Seoul National University

Discrete Random Variable

Date

Name: Chong-kwon Kim

SCONE Lab.

To Learn

- Concept of random variable
- Expectation
- Conditional expectation
- Several important discrete random variables (distribution)
  - Bernoulli
  - Binomial
  - Geometric
  - Poisson

Random Variable

- A **Random Variable** X is a real-valued function defined on sample space

  $$X: \Omega \rightarrow \mathbf{R}$$

- Discrete random variable
  - Takes finite or countably infinite number of values
- For a discrete rv X and value *a*
  - "X=*a*" is a set of the basic events in the sample space in which X is *a*
  - Set $\{s \in \Omega \mid X(s) = a\}$
  - $\Pr(X = a) = \sum_{s: X(s)=a} \Pr(s)$

Examples

- Flip a coin three times
  - Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
  - Define X = Number of Heads in the three trials

X(HHH) = 3
X(HTH) = 2

  - X = 1, {HTT, THT, TTH} ➔ Pr(X = 1)=3/8
  - X ≤ 1, {TTT, HTT, THT, TTH} ➔ Pr(X ≤ 1)=1/2
  - On the same sample space, we define X= # Heads - # Tails

X(HHH) = ?
X(HTH) = ?

  - X=-1, {HTT, TTH, THT} ➔ Pr(X=-1)=3/8

Examples

- Coin flips, X = Number of flips until the first heads

  H ➔ X = 1

  TH ➔ X = 2

  TTH ➔ X = 3

  …

  Pr(X=n) = ?

- Flip a coin N times, X = Number of heads in N trials

  HTTH ➔ X = 2

  Pr(X=k) = ?

- # babies born in a day, X = Number of babies born on March 19

  Pr(X=k) = ?

Independent Random Variable

- Definition: Two random variables X and Y are independent iff

  $\Pr((X=a) \cap (Y=b)) = \Pr(X=a)\Pr(Y=b)$ for all a and b

- Random variables $X_1, X_2, \ldots, X_k$ are independent iff

  for all subset $I \subseteq [1,k]$ and any values x_i, $i \in I$

  $\Pr(\bigcap_{i \in I}(X_i = x_i)) = \prod_{i \in I} \Pr(X_i = x_i)$

Expectation

- E[X]: Expectation of a rv X

  $\mathrm{E}[X] = \sum_i x_i \cdot \Pr(X = x_i)$

  - Weighted average of values that it assumes
  - Weight: probability that the rv assumes the value

- Examples
  - Flip a coin three times
  - Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
  - Define X = Number of Heads
  - E[X] = 0·Pr(X=0) + 1·Pr(X=1) + 2·Pr(X=2) + 3·Pr(X=3)
  - On the same sample space, we define X= # Heads - # Tails
  - E[X] = -3·Pr(X=-3) + … + 3·Pr(X=3)

Notations:
p(a) = Pr(X=a),
p_i= Pr(X=x_i)

Beating Casinos

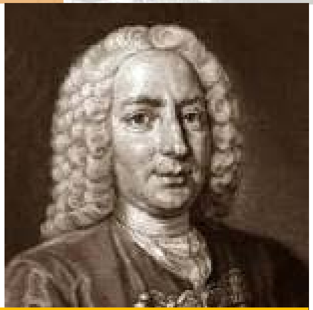
- One famous strategy to beat Casinos is **double betting**
  - Suppose you win \$Y with probability 2/5 and lose \$Y with 3/5 probability
  - Start from Y=1, every time you lose, double the bet
    1. Y=\$1
    2. Bet Y
    3. If Win, Stop
    4. If Loss, Y=2*Y and goto 2
  - Z: Result at the stop
  - E[Z]=(2/5)1 + (3/5)(2/5)(2-1) + (3/5)(3/5)(2/5)(4-2-1) + …

$$= \sum_{i=0}^{\infty} \left(\frac{3}{5}\right)^{i} \cdot \left(\frac{2}{5}\right) \cdot 1$$

- $E[Z] \geq 0$

Unbounded Expectation

Several (Most) random variables have bounded expectations
Some has unbounded expectations and/or variances
Ex: **Power Law** distribution



Daniel Bernoulli was a Dutch born Swiss mathematician, one of many in his family.

- St. Petersburg Paradox (By Daniel or Nicolas Bernoulli)
  - A player flips a fair coin repeatedly until the first tails comes up
  - If the first tails comes up at the i-th flip, then the player receives $\$2^i$
  - How much will you pay to enter the game?
  - X:Your winnings
  - E[X] = $(1/2) \cdot 2^1 + (1/2)^2 \cdot 2^2 + (1/2)^3 \cdot 2^3 + \cdots$
    $= \sum 1 = \infty$

Expansion

- Let X = # Heads - # Tails in flipping a fair coin three times
  - Pr(X = -3) = 1/8, Pr (X = -1) = 3/8, Pr (X = 1) = 3/8, Pr(X = 3) = 1/8
  - Compute E[X^2]
- One solution **E[X^2] = $\sum_i x_i{}^2$Pr$(X = x_i)$**
  - E[X^2] = $(-3)^2$Pr(X=-3) + $(-1)^2$Pr(X=-1) +1^2Pr(X=1) + 3^2Pr(X=3)

    =
- Another solution
  - Let Y = X^2

Y: Another Random Variable, (# Heads - # Tails)^2
Y = 1 ➔ {TTH, THT, HTT, HHT, HTH, THH}
Y = 9 ➔ { TTT, HHH}

Pr(Y=1) = Pr(X=-1) + Pr(X=1)
Pr(Y=9) = Pr(X=-3) + Pr(X=3)

E[Y] = $\sum_i y_i$ Pr$(Y = y_i)$

  - E[X^2] = E[Y] = 1·Pr(Y=1) + 9·Pr(Y=9)

    =

Note that E[X] = 0 and E[X^2]≠E$[X]^2$

Expansion

- Let Y=g(X), where g() is a real-valued function
- E[g(X)]=E[Y] = $\sum_j y_j \cdot (\Pr(Y = y_j))$

$$= \sum_j y_j \cdot (\sum_{i:g(x_i)=y_j} \Pr(x_i))$$

$$= \sum_j \sum y_j \Pr(x_i)$$

$$= \sum_j \sum g(x_i) \Pr(x_i)$$

Reconsider rv X in the previous slide
Define g(X) = $X^2 + X$
Compute E[g(X)]

- For any constant E[c·X] = c · E[X]
- n-th **moment** of X:

$$E[X^n] = \sum_i {x_i}^n Pr(X = x_i)$$

Linearity of Expectation

- For any finite collection of discrete rv $X_1, X_2, \cdots, X_n$

  $E[\sum_{i=1}^{n} X_i]$ = $\sum_{i=i}^{n} E[X_i]$

- Proof
  - For two rv X and Y, prove that

    $E[X+Y] = E[X] + E[Y]$
  - $E[X+Y] = \sum_i \sum_j (i+j) \cdot$ Pr$((X=i) \cap (Y=j))$

    $=$

Jensen's Inequality

- In general, E[X^2] ≠ E[X]^2
- Claim: E[X^2] ≥ E[X]^2
- Proof
  - Consider Y= $(X - E[X])^2$
  - 0 ≤ E[Y] = E[$(X - E[X])^2$]

    = E[X^2 - 2XE[X] + E[X]^2]

    = E[X^2] - E[X]^2

- Definition: Convex
  - A function f is convex if, for any x1 and x2 and 0≤λ≤1,

$$f(\lambda \cdot x_1 + (1-\lambda) \cdot x_2) \leq \lambda \cdot f(x_1) + (1-\lambda) \cdot f(x_2)$$

Convex function & Optimization
Optimization: Another important technique
Generally, we can easily find optimal points if functions are convex

Jensen's Inequality

- Theorem: If f is convex, then E[f(X)] $\geq$ f(E[X])

- Proof
  - Let $\mu = E[X]$
  - By Taylor's theorem, there is c such that

$$f(x) = f(\mu) + f'(\mu)\cdot(x-\mu) + \frac{f''(c)(x-\mu)^2}{2}$$
$$\geq f(\mu) + f'(\mu)\cdot(x-\mu)$$

Lemma: If f is convex, then $f''(x) \geq 0$

  - $E[f(X)] \geq E[f(\mu) + f'(\mu)\cdot(X-\mu)]$
    $= f(\mu) = f(E[X])$

Conditional Expectation

- Definition: E[Y | Z=z] = $\sum_y y \cdot \Pr(Y = y \mid Z = z)$

- Example:
  - Roll two dice
  - X_1: Number on the first die
  - X_2: Number on the second die
  - X: X_1+X_2
  - $E[X \mid X_1=2] = \sum_{x_2=1}^{6}(x_1+x_2)\Pr(x_2 \mid x_1 = 2)$
    $= \sum_{x_2=1}^{6}(2+x_2)\cdot\frac{1}{6}$
  - $E[X_1 \mid X=5] = \sum_{x_1} x_1 \Pr(X_1 = x_1 \mid X_1 + X_2 = 5)$
    $= \sum_{x_1=1}^{4} x_1 \Pr(X_1 = x_1 \mid X_1 + X_2 = 5)$
    $= \sum_{x_1=1}^{4} x_1 \frac{\Pr((X_1=x_1)\cap(X_1+X_2=5))}{\Pr(X_1+X_2=5)}$

Properties of Conditional Expectation

- Lemma: For any random variables X and Y,

$$\mathbf{E[X]} = \sum_y \mathbf{Pr}(Y = y) \cdot E[X|Y = y]$$

Important lemma
In many cases, E[X|Y=y] is easier to compute than E[X]

- Proof:
  - $E[X] = \sum_i x_i \cdot \Pr(X = x_i)$
    $= \sum_i x_i \cdot \sum_y \Pr(X = x_i | Y = y) \cdot \Pr(Y = y)$
    $= \sum_y \boxed{\sum_i x_i \cdot \Pr(X = x_i | Y = y)} \cdot \Pr(Y = y)$
    $\equiv E[X|Y = y]$

- **Linearity**: For any finite collection of rv X1, X2,…, Xn, and for any random variable Y,

$$E\left[\sum_i X_i \,|Y = y\right] = \sum_i E[X_i | Y = y]$$

RV Conditional Expectation

- Definition: Expression **E[Y | Z] is a r.v. f(Z)** that takes on the value E[Y | Z=z] when Z=z

Roll two dice
X_1: Number on the first die
X_2: Number on the second die
X: X_1+X_2

- Example
  - $\mathrm{E}[\mathrm{X}|X_1] = \sum_{x_2}(X_1 + x_2) \cdot \Pr(X = X_1 + x_2|X_1)$
    $= X_1 + \sum_{x_2} x_2 \cdot \Pr(X = X_1 + x_2|X_1)$
    $= X_1 + \frac{7}{2}$
  - Now $E[E[X|X_1]] = E[X_1 + 7/2] = E[X_1] + 7/2$

- Theorem: E[Y] = E[E[Y | Z]}
- Proof:
  - $\mathrm{E}[\mathrm{Y}|\mathrm{Z}] = \sum_i y_i \cdot \Pr(Y = y_i|Z)$
  - $\mathrm{E}[\mathrm{E}[\mathrm{Y}|\mathrm{Z}]] = \sum_j (\sum_i y_i \cdot \Pr\left(Y = y_i \middle| Z = z_j\right)) \cdot \Pr(Z = z_j)$
    $= \sum_j E[Y|Z = z_j] \cdot \Pr(Z = z_j)$
    $= E[Y]$

Bernoulli & Binomial RV

- Run an experiment
  - Success probability = p and Failure probability = (1-p)
- **Bernoulli (Indicator)** random variable Y is
  - Y = $\begin{cases} 1, & if\ success \\ 0, & if\ failure \end{cases}$
- E[Y] = p = Pr(Y=1)
- Now, perform the experiment n times. Random Variable X= the number of successes in n experiments
- Definition: **Binomial** random variable X with parameter n and p, **B(n,p)**, is

Pr(X=j) = $\binom{n}{j} \cdot p^j (1-p)^{n-j}$

E[X] of Binomial RV

- First prove that $\sum_{i=0}^{n} \Pr(X = i) = 1$
- $E[X] = \sum i \cdot \Pr(X = i)$

  $= \sum i \cdot \binom{n}{i} \cdot p^i (1-p)^{n-i}$

- Another method
  - $X = X_1 + X_2 + \cdots + X_n$ where X_i is the indicator function (Bernoulli rv) of i-th experiment

Geometric Distribution

- X: # coin flips until the first heads
- Definition: A **Geometric** random variable X with parameter p is given by the following probability distribution for n=1, 2,…

  Pr(X=n) = $(1-p)^{n-1} \cdot p$
- Properties
  - $\sum_{n\geq 1}$ Pr(X=n) = 1
  - **Memoryless property**: Given you tried k times w/o heads, how many more trials until the first success?
- Lemma: Pr(X=n+k | X>k) = Pr(X=n)
- Proof
  - Pr(X=n+k | X>k) = $\frac{\Pr(X=n+k \cap X>k)}{\Pr(X>k)}$

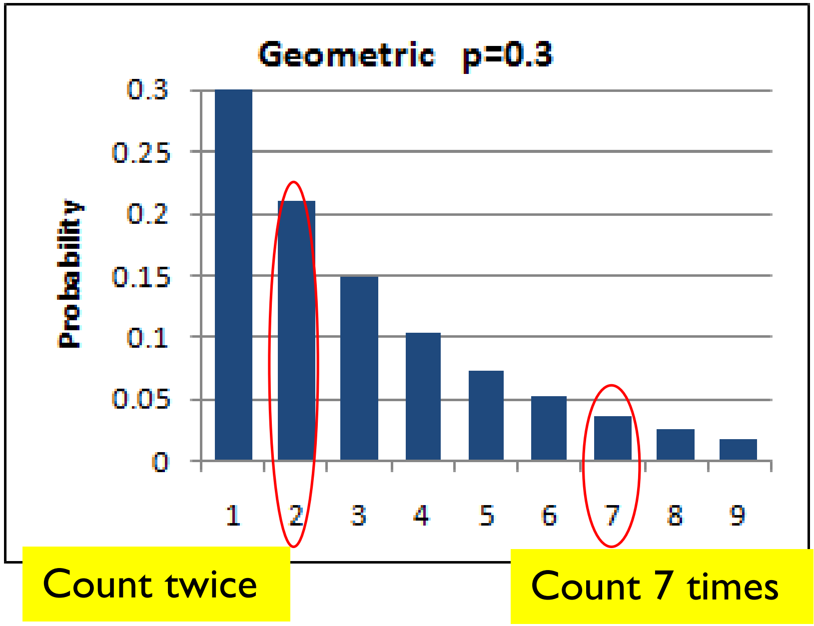
    $= \frac{(1-p)^{n+k-1} \cdot p}{\sum_{i=k}(1-p)^{i} \cdot p}$

Geometric - Expectation

- Claim: E[X] = $\sum_{i=1}^{\infty} \Pr(X \geq i)$
- Proof:
  - $\sum_{i=1}^{\infty} \Pr(X \geq i) = \sum_{i=1}^{\infty} \mathrm{i} \cdot \Pr(X = i)$
    $= E[X]$

- Note $\Pr(X \geq i) = \sum_{n=i}^{\infty} (1-p)^{n-1} \cdot p$
  $= (1-p)^{i-1}$

  ➔ E[X] = $\sum_{i=1}^{\infty} (1-p)^{i-1}$ = ${}^{1}/_{p}$



Geometric - Expectation

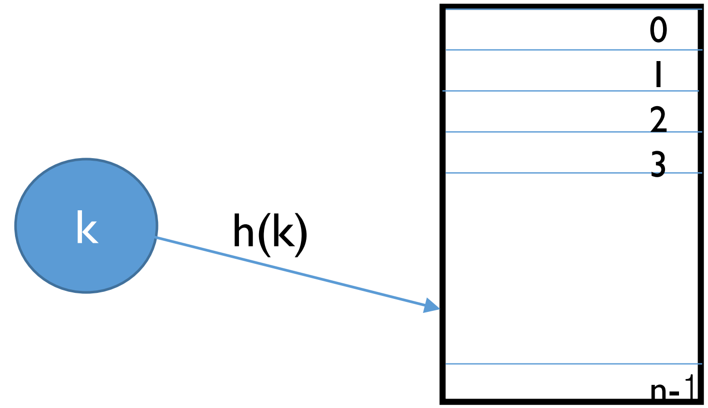
- Another Approach to Compute E[X]
  - Remember: E[X] = E[E[X|Y]]
  - Y: result of the first flip = {0, 1}

  - $E[X] = E[X \mid Y=0]\, Pr(Y=0) +$
    $E[X \mid Y=1]\, Pr(Y=1)$
    $= E[X+1] \cdot (1-p) + 1 \cdot p$

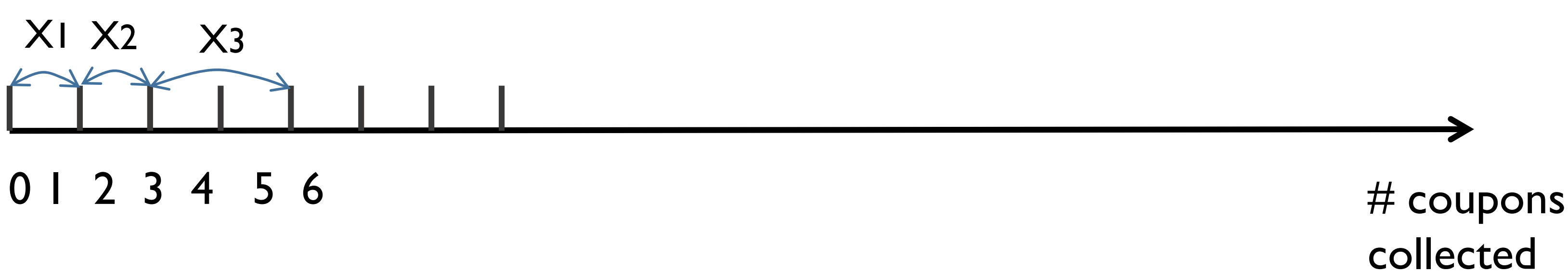
  ➔ $E[X] = {}^{1}/_{p}$

Coupon Collector's Problem

- Setting
  - There are N different types of coupon
  - Receive a coupon that is any one of N types
  - Any similar problems?
    ➔ Exactly same as "Hash Table"



- Interested in random variable T: # coupons need to be collected until at least one from every type of coupon is collected
  - E[T]??
  - Xi: Given that (i-1) types of coupon are collected, how many more to collect to obtain the i-th type



Coupon Collector's Problem

- Clearly, T = $X_1 + X_2 + \dots + X_N$
  - Xi: Geometric r. v. with $p_i = (1 - (i-1)/N)$
  - $E[X_i] = 1/p_i = N / (N-i+1)$
  - $E[T] = \sum_i E[X_i]$
    $= \sum_i \frac{N}{N-i+1}$
    $= N \cdot \sum_i {}^{1}/_{i}$

Harmonic number H(N) = ln N + Θ(1)

- Another Approach
  - Collect n coupons
  - Ai: Type i is not included in the n coupons
  - $Pr(A_i) = (\frac{N-1}{N})^n$
  - $Pr(T>n) = Pr(\cup_{j=1}^{N} A_j)$
    $= \dots$
    $= \sum_i^{N-1} \binom{N}{i} (\frac{N-i}{N})^n (-1)^{i+1}$

A_{j_1} and A_{j_2} indpendent?
No!!
$Pr(A_{j_1} \cap A_{j_2}) = (\frac{N-2}{N})^n$
$Pr(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_k}) = (\frac{N-k}{N})^n$

Coupon Collector's Problem

- … Continue

  Now, Pr(T=n) = Pr(T>n-1) - Pr(T>n)

- Another interesting random variable, Dn: # coupon types covered by n coupons
  - Pr(Dn=k)
  - Fix k types
  - Define A: each coupon is one of these k types, and

    B: each of these k types is represented
  - Pr(A) = $(\frac{k}{N})^n$
  - Now consider Pr(B | A): Same as probability Pr(T $\leq$ n) with k replacing N
  - Pr(B | A) = 1-$\sum_i^{k-1}\binom{k}{i}(\frac{k-i}{k})^n(-1)^{i+1}$
  - Pr(Dn=k) = $\binom{N}{k}$Pr(A$\cap$B) = $\binom{N}{k}$Pr(B | A) Pr(A)

QuickSort

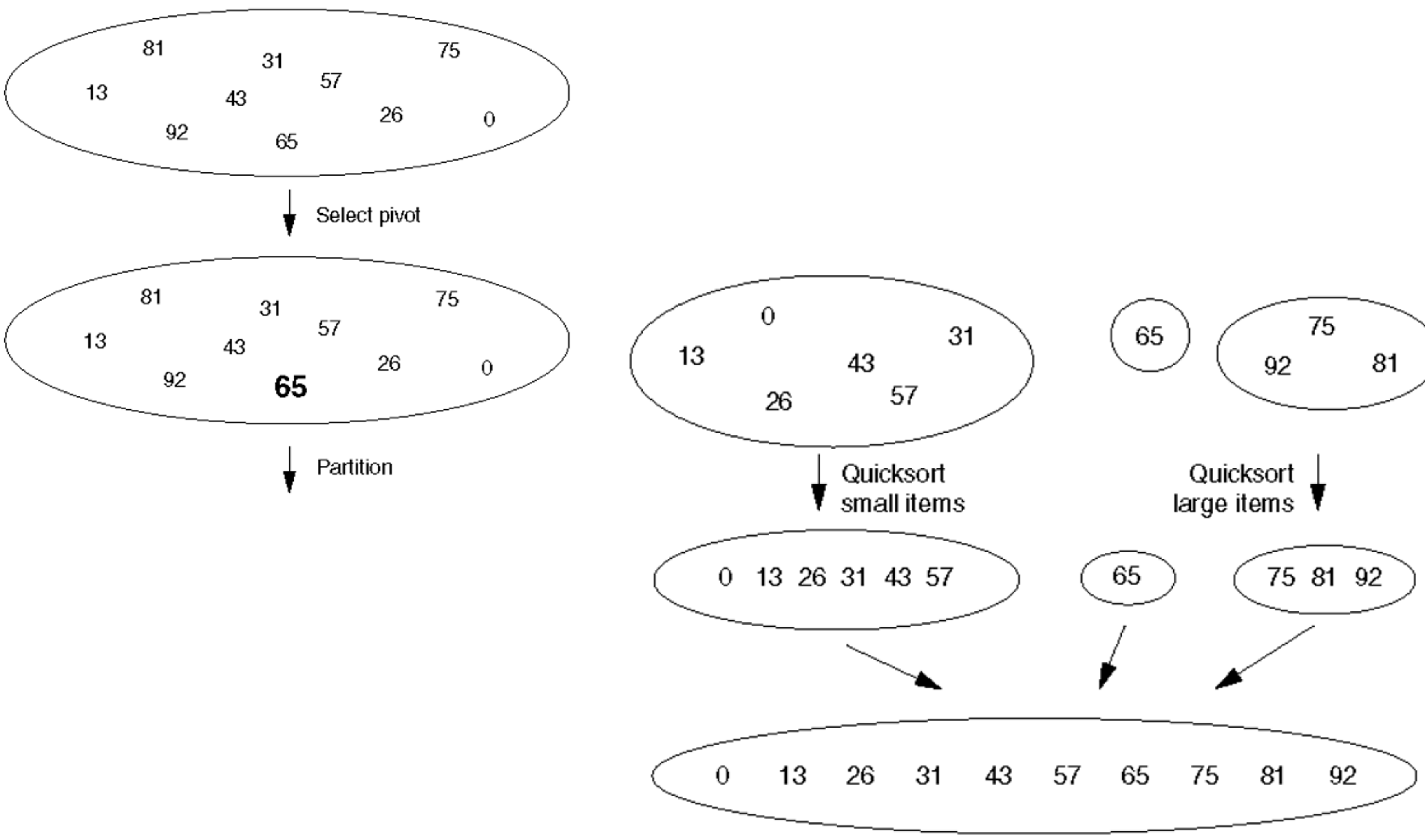
- Sorting problem
  - Given n comparable objects $x_1, x_2, \ldots, x_n$, arrange them in increasing order
    ➔ Let sorted result is $y_1, y_2, \ldots, y_n$

QuickSort Algorithm

Given objects $x_1, x_2, \ldots, x_n$

1. Pick a pivot element x_t, $1 \le t \le n$
2. Partition on x_t
   $S1 = \{x_i : x_i \le x_t\}$
   $S2 = \{x_i : x_i > x_t\}$
   $S1 \le x_t < S2$
3. Sort S1 & S2, respectively
4. Combine
   $y_1, y_2, \ldots, y_p, x_t, z_1, z_2, \ldots, z_q$

QuickSort - Example



Complexity of QuickSort

- Complexity of quick sort
  - $T(N) = T(|S1|) + T(|S2|) + O(N)$
  - Running time depends on the choice of the pivot
  - Worst case
    - $T(N) = T(N-1) + O(N)$
      $= O(N^2)$
  - Best case
    - $T(N) = 2\,T(N/2) + O(N)$
      $= O(N \log N)$
- Average case analysis (Probabilistic Analysis)
  - All N! permutations of the sorted order are equally likely
  - Always pick an element with a fixed index, say x_1, as a pivot
    - P_i = probability that x_1 is the i-th element in the sorted order
      $= 1/N$
    - C_n = Average number of operations for sorting a table of size n
      $= 1/N \sum(C_{i-1} + C_{n-i}) + a N$
      $= 2/N \sum C_i + a N$
      $= O(N\log N)$

Randomized QuickSort

- Randomized Algorithm
  - Select pivot numbers uniformly at random among the candidates
- Theorem: For any input, the expected number of comparisons made by randomized QuickSort is $2n \cdot \ln n + \Theta(n)$
- Proof
  - Let $y_1, y_2, \ldots, y_n$ be the sorted sequence
  - For $i < j$, define random variable X_{ij} such that
  - $X_{ij} = \begin{cases} 1, & \text{if } yi \text{ and } yj \text{ are compared} \\ 0, & ow \end{cases}$
  - Total number of comparisons $X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$
  - Pr(Xij) ??
  - $E[X] = E[\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}]$

    $= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[X_{ij}]$

What is the probability that y1 and yn are compared?
How about yi and yi+1?

Randomized QuickSort

– $\Pr(X_{ij} = 1) = \frac{2}{(j-i+1)}$

– $E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{(j-i+1)}$

$=$