$X: S \longrightarrow \mathbb{R}$

$(, , , \dots)$

data

Lecture 4: Expectation

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

October 22, 2024

Outline

Outline

Expectation of A Discrete R.V.

Definition

The *expected value* (also called the *expectation* or *mean*) of a discrete r.v. X whose distinct possible values are $x_1, x_2, \cdots$ is defined by

$$E(X) = \sum_{j=1}^{\infty} x_j P(X = x_j)$$

If the support is finite, then this is replaced by a finite sum. We can also write

$$E(X) = \sum_{x} \underbrace{x}_{\text{value}} \underbrace{P(X = x)}_{\text{PMF at } x}$$

where the sum is over the support of X.

Distribution

$[-1, \infty]$

if $E(X) = E(Y) \not\Rightarrow X \sim Y$

Theorem

If X and Y are discrete r.v.s with the same distribution, then $E(X) = E(Y)$ (if either side exists).

$$X = \begin{cases} 100 & \text{w.p. } \frac{1}{2} \\ 0 & \text{w.p. } \frac{1}{2} \end{cases} \qquad E(X) = E(Y) = 50.$$

$$Y = \begin{cases} 70 & \text{w.p. } \frac{1}{2} \\ 30 & \text{w.p. } \frac{1}{2} \end{cases}$$

Linearity

The expected value of a sum of r.v.s is the sum of the individual expected values.

Theorem

For any r.v.s X, Y and any constant c,

$$E(X + Y) = E(X) + E(Y)$$

$$E(cX) = cE(X)$$

Monotonicity of Expectation

$Z = X - Y.$ $2^\circ.$ $E(Z) \geq 0,$

$1^\circ.$ $Z \geq 0$ w.p.1 a.s.

$= E(X-Y)$

$= E(X) - E(Y)$

Theorem

Let X and Y be r.v.s such that $X \geq Y$ with probability 1. Then $E(X) \geq E(Y)$, with equality holding if and only if $X = Y$ with probability 1.

$3^\circ \Rightarrow E(X) \geq E(Y)$

if X and Y are independent

$E(X \cdot Y) = E(X) \cdot E(Y)$

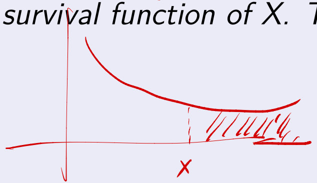
Expectation via Survival Function

Tail distribution

$F(x) = P(X \le x)$

Theorem

Let X be a nonnegative integer-valued r.v. Let F be the CDF of X, and $G(x) = 1 - F(x) = P(X > x)$. The function G is called the survival function of X. Then



$$E(X) = \sum_{n=0}^{\infty} G(n)$$

$= \sum_{n=0}^{\infty} P(X>n)$

$= \sum_{n=0}^{\infty} P(X \ge n+1)$

$= \sum_{n=1}^{\infty} P(X \ge n)$

That is, we can obtain the expectation of X by summing up the survival function (or, stated otherwise, summing up tail probabilities of the distribution).

$m = n+1$

$= \sum_{m=1}^{\infty} P(X \ge m)$

Proof

$E(X) \overset{?}{=} \underbrace{\sum_{n=1}^{\infty} P(X \geq n)}$

$$\sum_{n=0}^{\infty} G(n) = \sum_{n=0}^{\infty} P(X > n)$$

$$= \underbrace{\sum_{n=1}^{\infty} P(X \geq n)} = \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} P(X = m)$$

$1 \leq n ;\ n \leq m ;$

$$= \sum_{m=1}^{\infty} \left(\underbrace{\sum_{n=1}^{m} P(X = m)} \right)$$

$m \geq 1$

$1 \leq n \leq m$

$$= \sum_{m=1}^{\infty} m \cdot P(X = m) \quad (0 \cdot P(X=0) = 0)$$

$$= \sum_{m=0}^{\infty} m \cdot P(X = m) = E(X)$$

$P(X \geq 1) = P(X=1) + P(X=2) + P(X=3) + \cdots$

$+$

$P(X \geq 2) = P(X=2) + P(X=3) + \cdots$

$+$

$P(X \geq 3) = P(X=3) + \cdots$

$+$

$\vdots$

$$1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) + \cdots = \sum_{n=1}^{\infty} n \cdot P(X=n)$$

$$= \sum_{n=0}^{\infty} n \cdot P(X=n) = E(X)$$

Law Of The Unconscious Statistician (LOTUS)

$g(X)$ is a r.v. $\rightarrow$ distribution PMF of $g(X)$

$P(g(X)=y)$; $\Rightarrow$

Theorem

If X is a discrete r.v. and g is a function from $\mathbb{R}$ to $\mathbb{R}$, then

$$E(g(X)) = \sum_{x} g(x) P(X = x)$$

where the sum is taken over all possible values of X.

$$E[g(X)] = \sum_{y} y \cdot P[g(X)=y]$$

Variance and Standard Deviation

$\frac{1}{2}$ $\frac{1}{2}$

X : 49, 51

Y : 0, 100

$E(X) = E(Y) = 50$

Definition

The variance of an r.v. X is

$$\mathrm{Var}(X) = E(X - EX)^2 .$$

$\mathrm{Var}(X) < \mathrm{Var}(Y)$

The square root of the variance is called the *standard deviation (SD)*:

$$\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}.$$

Properties of Variance

$E(X^2) \geq (EX)^2$

- For any r.v. X, $\mathrm{Var}(X) = E(X^2) - (EX)^2$.
- $\mathrm{Var}(X + c) = Var(X)$ for any constant c.
- $\mathrm{Var}(cX) = c^2 Var(X)$ for any constant c.
- If X and Y are independent, then $\mathrm{Var}(X + Y) = Var(X) + Var(Y)$.
- $\mathrm{Var}(X) \geq 0$ with equality if and only if $P(X = a) = 1$ for some constant a.

Properties of Variance

Outline

Story: Geometric Distribution

$k \geq 0$ $k = 0, 1, 2, \ldots$

$k+1$ trials.

First k trials ✗

Last trial ✓

$$P(X=k) = (1-p)^k \cdot p = q^k \cdot p \quad (q = 1-p)$$

Consider a sequence of independent Bernoulli trials, each with the same success probability $p \in (0, 1)$, with trials performed until a success occurs. Let X be the number of **failures** before the first successful trial. Then X has the Geometric distribution with parameter p; we denote this by $X \sim Geom(p)$.

Geometric PMF

Method 1: $E(X) = \sum_{k=0}^{\infty} k \cdot P(X=k) = \sum_{k=0}^{\infty} k\, q^k \cdot p$

$= pq \sum_{k=0}^{\infty} k \cdot q^{k-1}$

$= pq \sum_{k=0}^{\infty} (q^k)'$

Fubini: $= pq \left(\sum_{k=0}^{\infty} q^k\right)'$

$= pq \cdot \left(\frac{1}{1-q}\right)'$

Method 2: $P(X \geq 0) = 1$;

$k \geq 1$; $P(X \geq k) = 1 - P(X < k)$

$= 1 - P(X \leq k-1) = 1 - \sum_{j=0}^{k-1} P(X=j)$

$= 1 - \sum_{j=0}^{k-1} q^j \cdot p = 1 - p \cdot \sum_{j=0}^{k-1} q^j$

$= 1 - p \cdot \frac{1-q^k}{1-q}$

$= q^k$

Theorem

If $X \sim \text{Geom}(p)$, *then the PMF of X is*

$$P(X = k) = q^k p$$

for $k = 0, 1, 2, \ldots$, *where* $q = 1 - p$. $p = 1-q$

$\Rightarrow E(X) = \sum_{k=0}^{\infty} P(X > k) = \sum_{k=1}^{\infty} P(X \geq k)$

$= \sum_{k=1}^{\infty} q^k = \frac{q}{1-q} = \frac{1-p}{p}$

Memoryless Property

1^0. $k=0$; $P(X \geq n \mid X \geq 0) = P(X \geq n)$

2^0. $k \geq 1$; $P(X \geq n+k \mid X \geq k)$

$= \frac{P(X \geq n+k, X \geq k)}{P(X \geq k)} = \frac{P(X \geq n+k)}{P(X \geq k)}$

Theorem

If $X \sim \mathrm{Geom}(p)$, then for any positive integer n,

$$P(X \geq n+k \mid X \geq k) = P(X \geq n)$$

for $k = 0, 1, 2, \ldots$

$= \frac{q^{n+k}}{q^k}$

$= q^n$

$= P(X \geq n)$

$k = 80$; $n = 20$

$P(X \geq 100 \mid X \geq 80) = P(X \geq 20)$

$>$

$<$

Memoryless Property

1°. $P(X \geq n+k \mid X \geq k) = \dfrac{P(X \geq n+k)}{P(X \geq k)} = P(X \geq n)$

$\Rightarrow P(X \geq n+k) = \underline{P(X \geq n) \cdot P(X \geq k)}$

2°. $k = 0 \Rightarrow P(X \geq n) = P(X \geq n) \cdot P(X \geq 0)$

$\Rightarrow P(X \geq 0) = 1$

$\forall n \in \mathbb{N}$.

Theorem

3°. $G(n) = P(X \geq n)$; $G(0) = P(X \geq 0) = 1$;

Suppose for any positive integer n, discrete random variable X satisfies

$$\boxed{P(X \geq n + k | X \geq k) = P(X \geq n)}$$

$G(1) = P(X \geq 1) = q \quad (0 \leq q \leq 1)$

for $k = 0, 1, 2, \ldots$, *then* $X \sim \mathrm{Geom}(p)$.

4°. $\underline{G(n+k) = G(n) \cdot G(k)}$; $n = k = 1 \Rightarrow G(2) = G^2(1) = q^2$;

$n = 2,\ k = 1$; $\Rightarrow G(3) = G(2) \cdot G(1) = G^3(1) = q^3$,

$\cdots\quad G(n) = G^n(1) = q^n \Rightarrow \underline{P(X \geq n) = q^n = (1-p)^n}$

Memoryless Property

$\forall n \in \mathbb{N}$.

$P(X \geq n+k \mid X \geq k) = P(X \geq n)$

$k = 1, 2, \ldots$

Theorem

Geometric distribution is the one and the only one discrete distribution that is memoryless.

First Successful Distribution

First Success Distribution

50% Textbook

$X \sim \text{Geom}(p)$

$Y \sim FS(p)$

$Y = 1 + X$

Definition

In a sequence of independent Bernoulli trials with success probability p, let Y be the number of trials until the first successful trial, including the success. Then Y has the First Success distribution with parameter p; we denote this by $Y \sim \mathrm{FS}(p)$.

50% Textbook

$FS(p)$ ours $\longrightarrow$ $\text{Geom}(p)$ Theirs.

Support of X

Example: Geometric & First Success Expectation

$P(Y=k)$

$= (1-p)^{k-1} \cdot p$

$k \geq 1$

$1^\circ.$ $P(X \geq k) = q^k$

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

Let $X \sim Geom(p)$ and $Y \sim \mathrm{FS}(p)$, find $E(X)$ and $E(Y)$.

$2^\circ.$ $Y = 1 + X$

$E(Y) = 1 + E(X)$

$= 1 + \frac{1}{p} - 1 = \frac{1}{p}$

Story: Negative Binomial Distribution

n ✗ before rth ✓

$n+r$ trials

the last trial ✓

$P(X=n)$

$n+r-1$ trials: n ✗; $r-1$ ✓

In a sequence of independent Bernoulli trials with success probability p, if X is the number of failures before the r^{th} success, then X is said to have the Negative Binomial distribution with parameters r and p, denoted $X \sim NBin(r, p)$. $q=1-p$

$r=1$ $X \sim Geom(p)$

$$\binom{n+r-1}{n} q^n \cdot p^{r-1} \cdot p$$

$$= \binom{n+r-1}{n} q^n \cdot p^r$$

Negative Binomial PMF

$(1+x)^r = \cdots$

$(1+x)^{-r} = \cdots$

$\binom{-r}{n}$

Theorem

If $X \sim \mathrm{NBin}(r, p)$, then the PMF of X is

$$P(X = n) = \binom{n+r-1}{r-1} p^r q^n$$

for $n = 0, 1, 2 \cdots$, where $q = 1 - p$.

$$\binom{-r}{n} = \frac{(-r)!}{n!\,(-r-n)!} = \frac{(-r-n+1)\cdots(-r)}{n!}$$

Geometric & Negative Binomial

k

$X_1 \sim \text{Geom}(p)$ X_1 : # of × before the 1^{st} ✓

n

$X_2 \sim \text{Geom}(p)$ X_2 : # of × between 1^{st} ✓ and 2^{st} ✓

Theorem

Let $X \sim \mathrm{NBin}(r, p)$, viewed as the number of failures before the rth success in a sequence of independent Bernoulli trials with success probability p. Then we can write $X = X_1 + \cdots + X_r$ where the X_i are i.i.d. $\mathrm{Geom}(p)$.

$P(X \geq n+k \mid X \geq k) = P(X \geq n)$

Example: Expectation

Method 1 : $P(X=n) = \binom{n+r-1}{n} p^r q^n$

$\Rightarrow E(X) = \sum_{n=0}^{\infty} n \cdot P(X=n)$

$= \sum_{n=0}^{\infty} n \binom{n+r-1}{n} p^r q^n$

Let $X \sim NBin(r, p)$, find $E(X)$.

Method 2 :

$X = X_1 + \cdots + X_r$

$X_i \sim Geom(p)$

$E(X_i) = \frac{1-p}{p}$

$\Rightarrow E(X) = E(X_1) + \cdots + E(X_r)$

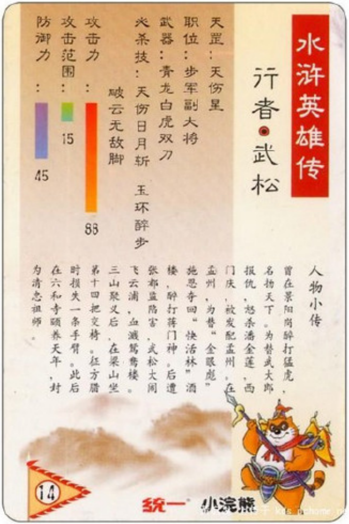
$= r \cdot \frac{1-p}{p}$;

Example:


统一
President
羊肉串口味
小浣熊
干脆面
本包装已申请专利 伪冒必究
油炸面饼净含量：48g
总净含量：50g

Example:

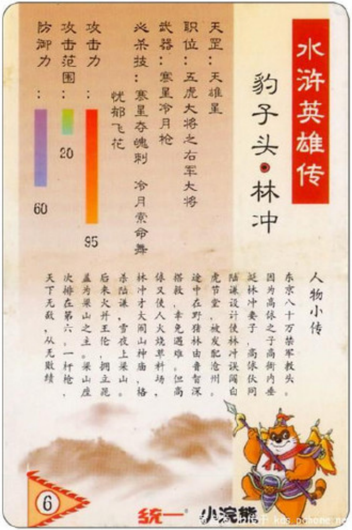
行者·武松
K



水浒英雄传
行者·武松
天罡：天伤星
职位：步军副大将
武器：青龙白虎双刀
必杀技：天伤日月斩 玉环醉步
破云无敌脚
攻击力：88
攻击范围：15
防御力：45
人物小传
曾在景阳冈醉打猛虎，
名扬天下。为替武大郎
报仇，怒杀潘金莲、西
门庆，被发配孟州。在
孟州，为替"金眼彪"
施恩夺回"快活林"酒
楼，醉打蒋门神。后遭
张都监陷害，武松大闹
飞云浦，血溅鸳鸯楼。
三山聚义后，在梁山坐
第十四把交椅。征方腊
时损失一条手臂。此后
在六和寺颐养天年，封
为清忠祖师。
14
统一 小浣熊

Example:

豹子头·林冲
A
♠



水浒英雄传
豹子头·林冲
天罡：天雄星
职位：五虎大将之右军大将
武器：寒星冷月枪
必杀技：寒星夺魄刺　冷月索命舞
忧郁飞花
攻击力：95
攻击范围：20
防御力：60
人物小传
东京八十万禁军教头。因为高俅之子高衙内垂涎林冲妻子，高俅伏同陆谦设计使林冲误闯白虎节堂，被发配沧州。途中在野猪林由鲁智深搭救，幸免遇难。但高俅又使人火烧草料场，林冲才大闹山神庙，格杀陆谦，雪夜上梁山。后来火并王伦，拥立晁盖为梁山之主。梁山座次排在第六，一杆枪，天下无敌，从无败绩。
6
统一
小浣熊

Example:

108

Example:

Example:

Model: Coupon Collector

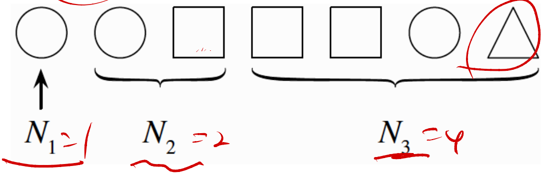
Suppose there are n types of toys, which you are collecting one by one, with the goal of getting a complete set. When collecting toys, the toy types are random (as is sometimes the case, for example, with toys included in cereal boxes or included with kids' meals from a fast food restaurant). Assume that each time you collect a toy, it is equally likely to be any of the n types. Let N denote the number of toys needed until you have a complete set. Find $E(N)$ and $\text{Var}(N)$.

Solution: Coupon Collector

1°. N : # of toys needed to obtain all types of toys.

$$N = N_1 + N_2 + N_3 + \cdots N_n$$

$N_1 = 1$



$p = \frac{n-1}{n}$

2°. N_2. collect the new type (2nd type) of toy w.p. $\frac{n-1}{n}$

$$N_2 \sim FS\left(1-\frac{1}{n}\right)$$

3°. $N_3 \sim FS\left(\frac{n-2}{n}\right)$ ... $N_j \sim FS\left(1-\frac{j-1}{n}\right)$

Solution: Coupon Collector

$X \sim Fs(p)$; $E(X) = \frac{1}{p}$

$N_j \sim Fs\left(\frac{n-(j-1)}{n}\right)$; $j = 1, 2, \ldots, N$.

$$E(N_j) = \frac{n}{n-(j-1)} = \frac{n}{n-j+1}$$

4°. $N = N_1 + \cdots + N_n \Rightarrow E(N) = E(N_1 + \cdots + N_n)$

$n = 108$;

$E(N) \approx 568$

$$= E(N_1) + \cdots + E(N_n)$$

$$= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + n$$

$$= n\left[\frac{1}{n} + \frac{1}{n-1} + \cdots + 1\right] = n\sum_{j=1}^{n}\frac{1}{j}$$

$n >> 1$

$$\approx n(\ln n + 0.577)$$

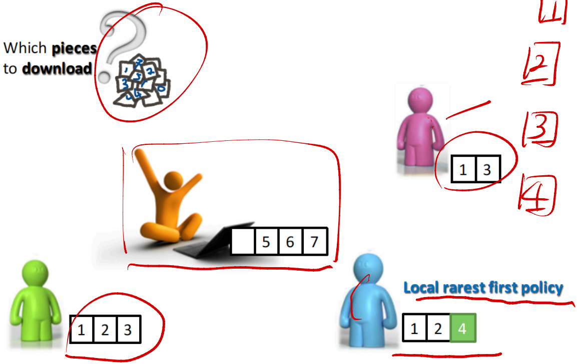
Application: Peer-to-Peer System

BT

- Target file is decomposed into n pieces. block
- Each peer randomly downloads pieces and uploads pieces from its neighbors.
- $\Theta(n \ln n)$ downloads to complete the downloading file.
- The last block problem: missing the last piece (stop at 99% downloading progress)

Application: Peer-to-Peer System

- Solution adopted by BitTorrent:
  - tries to download a block that is least replicated among its neighbors
  - maximize the diversity of content in the system, i.e., make the number of replicas of each block as equal as possible.



$\boxed{1}$ 3
$\boxed{2}$ 2
$\boxed{3}$ 2
$\boxed{4}$ 1

Outline

Properties of Indicator R.V.

$I_A = \begin{cases} 1 & \text{if event } A \text{ occurs.} \\ 0 & \text{otherwise.} \end{cases}$

Let A and B be events. Then the following properties hold.

1. $(I_A)^k = I_A$ for any positive integer k. $1^k = 1$; $0^k = 0$; $k \geq 1$
2. $I_{A^c} = 1 - I_A$.
3. $I_{A \cap B} = I_A I_B$.
4. $I_{A \cup B} = I_A + I_B - I_A I_B$.

③ $I_{A \cap B} = 1 \Rightarrow A$ and B both occurs

$I_A = 1$, $I_B = 1$, $I_A \cdot I_B = 1$

$I_{A \cap B} = 0 \Rightarrow A$ or B or (A both B) does not occur.

$I_A \cdot I_B = 0$

④ $I_{A \cup B} \overset{②}{=} 1 - I_{(A \cup B)^c} = 1 - I_{A^c \cap B^c}$

$\overset{③}{=} 1 - I_{A^c} \cdot I_{B^c} \overset{②}{=} 1 - (1 - I_A)(1 - I_B)$

$= 1 - (1 - I_A - I_B + I_A \cdot I_B) = I_A + I_B - I_A \cdot I_B$

Fundamental Bridge Between Probability and Expectation

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

$$E(I_A) = 1 \cdot P(A) + 0 \cdot [1 - P(A)]$$
$$= P(A)$$

Theorem

There is a one-to-one correspondence between events and indicator r.v.s, and the probability of an event A is the expected value of its indicator r.v. I_A:

$$\boxed{P(A) = E(I_A).}$$

Example: Booler's Inequality

$X \le Y \Rightarrow E(X) \le E(Y)$

<1> $\underbrace{I(A_1 \cup \cdots \cup A_n)}_{LHS} \le \underbrace{I(A_1) + \cdots + I(A_n)}_{RHS}$

1°. if $LHS = 0$ ✓ $RHS \ge 0$

2°. if $LHS = 1$; at least A_j occurs.

at least $I(A_j) = 1$;

$RHS \ge 1$

For any n events $A_1, A_2, \ldots, A_n$,

$$P(\bigcup_{i=1}^{n} A_i) \le \sum_{i=1}^{n} P(A_i)$$

<2> $E[I(A_1 \cup \cdots \cup A_n)] \le E(I(A_1) + \cdots + I(A_n))$

$P(A_1 \cup \cdots \cup A_n) \le P(A_1) + \cdots + P(A_n)$

Solution: Booler's Inequality

Example: Inclusion-Exclusion Formula

1°. $\cancel{1-} I(A_1 \cup \cdots \cup A_n) = I(\overline{\overline{A_1 \cup \cdots \cup A_n}}) = I(\overline{A_1} \cap \cdots \overline{A_n})$

$= I(\overline{A_1}) \cdots I(\overline{A_n}) = (1 - I(A_1)) \cdot (1 - I(A_2)) \cdots (1 - I(A_n))$

$= \cancel{1+} \sum_i I(A_i) + \sum_{i<j} I(A_i) I(A_j) \cdots + (-1)^n I(A_1) \cdots I(A_n)$

For any events $A_1, \ldots, A_n$:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k)$$

$$- \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n).$$

2°. $I(A_1 \cup \cdots \cup A_n) = \sum_i I(A_i) - \sum_{i<j} I(A_i) I(A_j) \cdots + (-1)^{n+1} I(A_1) \cdots I(A_n)$

$= \sum_i I(A_i) - \sum_{i<j} I(A_i \cap A_j) \cdots + (-1)^{n+1} I(A_1 \cap \cdots \cap A_n)$

3°. Taking Expectation to both sides.

Solution: Inclusion-Exclusion Formula

Outline

Moments of Indicator Methods

$\binom{X}{2} = \frac{X(X-1)}{2}$.

- Given n events $A_1, \ldots, A_n$ and indicators $I_j, j = 1, \ldots, n$.
- $X = \sum_{j=1}^{n} I_j$: the number of events that occur
- $\binom{X}{2} = \sum_{i<j} I_i I_j$: the number of pairs of distinct events that occur
- $E(\binom{X}{2}) = \sum_{i<j} P(A_i \cap A_j)$ $\quad I_i I_j = I(A_i) \cdot I(A_j) = I(A_i \cap A_j)$
  - $E(X^2) = 2\sum_{i<j} P(A_i \cap A_j) + E(X)$. $\quad E(I_i I_j) = P(A_i \cap A_j)$
  - $\text{Var}(X) = 2\sum_{i<j} P(A_i \cap A_j) + E(X) - (E(X))^2$.

$\Rightarrow E\left[\frac{X^2-X}{2}\right] = \frac{1}{2}E[X^2] - \frac{1}{2}E[X] = \sum_{i<j} P(A_i \cap A_j)$

$\Rightarrow E[X^2] = E[X] + 2\sum_{i<j} P(A_i \cap A_j)$

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

Moments of Binomial Random Variables

$X \sim Bin(n,p)$

$E[X^k]$

1°. Consider n independent Bernoulli trials, each ✓ w.p. p.

Event A_i: the i^{th} trial ✓, $I_j = I(A_j) \sim Bern(p)$.

$P(A_i) = p$.

2°. # of successful trials. $X = \sum_{j=1}^{n} I_j$

① $E[X] = \sum_{j=1}^{n} E(I_j) = \sum_{j=1}^{n} p = np$

② $E\left[\binom{X}{2}\right] = \sum_{i<j} P(A_i \cap A_j) \overset{A_i, A_j \text{ independent}}{=} \sum_{i<j} P(A_i) \cdot P(A_j) = \sum_{i<j} p^2 = \binom{n}{2} \cdot p^2$ $\quad i, j \in \{1, \dots, n\}$

$\Rightarrow E[X(X-1)] = n(n-1)p^2 \quad \Rightarrow E[X^2] = E[X] + n(n-1)p^2 = np + n(n-1)p^2$

$\Rightarrow Var(X) = E(X^2) - E^2(X) = np(1-p)$ $\quad Var(X) = \sum_{j=1}^{n} Var(I_j) = n \cdot p(1-p)$

③ $E\left[\binom{X}{3}\right] = \sum_{i<j<k} P(A_i \cap A_j \cap A_k) = \sum_{i<j<k} P(A_i) \cdot P(A_j) \cdot P(A_k) = \sum_{i<j<k} p^3 = \binom{n}{3} p^3$

$E\left[\binom{X}{k}\right] = \binom{n}{k} p^k$

Outline

Poisson Distribution

① valid PMF ✓

$$\sum_{k=0}^{\infty} P(X=k) \overset{?}{=} 1$$

$$\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} \overset{?}{=} 1$$

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \overset{?}{=} e^{\lambda}$$

Definition

An r.v. X has the *Poisson distribution* with parameter λ if the PMF of X is

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

We write this as $X \sim \mathrm{Pois}(\lambda)$.

② $E(X) = Var(X) = \lambda$.

Example: Poisson Expectation & Variance

① $E(X) = \sum_{k=0}^{\infty} k \cdot P(X=k) = \sum_{k=1}^{\infty} k \cdot P(X=k)$

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

$= e^{-\lambda} \cdot \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \cdot \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$ e^{λ}.

$= e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda$.

② $Var(X) = \lambda$.

Poisson Approximation

Let $A_1, A_2, \cdots, A_n$ be events with $p_j = P(A_j)$, where n is large, the p_j are small, and the A_j are independent or weakly dependent. Let

$$X = \sum_{j=1}^{n} I(A_j)$$

count how many of the A_j occur. Then X is approximately $\mathrm{Pois}(\lambda)$, with $\lambda = \sum_{j=1}^{n} p_j$.

Example: Birthday Problem Revisited

1°. m people ; $\binom{m}{2}$ pairs of people index $j = 1, 2, \dots \binom{m}{2}$

A_j : "the jth pair of people have the same birthday".

$P(A_j) = \frac{365}{365 \times 365} = \frac{1}{365}$, $j = 1, 2, \dots \binom{m}{2}$.

2°. $I_j = I(A_j)$; $n = \binom{m}{2}$; $X \triangleq$ # of birthday match.

$= \sum_{j=1}^{n} I_j$

3°. Poisson Approximation $X \;\dot\sim\; \text{Pois}(\lambda)$, $\lambda = \underline{n \cdot \frac{1}{365}} = \underline{\binom{m}{2} \cdot \frac{1}{365}}$.

4°. Prob (At least 1 birthday match) $= P(X \geq 1) = 1 - P(X < 1)$

$= 1 - P(X = 0) = \underline{1 - e^{-\lambda}}$

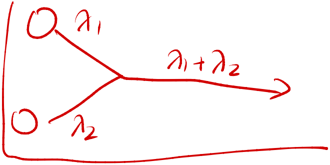
$m = 23$; $\lambda = \binom{23}{2} \cdot \frac{1}{365} = \frac{253}{365}$;

$1 - e^{-\lambda} \approx \underline{0.5002}$

Poisson & Binomial

- Poisson $\implies$ Binomial : **conditioning**
- Binomial $\implies$ Poisson: **taking a limit**

Sum of Independent Poissons



$$P(X+Y=k) \overset{\text{LOTP}}{=} \sum_{j=0}^{k} P(X+Y=k \mid X=j)\cdot P(X=j)$$

$$= \sum_{j=0}^{k} P(Y=k-j \mid X=j)\cdot P(X=j)$$

$$= \sum_{j=0}^{k} P(Y=k-j)\cdot P(X=j) = \sum_{j=0}^{k} \frac{e^{-\lambda_2}\cdot\lambda_2^{k-j}}{(k-j)!}\cdot\frac{e^{-\lambda_1}\cdot\lambda_1^{j}}{j!}$$

Theorem

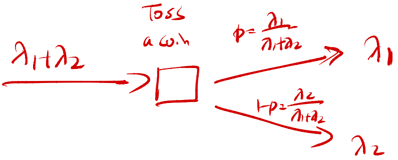
If $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$, and X is independent of Y, then $X + Y \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$.

$$= e^{-(\lambda_1+\lambda_2)}\cdot \sum_{j=0}^{k} \frac{\lambda_2^{k-j}\cdot\lambda_1^{j}}{(k-j)!\,j!} = \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{j=0}^{k}\frac{k!}{(k-j)!\,j!}\cdot\lambda_1^{j}\cdot\lambda_2^{k-j}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{j=0}^{k}\binom{k}{j}\cdot\lambda_1^{j}\cdot\lambda_2^{k-j} \quad (\lambda_1+\lambda_2)^k \quad \sim \mathrm{Pois}(\lambda_1+\lambda_2)$$

Poisson Given A Sum of Poissons

$0 \le k \le n$



$$P(X=k \mid X+Y=n) = \frac{P(X=k,\ X+Y=n)}{P(X+Y=n)} = \frac{P(X=k)\cdot P(Y=n-k)}{P(X+Y=n)}$$

$X+Y \sim \mathrm{Pois}(\lambda_1+\lambda_2)$

Theorem

If $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$, and X is independent of Y, then the conditional distribution of X given $X+Y=n$ is $\mathrm{Bin}(n, \lambda_1/(\lambda_1+\lambda_2))$.

$$= \frac{\frac{\cancel{e^{-\lambda_1}}\cdot\lambda_1^k}{k!}\cdot\frac{\cancel{e^{-\lambda_2}}\cdot\lambda_2^{n-k}}{(n-k)!}}{\cancel{e^{-(\lambda_1+\lambda_2)}}\cdot\frac{(\lambda_1+\lambda_2)^n}{n!}} = \frac{n!}{k!(n-k)!}\cdot\frac{\lambda_1^k}{(\lambda_1+\lambda_2)^k}\cdot\frac{\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^{n-k}}$$

$$= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\cdot\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

Poisson Approximation to Binomial

$\lambda = np$; Given k $(0 \le k \le n)$, $X \sim Bin(n,p)$

$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$; $p = \lambda/n$

PMF of X.

$= \frac{n(n-1)\cdots(n-k+1)}{k!} (\lambda/n)^k (1-\lambda/n)^{n-k}$.

Theorem

If $X \sim \mathrm{Bin}(n,p)$ and we let $n \to \infty$ and $p \to 0$ such that $\lambda = np$ remains fixed, then the PMF of X converges to the $\mathrm{Pois}(\lambda)$ PMF. More generally, the same conclusion holds if $n \to \infty$ and $p \to 0$ in such a way that np converges to a constant λ. $n \cdot p \to \lambda$.

$= \frac{\lambda^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n \cdot n \cdots n} \cdot (1-\lambda/n)^{n-k}$

fixed λ, k

$= \frac{\lambda^k}{k!} [1 \cdot (1-\frac{1}{n}) \cdots (1-\frac{k-1}{n})] \cdot (1-\lambda/n)^n \cdot (1-\lambda/n)^{-k}$

$n \to \infty$

$\to \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1 = e^{-\lambda} \frac{\lambda^k}{k!}$ Pois(λ)

Proof

Visitors to A Website

each day.
# of visitor to the site X

$X \sim Bin(n, p)$. $n = 10^6$, $p = 2\times10^{-6}$

$P(X \geq 3)$

Y : $Y \sim Pois(\lambda)$, $\lambda = n \cdot p = 2$

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

The owner of a certain website is studying the distribution of the number of visitors to the site. Every day, a million people independently decide whether to visit the site, with probability $p = 2 \times 10^{-6}$ of visiting. Give a good approximation for the probability of getting *at least three* visitors on a particular day.

$P(X \geq 3) \approx P(Y \geq 3) = 1 - P(Y<3) = 1 - P(Y=0) - P(Y=1) - P(Y=2)$

$= 1 - 5e^{-2} \approx 0.3233$

Outline

Typical Distance Measures

- Total Variation Distance
- Kullback–Leibler Divergence
- Jensen–Shannon Divergence
- Bhattacharyya Distance
- Wasserstein Distance (or called "Kantorovich–Rubinstein")

Total Variation Distance

- Distance measure between two probability distributions
- Apply such measure to characterize the accuracy of Poisson approximation

Definition

The **total variation distance** between two discrete distributions μ and ν on a countable set Ω is

$$d_{TV}(\mu, \nu) \overset{\Delta}{=} \| \mu - \nu \|_{TV}$$
$$= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

$$\leq \frac{1}{2}\left(\sum_{x\in\Omega} \mu(x) + \sum_{x\in\Omega} \nu(x)\right) = \frac{1}{2}(1+1) = 1$$

Example

① $\mu \sim \mathrm{Bern}(p)$. $\mu(1)=p$; $\mu(0)=1-p$; $\mu(n)=0$, $n\geq 2$;

$\nu \sim \mathrm{Pois}(p)$. $\nu(n) = \frac{e^{-p}\cdot p^n}{n!}$ $(n\geq 0)$ $\sum_{n=0}^{\infty} \nu(n) = 1$

② $2d_{TV}(\mu,\nu) = \sum_{x\in N} |\mu(x)-\nu(x)| = |\mu(0)-\nu(0)| + |\mu(1)-\nu(1)| + \sum_{n\geq 2} |\mu(n)-\nu(n)|$

$= |1-p-e^{-p}| + |p-pe^{-p}| + \sum_{n\geq 2}\nu(n)$ $[1-\nu(0)-\nu(1)] = [1-e^{-p}-pe^{-p}]$

Let μ be the distribution with $\mu(1) = p$ and $\mu(0) = 1 - p$. Let ν be a Poisson distribution with mean p. Then we have $d_{TV}(\mu,\nu) \leq p^2$.

$e^{-p} \geq 1-p$

$\Rightarrow 1-e^{-p} \leq p$

$= e^{-p}-(1-p) + p - pe^{-p} + [1-e^{-p}-pe^{-p}]$

$= 2p(1-e^{-p}) \leq 2p\cdot p = 2p^2$

③ $d_{TV}(\mu,\nu) \leq p^2$ $\mu \sim \mathrm{Bern}(p)$,

$\nu \sim \mathrm{Pois}(p)$,

The Law of Small Numbers

Law of Rare Events.

Theorem

Given independent random variables $Y_1, \cdots, Y_n$ such that for any $1 \le m \le n$, $\mathbb{P}(Y_m = 1) = p_m$ and $\mathbb{P}(Y_m = 0) = 1 - p_m$. Let $S_n = Y_1 + \cdots + Y_n$. Suppose

$$\sum_{m=1}^{n} p_m \to \lambda \in (0, \infty) \quad \text{as } n \to \infty,$$

and

$$\max_{1 \le m \le n} p_m \to 0 \quad \text{as } n \to \infty,$$

then

$$d_{TV}(S_n, Poi(\lambda)) \to 0 \quad \text{as } n \to \infty.$$

Gap of Poisson Approximation

- A bound on the gap due to Hodges and Le Cam (1960):

$$d_{TV}(S_n, Poi(\lambda)) \le \sum_{m=1}^{n} p_m^2,$$

- by Stein-Chen method (C.Stein 1987) we can have a tighter bound on the gap:

$$d_{TV}(S_n, Poi(\lambda)) \le \min(1, \frac{1}{\lambda}) \sum_{m=1}^{n} p_m^2.$$

Outline

Probability Generating Function

encoding

polynomial t^k

coefficient

Definition

The *probability generating function* (PGF) of a nonnegative integer-valued r.v. X with PMF $p_k = P(X = k)$ is the generating function of the PMF. By LOTUS, this is

$$E\left(t^X\right) = \sum_{k=0}^{\infty} p_k t^k$$

$f(t)$

The PGF converges to a value in $[-1, 1]$ for all t in $[-1, 1]$ since $\sum_{k=0}^{\infty} p_k = 1$ and $\left|p_k t^k\right| \le p_k$ for $|t| \le 1$.

Example: Generating Dice Probabilities $6 \le X \le 36$

① $E[t^X] = \sum_{k=0}^{\infty} P(X=k)t^k = f(t)$ t^{18}

② $X = X_1 + \cdots + X_6$ i.i.d. X_i $\Rightarrow E[t^X] = E[t^{X_1 + \cdots + X_6}]$

$= E[t^{X_1}] \cdot E[t^{X_2}] \cdots E[t^{X_6}] = (E[t^{X_1}])^6$

Let X be the total from rolling 6 fair dice, and let $X_1, \ldots, X_6$ be the individual rolls. What is $P(X = 18)$?

$E[t^X] = f(t)$ t^{18}

$P(X=17)$ t^{17}

$P(X=25)$. t^{25}

③ $E[t^{X_1}] = \sum_{j=1}^{6} P(X_1 = j) \cdot t^j = \frac{1}{6}(t + t^2 + \cdots + t^6)$

④ $E[t^X] = \frac{1}{6^6}(t + t^2 + \cdots + t^6)^6$

$P(X=18) = \frac{3431}{6^6}$

Solution

$E[t^X] = f(t) =$

$\frac{1}{6^6}[\, t^6 + 6t^7 + 21t^8 + 56t^9 + 126t^{10} + 252t^{11} + 456t^{12}$

$+ 756t^{13} + 1161t^{14} + 1666t^{15} + 2247t^{16} + 2856t^{17}$

$+ 3431t^{18} + 3906t^{19} + 4221t^{20} + 4332t^{21} + 4221t^{22}$

$+ 3906t^{23} + 3431t^{24} + 2856t^{25} + 2247t^{26}$

$+ 1666t^{27} + 1161t^{28} + 756t^{29} + 456t^{30} + 252t^{31}$

$+ 126t^{32} + 56t^{33} + 21t^{34} + 6t^{35} + t^{36}\,]$

$P(X=a)$
$= P(X=42-a)$

$6 \le a \le 36$?

$P(X=18) = \frac{3431}{6^6}$

$P(X=18) = P(X=24)$?

PGF and Moments

PGF of $X \to E[X^k]$, $k \geq 1$

① $g(t) = \sum_{k=0}^{\infty} p_k t^k = p_0 + \sum_{k=1}^{\infty} p_k t^k$;

$g'(t) = \sum_{k=1}^{\infty} p_k \cdot k \cdot t^{k-1}$ $\qquad g'(t)|_{t=1} = \sum_{k=1}^{\infty} p_k \cdot k = \sum_{k=0}^{\infty} p_k \cdot k$

$= E[X]$;

Let X be a nonnegative integer-valued r.v. with PMF $p_k = P(X = k)$, and the PGF of X is $g(t) = \sum_{k=0}^{\infty} p_k t^k$, we have

- $E(X) = g'(t)|_{t=1}$
- $E(X(X-1)) = g''(t)|_{t=1}$

② $g'(t) = p_1 + \sum_{k=2}^{\infty} p_k \cdot k t^{k-1}$

$g''(t) = \sum_{k=2}^{\infty} p_k \cdot k \cdot (k-1) \cdot t^{k-2}$

$k(k-1) = 0$; $k = 0, 1$;

$\Rightarrow g''(t)|_{t=1} = \sum_{k=2}^{\infty} p_k \cdot k(k-1) = \sum_{k=0}^{\infty} p_k \cdot k(k-1)$

$= E[X(X-1)]$

$= E(X^2) - E(X)$

PGF and Moments

PGF of X $\xrightarrow{\text{decoding}}$ PMF of X

① $g(t) = \sum_{k=0}^{\infty} p_k \cdot t^k = p_0 + \underbrace{\sum_{k=1}^{\infty} p_k \cdot t^k}$;

$g(0) = p_0 = P(X=0)$

② $g'(t) = \sum_{k=1}^{\infty} p_k \cdot k \cdot t^{k-1} = p_1 + \underbrace{\sum_{k=2}^{\infty} p_k \cdot k \cdot t^{k-1}}$

$g'(0) = p_1 = P(X=1)$, -

. . .

$P(X=k) = p_k = \dfrac{g^{(k)}(0)}{k!}$;

encoding — decoding

PGF and Moments

Binomial PMF

① $X \sim Bin(n,p)$. $X = X_1 + \cdots + X_n$ n i.i.d. Bern(p);

PGF of X

② $g_X(t) = E[t^X] = E[t^{X_1+\cdots+X_n}] = E[t^{X_1}] \cdots E[t^{X_n}]$

$= (E[t^{X_1}])^n = (pt+q)^n$

$X \perp\!\!\!\perp Y$

$h_1(X) \perp\!\!\!\perp h_2(Y)$

$$\boxed{E[t^{X_1}] = t^0 \cdot (1-p) + t^1 \cdot p = pt + q \quad (q = 1-p)}$$

③ $g_X(0) = q^n$; $g_X'(0) = n \cdot p \cdot q^{n-1}$; $g_X''(0) = \binom{n}{2} p^2 q^{n-2}$

$\cdots \; p_k = \frac{g_X^{(k)}(0)}{k!} = \binom{n}{k} p^k q^{n-k}$

Binomial Moments

PGF $g_X(t) = (pt+q)^n$, $\underline{p+q=1}$

① $g_X'(t) = \underline{np(pt+q)^{n-1}}$ $\quad g_X'(t)\big|_{t=1} = np = E[X]$

② $g_X''(t)\big|_{t=1} = n(n-1)p^2 = E[X(X-1)]$

$\underline{\frac{n}{2}(n-1)p^2} = E\left[\frac{X(X-1)}{2}\right]$

$\Rightarrow \binom{n}{2}p^2 = E\left[\binom{X}{2}\right]$

③ $E\left[\binom{X}{k}\right] = \binom{n}{k}p^k$. $k \geq 2$

Example: Pattern Matching

$q = 1-p$ T H H 1 2 3

① $p_k = P(N=k)$; $p_0 = 0$; $p_1 = 0$; $p_2 = p^2$; $p_3 = (1-p)\cdot p^2 = q\cdot p^2$

$p_4 =$ H or T T H H
1 2 3 4

② First-Step Method.

Suppose a coin with probability p for heads is tossed repeatedly, and we obtain a sequence of H and T (H denotes Head and T denotes Tail). Let N denote the number of toss to observe the first occurrence of the pattern "HH". Find $E(N)$ and $\text{Var}(N)$.

$k \geq 3$; S_1 : result of the first toss ; $S_1 = H$ or T

$p_k = P(N=k) = P(N=k, S_1 = H) + P(N=k, S_1 = T)$

Example: Pattern Matching

$k \geq 3$

③ 1°. $P(N=k;\ S_1=H)$

$= P(S_1=H)\cdot P(S_2=T)\cdot P(N=k-2)$

$= p\cdot q\cdot p_{k-2}$

H T
1 2 3 .... k

$(k-2)$ tosses to get HH for the first time.

2°. $P(N=k,\ S_1=T)$

$= P(S_1=T)\cdot P(N=k-1)$

$= q\cdot p_{k-1}$.

T
1 2 3 ... k

k-1 tosses to get HH for the first time.

$$\Rightarrow \begin{cases} p_k = p\cdot q\cdot p_{k-2} + q\cdot p_{k-1} & k\geq 3 \\ p_0 = 0;\ p_1 = 0;\ p_2 = p^2 \end{cases}$$

$q = 1-p$

Method 1: Find p_k $\Rightarrow E(N)$; $Var(N)$;

Example: Pattern Matching

Method 2: Find PGF $\to E[N]$, $Var[N]$.

$$\begin{cases} p_k = p_{k-1}\cdot q + p_{k-2}\cdot pq, \quad k\ge 3 \\ p_0 = 0;\ p_1 = 0;\ p_2 = p^2 \end{cases}$$

④ PGF of N: $g(t) = E[t^N]$

$$= \sum_{k=0}^{\infty} p_k \cdot t^k = \sum_{k=1}^{\infty} p_k \cdot t^k = \sum_{k=2}^{\infty} p_k \cdot t^k = p_2 \cdot t^2 + \sum_{k=3}^{\infty} p_k \cdot t^k = p^2 t^2 + \sum_{k=3}^{\infty} p_k \cdot t^k$$

$g(t)$

on the other hand; $p_k = p_{k-1}\cdot q + p_{k-2}\cdot pq \quad k \ge 3$

$$\sum_{k=3}^{\infty} p_k t^k = \sum_{k=3}^{\infty} (p_{k-1}\cdot q + p_{k-2}\, pq) t^k = \sum_{k=3}^{\infty} p_{k-1}\, q \cdot t^k + \sum_{k=3}^{\infty} p_{k-2}\cdot pq \cdot t^k$$

$$g(t) - p^2 t^2 = qt \sum_{k=3}^{\infty} p_{k-1}\cdot t^{k-1} + p\cdot q t^2 \sum_{k=3}^{\infty} p_{k-2}\cdot t^{k-2}$$

$$= qt \sum_{k=2}^{\infty} p_k \cdot t^k + p\cdot q t^2 \sum_{k=1}^{\infty} p_k \cdot t^k$$

$$= qt \cdot g(t) + p\cdot q t^2 \cdot g(t)$$

$$\Rightarrow g(t) = \frac{p^2 t^2}{1 - qt - pqt^2};$$

Example: Pattern Matching

⑤ PGF of N: $g(t) = \frac{p^2t^2}{1-qt-pqt^2}$

PMF of N

$E(N) = g'(t)\big|_{t=1} = g'(1) = \frac{1}{p} + \frac{1}{p^2}$

$Var(N) = g''(1) + g'(1) - [g'(1)]^2 = \frac{1-p^5-5qp^2}{q^2p^4}$

⑥ $p = \frac{1}{2}$; fair coins

$\Rightarrow E(N) = 6 > 4$

$Var(N) = 22$

$p' = \frac{1}{4}$

$E(\ ) = \frac{1}{p'} = 4$

Example: Pattern Matching

Outline

Probability Method

- Paul Erdős initiated this method: Erdős Method
- Widely used in information theory & combinatorics & theoretical computer science
- Main idea: to prove the existence of a structure with certain properties using probability or expectation

Principle I

- First we construct an appropriate probability space of structures.
- Then we show that a randomly chosen element in this space has the desired properties with positive probability

Theorem (The Possibility Principle)

Let A be the event that a randomly chosen object in a collection has a certain property. If $P(A) > 0$, then there exists an object with such property.

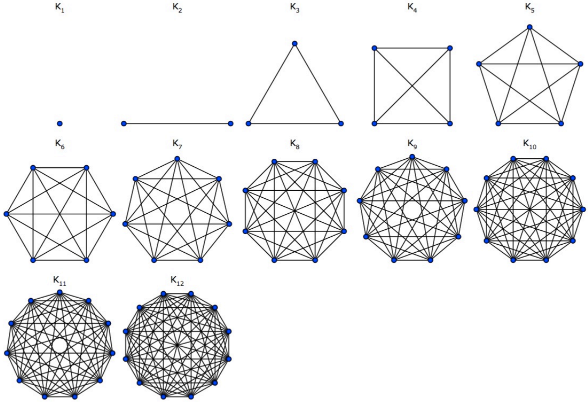
Principle II

Theorem (The Good Score Principle)

Let X be the score of a randomly chosen object. If $E(X) \geq c$, then there exists an object with a score of at least c.

Example: Graph Coloring

- Complete graph (clique): a simple undirected graph in which every pair of distinct vertices is connected by a unique edge.
- Complete graph K_n: a graph with n nodes and $\binom{n}{2}$ edges.



Example: Graph Coloring

Theorem

Given a complete graph K_n ($n \geq 3$), if $\binom{n}{m}2^{-\binom{m}{2}+1} < 1$, then it is possible to color the edges of K_n with two colors so that it has no monochromatic K_m subgraph ($1 < m < n$).

Testing Polynomial Identities

- Randomized algorithms can be dramatically more efficient than their best known deterministic counterparts.
- Input two polynomials Q and R over n variables, with coefficients in some field, and decides whether $Q \equiv R$.
- Example: $Q(x_1, x_2) = (1 + x_1)(1 + x_2)$, $R(x_1, x_2) = 1 + x_1 + x_2 + x_1 x_2$.
- n-variable polynomial $\prod_{i=1}^{n}(x_i + x_{i+1})$ expands into $O(2^n)$ monomials.

The Schwartz-Zippel Algorithm

- A Monte Carlo algorithm with a bounded probability of false positive and no false negative.
- Input polynomial $M(x_1, \ldots, x_n)$ and test whether $M \equiv 0$ $(M = Q - R)$.
- Assign values $r_1, \ldots, r_n$ chosen independently and uniformly at random from a finite set S to $x_1, \ldots, x_n$.
- Test if $M(r_1, \ldots, r_n) = 0$, outputting "Yes" if so and "No" otherwise.
- If "No", then $M \not\equiv 0$.
- If "Yes", it is possible that $M \not\equiv 0$ but $r_1, \ldots, r_n$ happens to be a zero of M.

Schwartz-Zippel Lemma

Lemma

Let $M \in F(x_1, x_2, \dots, x_n)$ be a non-zero polynomial of total degree $d \geq 0$ over a field F. Let S be a finite subset of F and let $r_1, r_2, \dots, r_n$ be selected at random independently and uniformly from S. Then

$$P[M(r_1, r_2, \dots, r_n) = 0] \leq \frac{d}{|S|}.$$

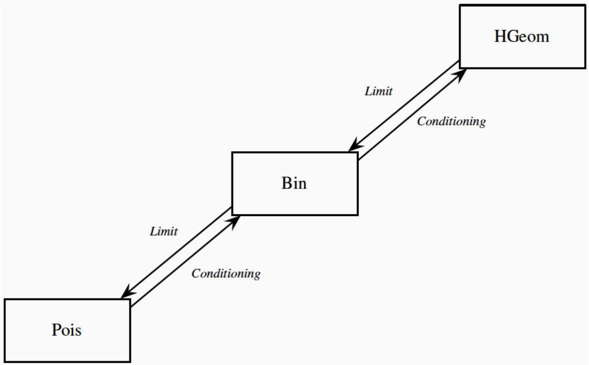
Remarks

- If we take the set S to have cardinality at least twice the degree of our polynomial ($|S| \geq 2d$), we can bound the probability of error (false positive) by $1/2$.
- This can be reduced to any desired small number by repeated trials.

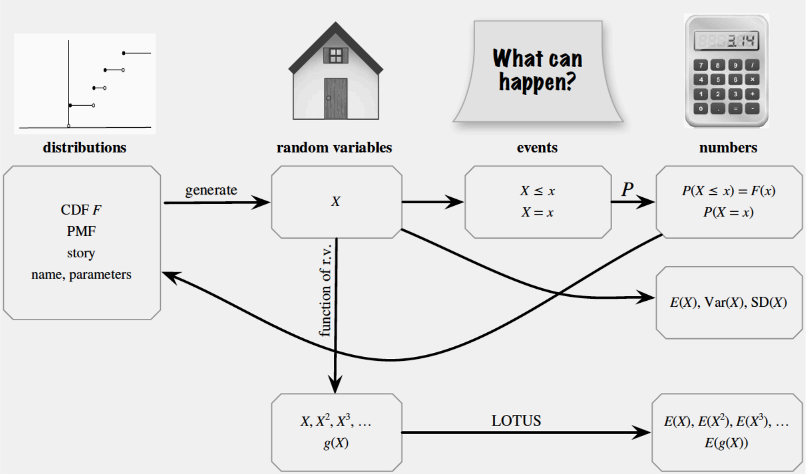
Summary 1

	With replacement	Without replacement
Fixed number of trials	Binomial	Hypergeometric
Fixed number of successes	Negative Binomial	Negative Hypergeometric

Summary 2



Summary 3



What can happen?
distributions
random variables
events
numbers
CDF F
PMF
story
name, parameters
generate
X
$X \le x$
$X = x$
P
$P(X \le x) = F(x)$
$P(X = x)$
function of r.v.
$E(X)$, $\mathrm{Var}(X)$, $\mathrm{SD}(X)$
$X, X^2, X^3, \ldots$
$g(X)$
LOTUS
$E(X), E(X^2), E(X^3), \ldots$
$E(g(X))$

References

- Chapters 4 & 6 of **BH**
- Chapter 2 of **BT**