Discrete Probability

No.
Date: / /

- 統計：描述過去

$N=50$ $\quad f_k = (\# x_i = k)$

$x_1 = 30$ $\quad f_0 = 2$
$x_2 = 50$ $\quad f_{10} = 3$
$x_i = 10$ $\quad f_{30} = 15$
$x_{50} = 30$ $\quad \vdots$
$\quad f_{100} = 1$

- 機率：預測未來

$p_k = \frac{f_k}{N} = P(X=k)$

- 平均值：$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i = \sum_{k=0}^{100} \frac{f_k}{N} k = \sum_{k=0}^{100} p_k k = E(X)$ ：期望值
- 標準差²：$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 = \sum_{k=0}^{100}\frac{f_k}{N}(k-\mu)^2 = \sum_{k=0}^{100} p_k(k-\mu)^2 = E((X-\mu)^2) = Var(X)$：變異數

$$= \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu^2 = \sum_k \frac{f_k}{N}k^2 - \mu^2 = \sum_k p_k k^2 - \mu^2 = E(X^2)-\mu^2$$

- Markov 不等式：$P(X \geq c\mu) \leq \frac{1}{c}$

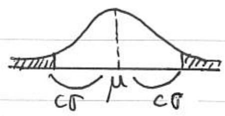
Chebyshev 不等式：$P((X-\mu)^2 \geq c^2\sigma^2) \leq \frac{1}{c^2}$

$P(|X-\mu| \geq c\sigma) \leq \frac{1}{c^2}$

$\mu = 30$分，50人 (總分$=1500$)

≥ 60分，≤ 25人，≥ 90分，$\leq \frac{50}{3}$人

$\geq c\mu$，$\leq \frac{N}{c}$



Probability Space (Ω, P, X)

- Sample space：$\Omega = \{\omega_1, \omega_2, \cdots\}$
  Probability：$P: \Omega \to [0,1]$，$\sum_{\omega\in\Omega} P(\omega) = 1$
  Random variable：$X: \Omega \to \mathbb{R}$，

Ω	⚀	⚁	⚂	⚃	⚄	⚅
P	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$
X	1	2	3	4	5	6
Y	5	-2	5	-2	5	-2
X^2	1	4	9	16	25	36

- $E(X) = \sum_{\omega\in\Omega} X(\omega)P(\omega) = \sum_{k\in X(\Omega)} k\,P(X=k) = \mu_X$
  $Var(X) = E((X-\mu)^2) = E(X^2) - \mu^2 = \sigma_X^2$

$E(X) = 1\frac{1}{6} + 2\frac{1}{4} + 3\frac{1}{12} + 4\frac{1}{4} + 5\frac{1}{6} + 6\frac{1}{12} = \frac{39}{12}$

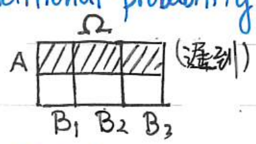
$E(Y) = 5\frac{1}{6} - 2\frac{1}{4} + 5\frac{1}{12} - 2\frac{1}{4} + 5\frac{1}{6} - 2\frac{1}{12} = \frac{11}{12}$

$= 5\left(\frac{1}{6}+\frac{1}{12}+\frac{1}{6}\right) - 2\left(\frac{1}{4}+\frac{1}{4}+\frac{1}{12}\right)$

- Event：$A \subset \Omega$，$P(A) = \sum_{\omega\in A} P(\omega)$

$P(X=a) = P(\{\omega\in\Omega \mid X(\omega)=a\}) = P(X^{-1}(a))$

- Conditional probability
  $P(A|B) = \frac{P(A\cap B)}{P(B)}$ $\quad \begin{cases} A = X^{-1}(a) \\ B = Y^{-1}(b) \end{cases}$
  $P(X=a|Y=b) = \frac{P(X=a, Y=b)}{P(Y=b)}$



$Var(X) = \left(1-\frac{39}{12}\right)^2\frac{1}{6} + \cdots + \left(6-\frac{39}{12}\right)^2\frac{1}{12} = \frac{363}{144}$

$= \left(1^2\frac{1}{6} + 2^2\frac{1}{4} + \cdots + 6^2\frac{1}{12}\right) - \left(\frac{39}{12}\right)^2 = //$

$G_X(z) = \frac{1}{6}z + \frac{1}{4}z^2 + \frac{1}{12}z^3 + \cdots + \frac{1}{12}z^6$

- A, B indep：$P(A\cap B) = P(A)P(B) \Leftrightarrow P(A|B) = P(A)$
  X, Y indep：$P(X=a, Y=b) = P(X=a)P(Y=b)$

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$

Y	5	-2
P	$\frac{5}{12}$	$\frac{7}{12}$

- (全機率) $P(A) = \sum_i P(B_i)P(A|B_i)$
  (貝氏定理) $P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_i P(B_i)P(A|B_i)}$

X	x_0 x_1 x_2 $\cdots$
P	p_0 p_1 p_2 $\cdots$

$X(\Omega)=\{0,1,2,\cdots\}$ $\Longrightarrow$

X	0 1 2 $\cdots$ k $\cdots$
P	p_0 p_1 p_2 $\cdots$ p_k $\cdots$

$P(X=k)=p_k,\quad k=0,1,2,\cdots$

- $$\begin{cases}\mu = E(X) = \sum_{\omega\in\Omega} X(\omega)P(\omega) = \sum_{k\in X(\Omega)} k\,P(X=k) = G'(1) = M'(0)\\ \sigma^2 = Var(X) = E\left((X-\mu)^2\right) = E(X^2)-\mu^2 = G''(1)+G'(1)-\mu^2 = M''(0)-\mu^2\end{cases}$$

- $$\begin{cases}\text{probability g.f.}\quad G_X(z) = E(z^X) = \sum_{\omega\in\Omega} P(\omega)\, z^{X(\omega)} = \sum_{k\in X(\Omega)} p_k z^k \qquad \left(p_k = P(X=k)\right)\\ \text{moment g.f.}\quad M_X(t) = G_X(e^t) = \sum_{k\in X(\Omega)} p_k e^{tk} \qquad G_X(1)=M_X(0)=1\end{cases}$$

- $G_X(z) = \sum_{b\in Y(\Omega)} P(Y=b)\, G_{X|b}(z)$

$\because G_X(z) = \sum_{k\in X(\Omega)} \underbrace{\sum_{b\in Y(\Omega)} P(Y=b)P(X=k|Y=b)}_{p_k\ (全機率)} z^k = \sum_{b\in Y(\Omega)} P(Y=b)\left(\sum_{k\in X(\Omega)} P(X=k|Y=b)\, z^k\right)$

(8.92)

- **定理 1** $X,Y:\Omega\to\mathbb{R}$

(1) $E(c) = c$ (2) $Var(c) = 0$

(3) $E(X+Y)=E(X)+E(Y)$ (4) $Var(X+Y) = Var(X)+Var(Y)+2\,Cov(X,Y)$

(5) $E(cX) = c\,E(X)$ (6) $Var(cX) = c^2 Var(X)$

$Cov(X,Y) = E\left((X-\mu_X)(Y-\mu_Y)\right) = E(XY)-\mu_X\mu_Y$

證 (3) $E(X+Y) = \sum_{\omega\in X(\Omega)} \left(X(\omega)+Y(\omega)\right)P(\omega) = \sum_\omega X(\omega)P(\omega) + \sum_\omega Y(\omega)P(\omega) = E(X)+E(Y)$

(4) $Var(X+Y) = E\left((X+Y)^2\right) - (\mu_X+\mu_Y)^2 = E\left(X^2+2XY+Y^2\right) - \left(\mu_X^2+2\mu_X\mu_Y+\mu_Y^2\right)$

- **定理 2** X,Y indep

(1) $E(XY) = E(X)E(Y) \Rightarrow Cov(X,Y)=0$

(2) $Var(X+Y) = Var(X)+Var(Y)$

(3) $\begin{cases}G_{X+Y}(z) = G_X(z)\,G_Y(z)\\ M_{X+Y}(t) = M_X(t)\,M_Y(t)\end{cases}$

證 $E(XY) = \sum_{\substack{a\in X(\Omega)\\ b\in Y(\Omega)}} ab\, \overbrace{P(X=a, Y=b)}^{P(X=a)P(Y=b)}$

$= \sum_{a\in X(\Omega)} a\,P(X=a) \sum_{b\in Y(\Omega)} b\,P(Y=b)$

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

常見 離散隨機分布	X: x_0 x_1 $\cdots$ x_k $\cdots$ P: p_0 p_1 $\cdots$ p_k $\cdots$	μ	σ^2	(1) $P(X=x_k)=p_k\geq 0$ (2) $\sum_{k=0}^{\infty} p_k = 1$
Bernoulli 分布 $X\sim Be(p)$	X: 0 1 P: q p $\quad (q=1-p)$	p	pq	$X=\begin{cases}1 & 成功\\ 0 & 失敗\end{cases}$ (只有兩種結果)
二項分布 $X\sim B(n,p)$	$P(X=k)=\binom{n}{k}p^k q^{n-k},\ k=0,1,\ldots,n$	np	npq	(1) $X=X_1+X_2+\cdots+X_n$ (2) i.i.d. $X_i\sim Be(p)$
Poisson 分布 $X\sim P(\lambda)$	$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!},\ k=0,1,2,\ldots$	λ	λ	(1) n 較大 (≥ 20), p 較小 (≤ 0.05) (2) $\lambda\approx np$ (3) $\binom{n}{k}p^k q^{n-k}\approx e^{-\lambda}\frac{\lambda^k}{k!}$
幾何分布 $X\sim G(p)$	$P(X=k)=q^{k-1}p,\ k=1,2,\ldots$	$\frac{1}{p}$	$\frac{q}{p^2}$	$k=$ 首次成功次數

- **Example 1** (Bernoulli), Flipping a coin, #H = ?

解

Ω	T	H
X	0	1
P	q	p

$(q = 1-p)$

(甲) $\begin{cases} E(X) = 0\cdot q + 1\cdot p = p, & E(X^2) = 0^2 q + 1^2\cdot p = p \\ Var(X) = E(X^2) - \mu^2 = p - p^2 = pq \end{cases}$

(乙) $G_X(z) = q + pz$, $\begin{cases} G'(z) = p, & E(X) = G'(1) = p \\ G''(z) = 0, & Var(X) = G''(1) + G'(1) - \mu^2 = p - p^2 \end{cases}$

- **Example 2** (Binomial) Flipping n coins, #H = ?

解

Ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
X	3	2	2	1	2	1	1	0
P	p^3	p^2q	p^2q	pq^2	p^2q	pq^2	pq^2	q^3

$p_k = P(X=k) = \binom{n}{k} p^k q^{n-k}$, $k = 0, 1, \cdots, n$

$\Rightarrow$

X	0	1	2	3
P	q^3	$\binom{3}{1}pq^2$	$\binom{3}{2}p^2q$	p^3

(甲) $G_X(z) = \sum_{k\geq 0} p_k z^k = \sum_{0\leq k\leq n} \binom{n}{k} p^k q^{n-k} z^k = (pz+q)^n, \xleftarrow[T \to q]{H \to pz} \Omega = (H+T)^n$

$\begin{cases} G'(z) = np(pz+q)^{n-1} \\ G''(z) = n(n-1)p^2(pz+q)^{n-2} \end{cases}$ $\begin{cases} E(X) = G'(1) = np \\ Var(X) = G''(1) + G'(1) - \mu^2 = npq \end{cases}$

(乙) $\begin{cases} X_i : \text{indep} \\ X_i \sim Be(p) \end{cases} \Rightarrow \begin{cases} E(X) = \sum_{i=1}^{n} E(X_i) = np, \\ Var(X) = \sum_{i=1}^{n} Var(X_i) = npq, \end{cases}$ $G_X(z) = G_{X_1}(z)\cdot G_{X_2}(z)\cdots G_{X_n}(z) = (q+pz)^n$

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

- **Example 3** (Geometric) # tosses to get H ?

解

Ω	$\{H,$	$TH,$	$T^2H,$	$\cdots,$	$T^{k-1}H,$	$\cdots\}$
X	1	2	3	$\cdots$	k	$\cdots$
P	p	qp	q^2p	$\cdots$	$q^{k-1}p$	$\cdots$

$= T^*H = \sum_{k\geq 1} T^{k-1}H$, $\begin{cases} H \to pz \\ T \to qz \end{cases}$

$p_k = P(X=k) = q^{k-1}p$, $k = 1, 2, \cdots$

$G_X(z) = pz + qpz^2 + q^2pz^3 + \cdots + q^{k-1}pz^k + \cdots = \frac{pz}{1-qz}$ $\begin{cases} \mu = \frac{1}{p} \\ \sigma^2 = \frac{q}{p^2} \end{cases}$

- **Example 4** # tosses to get HH ?

解 (甲) $\Omega = (T+HT)^*HH = \sum_{k=0}^{\infty} (T+HT)^k HH \xrightarrow[T \to qz]{H \to pz} \sum_{k=0}^{\infty} (qz + pqz^2)^k p^2 z^2 = \frac{p^2z^2}{1 - qz - pqz^2} = S(z)$

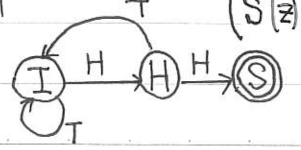
(乙) $\begin{cases} Ⓘ = \{1 + T + HT + \cdots\} \\ \quad = ⒾT + ⒽT + 1 \\ Ⓗ = ⒾH \\ Ⓢ = ⒽH \end{cases}$

(丙) n tosses

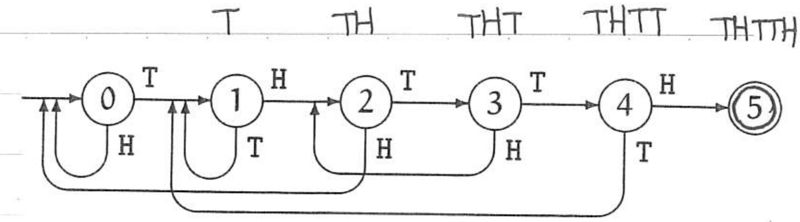
$\begin{cases} I_n = P(\text{at } I) = q I_{n-1} + q H_{n-1} + \delta_{n=0} \\ H_n = P(\text{〃} H) = p I_{n-1} \\ S_n = P(\text{〃} S) = p H_{n-1} \end{cases}$ $\begin{cases} I(z) = \sum I_n z^n \\ H(z) = \sum H_n z^n \\ S(z) = \sum_n S_n z^n \end{cases}$

$\xrightarrow[T \to qz]{H \to pz}$ $\begin{cases} I(z) = I(z)qz + H(z)qz + 1 \\ H(z) = I(z)pz \\ S(z) = H(z)pz \end{cases}$ $\Rightarrow S(z) = \frac{p^2z^2}{1 - qz - pqz^2}$

$\begin{cases} \mu = p^{-2} + p^{-1} \\ \sigma^2 = p^{-4} + 2p^{-3} - 2p^{-2} - p^{-1} \end{cases}$



- **Example 5** # tosses for THTTH ?



解 (甲)
$$\begin{cases} S_0 = 1 + S_0 H + S_2 H, \\ S_1 = S_0 T + S_1 T + S_4 T, \\ S_2 = S_1 H + S_3 H, \\ S_3 = S_2 T, \\ S_4 = S_3 T, \\ S_5 = S_4 H. \end{cases} \qquad \begin{cases} N = S_0 + S_1 + S_2 + S_3 + S_4 \\ S = S_5 \end{cases}$$

(乙)
$$\begin{cases} 1 + N(H+T) = N + S, \\ N\,THTTH = S + S\,TTH, \end{cases} \Rightarrow \begin{cases} 1 + N(z)(pz+qz) = N(z) + S(z) \\ N(z)\,p^2q^3z^5 = S(z) + S(z)\,pq^2z^3 \end{cases}$$

×××THTTH

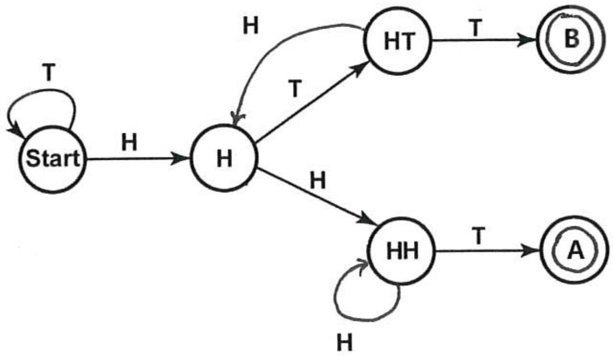
$$\Rightarrow S(z) = \frac{p^2q^3z^5}{p^2q^3z^5 + (1+pq^2z^3)(1-z)}$$

$$\Rightarrow \begin{cases} \mu = p^{-2}q^{-3} + p^{-1}q^{-1} \\ \sigma^2 = \mu^2 - 9p^{-2}q^{-3} - 3p^{-1}q^{-1} \end{cases}$$

General pattern: A = HTHTHHTHTH

$\Rightarrow$ N HTHTHHTHTH = S + S HTHTH + S THHTHTH + S THTHHTHTH

- **Example 6** $\begin{cases} A: \text{HHT} \\ B: \text{HTT} \end{cases}$ 求 $P(A \text{ wins}) = S_A = ?$



解 $(p = q = \frac{1}{2})$
$$\begin{cases} 1 + N(H+T) = N + S_A + S_B \\ N\,HHT = S_A \\ N\,HTT = S_A T + S_B \end{cases}$$

$$\xRightarrow[T=\frac{1}{2}]{H=\frac{1}{2}} \begin{cases} 1 + N = N + S_A + S_B \\ N\frac{1}{8} = S_A \\ N\frac{1}{8} = S_A\frac{1}{2} + S_B \end{cases} \Rightarrow \begin{cases} S_A = \frac{2}{3} \\ S_B = \frac{1}{3} \end{cases}$$

General pat. $\begin{cases} A = \text{HTTHTHTH} \\ B = \text{THTHTTTH} \end{cases} \Rightarrow \begin{cases} N\,\text{HTTHTHTH} = S_A\,\text{TTHTHTH} + S_A + S_B\,\text{TTHTHTH} + S_B\,\text{THTH} \\ N\,\text{THTHTTTH} = S_A\,\text{THTTH} + S_A\,\text{TTH} + S_B\,\text{THTTH} + S_B. \end{cases}$

- **定理** $\bar{\tau}_2 \tau_1 \tau_2 \cdots \tau_{\ell-1} > \tau_1 \tau_2 \cdots \tau_\ell$.

但 HTT > TTH > THH > HHT > HTT ! (De Bruijin sequence)

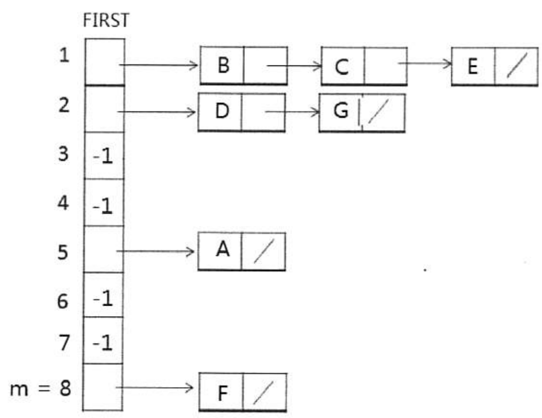
Analysis of Separate-chaining Hashing Algorithms

(#times executed)

unsucc	succ
1	1
$Y+1$	X
$Y+1$	X
Y	X
Y	$X-1$

```
Search(K) {
    i = hash(K); j = FIRST[i];
    while (1) {
        if (j <= 0) return(UNSUCC); ⇒
        if (KEY[j] == K) return(SUCC);
        i = j; j = NEXT[j];
    }
}
```

```
Insert(K) {
    n = n + 1;
    if (j < 0) FIRST[i] = n;
    else       NEXT[i] = n;
    NEXT[n] = 0; KEY[n] = K;
}
```



	KEY	NEXT
1	D	4
2	A	0
3	B	5
4	G	0
5	C	6
6	E	0
$n \to$ 7	F	0
8		
9		
10		

`HTable(2,5,1,2,1,1,8)` = hash table from inserting keys D, A, B, G, C, E, F with hash values 2, 5, 1, 2, 1, 1, 8

$$\begin{cases} \frac{1}{7}\sum\limits_{1\le k\le 7} X_7\big((2,5,1,2,1,1,8;\,k)\big) = \frac{1}{7}(1+1+1+2+2+3+1) = \frac{11}{7} \\ \frac{1}{8}\sum\limits_{1\le h\le 8} Y_7\big((2,5,1,2,1,1,8;\,h)\big) = \frac{1}{8}(3+2+0+0+1+0+0+1) = \frac{7}{8} \end{cases}$$

Sample Spaces:

$$\begin{cases} \Omega_{X_n} = \Big\{(h_1,h_2,\ldots,h_n;\,k)\ \Big|\ 1\le h_1,\ldots,h_n\le m,\ \ 1\le k\le n\Big\}, & |\Omega_{X_n}| = nm^n \\ \Omega_{Y_n} = \Big\{(h_1,h_2,\ldots,h_n;\,h)\ \Big|\ 1\le h_1,\ldots,h_n\le m,\ \ 1\le h\le m\Big\}, & |\Omega_{Y_n}| = m^{n+1} \end{cases}$$

`HTable`$(h_1,h_2,\ldots,h_n)$ = hash table from inserting keys with hash values $h_1,h_2,\ldots,h_n$

Random Variables: $X_n : \Omega_{X_n} \to \mathbb{R}; \quad Y_n : \Omega_{Y_n} \to \mathbb{R}$

$$\begin{cases} X_n\big((h_1, \ldots, h_n;\ k)\big) & = \ \# \text{ probes to } \textbf{succ} \text{ search } k^{\text{th}} \text{ key in } \texttt{HTable}(h_1, \ldots, h_n) \\ Y_n\big((h_1, \ldots, h_n;\ h)\big) & = \ \# \text{ probes to } \textbf{unsucc} \text{ search a key with hash value } h \text{ in } \texttt{HTable}(h_1, \ldots, h_n) \end{cases}$$

$$\begin{cases} \mathbf{E}(X_n) & = \# \text{ probes per } \textbf{succ} \text{ search in a random hash table with } n \text{ keys} \\ & = \displaystyle\sum_{\omega \in \Omega_{X_n}} P(\omega) X_n(\omega) = \frac{1}{nm^n} \sum_{\substack{1 \le h_1, \ldots, h_n \le m \\ 1 \le k \le n}} X_n(h_1, \ldots, h_n;\ k) \\ \mathbf{E}(Y_n) & = \# \text{ probes per } \textbf{unsucc} \text{ search in a random hash table with } n \text{ keys} \\ & = \displaystyle\sum_{\omega \in \Omega_{Y_n}} P(\omega) Y_n(\omega) = \frac{1}{m^{n+1}} \sum_{\substack{1 \le h_1, \ldots, h_n \le m \\ 1 \le h \le m}} Y_n(h_1, \ldots, h_n;\ h) \end{cases}$$

Theorem.

$$\begin{cases} \mathbf{E}(X_n) = 1 + \dfrac{n-1}{2m}; & \mathbf{VAR}(X_n) = \dfrac{(n-1)(6m+n-5)}{12m^2} \\ \mathbf{E}(Y_n) = \dfrac{n}{m}; & \mathbf{VAR}(Y_n) = \dfrac{n(m-1)}{m^2} \end{cases}$$

Proof:

(A)

$$\begin{aligned} \mathbf{E}(Y_n) &= \mathbf{E}(\ell_h) = \frac{n}{m}, \quad (\ell_h = \text{length of the } h^{\text{th}} \text{ chain}) \\ \mathbf{E}(\ell_{\text{hash}(K)}) &= 1 + \frac{n-1}{m} \\ \mathbf{E}(X_n) &= \frac{1 + \mathbf{E}(\ell_{\text{hash}(K)})}{2} = 1 + \frac{n-1}{2m} \qquad \blacksquare \end{aligned}$$

(B)

$$\begin{aligned} \mathbf{E}(Y_n) &= \frac{1}{m^{n+1}} \sum_{1 \le h_1, \ldots, h_n \le m} \sum_{1 \le h \le m} Y_n(h_1, \ldots, h_n;\ h) \\ &= \frac{1}{m^{n+1}} \sum_{1 \le h_1, \ldots, h_n \le m} n \\ &= \frac{1}{m^{n+1}} m^n n \\ &= \frac{n}{m} \end{aligned}$$

$$\begin{aligned} \mathbf{E}(X_n) &= \frac{1}{n}\Big[\big(1 + \mathbf{E}(Y_0)\big) + \big(1 + \mathbf{E}(Y_1)\big) + \cdots + \big(1 + \mathbf{E}(Y_{n-1})\big)\Big] \\ &= 1 + \frac{1}{n} \sum_{1 \le k \le n} \mathbf{E}(Y_{k-1}) \\ &= 1 + \frac{1}{n} \sum_{1 \le k \le n} \frac{k-1}{m} \\ &= 1 + \frac{n-1}{2m} \qquad \blacksquare \end{aligned}$$

$(2, 5, 1, 2, 1, 1, 8; 1)$

(C) $Y = Z_1 + Z_2 + \cdots + Z_n, \quad Z_j = \delta_{h_j = h} = \begin{cases} 1, & p = 1/m \\ 0, & q = 1 - 1/m \end{cases} \quad (1 \le j \le n) \quad \text{(i.i.d)}$

$$\Rightarrow G_Y(z) = (q + pz)^n = \left(\frac{m-1}{m} + \frac{1}{m}z\right)^n$$

$$\Rightarrow \begin{cases} \mathbf{E}(Y_n) & = & np & = & \dfrac{n}{m} \\ \mathbf{VAR}(Y_n) & = & npq & = & \dfrac{n(m-1)}{m^2} \end{cases}$$

$(2, 5, 1, 2, 1, 1, 8; 5)$

$$X \mid k = Z_1 + Z_2 + \cdots + Z_k, \quad Z_k = 1, \quad Z_j = \delta_{h_j = h_k} = \begin{cases} 1, & p = 1/m \\ 0, & q = 1 - 1/m \end{cases} \quad (1 \le j \le k-1)$$

$$\Rightarrow \begin{cases} G_{X|k}(z) = \left(\dfrac{m-1}{m} + \dfrac{1}{m}z\right)^{k-1} z \\ G_X(z) = \displaystyle\sum_{1 \le k \le n} \frac{1}{n} G_{X|k}(z) \qquad\qquad (8.92) \\ \qquad = \displaystyle\sum_{1 \le k \le n} \frac{1}{n} \left(\frac{m-1+z}{m}\right)^{k-1} z \\ \qquad = \dfrac{z}{n} \dfrac{1 - \left(\dfrac{m-1+z}{m}\right)^n}{1 - \dfrac{m-1+z}{m}} \\ \qquad = \dfrac{m}{n} \dfrac{z}{1-z} \left[1 - \left(\dfrac{m-1+z}{m}\right)^n\right] \end{cases}$$

$$\Rightarrow \begin{cases} \mathbf{E}(X_n) & = & 1 + \dfrac{n-1}{2m} \\ \mathbf{VAR}(X_n) & = & \dfrac{(n-1)(6m+n-5)}{12m^2} \end{cases} \qquad \blacksquare$$

Lemma. $\mathbf{E}(X_n) = 1 + \dfrac{1}{n} \displaystyle\sum_{1 \le k \le n} \mathbf{E}(Y_{k-1})$

Proof:
$$\begin{aligned} \mathbf{E}(X_n) &= \frac{1}{nm^n} \sum_{\substack{1 \le h_1, \ldots, h_n \le m \\ 1 \le k \le n}} X_n(h_1, \ldots, h_n;\ k) \\ &= \frac{1}{n} \sum_{1 \le k \le n} \frac{1}{m^n} \sum_{1 \le h_1, \ldots, h_n \le m} X_n(h_1, \ldots, h_n;\ k) \\ &= \frac{1}{n} \sum_{1 \le k \le n} \frac{1}{m^n} \sum_{1 \le h_1, \ldots, h_k \le m} \Big(1 + Y_k(h_1, \ldots, h_{k-1};\ h_k)\Big) m^{n-k} \\ &= \frac{1}{n} \sum_{1 \le k \le n} \frac{1}{m^k} \sum_{1 \le h_1, \ldots, h_k \le m} \Big(1 + Y_k(h_1, \ldots, h_{k-1};\ h_k)\Big) \\ &= 1 + \frac{1}{n} \sum_{1 \le k \le n} \mathbf{E}(Y_{k-1}) \qquad \blacksquare \end{aligned}$$

Analysis of Binary Search Trees (BST) Algorithms

Goal: to derive

$$\begin{cases} \mathbb{E}(X_n) & := \#\text{ probes per } \textbf{succ search} \text{ in } n\text{-node BST} \\ \mathbb{E}(Y_n) & := \#\text{ probes per } \textbf{unsucc search} \quad \text{"} \end{cases}$$

Sample Spaces:

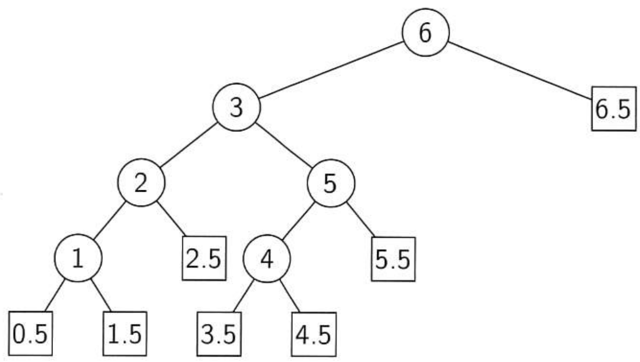
$$\begin{cases} \Omega_{X_n} &= \left\{(x_1, x_2, \dots, x_n; k) \,\middle|\, \begin{matrix}(x_1, x_2, \dots, x_n) \in S_n \\ 1 \le k \le n\end{matrix} \right\}, \quad |\Omega_{X_n}| = n!\, n \\ \Omega_{Y_n} &= \left\{(x_1, x_2, \dots, x_n; y) \,\middle|\, \begin{matrix}(x_1, x_2, \dots, x_n) \in S_n \\ y = 0.5,\, 1.5,\, \dots,\, n.5\end{matrix} \right\}, \quad |\Omega_{Y_n}| = (n+1)! \end{cases}$$

$$S_n \;=\; \text{set of all permutations (relative order of keys) of } \{1, 2, \dots, n\}$$
$$\Omega_{Y_n} \;=\; \big\{(x_1, x_2, \dots, x_n; y) \mid (x_1, x_2, \dots, x_n; y) \in S_{n+1}\big\}, \quad \text{by re-numbering}$$

Random Variables:

$$\begin{cases} X_n\big((x_1, x_2, \dots, x_n; k)\big) & := \#\text{ probes to } \textbf{succ} \text{ search } x_k \text{ in } T(x_1, x_2, \dots, x_n) \\ Y_n\big((x_1, x_2, \dots, x_n; y)\big) & := \#\text{ probes to } \textbf{unsucc} \text{ search } y \quad \text{"} \end{cases}$$

$$T(x_1, x_2, \dots, x_n) \;:=\; \text{BST formed by inserting keys } x_1, x_2, \dots, x_n$$
$$T(6, 3, 2, 1, 5, 4) \;=$$



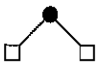
$$\begin{cases} \sum\limits_{1 \le k \le 6} X_6\big((6,3,2,1,5,4;k)\big) &= (1+2+3+4+3+4) &= 17 = 6 + I\big(T(6,3,2,1,5,4)\big) \\ \sum\limits_{y=0.5,\dots,6.5} Y_6\big((6,3,2,1,5,4;y)\big) &= (4+4+3+4+4+3+1) &= 23 = \quad E\big(T(\cdot)\big) \end{cases}$$

$$\begin{cases} I(T) &:= \sum\limits_{x\text{: int node of } T} \text{path length}(x) = 11 \\ E(T) &:= \sum\limits_{y\text{: ext node of } T} \text{path length}(y) = 23 \end{cases}$$

$$\mathbb{E}(X_6) = \frac{1}{6}\Big[\big(1 + \mathbb{E}(Y_0)\big) + \big(1 + \mathbb{E}(Y_1)\big) + \cdots + \big(1 + \mathbb{E}(Y_4)\big) + \big(1 + \mathbb{E}(Y_5)\big)\Big]$$

n	1	2	3	
$\mathbb{E}(X_n) = 2\frac{n+1}{n}H_n - 3$	1	$\frac{3}{2}$	$\frac{17}{9}$	$\cdots$
$\mathbb{E}(Y_n) = 2H_{n+1} - 2$	1	$\frac{5}{3}$	$\frac{13}{6}$	$\cdots$

$\boxed{n=1}$



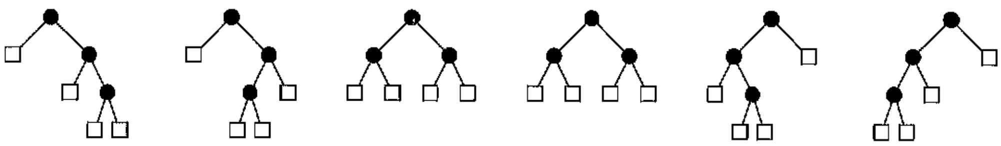
$$\begin{cases} \mathbb{E}(X_1) & = \frac{1}{1}\,[1] & = 1 \\ \mathbb{E}(Y_1) & = \frac{1}{2}\,[1+1] & = 1 \end{cases}$$

$\boxed{n=2}$

$(1,2)$ $(2,1)$

$$\begin{cases} \mathbb{E}(X_2) & = \frac{1}{2\cdot 2!}\,[(1+2)+(1+2)] & = \frac{3}{2} \\ \mathbb{E}(Y_2) & = \frac{1}{3\cdot 2!}\,[(1+2+2)+(2+2+1)] & = \frac{5}{3} \end{cases}$$

$\boxed{n=3}$



$(1,2,3)$ $(1,3,2)$ $(2,1,3)$ $(2,3,1)$ $(3,1,2)$ $(3,2,1)$

$$\begin{cases} \mathbb{E}(X_3) & = \frac{1}{3\cdot 3!}\,[(1+2+3)+(1+2+3)+(1+2+2)+(\cdot)+(\cdot)+(\cdot)] & = \frac{17}{9} \\ \mathbb{E}(Y_3) & = \frac{1}{4\cdot 3!}\,[(1+2+3+3)+(1+3+3+2)+(2+2+2+2)+(\cdot)+(\cdot)+(\cdot)] & = \frac{13}{6} \end{cases}$$

Lemma 1. n-node binary tree T has

(1) $2n$ edges,

(2) $n+1$ external nodes (leaves),

(3) $E(T) = I(T) + 2n$.

Lemma 2. $\mathbb{E}(X_n) = 1 + \dfrac{1}{n} \displaystyle\sum_{1 \le k \le n} \mathbb{E}(Y_{k-1})$

Theorem 3.

$$\begin{cases} \mathbb{E}(X_n) & = 2\frac{n+1}{n} H_n - 3 \quad \approx 2 \ln n \\ \mathbb{E}(Y_n) & = 2H_{n+1} - 2 \qquad \approx 2 \ln n \end{cases}$$

Proof:

$$\begin{cases} \mathbb{E}(X_n) & = \sum_{w \in \Omega_{X_n}} P(w) X_n(w) = \sum_{\substack{(x_1,\dots,x_n) \in S_n \\ 1 \le k \le n}} \frac{1}{n!\,n} X_n(x_1, \dots, x_n; k) \\ & = \frac{1}{n!\,n} \sum_{(x_1,\dots,x_n) \in S_n} \sum_{1 \le k \le n} X_n(x_1, \dots, x_n; k) \\ & = \frac{1}{n!\,n} \sum_T \big(n + I(T)\big) \\ & = \frac{1}{n} \Big[n + \frac{1}{n!} \sum_T I(T) \Big] \qquad (1) \\ \mathbb{E}(Y_n) & = \frac{1}{(n+1)!} \sum_{(x_1,\dots,x_n) \in S_n} \sum_{y = 0.5,\dots,n.5} Y_n(x_1, \dots, x_n; y) \\ & = \frac{1}{(n+1)!} \sum_T E(T) \\ & = \frac{1}{(n+1)n!} \sum_T \big(2n + I(T)\big) \\ & = \frac{1}{n+1} \Big[2n + \frac{1}{n!} \sum_T I(T) \Big] \qquad (2) \end{cases}$$

$(1) + (2) \qquad \Rightarrow (n+1)\mathbb{E}(Y_n) = n\mathbb{E}(X_n) + n$

$+$ **Lemma 2** $\Rightarrow$

$$\begin{aligned} (n+1)\,\mathbb{E}(Y_n) &= \sum_{1 \le k \le n} \mathbb{E}(Y_{k-1}) + 2n \\ -) \qquad n\,\mathbb{E}(Y_{n-1}) &= \sum_{1 \le k \le n-1} \mathbb{E}(Y_{k-1}) + 2(n-1) \\ \hline (n+1)\,\mathbb{E}(Y_n) &= (n+1)\,\mathbb{E}(Y_{n-1}) + 2 \end{aligned}$$

$$\Rightarrow \begin{cases} \mathbb{E}(Y_n) & = \mathbb{E}(Y_{n-1}) + \frac{2}{n+1} \\ \mathbb{E}(Y_{n-1}) & = \mathbb{E}(Y_{n-2}) + \frac{2}{n} \\ & \vdots \\ \mathbb{E}(Y_2) & = \mathbb{E}(Y_1) + \frac{2}{3} \\ \mathbb{E}(Y_1) & = \frac{2}{2} \end{cases}$$

$$\Rightarrow \begin{cases} \mathbb{E}(Y_n) & = 2H_{n+1} - 2 \\ \mathbb{E}(X_n) & = \frac{n+1}{n} E(Y_n) - 1 = 2\frac{n+1}{n} H_n - 3 \end{cases} \qquad \blacksquare$$

Lemma 2. $\mathbb{E}(X_n) = 1 + \frac{1}{n} \sum_{1 \le k \le n} \mathbb{E}(Y_{k-1})$

Proof:

$$\begin{aligned}
\mathbb{E}(X_n) &= \frac{1}{n!\,n} \sum_{(x_1,\ldots,x_n)} \sum_{1 \le k \le n} X_n(x_1,\ldots,x_n;k) \\
&= \frac{1}{n!\,n} \sum_{(x_1,\ldots,x_n) \in S_n} \sum_{1 \le k \le n} \Big(1 + Y_n(x_1,\ldots,x_{k-1};k)\Big) \\
&= 1 + \frac{1}{n!\,n} \sum_{1 \le k \le n} \sum_{(x_1,\ldots,x_k) \in S_k} n \cdots (k+1) Y_{k-1}(x_1,\ldots,x_{k-1};x_k) \\
&= 1 + \frac{1}{n} \sum_{1 \le k \le n} \left[\frac{1}{k!} \sum_{(x_1,\ldots,x_k) \in S_k} Y_{k-1}(x_1,\ldots,x_{k-1};x_k)\right] \\
&= 1 + \frac{1}{n} \sum_{1 \le k \le n} \mathbb{E}(Y_{k-1}) \qquad \blacksquare
\end{aligned}$$

Remark:

$$\begin{cases} \text{Dynamic model} & : \; n! \text{ trees;} \\ \text{Static} \quad \text{"} & : \; \frac{1}{n+1}\binom{2n}{n} \text{ trees,} \quad (\text{BST } 6,3,2,1,5,4 \equiv 6,3,5,4,2,1) \end{cases}$$