Lec 7

标记：r.v.

Random Variables

随机变量的定义：A function from sample space S to the real numbers R. 事件→实数的映射

Discrete Random Variable：

A r.v. X is said to be discrete if there is a finite list of values $a_1, a_2, \cdots, a_n$ or an infinite list of values $a_1, a_2, \cdots$ such that $P(X=a_j \text{ for some } j)=1$.

If X is a discrete r.v., then the finite or countably infinite set of values x such that $P(X=x)>0$ is called the support of X.

这里 $X = a_j$ 什么意思？映射是值？

~~In fact: 它是：$\{s: X(s) = a_j\}$ 的集合的标示~~ (事件！or 样本点集合)

概率质量函数：The probability mass function (PMF) of a discrete r.v. X is the function p_X given by $p_X(x) = P(X=x)$.

If x is in the support of X, $p_X(x) = 1$, otherwise 0.

Valid PMF:

Let X be a discrete r.v. with support $x_1, x_2 \cdots$ finite.

(假设它们 value 不同，且 support is countably infinite)

则 p_X of X 满足：

① 非负：$p_X(x) > 0$ if $x = x_j$ for some j, and $= 0$ otherwise

② $\sum_{j=1}^{\infty} p_X(x_j) = 1$

Lec 8 | Bernoulli and Binomial

Bernoulli Distribution: An r.v. X is said to be 伯努利分布 with p if $P(X=1) = p$ and $P(X=0) = 1-p$, where $0 < p < 1$. We write this as $X \sim \text{Bern}(p)$.

The symbol $\sim$ is read "is distributed as"

Indicator Random Variable: The ... of an event A is the r.v. which equals 1 if A occurs and 0 otherwise.

We will denote the indicator r.v. of A by I_A or $I(A)$.

有：$I_A \sim \text{Bern}(p)$ with $p = P(A)$

△: Indicator Function:

$$I_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}$$

，它不是随机变量

No.

Date

Binomial PMF:

Theorem: If $X \sim Bin(n,p)$, then the PMF of X is

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

for $k=0,1,\cdots,n$ ($P(X=k)=0$ otherwise)

Theorem: $X \sim Bin(n,p)$, $q=1-p$, then $n-X \sim Bin(n,q)$

△: 二项分布可认为是重复n次独立的伯努利实验

Hypergeometric 超几何分布

考虑一个w个白球 b个黑球的桶. 我取n个,并不放回. 则一共有$\binom{w+b}{n}$个 samples . X是sample中白球数量. 则认为X遵循超几何分布: $X \sim HGeom(w,b,n)$

$$P(X=k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} \quad (\text{PMF of } X).$$

Theorem: $HGeom(w,b,n)$ and $HGeom(n,w+b-n,w)$ 分布是一样的 (硬代公式暴力可得).

但应有更好的证明: Combinatory Proof

欲证: $\dfrac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} = \dfrac{\binom{n}{k}\binom{w+b-n}{w-k}}{\binom{w+b}{w}}$ 代公式直接等:

→ (给所有球).

理解: $HGeom(w,b,n)$: 先打 white & black label, 再掏n个球

$HGeom(n,w+b-n,w)$: 先给n个球打上"掏出 label",

则有: n个取, $(w+b-n)$个不取, 再给它们打上 white label.

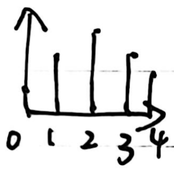
两个分布都在讨论: 取出的n个球中 白球有几个

* abbr: CDF. 累积分布函数

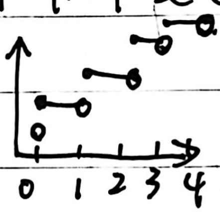
Cumulative Distribution Functions

Def. CDF of an r.v. X is the function F_X given by $F_X(x) = P(X \le x)$. 无歧义下，我们不加下标就写 F

Eg: $X \sim Bin(4, 1/2)$，则 PMF & CDF of X:



PMF



CDF

Valid CDFs:

① increasing: if $x_1 \le x_2$，则 $F(x_1) \le F(x_2)$

② Right-continuous: $F(a) = \lim_{x \to a^+} F(x)$

③ $\lim_{x \to -\infty} F(x) = 0$ & $\lim_{x \to \infty} F(x) = 1$

$$S \xrightarrow{X} R \xrightarrow{g} R$$

$$s \to X(s) \to g(X(s))$$

$g: R \to R$, $g(X)$ is the r.v. that maps s to $g(X(s))$ for all $s \in S$

则 $P(g(X)=y) = \sum_{x: g(x)=y} P(X=x)$