Lecture 4: Expectation

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

March 26, 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

if $E(X) = E(Y) \nRightarrow X \sim Y$.

$$X = \begin{cases} 100 & \text{w.p. } \frac{1}{2} \\ 0 & \text{w.p. } \frac{1}{2}. \end{cases} \qquad Y = \begin{cases} 70 & \text{w.p. } \frac{1}{2} \\ 30 & \text{w.p. } \frac{1}{2}. \end{cases}$$

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

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

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

$Z = X - Y \quad \Rightarrow \quad Z \geq 0$ w.p. 1.

$\Rightarrow E(Z) \geq 0 \quad \Rightarrow E(X-Y) \geq 0$

$\Rightarrow E(X) - E(Y) \geq 0 \quad \Rightarrow E(X) \geq E(Y)$

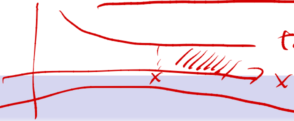
Theorem

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.

if X, 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

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

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

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

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).

$\{X>n\}$ $\{X \ge n+1\}$

Proof

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

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

$$1\cdot P(X=1) + 2\cdot P(X=2) + 3\cdot P(X=3) + \cdots$$
$$= \sum_{n=1}^{\infty} n\cdot P(X=n) = \sum_{n=0}^{\infty} n\cdot P(X=n)$$
$$= E(X)$$

$$(0\cdot P(X=0) = 0)$$

$$\sum_{n=0}^{\infty} G(n) = \sum_{n=0}^{\infty} P(X>n)$$
$$= \sum_{n=0}^{\infty} P(X\geq n+1) = \sum_{n=1}^{\infty} P(X\geq n)$$
$$= \sum_{n=1}^{\infty}\left(\sum_{m=n}^{\infty} P(X=m)\right)$$

Fubini theorem

$$\sum_{m=1}^{\infty}\left(\sum_{n=1}^{m} P(X=m)\right) = \sum_{m=1}^{\infty} m\cdot P(X=m)$$
$$= \sum_{m=0}^{\infty} m\cdot P(X=m) = E(X)$$

$\{X>n\} \Leftrightarrow \{X\geq n+1\}$, X, n integer.

Law Of The Unconscious Statistician (LOTUS)

$g(x)$ is a r.v. $\longrightarrow$ distribution of $g(x)$

$P(g(x)=y)$

Theorem

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

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

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

$\Big\downarrow$ definition

$E[g(x)]$

$= \sum_{y} y \cdot P(g(x)=y)$

Variance and Standard Deviation

X : 49 , 51 (each with probability $\frac{1}{2}$)

Y : 0 , 100

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

Definition

The variance of an r.v. X is

$$\mathrm{Var}(X) = E(X - EX)^2. \geq 0$$

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

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

Properties of Variance

- 1^o. For any r.v. X, $\mathrm{Var}(X) = E(X^2) - (EX)^2$.
- 2^o. $Var(X+c) = Var(X)$ for any constant c.
- 3^o. $Var(cX) = c^2 Var(X)$ for any constant c.
- 4^o. If X and Y are independent, then $Var(X+Y) = Var(X) + Var(Y)$.

  $\forall X, Y$, $E(X+Y) = E(X) + E(Y)$
- 5^o. $Var(X) \geq 0$ with equality if and only if $P(X = a) = 1$ for some constant a.

$Var(X) \geq 0 \iff E[(X-EX)^2] \geq 0$

$\iff E(X^2) - (E(X))^2 \geq 0$

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

Properties of Variance

Outline

Story: Geometric Distribution

$k+1$ trials.

$P(X=k)$

$k \geq 0.$

First k trial fail.

The last trial successful.

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 $X \sim Geom(p)$.

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

Geometric PMF

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

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

$(q^k)'$

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

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

$q = 1-p$

$(1-p)^k \cdot p$

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

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

$\left(\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \right)$

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

Memoryless Property

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

2° $k \geq 1$; $P(X \geq n+k \mid X \geq k)$

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

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

$$= \frac{P(X \geq n+k)}{P(X \geq k)} = \frac{q^{n+k}}{q^k} = q^n = P(X \geq n)$$ ✓

$k=80$; $n=20$;

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

$<$

$\geq$

Memoryless Property

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

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

2°. $k=0$. $P(X\geq n) = P(X\geq n)\cdot P(X\geq 0)$. $\forall n \in \mathbb{N}^+$.

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

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

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

$G(1) = P(X\geq 1)$

$0\leq \quad \leq 1$

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

Let $0\leq q = G(1) \leq 1$.

$\Rightarrow 4^\circ$. $G(n+k) = G(n)\cdot G(k)$; $n=k=1 \Rightarrow G(2) = G^2(1) = q^2$

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

$G(n) = G^n(1) = q^n \Rightarrow P(X\geq n) = q^n$.

Memoryless Property

Theorem

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

First Success Distribution

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

$Y \sim FS(p)$

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 $Y \sim \mathrm{FS}(p)$.

$$Y = 1 + X$$

Example: Geometric & First Success Expectation

1°. $P(X \geq k) = q^k$.

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

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

2° . $Y = 1 + X$

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

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

Story: Negative Binomial Distribution

$p(X=n)$

n failures before rth success.

$n+r$ trials.

the last trial is the rth successful trial

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 $X \sim NBin(r, p)$.

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

$n+r-1$ trials.

n failures, $r-1$ successful.

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

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

Negative Binomial PMF

$r>0$

$\binom{-r}{n}$

Theorem

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

$\binom{-r}{n}$

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

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

Geometric & Negative Binomial

$X_1 \sim \mathrm{Geom}(p)$ X_1 : # of failures before the 1st success.

$X_2 \sim \mathrm{Geom}(p)$ X_2 : # . . . . . . between 1st success and 2nd success.

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. $\mathrm{Geom}(p)$.

Example: Expectation

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

$\times$

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

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

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

Method 2 : $X = X_1 + \cdots + X_r$ $\quad X_i \sim Geom(p)$

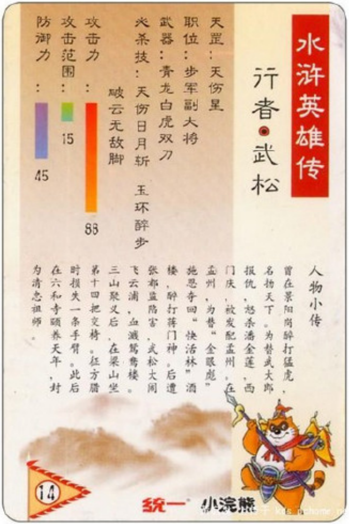
$E(X_i) = \frac{1-p}{p}$.

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

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

Example: 小浣熊干脆面与水浒英雄卡

Example: 小浣熊干脆面与水浒英雄卡



Example: 小浣熊干脆面与水浒英雄卡

Example: 小浣熊干脆面与水浒英雄卡

为了收集齐108张水浒英雄卡，平均而言你需要购买多少包小浣熊方便面？

Example: 盲盒收集

Example: 盲盒收集

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 $Var(N)$.

Solution: Coupon Collector

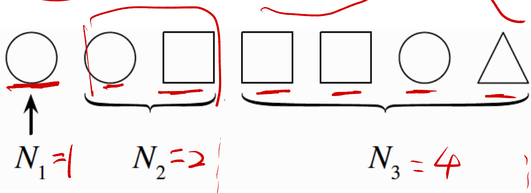
1°. N : # of toys needed to obtain all types of toys.

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

$N_1 = 1$;

N_2: additional # of toys until the second new type toy appeared.;

N_3: . . . . . .



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

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

Solution: Coupon Collector

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

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

4^0. $N = N_1 + \cdots + N_n$, $E(N) = E(N_1 + \cdots + N_n)$

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

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

$= n\left[\underbrace{\frac{1}{n} + \frac{1}{n-1} + \cdots + 1}\right]$

$= n \sum_{j=1}^{n} \frac{1}{j}$ $n \gg 1$

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

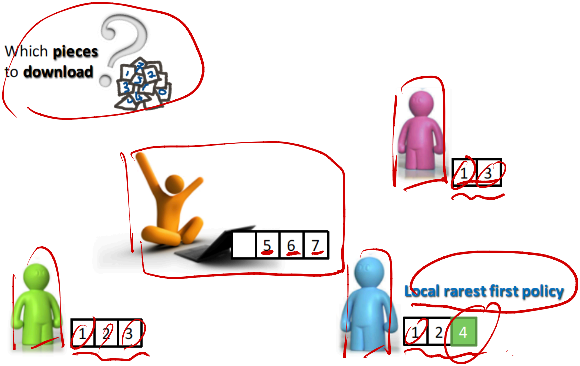
$n = 108$; $E(N) \approx 568$.

Application: Peer-to-Peer System

- Target file is decomposed into n pieces.
- 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.



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 : $I_{A\cap B} = 1 \Rightarrow A \cap B$ occurs.

A, B both occur

$I_A = 1$, $I_B = 1$;

$I_A I_B = 1$; ✓

$I_{A\cap B} = 0 \Rightarrow I_A = 0$ or $I_B = 0$

or $I_A, I_B = 0$

$I_A I_B = 0$; ✓

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

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

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

$= I_A + I_B - I_A I_B$

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:

$$P(A) = E(I_A).$$

Example: Booler's Inequality

(X, Y r.v. $X \le Y$; $\Rightarrow E(X) \le E(Y)$.)

$P(A) = E(I_A)$

<1> $I(A_1 \cup \dots \cup A_n) \le I(A_1) + \dots + I(A_n)$

LHS RHS.

1°. if LHS = 0; ✓ (RHS ≥ 0)

2°. if LHS = 1; ⇒ at least A_j occurs. $j \in \{1, 2, \dots, n\}$

⇒ at least $I(A_j) = 1$ ⇒ RHS ≥ 1

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

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

<2> Taking Expectation of both sides.

$E[I(A_1 \cup \dots \cup A_n)] \le E(I(A_1) + \dots + I(A_n))$

$P(A_1 \cup \dots \cup A_n) \le E(I(A_1)) + \dots + E(I(A_n))$

$P(A_1) + \dots + P(A_n)$. ✓

Solution: Booler's Inequality

Example: Inclusion-Exclusion Formula

1°. $\cancel{1-} I(A_1 \cup \cdots \cup A_n) = I(\overline{A_1 \cup \cdots \cup A_n}) = \underline{I(\bar{A}_1 \cap \cdots \cap \bar{A}_n)}$

$= I(\bar{A}_1) \cdots I(\bar{A}_n) = [1 - I(A_1)] \cdots [1 - I(A_n)]$

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

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

3°. Taking Expectation of both sides.

Solution: Inclusion-Exclusion Formula

Outline

Moments of Indicator Methods

$E[X^k]$: kth moment

$I_j = 1 \Rightarrow$ event A_j occurs.

- 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

  $I_i I_j = I(A_i) \cdot I(A_j) = I(A_i \cap A_j)$
- $E(\binom{X}{2}) = \sum_{i<j} P(A_i \cap A_j)$

  $\binom{X}{2} = \frac{X}{2}(X-1)$
  - $E(X^2) = 2\sum_{i<j} P(A_i \cap A_j) + E(X)$.
  - $Var(X) = 2\sum_{i<j} P(A_i \cap A_j) + E(X) - (E(X))^2$.

Moments of Binomial Random Variables

$X \sim Bin(n,p)$. $E[X^k]$

1°. Consider n independent Bernoulli trials, each with successful prob p.

event A_i : the ith trial is a success. $P(A_i) = p$.

$I_j = I(A_i) \sim Bern(p)$.

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

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

$i, j \in \{1, \dots, n\}$

A_i, A_j independent

② $E\binom{X}{2} = \sum_{i<j} P(A_i \cap A_j) = \sum_{i<j} P(A_i) \cdot P(A_j) = \sum_{i<j} p^2 = \binom{n}{2} p^2$

$\Rightarrow E(X(X-1)) = n(n-1)p^2 \Rightarrow E(X^2) = n(n-1)p^2 + E(X)$

$= n(n-1)p^2 + np$.

$\Rightarrow Var(X) = E(X^2) - (E(X))^2 = np(1-p)$

$E\binom{X}{3}$

$= \sum_{i<j<k} P(A_i \cap A_j \cap A_k)$

$= \sum_{i<j<k} P(A_i) P(A_j) P(A_k)$

$= \sum_{i<j<k} p^3 = \binom{n}{3} p^3$

$E\binom{X}{m}$

$= \binom{n}{m} p^m$

Outline

Poisson Distribution

① Valid PMF ✓

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

$$\Leftrightarrow \sum_{k=0}^{\infty} \frac{e^{-\lambda}\cdot\lambda^k}{k!} = 1 \checkmark$$

Definition

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

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

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

$$\Leftrightarrow \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}. \checkmark$$

② $E(X) = Var(X) = \lambda$

Example: Poisson Expectation & Variance

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

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

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

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

② $\underline{Var(X)}$ ..

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

$j = 1, 2, \dots \binom{m}{2}$

1^0. m people ; $\binom{m}{2}$ pairs of people ;

Prob (each pair of people have the same birthday) $= \frac{365}{365 \cdot 365} = \frac{1}{365}$.

2^0. A_j : " j th pair of people have the same birthday " . $P(A_j) = \frac{1}{365} = p$.

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

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

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

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

3^0. Poisson Approximation. ② $X \sim Pois(\lambda)$. $\lambda = np = \binom{m}{2} \cdot \frac{1}{365}$

4^0. Prob (at least one birthday match) $= P(X \geq 1) = 1 - P(X < 1)$

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

Poisson & Binomial

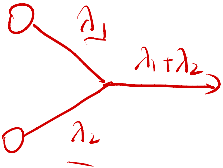
- Poisson $\Longrightarrow$ Binomial : **conditioning**
- Binomial $\Longrightarrow$ Poisson: **taking a limit**

Sum of Independent Poissons

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

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

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



Theorem

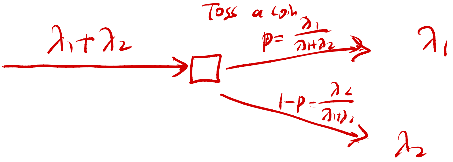
If $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$, and X is independent of Y, then $X + Y \sim \mathrm{Pois}(\lambda_1 + \lambda_2)$.

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

$\binom{k}{j}$.

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

Poisson Given A Sum of Poissons



$0 \le k \le n$

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

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 $\mathrm{Bin}(n, \lambda_1/(\lambda_1+\lambda_2))$.

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

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

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

Poisson Approximation to Binomial

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

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!\,(n-k)!} \cdot p^k (1-p)^{n-k} \qquad \Big|\ p = \lambda/n.$$

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

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.

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

λ: fixed.

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

$n \to \infty$

$$\longrightarrow \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1 = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \sim \mathrm{Pois}(\lambda)$$

Proof

Visitors to A Website

Y : $Y \sim pois(\lambda)$, $\lambda = np = 10^6 \times 2\times10^{-6} = 2$

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

$X \sim Bin(n,p)$

$n = 10^6$, $p = 2\times10^{-6}$

$P(X \geq 3)$

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.

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

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

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

(Discrete)

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

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

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

Example

① $\mu(1) = p$; $\mu(0) = 1-p$; $\mu(n) = 0$; $n \geq 2$

$v(n) = \frac{e^{-p} \cdot p^n}{n!}$ $(n \geq 0)$ $\left(\sum_{n=0}^{\infty} v(n) = 1\right)$ $\Omega = \{0,1,2,\dots\}$

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

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

$0 \leq x \leq 1$; lemma

$1 - x \leq e^{-x} \leq 1$.

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 $d_{TV}(\mu, \nu) \leq p^2$.

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

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

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

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

$\Rightarrow$ ③ $d_{TV}(\mu, v) \leq p^2$

$\mu \sim Bern(p)$.

$v \sim Pois(p)$.

The Law of Small Numbers

Law of Rare Events.

Theorem

Given independent random variables $Y_1, \cdots, Y_n$ *such that for any* $1 \leq m \leq n$, $\mathbb{P}(Y_m = 1) = p_m$ *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

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

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):

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

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

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

Outline

Probability Generating Function

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

$E(t^X) = \sum_{k=0}^{\infty} t^k \cdot p(X=k)$

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

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

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

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

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

Let X be the total from rolling 6 fair dice, and let $X_1, \ldots, X_6$ be the individual rolls. What is $P(X = 18)$? $P(X=17)$?

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 $E[t^X] = \frac{1}{6^6}(t + t^2 + \cdots + t^6)^6 = \frac{t^6}{6^6}(1 + t + \cdots + t^5)^6$

⑤ $P(X=18) = \frac{3431}{6^6}$

Solution

PGF and Moments

PGF of $X \longrightarrow E[X^n],\ n \geq 1.$

① $g(t) = \sum_{k=0}^{\infty} p_k t^k = p_0 + \sum_{k=1}^{\infty} p_k t^k$; $g'(t) = \sum_{k=1}^{\infty} p_k \cdot k\, t^{k-1}$

$$g'(t)\Big|_{t=1} = \sum_{k=1}^{\infty} p_k \cdot k = \sum_{k=0}^{\infty} p_k \cdot k = E[X]$$

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

- $E(X) = g'(t)|_{t=1}$
- $E(X(X-1)) = g''(t)|_{t=1}$ $\quad E(X^2) - E(X)$

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

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

$$= \sum_{k=0}^{\infty} p_k \cdot k(k-1) = E[X(X-1)]$$

PGF and Moments

PGF of X $\longrightarrow$ PMF of X.

① $g(t) = \sum_{k=0}^{\infty} p_k \cdot t^k = p_0 + \sum_{k=1}^{\infty} p_k \cdot t^k$. $g(0) = p_0 = P(X=0)$

② $g'(t) = \sum_{k=1}^{\infty} k \cdot p_k \cdot t^{k-1} = p_1 + \sum_{k=2}^{\infty} k \cdot p_k \cdot t^{k-1}$: $g'(0) = p_1 = P(X=1)$

....

$$P(X=k) = p_k = \underline{\frac{g^{(k)}(0)}{k!}}$$

PGF and Moments

Binomial PMF

① $X \sim \text{Bin}(n,p)$. $\quad X = \underline{X_1 + \cdots + X_n}$, $X_i \sim$ i.i.d. $\text{Bern}(p)$.

② $\overset{\text{PGF}}{g_X(t)} = E[t^X] = E[t^{X_1+\cdots+X_n}] = (E[t^{X_1}])^n$.

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

$\Rightarrow \; g_X(t) = \underline{(pt+q)^n}$

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

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

Binomial Moments

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

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

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

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

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

③ $$\underline{E\left[\binom{X}{k}\right] = \binom{n}{k}p^k, \quad k\geq 2} \checkmark$$

Example: Pattern Matching

$q = 1-p$.

T H H
1 2 3

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

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

② First-Step Method. H , T .

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 $Var(N)$.

$k \geq 3$; 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

S_2 :

1 2 3 . . . . k

H T $\underbrace{\qquad\qquad}_{k-2 \text{ tosses to get HH for the first time}}$

③ $P(N=k, S_1=H)$

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

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

1 2 3 . . . K .

T $\underbrace{\qquad\qquad}_{k-1 \text{ tosses to get HH for the first time.}}$

$P(N=k, S_1=T)$

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

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

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

$\boxed{\text{Find } p_k.}$

$E(N)$; $Var(N)$

Example: Pattern Matching

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

④ PGF of N; $g(t)=E[t^N] = \sum_{k=0}^{\infty} p_k\cdot t^k = \sum_{k=1}^{\infty} p_k\cdot t^k = \sum_{k=2}^{\infty} p_k\cdot t^k$

$$= p_2\cdot t^2 + \sum_{k=3}^{\infty} p_k\cdot t^k \quad = p^2t^2 + \sum_{k=3}^{\infty} p_k t^k = g(t)$$

On the other hand; $p_k = p_{k-1}\cdot q + p_{k-2}\cdot pq$ $(k\geq 3)$

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

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

$$= qt\sum_{k=2}^{\infty} p_k t^k + pqt^2\sum_{k=1}^{\infty} p_k t^k$$

$$= qt\cdot g(t) + pqt^2\cdot g(t) = (qt+pqt^2)\, g(t)$$

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

Example: Pattern Matching

⑤ PGF of N. $g_N(t) = g(t) = \boxed{\dfrac{p^2t^2}{1-qt-pqt^2}}$

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

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

⑥ $p = \frac{1}{2}$ fair coin ; $E[N] = 6 > 4$.

$Var[N] = 22$

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.

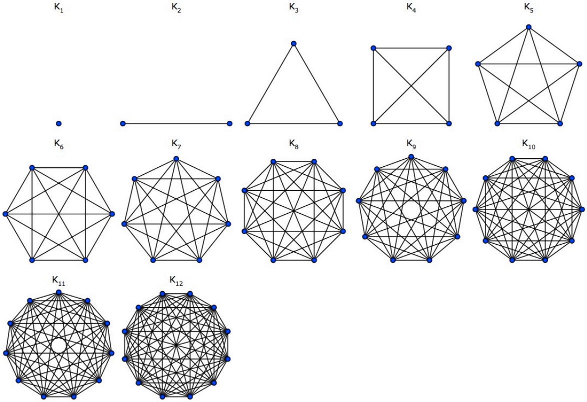
Principle II

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 $M \in F(x_1, x_2, \ldots, x_n)$ 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 $r_1, r_2, \ldots, r_n$ be selected at random independently and uniformly from S. Then

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

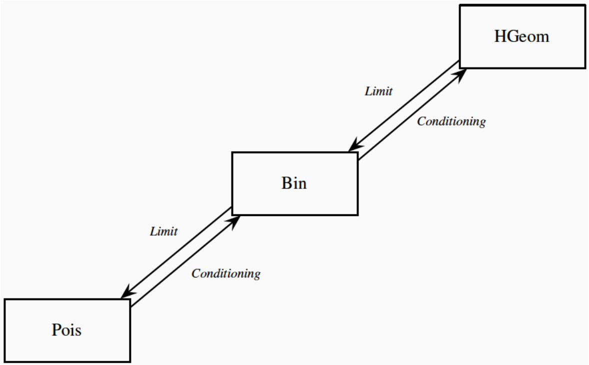
Remarks

- 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.

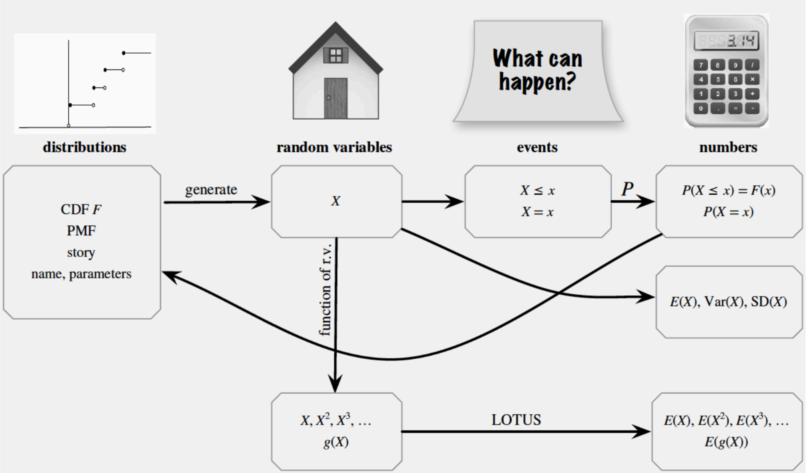
Summary 1

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

Summary 2



Summary 3


distributions
random variables
What can happen?
events
numbers
CDF F
PMF
story
name, parameters
generate
X
$X \leq x$
$X = x$
P
$P(X \leq x) = F(x)$
$P(X = x)$
function of r.v.
$E(X)$, Var(X), 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**