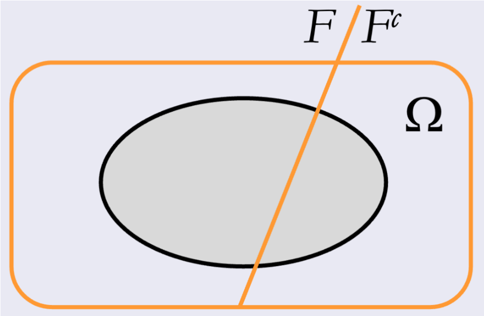
$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)\mathbb{P}(E_3|E_1 \cap E_2)\cdots\mathbb{P}(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}).$$

4.3 全概率公式

> Total Probability Theorem

For every event $E,\ F,\ F^c$,

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c) \\ &= \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|F^c)\mathbb{P}(F^c)\end{aligned}$$



More generally, if events $F_1, F_2, \cdots, F_n$ *partition* Ω (disjoint events and $F_1 \cup F_2 \cup \cdots \cup F_n = \Omega$), then the *Total Probability Theorem* says

$$\mathbb{P}(E) = \mathbb{P}(E|F_1)\mathbb{P}(F_1) + \mathbb{P}(E|F_2)\mathbb{P}(F_2) + \cdots + \mathbb{P}(E|F_n)\mathbb{P}(F_n)$$

> 注意条件：互斥，且全部的并集为 sample space
>
> 关于 $partition$ 的定义，见 `ENGG 2440 离散数学 6.1.5 集合划分`.

4.4 贝叶斯公式

> Bayes' Rule

Consider Events C and E. Then,

$$\begin{aligned}\mathbb{P}(C|E) &= \frac{\mathbb{P}(E|C)\mathbb{P}(C)}{\mathbb{P}(E)} \\ &= \frac{\mathbb{P}(E|C)\mathbb{P}(C)}{\mathbb{P}(E|C)\mathbb{P}(C) + \mathbb{P}(E|C^c)\mathbb{P}(C^c)}\end{aligned}$$

> 注意，分母用全概率公式进行变换.

More generally, if $C_1, C_2, \cdots, C_n$ **partition** the set of possible *causes* S,

$$\mathbb{P}(C_1|E) = \frac{\mathbb{P}(E|C_1)\mathbb{P}(C_1)}{\mathbb{P}(E|C_1)\mathbb{P}(C_1) + \mathbb{P}(E|C_2)\mathbb{P}(C_2) + \cdots + \mathbb{P}(E|C_n)\mathbb{P}(C_n)}$$

4.5 独立事件

> Independent Events

4.5.1 定义

We call Events A and B *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

or equivalently

$$\mathbb{P}(A|B) = \mathbb{P}(A)$$

4.5.2 性质

> Rules of Independent Events

If E, F are *independent events*, then events E^c, F will also be *independent*.

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F) \Rightarrow \mathbb{P}(E^c \cap F) = \mathbb{P}(E^c)\mathbb{P}(F)$$

证明：If $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$, then

$\mathbb{P}(E\cup E^c|F)=\mathbb{P}(E|F)+\mathbb{P}(E^c|F)=\mathbb{P}(E)+\mathbb{P}(E^c|F)=1\Rightarrow\mathbb{P}(E^c|F)=\mathbb{P}(E^c)$

> 第一个等号来自 `4.1.1 性质`，第二个等号来自题目条件，最后一个等号来自补集定义.

4.6 三事件独立

> Independence of Three Events

We call **three events** $A,\ B,\ C$ $\color{red}{independent\ events}$ if these $\color{blue}{four\ conditions}$ are satisfied:

1. A and B are $\color{red}{independent}$: $\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B)$
2. B and C are $\color{red}{independent}$: $\mathbb{P}(B\cap C)=\mathbb{P}(B)\mathbb{P}(C)$
3. A and C are $\color{red}{independent}$: $\mathbb{P}(A\cap C)=\mathbb{P}(A)\mathbb{P}(C)$
4. $\color{red}{AND}$ we require $\mathbb{P}(A\cap B\cap C)=\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$

> 前三个条件不能充分说明独立性，前三个都满足的情况下第四个条件也可能不满足.
>
> 四个都满足，才能证明三个事件是独立事件.
>
> 关于这么定义的原因，见 `4.7 多事件独立`.

4.7 多事件独立

> Independence of Several Events

We call n **events** $A_1,\ A_2,\ \cdots,\ A_n$ $\color{red}{independent\ events}$ if for $\color{red}{every\ subset\ of\ \{A_1,\ \cdots,\ A_n\}}$, the probability of **the intersection** is **the product** of their probabilities.

If n **events** $\{A_1,\ \cdots,\ A_n\}$ are $\color{red}{independent\ events}$, then the independence is $\color{blue}{preserved}$ when we **replace** some event(s) by their complements, intersections, unions.

> 交并补仍然独立，这是多事件独立性采用这种定义的出发点.

4.8 条件独立

> Conditional Independence

Events A and B are $\color{red}{independent\ conditioned\ on\ event\ F}$ if

> 事件 A,B 对于给定事件 F 是条件独立的

$\mathbb{P}(A\cap B|F)=\mathbb{P}(A|F)\mathbb{P}(B|F)$

Note that the above equation is **equivalent to**

$\mathbb{P}(A|B\cap F)=\mathbb{P}(A|F)$

> 在 F 条件下，B 发生与否对 A 的概率没有影响.

用条件概率的定义，这两个公式可以互推：

$\mathbb{P}(A\cap B|F)=\mathbb{P}(A|F)\mathbb{P}(B|F)\Leftrightarrow\mathbb{P}(A\cap B\cap F)=\mathbb{P}(A|F)\mathbb{P}(B|F)\mathbb{P}(F)$

$\mathbb{P}(A|B\cap F)=\mathbb{P}(A|F)\Leftrightarrow\mathbb{P}(A\cap B\cap F)=\mathbb{P}(A|F)\mathbb{P}(B|F)\mathbb{P}(F)$

注意：Independence does NOT imply conditional independence

> 两个事件独立，不意味着它们在某些条件下仍然独立.

> 例如，roll two dice，在和为 7 的条件下，"the second die is 3" 对 "the first die is 4" 的概率有影响（从 $\frac{1}{6}$ 变为 100%）

Lecture 5 随机变量

> Random Variables：概念介绍

5.1 离散随机变量

> Discrete random variable

A $\color{red}{discrete\ random\ variable}$ assigns a discrete value to every outcome in the sample space Ω.

For example: Random variable X is the number of Heads in tossing 3 coins.

> 扔三次硬币，正面出现的次数就是一个离散随机变量（$0, 1, 2, 3$，涵盖所有结果）.

5.2 概率质量函数 PMF

> Probability Mass Function，有时也叫离散密度函数（discrete density function）. 注意，不要和概率密度函数（Probability Density Function, PDF）混淆，后者用于连续随机变量.

The **Probability Mass Function (PMF)** $p : R \to [0, 1]$ of a discrete random variable X is the function

$$p(x) = \mathbb{P}(X = x)$$

We can describe the PMF by a $\color{blue}{table}$ or by a $\color{brown}{chart}$.

x	0	1	2	3
$p(x)$	$1/8$	$3/8$	$3/8$	$1/8$

Table describing the PMF

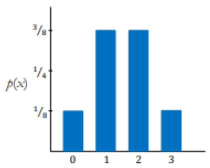


Chart describing the PMF

For every random variable X, its **probability mass function** satisfies the following (based on $\color{red}{the\ axioms\ of\ probability}$):

1. For every $x \in \mathbb{R}$, $p(x) \geq 0$. 【$\color{blue}{non\text{-}negative}$】
2. If χ is the set of all possible values of X, then the PMF values on χ will add up to 1:

$$\sum_{x \in \chi} p(x) = 1.$$

Lecture 6 二项/几何分布

> Random Variables：二项和几何分布

6.1 二项分布

> Binomial Random Variable

We call X a $Binomial(n,p)$ $Random\ Variable$ when X represents $the\ number\ of\ successes$ over $n\ independent\ trials$, each with a $success\ probability\ of\ p$.

> 隐藏条件：试验次数 n 和每次的成功率 p 恒定；独立实验（如果是抽东西，需要放回，否则不是二项分布）.
>
> 成功率 p 恒定，也暗示了各次 trial 之间独立.
>
> X 满足二项分布，简记：$X \sim Binomial(n,p)$.
>
> 这里的每次 $independent\ trial$ 有两种 $result$，成功记为 1，失败记为 0. 该 $result$ 即 $Lecture\ 8$ 会提到的 $Bernoulli\ Random\ Variable$.

Example 1: We toss n coins. The $number\ of\ Heads$ is $Binomial(n, \frac{1}{2})$.

Example 2: We toss n dice. The $number\ of\ 3's$ is $Binomial(n, \frac{1}{6})$.

Counter Example:

We draw a 10-card hand from a 52-card deck. Let

$N = number\ of\ Aces\ among\ the\ picked\ cards.$

Which one is the random variable N?

1. $Binomial(52, \frac{1}{4})$
2. $Binomial(52, \frac{1}{13})$
3. $Binomial(10, \frac{1}{4})$
4. $Binomial(10, \frac{1}{13})$
5. $None$

注意审题，看似是 4，实际上是 5，因为每次抽牌的概率不同（不独立）.

这实际上是超几何分布.

为什么不放回抽牌，各次抽牌之间是不独立的?

Proof:

首先证明均匀分布（任意一个位置抽到 A 的概率都是 $\frac{1}{13}$）.

考虑从 52 张的牌堆里连续抽 2 张. 设 A 表示 "第一次抽到 Ace"，B 表示 "第二次抽到 Ace".

$P(A) = \frac{4}{52} = \frac{1}{13}$，$P(B) = P(A) \cdot \frac{3}{51} + (1 - P(A)) \cdot \frac{4}{51} = \frac{1}{13}$

法二：选择 1 张 A 放在第 N 张的位置，有 4 种选择. 剩下 51 张牌任意排列，有 $51!$ 种排列方式.

$P(第N次抽到A) = \frac{4\times 51!}{52!} = \frac{1}{13}$

$\Rightarrow P(A) = P(B) = \frac{1}{13}$

然后证明不独立:

$P(AB) = P(A) \cdot P(B|A) = \frac{1}{13} \cdot \frac{3}{51} \neq P(A) \cdot P(B)$

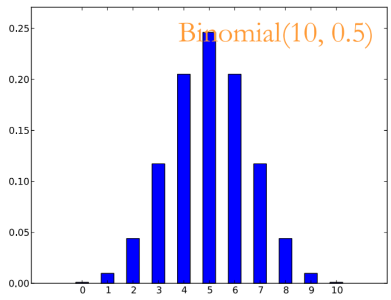
6.1.1 PMF

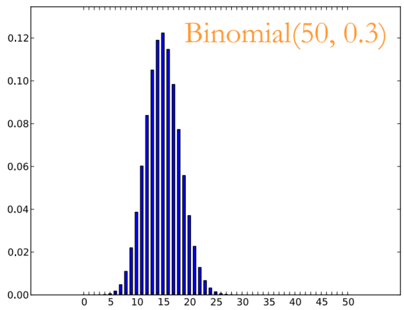
> PMF of Binomial Random Variables

The **probability mass function (PMF)** of a $Binomial(n, p)\ Random\ Variable$ is

$$p(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

> 二项分布才能用！各次 trial 不独立时（比如抽牌）不能这么算
>
> k 是从 0 到 n 的整数.





期望：np

6.1.2 期望/众数/方差

期望：np

> 见 `9.5 二项分布的期望`.

众数：$\lfloor (n+1)p \rfloor$

> 若 $(n+1)p$ 是整数，则有 $(n+1)p$，$(n+1)p-1$ 两个众数

方差：$np(1-p)$

> 见 `10.12 二项分布的方差`.

6.2 几何分布

> Geometric Random Variable（注意，不是超几何分布）

We call X a $Geometric(p)\ Random\ Variable$ when X represents $the\ first\ time\ of\ success$ over $a\ series\ of\ independent\ trials\ X_1, X_2, \cdots,$ each with a $success\ probability\ of\ p$.

$X =$ first (smallest) n such that $X_n = 1$ (success).

> 描述 n 次实验中，第一次成功在第 k 次的概率.
>
> X 满足几何分布，简记：$X \sim Geometric(p)$.

Example 1: We toss a coin until we see the first Heads. The $number\ of\ coin\ tosses\ to\ see\ the\ first\ Heads$ is $Geometric(\frac{1}{2})$.

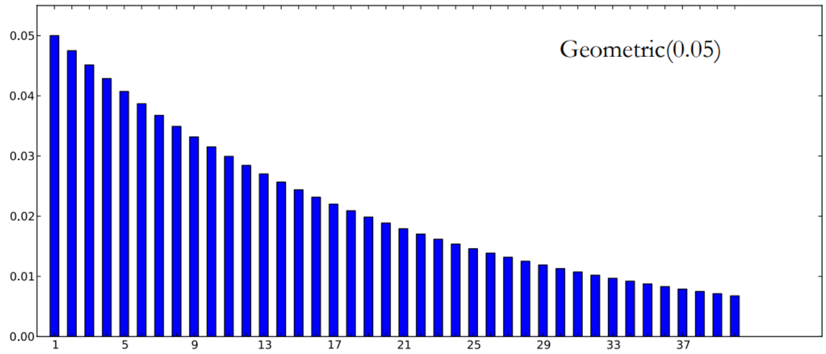
Example 2: We roll dice until we see the first 6. The $number\ of\ die\ rolls\ to\ see\ the\ first\ 6$ is $Geometric(\frac{1}{6})$.

6.2.1 PMF

> PMF of Geometric Random Variables

The **probability mass function (PMF)** of a $Geometric(p)\ Random\ Variable$ is

$p(k) = \mathbb{P}(X = k) = p(1-p)^{k-1}$



> 这是一个等比数列，英文又名几何数列，因此这种分布叫做几何分布.
>
> k 是从 1 到 ∞ 的整数.

6.2.2 期望/众数/方差

期望：$\frac{1}{p}$

> 见 `10.6 几何分布的期望`.

众数：1

方差：$\frac{1-p}{p^2}$

> 见 `10.7 几何分布的方差`.

Lecture 7 CDF 泊松 期望

> Random Variables：累积分布函数、泊松分布、期望

7.1 累积分布函数

> Cumulative Distribution Function

For a random variable X, its cumulative distribution function (CDF) $F(x)$ is:

$F(x) = \mathbb{P}(X \le x) = 1 - \mathbb{P}(X > x)$

- For a $Geometric(p)$ random variable X, the CDF will be

$F(k) = \mathbb{P}(X \le k) = 1 - \mathbb{P}(X > k) = 1 - (1-p)^k$

法二：$F(k) = p + p(1-p) + p(1-p)^2 + \cdots + p(1-p)^{k-1} = p \cdot \frac{1-(1-p)^k}{1-(1-p)} = 1 - (1-p)^k$

- For a $Binomial(n, p)$ random variable X, the CDF will be

$F(k) = \mathbb{P}(X \le k) = \binom{n}{0} p^0 (1-p)^{n-0} + \binom{n}{1} p^1 (1-p)^{n-1} + \cdots + \binom{n}{k} p^k (1-p)^{n-k}$

> 注意，如果这里 k 很接近 n，例如 $k=n-2$ 或 $k=n-1$，要用补集 $1-\mathbb{P}(X>k)$ 化简.

- Question: Alice randomly sprinkles 25 chocolate chips on 5 cookies. What is the random variable N on how many chips a cookie gets?
  - Answer: $N\sim \text{Binomial}(25,\frac{1}{5})$
- What is the probability a cookie gets no chips?
  - Answer: $\mathbb{P}(N=0)=\left(1-\frac{1}{5}\right)^{25}\approx 0.004$
  - 法二：$\mathbb{P}(N=0)=\frac{4^{25}}{5^{25}}\approx 0.004$
- What is the probability a cookie gets exactly 5 chips?
  - Answer: $\mathbb{P}(N=5)=\binom{25}{5}\left(\frac{1}{5}\right)^5\left(1-\frac{1}{5}\right)^{20}\approx 0.196$
  - 法二：$\mathbb{P}(N=5)=\frac{\binom{25}{5}4^{20}}{5^{25}}\approx 0.196$
- What is the probability a cookie gets at most 5 chips?
  - Answer: $\mathbb{P}(N\le 5)=\binom{25}{0}(\frac{1}{5})^0(\frac{4}{5})^{25}+\cdots+\binom{25}{5}(\frac{1}{5})^5(\frac{4}{5})^{20}\approx 0.617$

- Question: Alice randomly sprinkles 250 chocolate chips on 50 cookies. What is the Random variable N on how many chips a cookie gets?
  - Answer: $N\sim \text{Binomial}(250,\frac{1}{50})$
- What is the probability a cookie gets no chips?
  - Answer: $\mathbb{P}(N=0)=\left(1-\frac{1}{50}\right)^{250}\approx 0.0064$
- What is the probability a cookie gets exactly 5 chips?
  - Answer: $\mathbb{P}(N=5)=\binom{250}{5}\left(\frac{1}{50}\right)^5\left(1-\frac{1}{50}\right)^{245}\approx 0.177$
- What is the probability a cookie gets at most 5 chips?
  - Answer: $\mathbb{P}(N\le 5)\approx 0.616$

$$\sum_{i=0}^{5}\binom{250}{i}\cdot(\frac{1}{50})^i\cdot(1-\frac{1}{50})^{250-i}$$

> 计算机会溢出，需另辟蹊径：使用泊松分布近似.
>
> 2024.12.10：有个很尴尬的事，复习的时候用卡西欧重按了一次，直接算出来了...

7.2 泊松分布

> Poisson Random Variable

二项分布试验次数 n 足够大（相比期望）时，Binomial Random Variable X 可以近似为 $\textit{Poisson Random Variable}$ X. 这时 n 与 p 不再是变量，只有一个变量 $\lambda=np$.

> 试验次数足够大是相对的，换句话说就是每次成功率足够小（远小于 1）
>
> 变量只有 λ，意思是 $n=100$，$p=\frac{1}{100}$ 和 $n=1000$，$p=\frac{1}{1000}$ 的二项分布会有差异，但是泊松分布的概率质量函数是完全相同的. 这里的 λ 是期望.
>
> 注意，当使用泊松分布来近似二项分布，可能有不合常理的情况，例如能算出 $k>n$ 有非零概率（虽然很小，但也不符合实际情况，实际的二项分布是严格等于 0）
>
> 泊松分布继承二项分布的性质，因此也有隐藏条件：n,p,λ 恒定，独立实验.
>
> X 近似满足泊松分布，简记：$X\sim Poisson(\lambda)$.

7.2.1 PMF

> PMF of Poisson Random Variables

A $Poisson(\lambda)\ random\ variable$ X has the PMF:

$p(k)=e^{-\lambda}\frac{\lambda^k}{k!}$

> k 是从 0 到正无穷的整数.

n 充分大时，为什么可以近似成泊松?

Solution:

考虑二项分布 $X\sim Binomial(n,p)$.

$p(k)=\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

期望 $\lambda=np\Rightarrow p=\frac{\lambda}{n}$

$k=0$ 时，$\lim\limits_{n\to\infty}p(k)=\lim\limits_{n\to\infty}(1-\frac{\lambda}{n})^n=e^{-\lambda}$

$k\geq 1$ 时，

$$\begin{aligned}\lim_{n\to\infty}p(k)&=\lim_{n\to\infty}\frac{n!}{k!(n-k)!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}\\&=\lim_{n\to\infty}\frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{\lambda^k}{k!}(1-\frac{\lambda}{n})^{n-k}\\&=\lim_{n\to\infty}1\cdot(1-\frac{1}{n})\cdots(1-\frac{k-1}{n})\frac{\lambda^k}{k!}(1-\frac{\lambda}{n})^{n-k}\\&=e^{-\lambda}\frac{\lambda^k}{k!}\end{aligned}$$

7.2.2 期望/众数/方差

期望：λ

> 见 `9.6 泊松分布的期望`.

众数：$\lfloor\lambda\rfloor$

> 若 λ 是整数，则有 λ，$\lambda-1$ 两个众数.

方差：λ

> 见 `10.13 泊松分布的方差`.

7.3 期望

> 数学期望，Expected Value of a Random Variable
>
> 这里讨论的

概率论中称为期望，统计学中一般称为均值 ($Mean$)，都是一种加权平均操作.

The expected value (expectation) of a random variable X with PMF $p(x)$ is

$\mathbb{E}[X]=\sum\limits_{x}x\,p(x)$

> $\mathbb{E}$ 使用中括号，因为里面是一个数组（随机变量 X）

The expectation is the **average value** the random variable takes when the experiment is done $\textit{many times}$.

7.4 单随机变量的函数

> Function of a Random Variable

$Y = f(X)$

随机变量的函数也是随机变量.

> 函数的意思是，把 X 的每个结果进行 f 映射，得到新的分布 Y.

7.4.1 PMF

> PMF of a function of Random Variable

If X is a random variable with PMF p_X , then $Y = f(X)$ will also be a random variable with PMF p_Y :

$$p_Y(y) = \sum_{x:f(x)=y} p_X(x)$$

> 可能有多个 x 映射到同一个 y 上，需要把它们的概率求和得到新的概率.

Example:

x	0	1	2
$p(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Let $y = f(x) = (x-1)^2$

y	0	1
$p(y)$	$\frac{1}{3}$	$\frac{2}{3}$

7.4.2 期望

> Expectation of a Function of Random Variable

The Expected Value of $f(x)$ for a function f and random variable X is

$$\mathbb{E}[f(X)] = \sum_x f(x)p(x)$$

> 理解：$p(x)$ 从原本对 x 做贡献映射到对 $f(x)$ 做贡献，因此乘一个 $f(x)$ 因子.

Example

PMF of X:

$$\begin{array}{c|ccc} x & 0 & 1 & 2 \\ \hline p(x) & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}$$

$$\mathbb{E}[X] = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} = 1$$

$$\mathbb{E}[X-1] = (0-1) \cdot \frac{1}{3} + (1-1) \cdot \frac{1}{3} + (2-1) \cdot \frac{1}{3} = 0$$

$\mathbb{E}[(X-1)^2] = (0-1)^2 \cdot \frac{1}{3} + (1-1)^2 \cdot \frac{1}{3} + (2-1)^2 \cdot \frac{1}{3} = \frac{2}{3}$

For a general f, $\mathbb{E}[f(X)] \neq f(\mathbb{E}[X])$

> 例如上述例子，$\mathbb{E}[(X-1)^2] \neq (\mathbb{E}[X]-1)^2$
>
> 线性映射可以取等，但最好还是认真计算.

Lecture 8 复习周

> 期中复习

Lecture 9 联合分布

> 注意，这里讨论的是离散.

复习：期望

The expected value (expectation) of a random variable X with PMF $p(x)$ is

$\mathbb{E}[X] = \sum_x x\, p(x)$

The Expected Value of $f(x)$ for a function f and random variable X is

$\mathbb{E}[f(X)] = \sum_x f(x) p(x)$

9.1 联合 PMF

> Joint Probability Mass Function

The joint PMF of random variables X, Y is the bivariate function

$p(x, y) = \mathbb{P}(X = x,\ Y = y)$

> $\mathbb{P}(X = x,\ Y = y)$ 指 $X = x$ 和 $Y = y$ 同时发生的概率.
>
> 例：摇多个骰子，多张卡抽卡
>
> 注意，此处并没有说 $X = x$ 和 $Y = y$ 这两个事件是独立的.

9.2 双随机变量的函数

> PMF and Expectation of a Function of Two Random Variables

If $X,\ Y$ are two random variables with Joint PMF p_{XY}, then $Z = f(X, Y)$ will also be a random variable with PMF p_Z:

$p_Z(z) = \sum_{x,y: f(x,y)=z} p_{XY}(x, y)$

> 一个很好理解的例子是，投两个骰子，第一个 X, 第二个 Y, $Z = f(X, Y) = X + Y$ 是两个骰子的和.
> 那么 $p_Z(3) = p_{XY}(1, 2) + p_{XY}(2, 1)$

The Expected Value of $f(X,Y)$ for a function f and random variables X, Y is

$$\mathbb{E}[f(X,Y)] = \sum_{x,y} f(x,y)\, p_{XY}(x,y)$$

> 这和上面的 PMF 不冲突，对于可能映射到同一个 z 的不同 x, y 组合，这个求和都考虑在内.
>
> 不应该用 $p_{f(X,Y)}(f(x,y))$ 来算，因为求和符号的下标是 x, y，而不是 $f(x,y)$.

当然也可以写成这样：

$$\mathbb{E}[f(X,Y)] = \sum_{f(x,y)} f(x,y)\, p_{f(X,Y)}(f(x,y))$$

> 注意，$f(x,y)$ 如果有重复，只计 1 次.

或

$$\mathbb{E}[Z] = \sum_{z} z\, p_Z(z)$$

9.3 期望的线性

> Linearity of Expectation

The Expected Value of $X + Y$, i.e. the **sum** of random variables X, Y satisfies

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

> 注意：期望的线性对任意变量的线性组合都成立，但方差的线性要满足不相关前提.

9.4 伯努利分布

> Bernoulli Random Variable，又名两点分布，0-1分布

A $Bernoulli(p)\ random\ variable$ X shows the result of a $trial$ where $X = 1$ for the $success\ outcome$ with $probability\ p$ and $X = 0$ for the $failure\ outcome$ with $probability\ 1-p$.

> 注意只有一个参数 p，因为只进行一次实验.

The PMF of a $Bernoulli(p)$ random variable is

x	0	1
$p(x)$	$1-p$	p

The expected value of a $Bernoulli(p)$ random variable is

$$\mathbb{E}[X] = 0 \times (1-p) + 1 \times p = p$$

方差：$p(1-p)$

> 见 `10.12 二项分布的方差`.

9.5 二项分布的期望

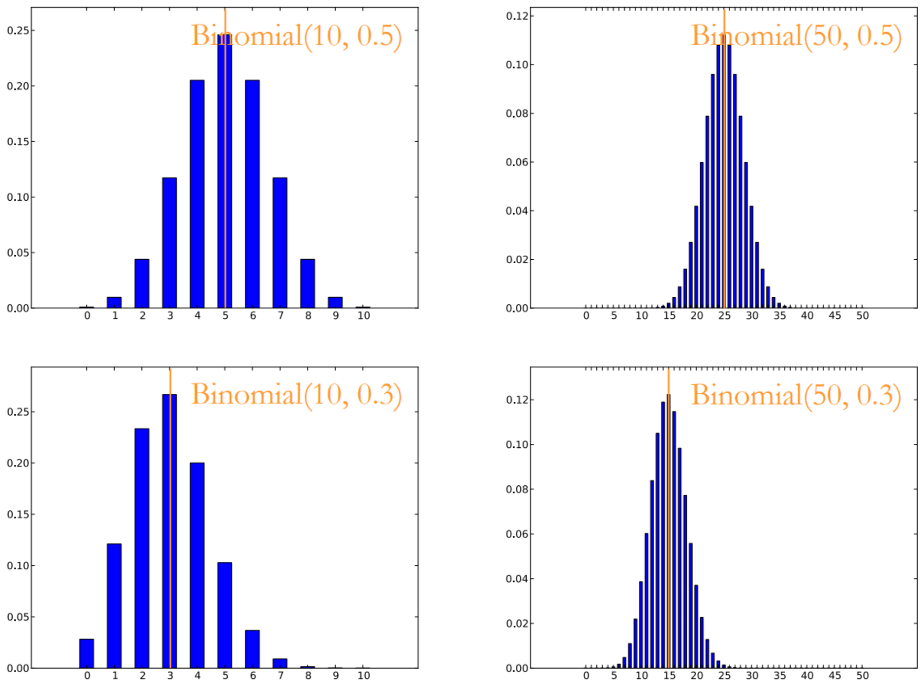
> Mean of Binomial Random Variable

A $\color{red}{Binomial(n,p)}$ random variable is the $\color{red}{sum}$ of n independent $Bernoulli(p)$ random variables $X_1,\ X_2,\ \cdots,\ X_n$. 因此,

$X = X_1 + X_2 + \cdots + X_n.$

The expected value of a $\color{red}{Binomial(n,p)}$ random variable is

$\mathbb{E}[X] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = np.$



> 注意：均值不一定等于众数！这四张图只是恰好相等.
>
> 期望：np
>
> 众数：$\lfloor (n+1)p \rfloor$
>
> 方差：$np(1-p)$

9.6 泊松分布的期望

> Mean of Poisson Random Variable

The **expected value** of a $\color{red}{Poisson(\lambda)\ random\ variable\ X}$ is

$\mathbb{E}[X] = \lambda$

> Informal Proof: $\lim\limits_{n\to\infty} Binomial(n, \frac{\lambda}{n}) = Poisson(\lambda)$ and so
>
> $\mathbb{E}[X] = \lim\limits_{n\to\infty} n \cdot \frac{\lambda}{n} = \lambda$

Lecture 10 方差

10.1 方差和标准差

> Variance and Standard Deviation

Consider random variable X with $\textit{expected value } \mu = \mathbb{E}[X]$. Then, we define the $\textit{variance}$ of X as

$Var(X) = \mathbb{E}[(X-\mu)^2]$

Furthermore, we define the $\textit{standard deviation}$ of X to be

$\sigma = \sqrt{Var(X)} = \sqrt{\mathbb{E}[(X-\mu)^2]}$

> Note that $\textit{variance}$ measures how close X and $\mathbb{E}[X]$ are for a $\textit{typical outcome of } X$.

10.2 方差公式

> Another Formula for Variance

$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

> 注：X^2 是 random variable，每个 possible value 是 X 对应取值的平方.

证明：由期望的线性，

$$\begin{aligned} Var[X] &= \mathbb{E}[(X-\mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2 + \mathbb{E}[X]^2 - 2X\mathbb{E}[X]] \\ &= \mathbb{E}[X^2] + \mathbb{E}[X]^2 - 2\mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \end{aligned}$$

Example: 投一个六面骰子

$\mathbb{E}[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$

$Var[X] = \frac{35}{12} \approx 2.9167$

> 法一：$Var(X) = \mathbb{E}[(X-\mu)^2] = (1-3.5)^2 \times \frac{1}{6} + (2-3.5)^2 \times \frac{1}{6} + \cdots + (6-3.5)^2 \times \frac{1}{6}$
>
> 法二：$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = (1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \cdots + 6^2 \times \frac{1}{6}) - 3.5^2$

10.3 条件 PMF

> Conditional PMF

The $\textit{conditional PMF } p_{X|Y}(x|y) \textit{ of } X \textit{ given } Y$ is defined as

$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{\mathbb{P}(X=x,\, Y=y)}{\mathbb{P}(Y=y)}$

> For a fixed y, $p_{X|Y}(x|y)$ is a PMF as a function of x.

Example: Roll two 3-sided dice.

$X \in \{1,2,3\}$

$Y \in \{1,2,3\}$

$S = X + Y \in \{2,3,4,5,6\}$

(1) What is the PMF of the sum given the first roll?

The PMF of the first roll X:

x	$p(x)$
1	$\frac{1}{3}$
2	$\frac{1}{3}$
3	$\frac{1}{3}$

The Joint PMF of the first roll X and the sum S:

$X \setminus S$	2	3	4	5	6
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0	0
2	0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0
3	0	0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

The Conditional PMF of S given X:

$X \setminus S$	2	3	4	5	6
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	0
2	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
3	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

(2) What is the PMF of the first roll given the sum?

待补充.

(2024.12.13) 复习补充

The PMF of the sum S:

s	$p(s)$
2	$\frac{1}{9}$
3	$\frac{2}{9}$
4	$\frac{3}{9}$
5	$\frac{2}{9}$
6	$\frac{1}{9}$

The Joint PMF of the first roll X and the sum S:

$X \setminus S$	2	3	4	5	6
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0	0
2	0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0
3	0	0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

The Conditional PMF of X given S:

$X \setminus S$	2	3	4	5	6
1	1	$\frac{1}{2}$	$\frac{1}{3}$	0	0
2	0	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$	0
3	0	0	$\frac{1}{3}$	$\frac{1}{2}$	1

求解这类 Conditional PMF 问题，只需要按顺序列出三个表：

① 条件事件的 PMF（如果只有一个 possible value 就不用）

② 条件事件和待求事件的 Joint PMF

③ 待求事件在条件事件发生后的 Conditional PMF

> Conditional PMF 表的画法：条件事件本身可能有分布. 定住 Joint PMF 表（表 ②）中条件事件的某个 possible value，用这个 value 去查条件事件的 PMF 表（表 ①），查出条件事件取到该 value 的概率. 将表 ② 中 possible value 所在行（或列）的所有单元格除以这个查到的概率. 遍历条件事件的所有可能取值，得到 Conditional PMF 表.

10.4 条件期望

> Conditional Expectation

The conditional expectation $\mathbb{E}[X|Y=y]$ of X given $Y=y$ is defined as

$$\mathbb{E}[X|Y=y]=\sum_x x\cdot p_{X|Y}(x|y)$$

> For a fixed y, $p_{X|Y}(x|y)$ is a PMF as a function of x.

10.5 全期望定理

> Total Expectation Theorem

For random variables X,Y, the following holds:

$$\mathbb{E}[X]=\sum_y \mathbb{P}(Y=y)\mathbb{E}[X|Y=y]$$

> 加权平均操作，类似全概率公式.
>
> 等式右边相当于 $\mathbb{E}[\mathbb{E}[X|Y]]$
>
> $\mathbb{E}[X|Y=y]$ 是一个值（条件期望值），$\mathbb{E}[X|Y]$ 是一个分布，对于 Y 的不同取值 y, $\mathbb{E}[X|Y]$ 可能有不同值 $\mathbb{E}[X|Y=y]$.

证明：

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_x x \cdot p(x) \\
&= \sum_x x \cdot \mathbb{P}(X = x) \\
&= \sum_x x \cdot (\sum_y \mathbb{P}(Y = y)\mathbb{P}(X = x|Y = y)) \qquad (Total\ Probability\ Theorem) \\
&= \sum_y \sum_x x \cdot \mathbb{P}(Y = y) \cdot p_{X|Y}(x|y) \\
&= \sum_y \mathbb{P}(Y = y) \sum_x x \cdot p_{X|Y}(x|y) \\
&= \sum_y \mathbb{P}(Y = y)\mathbb{E}[X|Y = y]
\end{aligned}
$$

Total expectation theorem can be **equivalently** shown for **disjoint** events $A_1, A_2, \cdots, A_k$ partitioning the sample space $A_1 \cup \cdots \cup A_k = \Omega$ as

$$\mathbb{E}[X] = \sum_{i=1}^{k} \mathbb{P}(A_i)\mathbb{E}[X|A_i]$$

10.6 几何分布的期望

> Expected Value of Geometric Random Variable

Consider a $Geometric(p)$ random variable X. Then, $\mathbb{E}[X] = \frac{1}{p}$

证明：use the Total Expectation Theorem

$A = \{X > 1\}, A^c = \{X = 1\}$

$$
\begin{aligned}
\mathbb{E}[X] &= \mathbb{P}(A)\mathbb{E}[X|A] + \mathbb{P}(A^c)\mathbb{E}[X|A^c] \\
&= (1 - p) \cdot (1 + \mathbb{E}[X]) + p \cdot 1
\end{aligned}
$$

$\Rightarrow \mathbb{E}[X] = \frac{1}{p}$

> $\mathbb{E}[X|A] = 1 + \mathbb{E}[X]$，并没有给出证明.
>
> 事实上，我们应该用法二先严格证明得到结论，然后反推出 $\mathbb{E}[X|A] = 1 + \mathbb{E}[X]$（非记忆性，Memoryless Property）

法二（严谨）：

$$
\begin{aligned}
\mathbb{E}[X] &= \lim_{n\to\infty} 1\cdot p + 2\cdot(1-p)\cdot p + \cdots + n\cdot(1-p)^{n-1}\cdot p \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} i\cdot(1-p)^{i-1}\cdot p \\
&= -p\lim_{n\to\infty} \sum_{i=1}^{n} -i\cdot(1-p)^{i-1} \\
&= -p\lim_{n\to\infty} \sum_{i=1}^{n} [(1-p)^{i}]' \\
&= -p\lim_{n\to\infty} [\sum_{i=1}^{n} (1-p)^{i}]' \\
&= -p\lim_{n\to\infty} [(1-p)\cdot\frac{1-(1-p)^n}{1-(1-p)}]' \\
&= -p\lim_{n\to\infty} [\frac{(1-p)-(1-p)^{n+1}}{p}]' \\
&= -p\lim_{n\to\infty} [\frac{1}{p}-1-\frac{(1-p)^{n+1}}{p}]' \\
&= -p\lim_{n\to\infty} [-\frac{1}{p^2}-\frac{-(n+1)(1-p)^n\cdot p-(1-p)^{n+1}}{p^2}] \\
&= \frac{1}{p}
\end{aligned}
$$

> 求导那步，严格写应该是 $f(x)$ 求导后代入 p. p 不是变量，只是个常数，不然一开始就不能移 1 个 p 到 $\sum$ 外面.

10.7 几何分布的方差

> Variance of Geometric Random Variable

Consider a $Geometric(p)$ random variable X. Then, $Var[X] = \frac{1-p}{p^2}$

证明：

$$
\begin{aligned}
Var[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \frac{2-p}{p^2} - \frac{1}{p^2} \\
&= \frac{1-p}{p^2}
\end{aligned}
$$

其中，设 $A = \{X > 1\}$, $A^c = \{X = 1\}$

$$
\begin{aligned}
\mathbb{E}[X^2] &= \mathbb{P}(A^c)\mathbb{E}[X^2|A^c] + \mathbb{P}(A)\mathbb{E}[X^2|A] \\
&= p\cdot 1^2 + (1-p)\cdot\mathbb{E}[X^2|X>1] \\
&= p\cdot 1^2 + (1-p)\cdot\mathbb{E}[(X+1)^2] \\
&= p\cdot 1^2 + (1-p)\cdot(\mathbb{E}[X^2] + \frac{2}{p} + 1)
\end{aligned}
$$

$\Rightarrow \mathbb{E}[X^2] = \frac{2-p}{p^2}$

> 此处应该有一个定义证明，时间不够还没写. 下次看到的时候，有空就写一下.

(2024.12.11 复习补充) 法二（严谨）：

已知 $a \neq 1$ 时，$\sum_{x=1}^{n} xa^x = \frac{a\left(1-(n+1)a^n+na^{n+1}\right)}{(1-a)^2}$

$$
\begin{aligned}
\mathbb{E}[X] &= \lim_{n\to\infty}\sum_{i=1}^{n} i^2\cdot(1-p)^{i-1}\cdot p \\
&= -p\lim_{n\to\infty}\sum_{i=1}^{n} -i^2\cdot(1-p)^{i-1} \\
&= -p\lim_{n\to\infty}\sum_{i=1}^{n}[i\cdot(1-p)^i]' \\
&= -p\lim_{n\to\infty}[\sum_{i=1}^{n} i\cdot(1-p)^i]' \\
&= -p\lim_{n\to\infty}[\frac{(1-p)\left(1-(n+1)(1-p)^n+n(1-p)^{n+1}\right)}{p^2}]' \\
&= \frac{2-p}{p^2}
\end{aligned}
$$

> 小技巧：含概率 n 次幂的项求导后数量级还是概率 n 次幂，趋于无穷后为 0，因此不妨在求导前就忽略这些项，减小计算量.

10.8 独立分布

> Independent Random Variables
>
> 注意和后文的"独立同分布"区分.

X and Y are called **independent random variables** if every outcome pair $X=x$ and $Y=y$ are **independent events** for all x, y values:

$\mathbb{P}(X=x,\ Y=y)=\mathbb{P}(X=x)\mathbb{P}(Y=y)$

> 注意要对所有可能的 x, y 都成立.

Note that the $joint\ PMF\ p_{X,Y}$ of independent random variables X, Y can be written as the **product** of their $marginal\ PMF\ p_X,\ p_Y$:

$p_{X,Y}(x,y)=p_X(x)\cdot p_Y(y)$

> marginal 边缘，边际. 边缘分布和联合分布是一组相对概念，如果只有一个变量就不用强调边缘，直接说 PMF 即可.

① 判定：条件 PMF

> Independent Random Variables and Conditional PMF

Random Variables $X,\ Y$ are independent if and only if for every outcome y the conditional PMF $p_{X|Y}(x|y)$ is the same as X's PMF $p_X(x)$:

$$p_{X|Y}(x|y)=p_X(x)$$

> 这里需要同时满足任意的 $y\in Y$ 和任意的 $x\in X$.

② 判定：期望

> Independence and Expectation

X and Y are *independent* **if and only if** for *every function* f, g we have

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$$

> 证明待补充.

(2024.12.12) 直观理解：

独立性是全局性质，X 和 Y 独立意味着它们的所有可能行为（包括分布）完全解耦，即 X 的任何信息都不能影响 Y，反之亦然. 这种全局性质需要通过对所有函数 f 和 g 的验证来确保.

数学证明：

$$\mathbb{E}[f(X)g(Y)] = \sum_{x,y} f(X)g(Y)\mathbb{P}_{XY}(x,y)$$
$$\mathbb{E}[f(X)]\mathbb{E}[g(Y)] = [\sum_{x} f(X)\mathbb{P}_X(x)][\sum_{y} g(Y)\mathbb{P}_Y(y)]$$

左边相等意味着 X 与 Y 的任意函数都是解耦的.

> 其实我还是没太搞懂.

Note: $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ is **not enough** to guarantee X and Y are independent.

> 只能说明不相关.

For example, for random variable X with $\mathbb{P}(X=-1) = \mathbb{P}(X=0) = \mathbb{P}(X=1) = \frac{1}{3}$, X and $Y = X^2$ satisfy

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = 0 = \mathbb{E}[X]\mathbb{E}[Y]$$

However, X and Y are **not independent** in this example.

10.9 协方差

> Covariance

The *covariance* of random variables X, Y with expected values $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$ is defined as

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

The covariance can be also found using the formula

$$Cov[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

> 证明待补充.

(2024.12.12) 复习补充：

$$\begin{aligned} Cov[X,Y] &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\ &= \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

10.10 不相关分布

> Uncorrelated Random Variables

We call random variables X and Y $uncorrelated$ if their $covariance\ is\ zero$ $(Cov[X,Y]=0)$ or equivalently $\mathbb{E}[XY]=\mathbb{E}[X]\mathbb{E}[Y]$.

Therefore, every $independent$ $X,\ Y$ will be $uncorrelated$, but the converse of this statement is **not always true**.

> 独立一定不相关，不相关不一定独立.

10.11 方差的线性

> Variance of Sum of Independent Random Variables

Suppose X,Y are **independent**. Then,

$$Var[X+Y]=Var[X]+Var[Y]$$

Note: The above identity holds only for $uncorrelated$ X,Y, and it does **not** hold for $correlated$ random variables.

> 等价于协方差为 0.
>
> 只要不相关就成立. 独立是更强的条件，只要独立就不相关. 自然也成立.

证明：

$$\begin{aligned} Var[X+Y] &= \mathbb{E}[(X+Y)^2]-\mathbb{E}[X+Y]^2 \\ &= \mathbb{E}[X^2]+\mathbb{E}[Y^2]+2\mathbb{E}[XY]-(\mathbb{E}[X]+\mathbb{E}[Y])^2 \\ &= Var[X]+Var[Y]+2Cov[X,Y] \end{aligned}$$

当且仅当 $Cov[X,Y]=0$ 时（不相关），$Var[X+Y]=Var[X]+Var[Y]$.

> 注意，$Var[aX+bY]=a^2\,Var[X]+b^2\,Var[Y]+2ab\,Cov[X,Y]$.

$Var[aX]=a^2\,Var[X]$ 而非 $a\,Var[X]$，是因为 aX 相当于 a 个 X 相加，而 X 与 X 显然是相关的，此时方差的线性不适用. 特别地，$Cov[X,X]=\mathbb{E}[X^2]-\mathbb{E}[X]^2=Var[X]$，因此
$Var[2X]=Var[X+X]=Var[X]+Var[X]+2Cov[X,X]=4Var[X]$.

推广：

Suppose $X_1,\ X_2,\ \ldots,\ X_n$ are **pairwise independent** random variables, meaning for every $i\neq j$, X_i and X_j are independent. Then,

$$Var[X_1+X_2+\cdots+X_n]=Var[X_1]+Var[X_2]+\cdots+Var[X_n]$$

10.12 二项分布的方差

> Variance of Binomial Random Variables

Suppose X is a $Binomial(n,p)$ random variable. Then,

$$Var[X]=np(1-p)$$

Proof: We know $X=X_1+\ldots+X_n$ is the **sum of n independent $Bernoulli(p)$ random variables**. Therefore,

$$Var[X]=Var[X_1]+\cdots+Var[X_n]=np(1-p)$$

证明：$X \sim Bernoulli(p)$ 的方差为 $p(1-p)$

$\mathbb{E}[X] = p$

$$
\begin{aligned}
Var[X] &= \mathbb{E}[(X-\mu)^2] \\
&= p(1-p)^2 + (1-p)(0-p)^2 \\
&= p(1-p)
\end{aligned}
$$

10.13 泊松分布的方差

> Variance of Poisson Random Variables

Suppose Y is a $Poisson(\lambda)$ random variable. Then,

$$Var[Y] = \lambda$$

Informal Proof:

$\lim\limits_{n\to\infty} Binomial(n, \frac{\lambda}{n}) = Poisson(\lambda)$ and so

$$Var[Y] = \lim_{n\to\infty} n \cdot \frac{\lambda}{n} \cdot (1 - \frac{\lambda}{n}) = \lambda.$$

Lecture 11 连续分布

> Continuous Random Variables

11.1 均匀分布

> Uniform Random Variable

A uniform random variable T over internal $[0,1]$ satisfies

$$\mathbb{P}(T \le t) = \begin{cases} 0 & \text{if } t \le 0, \\ t & \text{if } 0 < t \le 1, \\ 1 & \text{if } 1 < t \end{cases}$$

> 基于这个定义，无法计算 T 大于等于某个值的概率.
>
> 但是可以计算左开右闭区间，例如 $\mathbb{P}(\frac{1}{2} < T \le \frac{3}{4})$.
>
> 结合下面的点概率为 0，可知加不加等号无所谓.

The **probability mass function (PMF)** $does\ not\ make\ much\ sense$ because $\mathbb{P}(T = t) = 0$ for every t

11.2 累积分布函数

> Cumulative Distribution Function.
>
> 注意这里不仅是均匀分布，其他分布也涵括在内. 之前研究离散变量已提过 CDF，见 `7.1 累积分布函数`.

The $cumulative\ distribution\ function\ (CDF)\ F$ of random variable X is $F(x) = \mathbb{P}(X \le x)$

Every CDF $F : \mathbb{R} \to [0,1]$ must satisfy the following properties:

① F is monotonically increasing.

② $\lim_{x\to-\infty} F(x) = 0$

③ $\lim_{x\to\infty} F(x) = 1$

Exercise: Find the CDF of a $Geometric(p)$ random variable.

$F(k) = \mathbb{P}(X \le k) = 1 - \mathbb{P}(X > k) = 1 - (1-p)^k$

> 前面已写过.

11.3 均匀分布 CDF

> CDF of a Uniform Random Variable

The CDF of a Uniform random variable over interval $[0, 1]$ is

$$F(x) = \mathbb{P}(X \le x) = \begin{cases} 0 & \text{if } x \le 0, \\ x & \text{if } 0 < x \le 1, \\ 1 & \text{if } 1 < x \end{cases}$$

> 实际上，均匀随机变量就是用累积分布函数定义的.

11.4 概率密度函数

> Probability Density Function (PDF)，和离散变量的 PMF 对应的概念.

For a $continuous\ random\ variable\ X$, we define its **probability density function (PDF)** as the $derivative\ of\ its\ CDF$:

$$f(x) = \lim_{\delta\to 0} \frac{F(x+\delta) - F(x)}{\delta} = \frac{dF(x)}{dx}$$

Note: For a small $\delta > 0$, we know $F(x+\delta) - F(x) \approx f(x)\delta$ and so for small δ

$$\mathbb{P}(x \le X \le x+\delta) \approx f(x)\delta$$

The PDF of the $Uniform([0, 1])$ random variable is

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{if } 1 < x \end{cases}$$

A PDF can take **any non-negative value** $f(x) \ge 0$ and its value can be $greater\ than\ 1$.

11.5 对 PDF 积分

> Integral of PDF and Probability Calculation

For a continuous random variable X, the probability of an event E can be calculated using $the\ integral\ of\ PDF\ f$. Therefore,

$\mathbb{P}(E) = \int_E f(x)dx$

$\mathbb{P}(X \le t) = \int_{-\infty}^{t} f(x)dx$

$\int_{-\infty}^{\infty} f(x)dx = 1$

> 这条对应离散变量中的概率总和为 1. 见 `1.3 概率公理（一）`.

11.6 连续分布的期望

> Expected Value and Variance of Continuous Random Variables
>
> 注意，这里讨论的是连续变量，关于离散变量的期望，见 `7.3 期望`.

求和变成积分，待补充.

(2024.12.12) 复习补充：

$\mathbb{E}[X] = \int_x xf(x)dx$

> 这里 x 积分范围一般是 $(-\infty, +\infty)$，但对于某些特定分布，有大量 PDF 为 0 的区域可以略去不算.

11.7 连续分布的方差

求和变成积分，待补充.

(2024.12.12) 复习补充：

$Var[X] = \int_x (x - \mathbb{E}[X])^2 f(x)dx$

注意，`10.2 方差公式` 可以拓展到连续变量：

$\mathbb{E}[X] = \int_x xf(x)dx$

$\mathbb{E}[X^2] = \int_x x^2 f(x)dx$

$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

11.8 均匀分布的期望

> Expected Value and Variance of Uniform Random Variables
>
> $X \sim Uniform([a, b])$

Consider a Uniform random variable X over internal $[a, b]$. Then,

The PDF of X is

$$f(x) = \begin{cases} 0 & \text{if } x < a, \\ \frac{1}{b-a} & \text{if } a \le x \le b, \\ 0 & \text{if } b < x \end{cases}$$

期望：$\mathbb{E}[X] = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$

11.9 均匀分布的方差

方差：$Var[X] = \int_a^b (x - \mathbb{E}[X])^2 \cdot \frac{1}{b-a} \cdot dx = \frac{1}{3(b-a)}(x - \frac{a+b}{2})^3 \Big|_a^b = \frac{(b-a)^2}{12}$

(2024.12.12) 法二：

$\mathbb{E}[X] = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$

$\mathbb{E}[X^2] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3}(a^2 + b^2 + ab)$

$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{(b-a)^2}{12}$

Lecture 12 指数/正态分布

> Continuous Random Variables

12.1 指数分布

> Exponential Random Variable

Rain is falling on your head at a rate of λ drops/sec. How long do we wait until the next drop?

将 1 秒分为 n 个小区间，下一个雨滴落在第 t 个区间（t 可以大于 n）的概率是：

$\mathbb{P}(T=\frac{t}{n})=\frac{\lambda}{n}(1-\frac{\lambda}{n})^{ns-1}$

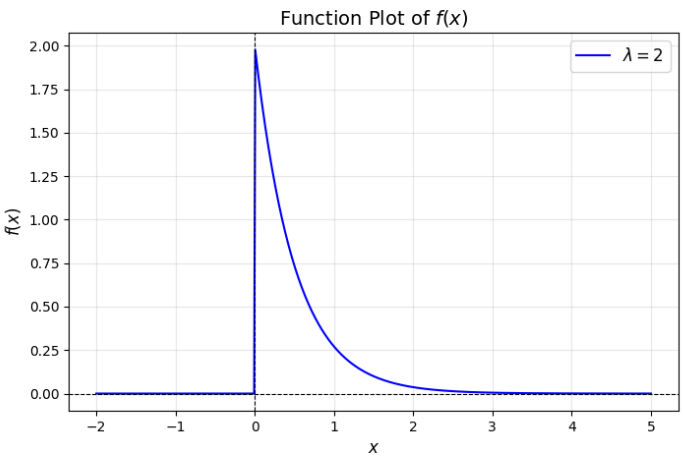
> 其中 $s=\frac{t}{n}$

将离散概率值转换成连续概率密度. 除以步长 $\frac{1}{n}$

$f(s)=\lim\limits_{n\to\infty}\frac{\mathbb{P}(T=\frac{t}{n})}{\frac{1}{n}}=\lambda e^{-\lambda s}$

An $Exponential(\lambda)$ random variable has the following PDF:

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{if } x\geq 0\\ 0 & \text{if } x<0\end{cases}$$



The $Exponential(\lambda)$ random variable X satisfies:

The CDF of X is $F(x)=1-e^{-\lambda x}$

> $\mathbb{P}(X\geq x)=e^{-\lambda x}$.

The expected value of X is $\mathbb{E}[X]=\frac{1}{\lambda}$

> $\mathbb{E}[X]=\int_0^\infty x\lambda e^{-\lambda x}dx=\int_0^\infty -xd(e^{-\lambda x})=-xe^{-\lambda x}|_0^\infty-\int_0^\infty e^{-\lambda x}d(-x)=\int_0^\infty e^{-\lambda x}dx=\frac{1}{\lambda}$

The variance of X is $Var[X]=\frac{1}{\lambda^2}$

> $\mathbb{E}[X^2]=\int_0^\infty x^2\lambda e^{-\lambda x}dx=\int_0^\infty -x^2d(e^{-\lambda x})=-x^2e^{-\lambda x}|_0^\infty-\int_0^\infty e^{-\lambda x}d(-x^2)=\frac{2\mathbb{E}[X]}{\lambda}=\frac{2}{\lambda^2}$
>
> $Var[X]=\mathbb{E}[X^2]-\mathbb{E}[X]^2=\frac{1}{\lambda^2}$

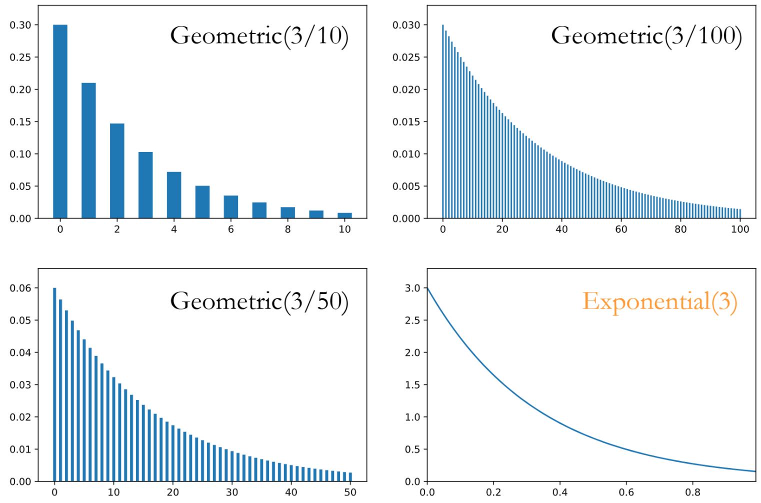
The standard deviation of X is $\sigma=\frac{1}{\lambda}$

12.1.1 指数 vs 几何

> Exponential vs. Geometric Random Variables

待补充.

$\lim\limits_{n\to\infty} \frac{1}{n} Geometric(\frac{\lambda}{n}) = Exponential(\lambda)$



> 这个图有点问题，几何分布应该从 1 开始. 不过无伤大雅.

注意这里 $\frac{1}{n}$ 是一个缩放因子，横向压缩为原图的 $\frac{1}{n}$. 至于纵坐标貌似不同，因为 $Geometric$ 是离散分布，给的是 PMF 图像. n 足够大时，PMF 值除以单位区间长度 $\frac{1}{n}$ 就得到右下角 $Exponential$ 的 PDF 图像.

这个 $\frac{1}{n}$ 作用在变量上，而不是 PMF 函数值上：

$$\begin{aligned}\mathbb{P}(X=x) &= \frac{\lambda}{n}(1-\frac{\lambda}{n})^{x-1} \\ &= \frac{\lambda}{n}(1-\frac{\lambda}{n})^{n\cdot\frac{x}{n}-1}\end{aligned}$$

除以单位区间，并令 $n\to\infty$，得到 PDF：

$$f_{X\sim Geometric(\frac{\lambda}{n})}(x) = \lim_{n\to\infty}\frac{\frac{\lambda}{n}(1-\frac{\lambda}{n})^{n\cdot\frac{x}{n}-1}}{\frac{1}{n}} = \lambda e^{-\lambda\frac{x}{n}}$$

> 注意这个值为 0.

和指数分布的 PDF 对比：

$$f_{X\sim Exponential(\lambda)}(x) = \lambda e^{-\lambda x}$$

发现只要在对几何分布抽样的时候，把变量乘 $\frac{1}{n}$，当 $n\to\infty$ 时就与指数分布表现一致了.

通过均值和方差的计算，也可以说明 $\frac{1}{n}$ 的正确性. 设 $X\sim Geometric(\frac{\lambda}{n})$，则

$\mathbb{E}[X] = \frac{n}{\lambda}$

$Var[X] = \frac{1-\frac{\lambda}{n}}{(\frac{\lambda}{n})^2}$

$\lim\limits_{n\to\infty}\mathbb{E}[\frac{X}{n}] = \frac{1}{\lambda}$

$\lim\limits_{n\to\infty}Var[\frac{X}{n}] = \frac{1}{\lambda^2}$

> 下面两个正是指数分布的均值和方差.

12.1.2 指数 vs 泊松

> Exponential vs. Poisson Random Variables

泊松分布算单位时间内事件发生频率，指数分布算事件下一次发生的时间.

	$Poisson(\lambda)$	$Exponential(\lambda)$
description	number of events within time unit	time until first event happens
expectation	λ	$\frac{1}{\lambda}$
variance	λ	$\frac{1}{\lambda^2}$

12.1.3 无记忆性

> Memoryless Property of Exponential Distribution

The Exponential distribution is memoryless (the past has no effects on its future), which means for every expotential random variable X and constants $s, t \geq 0$:

$$\mathbb{P}(X \geq s+t \mid X \geq t) = \mathbb{P}(X \geq s)$$

同理，

$$\mathbb{P}(X \leq s+t \mid X \geq t) = \mathbb{P}(X \leq s)$$

12.2 正态分布

> Normal Distribution as the Limit of Binomial Distribution，也叫 $Gaussian\ distribution$.

We define the normal (Gaussian) probability density function (PDF) with parameters μ and σ^2 as

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

Note that the parameters μ and σ^2 in the above definition are equal to the $mean$ and $variance$ of the **normal random variable** X:

$$\mathbb{E}[X] = \mu, \qquad Var[X] = \sigma^2$$

注意，泊松分布和正态分布都是二项分布的近似，但有诸多不同：

① 近似条件不同. 泊松分布：$n \to \infty$ 且 $p \to 0$，但 $np = \lambda$ 为常数；正态分布：$n \to \infty$ 且 p 不取极端值.

> 泊松分布是二项分布的稀疏极限，正态分布是二项分布的大样本极限.

② 适用场景不同. 泊松分布：描述稀有事件的发生频率；正态分布：描述大量随机变量的总体分布.

③ 变量类型不同. 泊松分布：离散分布，只能取非负整数；正态分布：连续分布，可以取任意实数.

④ 数学性质不同. 泊松分布：均值 λ，方差 λ，不对称；正态分布：均值 μ，方差 σ^2，对称.

12.2.1 标准正态分布

> Standard Normal Distribution

Standard normal is the distribution of a normal random variable Z with parameters $\mu = 0$ and $\sigma^2 = 1$, i.e. with $zero\ mean$ and $unit\ variance$, as

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

CDF 表

标准正态分布 $X \sim N(0,1)$ 的 CDF 记作 $\Phi(x)$. $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ 没有解析解，实际计算依赖于数值方法或近似公式.

同时，查表也是一种常用方法：

Table 5.1 Area $\Phi(x)$ Under the Standard Normal Curve to the Left of X.

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

12.3 标准化

> Normalizing a Normal Random Variable

Suppose a random variable X has expected value $\mathbb{E}[X] = \mu$ and variance $Var[X] = \sigma^2$. Then, random variable $Z = \frac{X-\mu}{\sigma}$ has $zero\ expected\ value$ and $unit\ variance$, i.e.,

$$\mathbb{E}[Z] = 0, \qquad Var[Z] = 1.$$

> 标准化对任意分布都适用，但是有些没什么意义.
>
> 例如，$X \sim Exp(1)$，$\mathbb{E}[X] = 1$，$Var[X] = 1$
>
> $Z = \frac{X-1}{1} = X - 1$，但 Z 不再是 $Exponential\ Random\ Variable$.

If random variable X has a **normal distribution** with parameters μ and σ^2, then random variable $Z = \frac{X-\mu}{\sigma}$ has a **standard normal distribution**.

12.4 正态近似

> Normal Approximation of Binomial Distribution

① 棣莫弗-拉普拉斯

> DeMoivre Laplace Limit Theorem

If S_n is a $Binomial(n, p)$ $random\ variable$ with $mean\ np$ and $variance\ np(1-p)$, then for every $a < b$:

$$\lim_{n\to\infty} \mathbb{P}(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b) = \Phi(b) - \Phi(a)$$

where Φ is the CDF of **standard normal distribution**.

> 注意，$p = 0.5$ 的二项分布，标准化后是最接近标准正态分布的钟形曲线的，但是 $p \ne 0.5$ 时，只要 n 足够大，仍然可以很好地用正态分布近似，能否近似的重点不是 p 等不等于 0.5，而是 n 够不够大. 即使 $p = 0.5$，只要 n 不够大，看起来仍然是离散的点.
>
> 对于 $p \ne 0.5$ 的情况，二项分布的正态近似在均值附近区域的效果更好，在尾部近似误差会增大（偏态）.
>
> 注意，题目求的可能是 $r \le S_n \le q$ 的概率，不能直接把 r, q 当作 a, b，要先标准化再代入 Φ 函数.
>
> 查表得到 Φ 函数值. 见 `12.2.1 标准正态分布 CDF 表`.

Note that $Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}$ has $zero\ mean$ and $unit\ variance$ for every n, and the above theorem shows the CDF of Z_n **converges to** the CDF of standard normal.

$\mathbb{P}(r \le S_n \le q) = \mathbb{P}(\frac{r-np}{\sqrt{np(1-p)}} \le \frac{S_n-np}{\sqrt{np(1-p)}} \le \frac{q-np}{\sqrt{np(1-p)}})$

② 连续性校正

上面的公式是 $n \to \infty$ 的情况，通过二项分布的收敛特性，可以直接取等号. 但是，当 n 是一个有限大的值时，即使我们用约等号，也不能直接约等于 $\Phi(b) - \Phi(a)$，而是需要一步连续性校正. 理由如下:

离散分布中，$\mathbb{P}(r \le X \le q) = \mathbb{P}(X = r) + \mathbb{P}(X = r+1) + \cdots + \mathbb{P}(X = q)$

近似转换为连续变量:

$\mathbb{P}(X = r) \approx \mathbb{P}(r - 0.5 \le X \le r + 0.5)$

$\mathbb{P}(X = r+1) \approx \mathbb{P}(r + 1 - 0.5 \le X \le r + 1 + 0.5)$

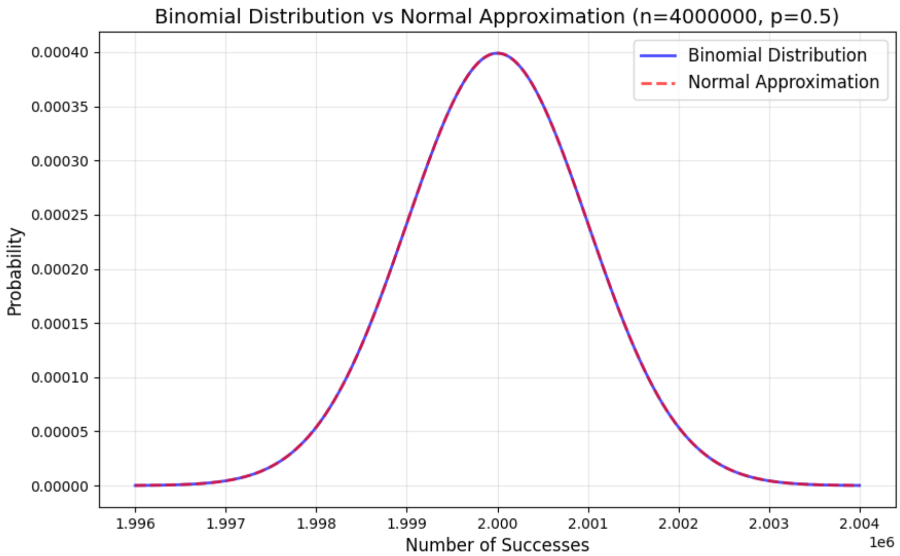
$\vdots$

$\mathbb{P}(X = q) \approx \mathbb{P}(q - 0.5 \le X \le q + 0.5)$

$\mathbb{P}(r \le S_n \le q)$ 转换为连续变量时要进行校正:

$\mathbb{P}(r \le S_n \le q) \approx \mathbb{P}(r - 0.5 \le S_n \le q + 0.5)$

> 这里默认 r, q 是整数，如果不是，也是校正到 .5
>
> 如果 $n \to \infty$ 且 $np \to \infty$ (例如 p 是常量)，且 r 和 q 和均值的距离足够大时，则校正是无关紧要的，因为加减 0.5 的影响相较于 r 和 q 和均值的距离可以忽略不计. 例如，均值 2000000，$p = 0.5$，则 $\mathbb{P}(1999000 \le X \le 2001000)$ 近似成连续变量，严格来说要校正为 $\mathbb{P}(1999000 - 0.5 \le X \le 2001000 + 0.5)$，但标准化后为 $\mathbb{P}(\frac{1999000-0.5-2000000}{1000} \le \frac{X-2000000}{1000} \le \frac{2001000+0.5-2000000}{1000}) = \mathbb{P}(-1.0005 \le X \le 1.0005)$，误差已经小于 CDF 表的精度（0.01），和不校正，用 $\mathbb{P}(-1 \le X \le 1)$ 计算没什么区别.



12.5 多连续变量

> Multiple Continuous Random Variables

12.5.1 双离散变量

> Formulas for Two Discrete Random Variables

joint PMF	$f_{XY}(x,y) = \mathbb{P}(X=x,\ Y=y)$
Probability of A	$\sum\limits_{(x,y)\in A} f_{XY}(x,y)$
Marginals	$f_X(x) = \sum\limits_y f_{XY}(x,y)$
Independence	$f_{XY}(x,y) = f_X(x)\cdot f_Y(y)$
Expectation of $Z=g(X,Y)$	$\sum\limits_{x,y} g(x,y) f_{XY}(x,y)$

12.5.2 联合分布

> 注意，这里讨论的是连续变量.

① CDF

> Joint CDF and PDF of Random Variables

A pair of random variables X, Y can be described by their $\color{red}{joint\ CDF\ F_{XY}}$ which is defined as:

$$F_{XY}(x,y) = \mathbb{P}(X \le x,\ Y \le y)$$

② PDF

Also, we define the $\color{blue}{joint\ PDF\ f_{XY}}$ using the $\color{blue}{derivative\ of\ joint\ CDF}$ F_{XY}:

$$
\begin{aligned}
f_{XY}(x,y) &= \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{XY}(x,y) \\
&= \lim_{\epsilon,\sigma\to 0}\frac{\mathbb{P}(x\le X\le x+\epsilon,\ y\le Y\le y+\delta)}{\epsilon\cdot\sigma}
\end{aligned}
$$

> 可以理解为面密度.

Example: Rain drops at a rate of λ drop/sec. Let X and Y be the arrival times of the first and second raindrops. Find F_{XY} and f_{XY}.

> 有争议，答案有点问题，应该用条件概率.
>
> 待补充.

(2024.12.12) 复习补充

当 $0\le x\le y$，

$$
\begin{aligned}
F_{XY}(x,y) &= \mathbb{P}(X\le x, Y\le y) \\
&= \int_0^x \lambda e^{-\lambda x}dx\ \mathbb{P}(Y\le x+(y-x)|Y\ge x) \\
&= \int_0^x \lambda e^{-\lambda x}(1-e^{-\lambda(y-x)})dx \\
&= 1-e^{-\lambda x}-\lambda x e^{-\lambda y} \\
f_{XY}(x,y) &= \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{XY}(x,y) \\
&= \lambda^2 e^{-\lambda y}
\end{aligned}
$$

当 $0\le y\le x$，

$$
\begin{aligned}
F_{XY}(x,y) &= \mathbb{P}(X\le x, Y\le y) \\
&= \int_0^y \lambda e^{-\lambda x}dx\ \mathbb{P}(Y\le x+(y-x)|Y\ge x) \\
&= \int_0^y \lambda e^{-\lambda x}(1-e^{-\lambda(y-x)})dx \\
&= 1-e^{-\lambda y}-\lambda y e^{-\lambda y} \\
f_{XY}(x,y) &= \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{XY}(x,y) \\
&= 0
\end{aligned}
$$

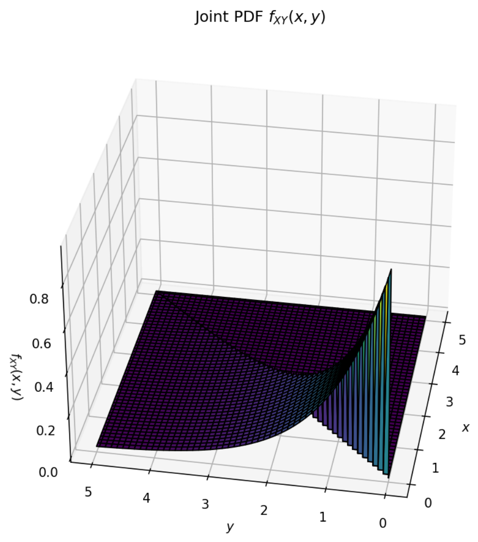
当 $x\le 0$ 或 $y\le 0$，$F_{XY}(x,y)=0$，$f_{XY}=0$.

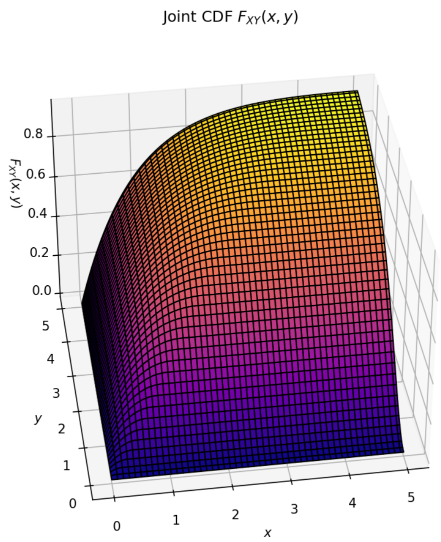
综上所述，

$$
F_{XY}(x,y)=\begin{cases}1-e^{-\lambda x}-\lambda x e^{-\lambda y} & \text{if } 0\le x\le y, \\ 1-e^{-\lambda y}-\lambda y e^{-\lambda y} & \text{if } 0\le y\le x, \\ 0 & \text{otherwise.}\end{cases}
$$

$$
f_{XY}(x,y)=\begin{cases}\lambda^2 e^{-\lambda y} & \text{if } 0\le x\le y, \\ 0 & \text{otherwise.}\end{cases}
$$

图像（$\lambda=1$）：





12.5.3 边缘分布

> Marginal and Joint Distributions
>
> 注意，这里讨论的是连续变量.

① CDF

Suppose F_{XY} is the $\textit{joint CDF}$ of X, Y. Then, the marginal CDFs $F_X,\ F_Y$ can be found as

$$F_X(x) = F_{XY}(x, +\infty) = \lim_{y \to +\infty} F_{XY}(x, y)$$
$$F_Y(y) = F_{XY}(+\infty, y) = \lim_{x \to +\infty} F_{XY}(x, y)$$

② PDF

Suppose f_{XY} is the $\textit{joint PDF}$ of X, Y. Then, the **marginal PDFs** f_X, f_Y can be found by the integration of joint PDF

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

> 两边同乘一个 dx，然后用面积来理解.

$$f_Y(y) = \int_{x=-\infty}^{+\infty} f_{XY}(x, y)dx$$

12.5.4 公式表

> Cheat Sheet for Two Continuous Random Variables

待补充.

(2024.12.13) 复习补充

X, Y continuous with joint PDF $f_{XY}(x, y)$. Then,

Probability of $\forall$	[illegible]
Marginals	$f_X(x) = \int_{y=-\infty}^{\infty} f_{XY}(x, y)dy$
Independence	$f_{XY}(x, y) = f_X(x)f_Y(y)$ **for all** x, y

Probability of A	$\mathbb{P}(A) = \iint_{(x,y)\in A} f_{XY}(x,y)dxdy$
Expectation of $Z = g(X,Y)$	$\mathbb{E}[Z] = \int_x \int_y g(x,y) f_{XY}(x,y)dxdy$

Lecture 13 中心极限定理

> Multiple Continuous Random Variables, Chebyshev's Inequality and Law of Large Numbers

13.1 连续独立分布

> Independent Continuous Random Variables

Random variables X, Y are called **independent** if their $joint\ CDF$ F_{XY} is the $product\ of\ marginal\ CDFs$:

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

The above definition is equivalent to the $joint\ PDF$ f_{XY} being the $product\ of\ marginal\ PDFs$:

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

Example: The joint PDF of independent normal random variables X, Y with mean μ and variance σ^2:

$f_{XY}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2+(y-\mu)^2}{2\sigma^2}}$

Random variables X, Y are called **uncorrelated** if they satisfy

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

If the above does not hold, we call X and Y **correlated**.

> 这里把相关概念拓展到连续变量. 关于相关性，见 `10.10 不相关分布`.

Note: If X, Y are $independent$, then they will be $uncorrelated$. However, $uncorrelated$ random variables **may not** be $independent$.

13.2 条件 PDF 与独立

> Conditional PDF and Independence

For random variables X, Y, we define the **conditional PDF** $f_{X|Y}$ as

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

> 此定义的合理性证明，见 `ESTR 2020 统计 1.2.2 贝叶斯推断` .

① Independent Random Variables:

$f_{X|Y}(x|y) = f_X(x)$ for all $x, y \in \mathbb{R}$

② Total Probability Theorem:

$f_X(x) = \int_{y=-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy$

③ Total Expectation Theorem:

$\mathbb{E}[X] = \int \mathbb{E}[X|Y=y] f_Y(y) dy$

13.3 连续分布贝叶斯

> Continuous Bayes Rule

回忆离散事件的贝叶斯公式:

Consider Events C and E. Then,

$$\begin{aligned}\mathbb{P}(C|E) &= \frac{\mathbb{P}(E|C)\mathbb{P}(C)}{\mathbb{P}(E)} \\ &= \frac{\mathbb{P}(E|C)\mathbb{P}(C)}{\mathbb{P}(E|C)\mathbb{P}(C) + \mathbb{P}(E|C^c)\mathbb{P}(C^c)}\end{aligned}$$

More generally, if $C_1, C_2, \cdots, C_n$ partition the set of possible causes S,

$\mathbb{P}(C_1|E) = \frac{\mathbb{P}(E|C_1)\mathbb{P}(C_1)}{\mathbb{P}(E|C_1)\mathbb{P}(C_1)+\mathbb{P}(E|C_2)\mathbb{P}(C_2)+\cdots+\mathbb{P}(E|C_n)\mathbb{P}(C_n)}$

推广到连续变量:

For random variables X, Y, the conditional PDF $f_{X|Y}$ and $f_{Y|X}$ satisfy

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$$

此处 $f_Y(y)$ 可以仿照贝叶斯的做法代入全概率:

$$f_Y(y) = \int_{x=-\infty}^{+\infty} f_{Y|X}(y|x) f_X(x) dx$$

13.4 卷积

> PDF of Sum of Independent Random Variables

The PDF of the $Z = X + Y$, the $summation$ of **independent** random variables $X,\ Y$, is the $convolution$ of the marginal $PDFs$:

$$f_Z(z) = \int_{x=-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

Corollary: The sum $Z = X + Y$ of $independent\ normal\ random\ variables$ $X \sim N(\mu_x, \sigma_x^2),\ Y \sim N(\mu_y, \sigma_y^2)$ is a $normal\ random\ variable$ with mean $\mu_x + \mu_y$ and variance $\sigma_x^2 + \sigma_y^2$:

$$X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

> 注意前提：独立.

(2024.12.13) 复习补充

卷积公式的严格推导，需要引入狄拉克 δ 函数.

狄拉克 δ 函数是一种广义函数（或称为分布），具有以下两个主要特性:

① 在 $x=0$ 处取无限大值.

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

② 在整个实数轴上积分为 1.

$$\int_{-\infty}^{+\infty} \delta(x)dx = 1$$

> 可以理解为一个无限窄且无限高的矩形面积为 1.

考虑 Z 落在一个小区间 $[z, z+dz]$ 的概率:

$$\mathbb{P}(z \leq Z \leq z+dz) \approx f_Z(z)dz$$

这个概率也可以用指示函数表示:

$$\mathbb{P}(z \leq Z \leq z+dz) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} 1_{[z,z+dz]}(x+y) f_X(x) f_Y(y) dx dy$$

> $1_{[z,z+dz]}(\cdot)$ 是指数函数，变量 $x+y$ 位于 $[z, z+dz]$ 时为 1，否则为 0.

当 $dz \to 0$ 时，指示函数可近似为狄拉克函数和 dz 的乘积:

$$1_{[z,z+dz]}(Z) \approx \delta(z-Z)dz$$

> 这里的狄拉克和严格的 δ 有些区别，这里可以认为 Z 在 z 到 $z+dz$ 时 $\delta(z-Z)$ 取一个很大的值 $\frac{1}{dz}$（这样 $dz \to 0$ 时就能满足 δ 函数定义）.

代入指示函数,

$$\mathbb{P}(z \leq Z \leq z+dz) \approx \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(z-x-y)dz \cdot f_X(x) f_Y(y) dx dy$$

消去 dz,

$$f_Z(z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(z-x-y) f_X(x) f_Y(y) dx dy$$

利用狄拉克 δ 函数消除变量 y,

$$\int_{-\infty}^{+\infty} \delta(z-x-y) f_Y(y) dy = \int_{-\infty}^{+\infty} \delta(y-(z-x)) f_Y(y) dy = f_Y(z-x)$$

> 狄拉克 δ 函数是偶函数.

代入 $f_Z(z)$,

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx$$

这正是卷积的定义.

> (2024.12.13) 以下关于卷积的知识不考，仅作补充.

定义

① 连续卷积

对于两个实数函数 $f(t)$ 和 $g(t)$，它们的卷积记作 $(f*g)(t)$，定义为:

$$(f*g)(t) = \int_{-\infty}^{+\infty} f(\tau) \cdot g(t-\tau)\, d\tau$$

解释:
卷积操作相当于“滑动”一个函数 g（通常称为卷积核或滤波器）沿着另一个函数 f，在每个位置 t 处计算这两个函数的重叠程度.

> $-\tau$ 体现翻转，t 体现平移.

> 关于为什么要翻转后再平移，而不是直接 $g(t+\tau)$：信号处理中，卷积用于描述信号的「系统响应」或「滤波过程」，就此目的而言，翻转更有物理意义，若不翻转，则变为互相关的定义（互相关用于衡量信号的相似性，自行查资料）. 此外，翻转确保了卷积的数学对称性，即交换律 $f*g=g*f$.

② 离散卷积

对于离散值的序列 $f[n]$ 和 $g[n]$，它们的卷积定义为：

$$(f*g)[n]=\sum_{k=-\infty}^{+\infty} f[k]\cdot g[n-k]$$

解释：
类似连续卷积，离散卷积在每个离散点 n 处计算两个序列的重叠和.

应用

① 概率论

概率论中，卷积用于计算两个独立随机变量之和的概率分布.

(定理) 如果 X 和 Y 是两个独立的连续随机变量，且它们的概率密度函数分别为 $f_X(x)$ 和 $f_Y(y)$，那么它们的和 $Z=X+Y$ 的概率密度函数 $f_Z(z)$ 就是 f_X 和 f_Y 的卷积：

$$f_Z(z)=(f_X*f_Y)(z)=\int_{-\infty}^{+\infty} f_X(x)\cdot f_Y(z-x)\,dx$$

> 这表示 Z 在 z 处取值的概率密度是 X 和 Y 不同值的所有可能组合中，满足 $x+y=z$ 的概率密度之和.

例 1：

假设 X 和 Y 都服从均匀分布，X 在 $[0,1]$ 上均匀分布，Y 也在 $[0,1]$ 上均匀分布. 则 $Z=X+Y$ 的 PDF 为：

$$f_Z(z)=\begin{cases} z, & 0\le z\le 1 \\ 2-z, & 1<z\le 2 \\ 0, & \text{其他} \end{cases}$$

这个结果可以通过对 f_X 和 f_Y 进行卷积计算得到.

例 2：

X 服从指数分布，参数为 λ：

$$f_X(x)=\lambda e^{-\lambda x},\quad x\ge 0$$

Y 也服从相同的指数分布，参数为 λ：

$$f_Y(y)=\lambda e^{-\lambda y},\quad y\ge 0$$

X 和 Y 相互独立. 求 $Z=X+Y$ 的概率密度函数 $f_Z(z)$.

根据卷积定理，

$$f_Z(z)=(f_X*f_Y)(z)=\int_{-\infty}^{+\infty} f_X(x)\cdot f_Y(z-x)\,dx$$

积分区间简化为 $0\le x\le z$：

$$f_Z(z)=\int_0^z \lambda e^{-\lambda x}\cdot\lambda e^{-\lambda(z-x)}\,dx=\lambda^2 e^{-\lambda z}\int_0^z dx=\lambda^2 e^{-\lambda z}\cdot z$$

因此，

$$f_Z(z)=\lambda^2 z e^{-\lambda z},\quad z\ge 0$$

> 此结果符合伽玛分布. 两个独立的指数分布随机变量之和服从参数为 λ 的伽玛分布，形状参数 $k = 2$.

② 信号处理

在信号处理领域，卷积用于描述一个信号通过线性时不变系统（Linear Time-Invariant System, LTI）的响应.

如果 $x(t)$ 是输入信号，$h(t)$ 是系统的脉冲响应（即系统对单位脉冲的响应），则系统的输出信号 $y(t)$ 是 $x(t)$ 和 $h(t)$ 的卷积：

$$y(t) = (x * h)(t) = \int_{-\infty}^{+\infty} x(\tau) \cdot h(t - \tau)\, d\tau$$

> 表示系统的输出信号可以看作输入信号与系统特性的一个叠加、平移后的"混合".

例：假设输入信号 $x(t)$ 是一个矩形脉冲，脉冲响应 $h(t)$ 是一个单位脉冲. 则输出信号 $y(t)$ 仍然是矩形脉冲，因为单位脉冲不改变形状.

③ 图像处理

在图像处理中，卷积用于应用滤镜（如模糊、锐化、边缘检测等）到图像上.

对于二维图像，卷积操作扩展到两个维度. 设 $f(x, y)$ 是输入图像，$g(x, y)$ 是卷积核（滤镜），则卷积 $(f * g)(x, y)$ 定义为：

$$(f * g)(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\tau, \sigma) \cdot g(x - \tau, y - \sigma)\, d\tau\, d\sigma$$

> 卷积核在图像上滑动，对每个位置的局部区域应用加权和，生成输出图像.

例：使用高斯模糊核对图像进行卷积，可以有效模糊图像，减轻噪点.

性质

① 交换律

$$f * g = g * f$$

② 结合律

$$f * (g * h) = (f * g) * h$$

③ 分配律

$$f * (g + h) = f * g + f * h$$

④ 傅里叶变换

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\}$$

> 傅里叶变换把信号从时域转到频域，时域中的卷积对应于频域中的乘法.

13.5 马尔可夫不等式

> Markov's Inequality

Suppose that random variable $X \geq 0$ only takes non-negative values. Then, for every $t > 0$

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$

Proof: The proof follows from the **total expectation theorem**:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{P}(X \geq t)\mathbb{E}[X|X \geq t] + \mathbb{P}(X < t)\mathbb{E}[X|X < t] \\ &\geq \mathbb{P}(X \geq t)\mathbb{E}[X|X \geq t] \\ &\geq t \cdot \mathbb{P}(X \geq t).\end{aligned}$$

> However, this inequality only applies to non-negative random variables.
>
> 非负值是为了保证 $\mathbb{E}[X|X < t] \geq 0$.
>
> 注意，这里要求 $t > 0$.

13.6 切比雪夫不等式

> Chebyshev's Inequality

Suppose that random variable X has a mean $\mathbb{E}[X] = \mu$ and variance $Var[X] = \sigma^2$. Then, for every $t > 0$

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

Proof: The proof follows from Markov's inequality for random variable $Y = (X - \mu)^2 \geq 0$. We know that $\mathbb{E}[Y] = Var[X] = \sigma^2$ and so

$$\mathbb{P}(|X - \mu| \geq t) = \mathbb{P}((X - \mu)^2 \geq t^2) = \mathbb{P}(Y \geq t^2) \leq \frac{\sigma^2}{t^2}$$

> Chebyshev's inequality applies to every random variable, even if it takes negative values.

Example:

(a) Find a bound on the probability of ≤ 24 heads in 64 coin flips?

Solution:

$X \sim Binomial(64, \frac{1}{2})$

$\Rightarrow \mathbb{E}[X] = np = 32,\ Var[X] = np(1-p) = 16$

$\mathbb{P}(X \leq 24) = \frac{1}{2}\mathbb{P}(|X - 32| \geq 8) \leq \frac{1}{2} \cdot \frac{16}{64} = \frac{1}{8}$

找到一个上界 $\frac{1}{8}$，即概率不大于 $\frac{1}{8}$.

> 注意切比雪夫不等式覆盖了双尾，如果题目求单尾可以结合相应分布的性质进一步缩紧不等式.

13.7 大数定律

> Law of Large Numbers

Imagine you observe **Independent and Identically Distributed (IID)** random variables $X_1,\ X_2,\ \ldots,\ X_n$ distributed according to $PMF\ p$ with $expected\ value\ \mathbb{E}[X_i] = \mu$. The goal in the Law of Large Numbers is to show:

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{n \to \infty} \mu$$

13.8 独立同分布

> Mean and Variance of the Empirical Mean of IID Variables

Suppose $X_1,\ X_2,\ \ldots,\ X_n$ are IID random variables with mean $\mathbb{E}[X_i]=\mu$ and variance $Var[X_i]=\sigma^2$. Then,

What is the $\textcolor{blue}{expected\ value}$ of $\overline{X}_n=\frac{X_1+\cdots+X_n}{n}$?

$$\mathbb{E}[\overline{X}_n]=\mathbb{E}[\frac{X_1}{n}]+\cdots+\mathbb{E}[\frac{X_n}{n}]=n\times\frac{\mu}{n}=\mu$$

What is the $\textcolor{red}{variance}$ of $\overline{X}_n$?

Because the variance of the $\textcolor{red}{sum\ of\ independent\ random\ variables}$ is the $\textcolor{red}{sum\ of\ the\ variances}$:

$$\begin{aligned}Var[\overline{X}_n]&=Var(\frac{X_1}{n}+\cdots+\frac{X_n}{n})\\&=Var(\frac{X_1}{n})+\cdots+Var(\frac{X_n}{n})\\&=n\times\frac{\sigma^2}{n^2}\\&=\frac{\sigma^2}{n}\end{aligned}$$

13.9 弱大数定律

> Weak Law of Large Numbers (WLLN)

Suppose $X_1,X_2,\ldots,X_n$ are **IID** with mean $\mathbb{E}[X_i]=\mu$ and variance $Var[X_i]=\sigma^2$. Then, **Chebyshev's inequality** shows

$$\mathbb{P}(|\frac{X_1+\cdots+X_n}{n}-\mu|\geq\epsilon)\leq\frac{\sigma^2}{\textcolor{red}{n}\epsilon^2}$$

① WLLN

Suppose $X_1,X_2,\ldots$ is a sequence of IID random variables with $\textcolor{blue}{expected\ value\ \mathbb{E}[X_i]=\mu}$ and $\textcolor{red}{finite\ variance\ Var[X_i]<\infty}$. Define $\overline{X}_n=\frac{1}{n}(X_1+\ldots+X_n)$. Then, for every $\epsilon>0$:

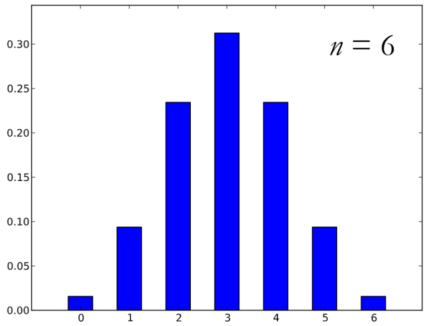
$$\lim_{n\to\infty}\mathbb{P}(|\overline{X}_n-\mu|\geq\epsilon)=0$$

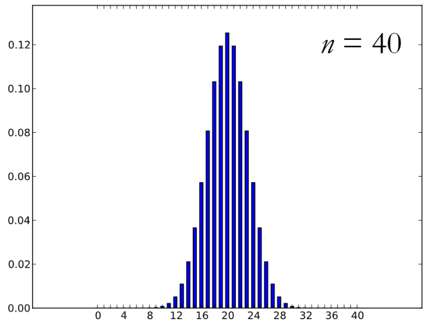
② 例子

> Examples of the Distribution of Sum of Independent RVs

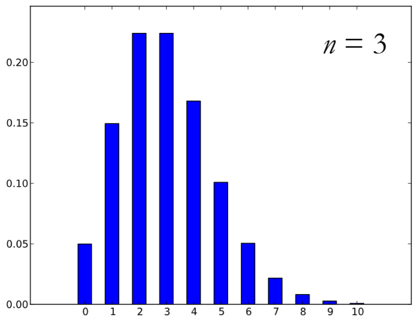
几个图像，待补充.

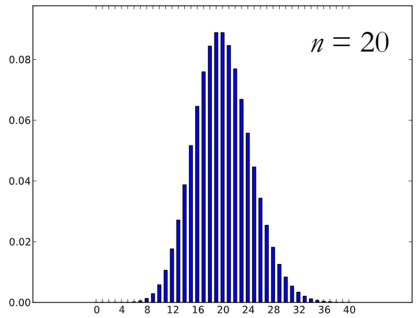
$X = X_1 + \ldots + X_n$ $\quad$ X_i **independent** Bernoulli$(1/2)$



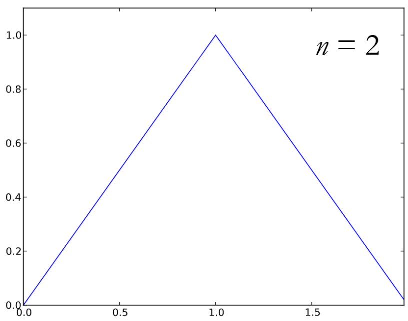


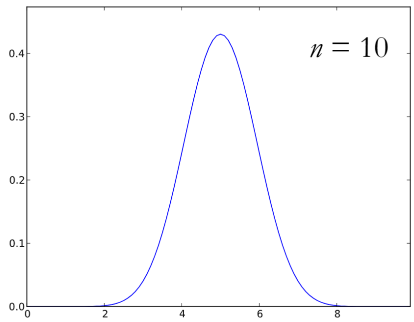
$X = X_1 + \ldots + X_n$ $\quad$ X_i **independent** Poisson(1)



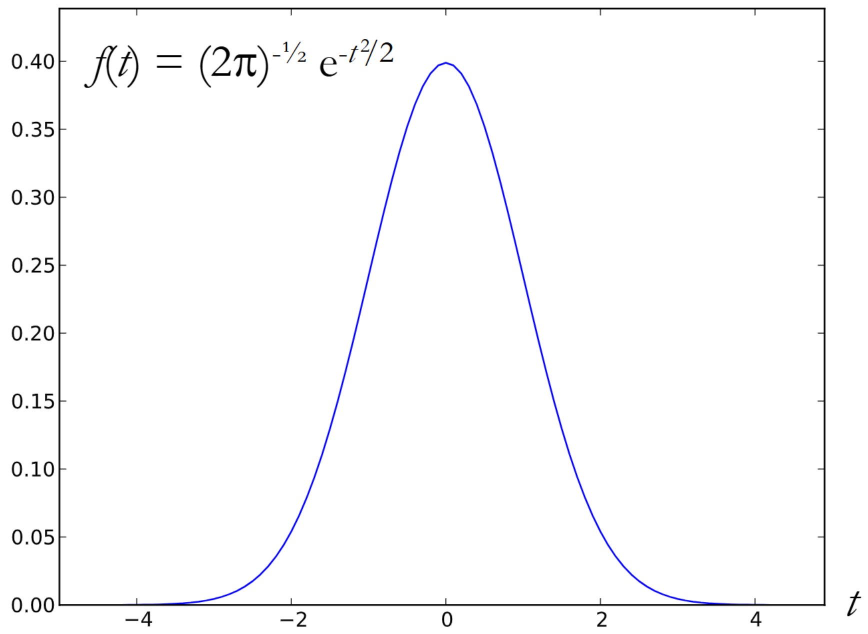


$X = X_1 + \ldots + X_n$ $\quad$ X_i **independent** Uniform$(0, 1)$





上述图像随 n 增大，形状都趋于正态分布的钟形曲线：



> 揭示了更深层次的性质——见 `13.11 中心极限定理`.

13.10 标准化

> Normalizing the Empirical Mean Random Variable. 关于标准化，`12.3 标准化` 亦有提及.

Suppose a random variable X has expected value $\mathbb{E}[X] = \mu$ and variance $Var[X] = \sigma^2$. Then, random variable $Z = \frac{1}{\sigma}(X - \mu)$ has $\color{blue}{zero\ expected\ value}$ and $\color{red}{unit\ variance}$:

$$\mathbb{E}[Z] = 0,\ Var[Z] = 1.$$

Consider IID random variables $X_1, X_2, \ldots, X_n$ has expected value $\mathbb{E}[X_i] = \mu$ and variance $Var[X_i] = \sigma^2$ and define

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

Then, random variable $Z_n = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ has $\color{blue}{zero\ expected\ value}$ and $\color{red}{unit\ variance}$.

13.11 中心极限定理

> Central Limit Theorem (CLT)
>
> 中心极限定理说明了一组独立同分布随机变量的平均值在 $n \to \infty$ 时趋于正态分布，无论这些随机变量的原始分布是什么.

Consider IID random variables $X_1, X_2, \ldots$ with expected value $\mathbb{E}[X_i] = \mu$ and variance $Var[X_i] = \sigma^2$. Then, if we define

$$Z_n = \frac{\frac{1}{n}(X_1 + \cdots + X_n) - \mu}{\frac{\sigma}{\sqrt{n}}}$$

the CDF of random variable Z_n will converge to the CDF of a standard normal distribution $\phi(z)$, i.e. for every $z \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}(Z_n \le z) = \phi(z).$$

Therefore, according to the central limit theorem, the sum of $many\ independent\ random\ numbers$ will approximately have a $normal\ distribution$.

Extended class

Class 1

1. return to Origin

   odd：0%

   even: head & tail 个数相同

   n/2 head 和 n/2 tail

   取head 分析

   总个数：2^n

   head n/2 个数:

   $C_n^{\frac{n}{2}} = \frac{n!}{\frac{n}{2}!\frac{n}{2}!}$

概率：$\frac{\frac{n!}{\frac{n}{2}!\frac{n}{2}!}}{2^n}$

注意外国教材上下角标是相反的。

验证:

when n=2 , $x_2 = \frac{1}{2}$

when n=4, $x_4 = \frac{3}{8}$

Teacher:

1. 列样本空间 the size of $\Omega = 2^n$

enhanced: 第一次回到原点

recursion: 考虑 t=1 和 t=n-1

Class 2

$P(n)$ 可以回到中间，且单侧

$F(n)$ 不可回到中间（0），且有双侧情况

$|P_n| = \frac{1}{2}(|F_2||P_{n-2}| + \cdots + |F_n||P_0|)$

Class 3

Week 4

康托尔集

Class 4

Week 5: Bayes' Rule in Machine Learning

Bayes' Rule (Lecture 4)

Consider Events C and E. Then,

$$\begin{aligned}\mathbb{P}(C|E) &= \frac{\mathbb{P}(E|C)\mathbb{P}(C)}{\mathbb{P}(E)} \\ &= \frac{\mathbb{P}(E|C)\mathbb{P}(C)}{\mathbb{P}(E|C)\mathbb{P}(C) + \mathbb{P}(E|C^c)\mathbb{P}(C^c)}\end{aligned}$$

Class 5

Week 6: Independence of Multiple Events and Secret Sharing

回顾：Independent Events（Lecture 4）

We call Events A and B independent if

$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

or equivalently

$\mathbb{P}(A|B) = \mathbb{P}(A)$

回顾：Three Independent Events（Lecture 4）

We call three events $A,\ B,\ C$ independent events if these four conditions are satisfied:

1. A and B are independent: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
2. B and C are independent: $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$
3. A and C are independent: $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$
4. AND we require $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$

> 前三个条件不能充分说明独立性，前三个都满足的情况下第四个条件也可能不满足。
>
> 四个都满足，才能证明三个事件是独立事件。

Can we design 4 events $E_1,\ E_2,\ E_3,\ E_4$ such that every two of the events E_i, E_j are independent, but every three of the events E_i, E_j, E_l are not independent?

The answer to this question could be useful for secret sharing. Each event E_i can be thought of as a part of a secret given to individual i.

We consider the modular arithmetic with mod 5, i.e. we replace every integer in our calculation with its remainder to 5:

$x\ mod\ 5 =$ remainder when x is divided by 5

Solution:

We toss two 5-sided dice and suppose we have equally-likely outcomes in the sample space $\Omega = \{(1,1),(1,2),\cdots,(5,5)\}$

Define Event E_i as the occurrence of $i \times D_1 + D_2\ mod\ 5 = 0$ for the output of die 1 (D_1) and die 2 (D_2)

可以证明：

1. For every $i \neq j \in \{1,2,3,4\}$, E_i and E_j are independent.
2. For every triple $i,j,k \in \{1,2,3,4\}$, E_i, E_j, E_k are not independent.

Class 6

> Week 7: Independence and Graphical Models

Class 7

> Week 9: Independent Events and Fairness Notions

Class 8

> Week 10: Probability and Game Theory

Class 9

> Week 11: Inequalities in Probability

Homework & Quiz

Homework 1

Q1. In how many ways can we roll four dice so that

1. The face values of the dice are all different?
2. The face values of the dice are increasing (e.g., 2356 but not 3516, 1224)?

Solution:

1. The first die has 6 possible outcomes. For each of them, there are 5 possibilities for the second die that are different from the first, 4 possibilities for the third die different from the first two, and 3 possibilities for the last die different from the first 3. The total number of possibilities is therefore $6 \times 5 \times 4 \times 3 = 360$.

(法二) $A_6^4 = \frac{6!}{(6-4)!} = 360$.

2. We are choosing 4 distinct face values out of the set $\{1, 2, 3, 4, 5, 6\}$ and then writing them down from smallest to largest. This can be done in $\binom{6}{4} = 15$ possible ways.

Q2. A bin contains 10 black balls and 10 white balls. You draw three balls at random without replacement. What is the probability that all three balls are black?

Solution: The sample space Ω consists of all $\binom{20}{10}$ arrangements of the balls. **We assume equally likely outcomes**. The event A consists of those arrangements in which three balls appear in the first three positions. As the other 7 black balls can appear anywhere in the other 17 positions, A has size $\binom{17}{7}$. **By the equally likely outcomes formula**,

$$\mathbb{P}(A)=\frac{\binom{17}{7}}{\binom{20}{10}}=\frac{10\times 9\times 8}{20\times 19\times 18}\approx 0.1053.$$

> 想象 20 个空槽，把黑球放入其中（放完黑剩下全白，不用考虑白球）。
>
> 本质是给抽的顺序编号（第一次抽，第二次抽...）
>
> arrangements: 放法，排法。

(法二) $\mathbb{P}(A)=\frac{\binom{10}{3}}{\binom{20}{3}}=\frac{\frac{10!}{3!7!}}{\frac{20!}{3!17!}}\approx 0.1053$

> 本质是给 20 个球编号

(法三) $\mathbb{P}(A)=\frac{10}{20}\times\frac{9}{19}\times\frac{8}{18}\approx 0.1053$

> 条件概率

Q3. ENGG2760A has 109 students this year, including Alice and Bob. The students are randomly divided into four tutorial sections with 28, 27, 27, 27 students, respectively.

1. What is the probability that Alice and Bob are both assigned to the 28-student tutorial section?
2. What is the probability that Alice and Bob are assigned to the same tutorial section?

solution: In the following, we use the definition of the **multinomial coefficient** as

$$\binom{n}{k_1,k_2,\cdots,k_t}=\frac{n!}{k_1!k_2!\cdots k_t!}$$

The above multinomial coefficient counts the number of partitioning of n different objects into groups of size $k_1,\ k_2,\ \cdots,\ k_t$. Note that here we suppose $n=k_1+k_2+\cdots+k_t$.

1. The sample space Ω consists of all possible $partitions$ of the 105 students into four tutorials (subsets) of size $28,\ 27,\ 27$, and 27. There are $\binom{109}{28,27,27,27}=\frac{109!}{28!27!27!27!}$ such partitions. The event A_1 of interest consists of those partitions in which both Alice and Bob land in the 28-student tutorial. The size of A_1 is the number of ways to partition the rest of the students into the remaining tutorial slots, which is $\binom{107}{26,27,27,27}=\frac{107!}{26!27!27!27!}$. Therefore

$$\mathbb{P}(A_1)=\frac{\binom{107}{26,27,27,27}}{\binom{109}{28,27,27,27}}=\frac{28\times 27}{109\times 108}\approx 0.0642.$$

(法二) $\mathbb{P}(A_1)=\frac{28}{109}\times\frac{27}{108}\approx 0.0642.$

> 条件概率

2. Now the event A of interest is a union of four disjoint events $A_1,\ A_2,\ A_3$, and A_4 consisting of those outcomes in which Alice and Bob are assigned together into the first, second, third, and fourth tutorial, respectively. By a similar calculation as in the previous part, $|A_2|=|A_3|=|A_4|=\binom{107}{28,27,27,25}$. As $|A|=|A_1|+|A_2|+|A_3|+|A_4|$ we get that

$$\mathbb{P}(A)=\frac{|A|}{|\Omega|}=\frac{\binom{107}{26,27,27,27}+3\times\binom{107}{28,27,27,25}}{\binom{109}{28,27,27,27}}=\frac{28\times 27+3\times 27\times 26}{109\times 108}\approx 0.2431.$$

(法二) $\mathbb{P}(A)=\mathbb{P}(A_1)+3\times\frac{27}{109}\times\frac{26}{108}\approx 0.2431.$

> 全概率公式 + 条件概率

Q4. A six-sided die is rolled three times. Which one is more likely: A sum of 11 or a sum of 12?

solution: Let A and B be the events of a sum of 11 and a sum of 12, respectively. **As the outcomes are equally likely**, the probabilities of the two sums are $\frac{|A|}{6^3}$ and $\frac{|B|}{6^3}$ so we need to determine which of the sets A and B is bigger. The set A can be partitioned into A_1 up to A_6 depending on the first die roll. Similarly, B can be partitioned into B_1 up to B_6. Now A_1 has the same size as B_2 as they both share the same pairs of values for the second and third dice. By the same argument, $|A_2|=|B_3|$, $|A_3|=|B_4|$, $|A_4|=|B_5|$, and $|A_5|=|B_6|$. Comparing $|A|$ and $|B|$ therefore amounts to comparing $|A_6|$ and $|B_1|$. For an outcome to be in A_6 the remaining two rolls must add up to 5, while for B_1 they must add up to 11. Therefore $|A_6|=4$ and $|B_1|=2$, so A is the larger set and a sum of 11 is more likely.

> $|A_6|=4:$ (1,4), (2,3), (3,2), (4,1)
>
> $|B_1|=2:$ (5,6), (6,5)

ESTR 2018 Extra Question:

Consider a random walk on a line where we begin the walk at the origin 0. At each round, we toss a fair coin and move one step forward (i.e. we add $+1$ to the number) if we see Heads or move one step backward (i.e. we add -1 to the number) if we see Tails.

1. Which integers could possibly be the output of the random walk at Round n?
2. What is the probability that at the nth round we will go back to the origin 0?
3. What is the probability that at the nth round we will go to an arbitrary integer k?

solution:

1. The possible integers to arrive at Round n are those integers between $-n$ and n that share the same parity with n, i.e.,

$$\{-n,\ -n+2,\ -n+4,\ \cdots,\ n-4,\ n-2,\ n\}$$

2. Going back to 0 at Round n is only feasible for even n's. Here the sample space is $\Omega=\{(a_1,\cdots,a_n): a_i=-1\ or\ +1\}$. Then, for going back to 0 we require $a_1+a_2+\cdots+a_n=0$ i.e. we have exactly $(\frac{n}{2})$ $+1$'s and $(\frac{n}{2})$ -1's. The number of such outcomes is $\binom{n}{\frac{n}{2}}$ for even n's and so the probability will be:

$$\mathbb{P}(A_0)=\begin{cases}0 & \text{if } n \text{ is odd}\\ \frac{\binom{n}{\frac{n}{2}}}{2^n} & \text{if } n \text{ is even}\end{cases}$$

3. Considering the sample space $\Omega=\{(a_1,\cdots,a_n): a_i=-1\ or\ +1\}$, to arrive at integer k we need $a_1+a_2+\cdots+a_n=k$. This equation can be true only if n and k share the same parity, i.e., $n+k$ is an even number. Then, we need precisely $(\frac{n+k}{2})$ $+1$'s and $(\frac{n-k}{2})$ -1's. Therefore, if $n+k$ is odd we have no outcomes that the random walk arrives at k and if $n+k$ is even we have $\binom{n}{\frac{n+k}{2}}$ outcomes

that the random walk arrives at k. Therefore, the probability to arrive at k at Round n will be

$$\mathbb{P}(A_k) = \begin{cases} 0 & \text{if } n+k \text{ is odd} \\ \frac{\binom{n}{\frac{n+k}{2}}}{2^n} & \text{if } n+k \text{ is even} \end{cases}$$

Homework 2

Q1. Alice, Bob and Charlie hold a lucky draw for two tickets to a concert with the following odds:

- The probability that Alice gets one of the tickets is 60%.
- The probability that Bob gets one of the tickets is 70%.

What is the probability that Alice and Bob both get tickets?

Solution: The sample space consists of the three outcomes $\{ab,\ ac,\ bc\}$, where ab represents Alice and Bob getting the tickets, and so on. Denote their probabilities by $p_{ab},\ p_{ac}$, and p_{bc}. The event "Alice gets one of the tickets" is $\{ab,\ ac\}$ so $p_{ab} + p_{ac} = 0.6$. Similarly $p_{ab} + p_{bc} = 0.7$. Since the probabilities must add up to one,

$$p_{ab} = (p_{ab} + p_{ac}) + (p_{ab} + p_{bc}) - 1 = 0.6 + 0.7 - 1 = 0.3.$$

Q2. Alice flips seven fair coins. Let H be the event that the last flip is a head and A be the event that at least one flip is a head. Calculate

(a) $\mathbb{P}(A^c)$;

(b) $\mathbb{P}(H|A)$;

(c) $\mathbb{P}(A|H)$.

Solution: The sample space consists of all sequences of seven heads or tails, namely $\Omega = \{H,\ T\}^7$. All outcomes are equally likely.

(a) A^c is the event "all flips are tails", so $\mathbb{P}(A^c) = \frac{|A^c|}{2^7} = \frac{1}{128}$.

(b) From the complement rule $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = \frac{127}{128}$. Then

$$\mathbb{P}(H|A) = \frac{\mathbb{P}(H \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(H)}{\mathbb{P}(A)} = \frac{\frac{1}{2}}{\frac{127}{128}} = \frac{64}{127}.$$

(c) $\mathbb{P}(A|H) = \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)} = \frac{\mathbb{P}(H)}{\mathbb{P}(H)} = 1$. This is sensible: Given that the last flip is a head, at least one must be a head with probability 1.

Q3. The Los Angeles Lakers and the Boston Celtics play one game in each city. Each team wins their home game with 70% probability. There is a 40% probability that both win their home games. The Lakers win their home game. What is the probability that they win in Boston?

Solution: Let L and C be the events that the Lakers and Celtics win their home games, respectively. We are told that $\mathbb{P}(L) = \mathbb{P}(C) = 0.7$ and $\mathbb{P}(L \cap C) = 0.4$. We are interested in $\mathbb{P}(C^c|L) = \frac{\mathbb{P}(L \cap C^c)}{\mathbb{P}(L)}$. By the axioms of probability $\mathbb{P}(L \cap C^c) = \mathbb{P}(L) - \mathbb{P}(L \cap C) = 0.7 - 0.4 = 0.3$, so $\mathbb{P}(C^c|L) = \frac{0.3}{0.7} = \frac{3}{7} \approx 0.429$.

Q4. There are 6 red balls and 1 blue ball. Each ball is randomly placed in one of two bins.

1. What is the probability that the bin with the larger number of balls contains k balls ($k \in \{4,\ 5,\ 6,\ 7\}$)?
2. What is the probability that the bin with the larger number of balls contains 3 red balls and 1 blue ball?

Solution:

1. The sample space Ω consists of all sequences of length 7, where the value of each position can be either 1 or 2, denoting which bin the ball goes to. Ω has size 2^7. Let E_k denote the event the bin with the larger number of balls contains k balls. Then E_k consists of strings that contain k 1s and $7-k$ 2s, or k $2s$ and $7-k$ 1s. Therefore, by the equally likely outcomes formula,
$$\mathbb{P}(E_k) = \frac{|E_k|}{|\Omega|} = \frac{\binom{7}{k} + \binom{7}{7-k}}{2^7} = \frac{\binom{7}{k}}{2^6}.$$

> k=4567 分别代入

2. Let A be the event of interest. We represent the blue ball by the first position in the sequence. Then A consists of those sequences that start with a 1 and have exactly four 1s or start with a 2 and have exactly four 2s. By the generalized multiplication rule, $|A| = 2 \cdot \binom{6}{3} = 40$. By the equally likely outcomes formula $\mathbb{P}(A) = \frac{40}{2^7} = \frac{5}{16}$.

ESTR2018 Extra Quetion: Let A be the event "Peter will have installed a home alarm by the end of next year" and B be the event "Peter's home will be burglarized by the end of next year." The psychologists Amos Tversky and Daniel Kahneman carried out a survey in which 131 out of 162 people said that $\mathbb{P}(A|B) > \mathbb{P}(A|B^c)\ and\ \mathbb{P}(B|A) < \mathbb{P}(B|A^c)$. Prove that this is impossible. How do you explain the results of the survey? As a possible project you can read up on this and other experiments in which human intuition is at odds with probability theory. [Exercise 2.14 in Blitzstein-Hwang "Introduction to Probability" textbook]

Homework 3

Question 1

There are 5 red balls, 4 blue balls, and 3 green balls in a bin. You draw two balls from a bin. What is the probability that:

1. Ball 2 is red?
2. Ball 2 is red given that ball 1 is red, if the balls are drawn with replacement?
3. Ball 2 is red given that ball 1 is red, if the balls are drawn without replacement?
4. Ball 2 is not blue given that ball 1 is red, if the balls are drawn without replacement?
5. Ball 2 is red given that ball 1 is not blue, if the balls are drawn without replacement?

Solution: Let R_1 be the event "ball 1 is red" and B_1, G_1, R_2 be defined similarly,

1. $\mathbb{P}(R_2) = 5/(5+4+3) = 5/12$ as the outcomes are equally likely.
2. $P(R_2|R_1) = P(R_2) = 5/12$. Owing to the replacement the color of the first ball does not affect the color of the second ball.
3. $P(R_2|R_1) = 4/11$ because after the first ball is drawn there are 4 red balls to choose out of 11.
4. $P(B_2^c|R_1) = 7/11$ because after the first ball is drawn there are 7 non-blue balls to choose out of 11.
5. The conditioned sample space given B_1^c is partitioned by the events R_1 and G_1. By the total probability theorem,
   $P(R_2|B_1^c) = P(R_2|R_1)P(R_1|B_1^c) + P(R_2|G_1)P(G_1|B_1^c) = 4/11 \cdot 5/8 + 5/11 \cdot 3/8 = 35/88.$

Question 2

Alice, Bob, and Charlie are equally likely to have been born on any three days of the year. Let E_{AB} be the event that Alice and Bob were born on the same day. Define E_{BC} and E_{CA} analogously. Which of the following statements is true:

1. Any two of the three events E_{AB}, E_{BC}, E_{CA} are independent.
2. E_{AB}, E_{BC}, and E_{CA} are independent.
3. $E_{AB} \cup E_{BC}$ and E_{CA} are independent.

Solution: Our sample space will consist of all triples of possible birthdays (a, b, c) where a, b, and c are numbers between 1 and 365 (we exclude February 29 to keep things simple). We assume equally likely outcomes, so all triples occur with probability 365^{-3}.

1. True. The intersection of any two events is the event that all three were born on the same day. There are 365 such outcomes, each occurring with probability 365^{-3}, so
   $P(E_{AB} \cap E_{BC}) = P(E_{AB} \cap E_{CA}) = P(E_{BC} \cap E_{CA}) = 365^{-2}$.
   On the other hand, the probability that any two of them were born on the same day is
   $P(E_{AB}) = P(E_{BC}) = P(E_{CA}) = 365 \cdot 1/365^2 = 365^{-1}$.
   Since $P(E_{AB} \cap E_{BC}) = 365^{-2} = P(E_{AB}) \cdot P(E_{BC})$, the two events E_{AB} and E_{BC} are independent, and similarly for the other two pairs.
2. False. $E_{AB} \cap E_{BC} \cap E_{CA}$ is also the event that all three were born on the same day, so $P(EAB \cap EBC \cap ECA) = 365^{-2}$. On the other hand, $P(E_{AB}) \cdot P(E_{BC}) \cdot P(E_{CA}) = 365^{-3}$, so the three events are not independent.
3. False. The event $(E_{AB} \cup E_{BC}) \cap E_{AC}$ happens exactly when all of Alice, Bob, and Charlie have the same birthday, so $P((E_{AB} \cup E_{BC}) \cap E_{AC}) = 365^{-2}$. **Using the union rule**,
   $P(E_{AB} \cup E_{BC}) = P(E_{AB}) + P(E_{BC}) - P(E_{AB} \cap E_{BC}) = 2 \cdot 365^{-1} - 365^{-2}$.
   As $(2 \cdot 365^{-1} - 365^{-2}) \cdot 365^{-1} \neq 365^{-2}$, the events are not independent.

> union rule, **Inclusion-Exclusion Principle**, 容斥原理

Question 3

Cup 1 contains 3 blue balls and 2 red balls. Cup 2 contains 2 blue balls and 8 red balls. We choose a random cup and draw a ball from it.

1. What is the probability that it is blue?
2. The ball is blue. What is the probability that it came from cup 1?
3. I draw another ball from the same cup without replacement. What is the probability that it is also blue?

Solution: Let C_i be the event "cup i was chosen", B_1 be the event "the first ball is blue", and B_2 be the event "the second ball is blue".

1. **By the total probability theorem**:
   $P(B_1) = P(B_1|C_1) \cdot P(C_1) + P(B_1|C_2) \cdot P(C_2) = 3/5 \cdot 1/2 + 2/10 \cdot 1/2 = 2/5.$
2. **By Bayes' rule**:
   $P(C_1|B_1) = P(B_1|C_1) \cdot P(C_1)/P(B_1) = (3/10)/(2/5) = 3/4.$
3. **By the complement rule**:

   $P(C_2|B_1) = 1 - P(C_1|B_1) = 1/4.$

   We use the total probability theorem again, now conditioned on B_1:

   $P(B2|B1) = P(B_2|C_1 \cap B_1) \cdot P(C_1|B_1) + P(B_2|C_2 \cap B_1) \cdot P(C_2|B_1) = 2/4 \cdot 3/4 + 1/9 \cdot 1/4 = 29/72$

   .

Question 4

Computers a and b are linked through seven cables as in the picture. Each cable fails with probability 10 independently of the others. Let C be the event "there is a connection between a and b" and F be the event "the middle vertical cable fails".

1. What is the probability of C given F?

Solution: Let T and B be the events that a connects to b via the top and bottom paths respectively. Conditioned on F, events T and B are independent so
P(C|F) = P(T ∪ B|F) = 1 − P(Tc ∩ Bc|F) = 1 − P(Tc|F)P(Bc|F).
Both T and B are independent of F, so by the algebra of independent events, P(Tc|F) = P(Tc) and P(Bc|F) = P(Bc) and we get that
P(C|F) = 1 − P(Tc)P(Bc) = 1 − $(1 - 0.9^3)^2$.

2. What is the probability of C given Fc?

Solution: Let c be the top middle node. Conditioned on Fc, we can contract the top and bottom single nodes and picture the network like this:

Let L and R be the events "there is a connection from a to c" and "there is a connection from c to b", respectively. They are independent so
P(C|Fc) = P(L ∩ R|Fc) = P(L|Fc)P(R|Fc) = (1 − P(Lc|Fc))(1 − P(Rc|Fc)).
The complement of L (given Fc) happens when both of the connections from a to c fail. Since they are independent,
P(Lc|Fc) = 0.1 · $(1 - 0.9^2)$.
By symmetry, P(Rc|Fc) = 0.1 · $(1 - 0.9^2)$, and so
P(C|Fc) = $(1 - 0.1 \cdot (1 - 0.9^2))^2$.

3. What is the probability of C?

Solution: By the total probability theorem:
P(E) = $0.1 \cdot (1 - (1 - 0.9^3)^2) + 0.9 \cdot (1 - 0.1 \times (1 - 0.9^2))^2 \approx 0.959$.

ESTR2018 Extra Question

Can there be four events E_1, E_2, E_3, E_4 so that every pair E_i, E_j is independent but every triple E_i, E_j, E_k is not (i, j, k are distinct indices)?

Solution: We can use **modular arithmetic** with mode 5 to answer this question.

We **toss two 5-sided dice** and suppose we have **equally-likely outcomes** in the sample space $\Omega = \{(1,1),(1,2),\ldots,(5,5)\}$.

Define Event E_i as the occurrence of $i \times D_1 + D_2 \; mod \; 5 = 0$ for the output of die 1 (D_1) and output of die 2 (D_2).

The events satisfy the independence conditions stated in the question.

Homework 4

> 待补充.

Homework 5

Question 1. A point is chosen uniformly at random inside a circle with radius 1. Let X be the distance from the point to the center of the circle. What is the

(a) CDF

(b) PDF

(c) expected value

(d) variance

of X?

Solution:

The PDF of the point is uniform over the circle which has area π, so it has value $\frac{1}{\pi}$ inside the center and zero outside. The event $X \le x$ consists of all the points in the circle that are at distance less than or equal to x from the center, which is itself a circle of radius x.

(a)(b) Therefore the CDF is $\mathbb{P}(X \le x) = \frac{1}{\pi} \times \pi x^2 = x^2$, and the PDF is $f_X(x) = \frac{d\mathbb{P}(X \le x)}{dx} = 2x$ for $0 \le x \le 1$.

(c) $\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^1 2x^2 dx = \frac{2}{3}$

(d) $\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^1 2x^3 dx = \frac{1}{2}$

$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{18}$

Question 2.

Bob's arrival time at a meeting with Alice is X hours past noon, where X is a random variable with PDF

$$f(x)=\begin{cases}cx, & if\ 0\le x\le 1\\ 0, & otherwise.\end{cases}$$

1. Find the value of the constant c.

Solution: （归一化条件）By the axioms of probability,

$\int_{-\infty}^{+\infty} f(x)dx=\frac{1}{2}c=1\Rightarrow c=2$

2. What is the probability that Bob arrives by 12.30?

Solution: The CDF is $F(x)=\int_{-\infty}^{x} f(x)dx=x^2$. So $\mathbb{P}(X\le 0.5)=F(0.5)=0.25$

3. What is the expected time of Bob's arrival?

$\mathbb{E}[X]=\int_{-\infty}^{+\infty} xf(x)dx=\frac{2}{3}$. So Bob is expected to arrive at $12:40$.

4. Given that Bob hasn't arrived by $12:30$, what is the probability that he arrives by $12:45$?

$\mathbb{P}(X\le 0.75|X>0.5)=\frac{\mathbb{P}(0.5<X\le 0.75)}{\mathbb{P}(X>0.5)}=\frac{0.75^2-0.5^2}{1-0.5^2}=\frac{5}{12}$

5. Given that Bob hasn't arrived by $12:30$, what is the expected time of Bob's arrival?

$$f_{X|X>0.5}(x)=\frac{f_X(x)}{\mathbb{P}(X>0.5)}=\begin{cases}\frac{8}{3}x & 0.5<x\le 1\\ 0 & otherwise.\end{cases}$$

$\mathbb{E}[X|X>0.5]=\int_{-\infty}^{+\infty} xf_{X|X>0.5}(x)dx=\frac{7}{9}$

Bob's conditional expected arrival time is about $12:47$.

Question 3.

The joint PDF of X and Y is

$$f_{X,Y}(x,y)=\begin{cases}C(x+y+1)y & if\ 0\le x\le 2,\ 0\le y\le 2\\ 0 & otherwise.\end{cases}$$

Find

(a) the value of C

(b) The conditional PDF $f_{Y|X}(y|x)$.

Solution:

(a) 归一化条件

$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy=\int_0^2\int_0^2 C(x+y+1)ydxdy=\frac{40}{3}C=1$

$C=\frac{3}{40}$

(b) $f_X(x)=\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy=C(2x+\frac{14}{3})$

$f_{Y|X}(y|x)=\frac{f_{X,Y}(x,y)}{f_X(x)}=\frac{(x+y+1)y}{2x+\frac{14}{3}}$