# Fundamentals of Discrete Probability
## (Bertsimas & Freund 2004, Chapter 2)

**Howard Hao-Chun Chuang (莊皓鈞)**

**Professor**
**College of Commerce**
**National Chengchi University**

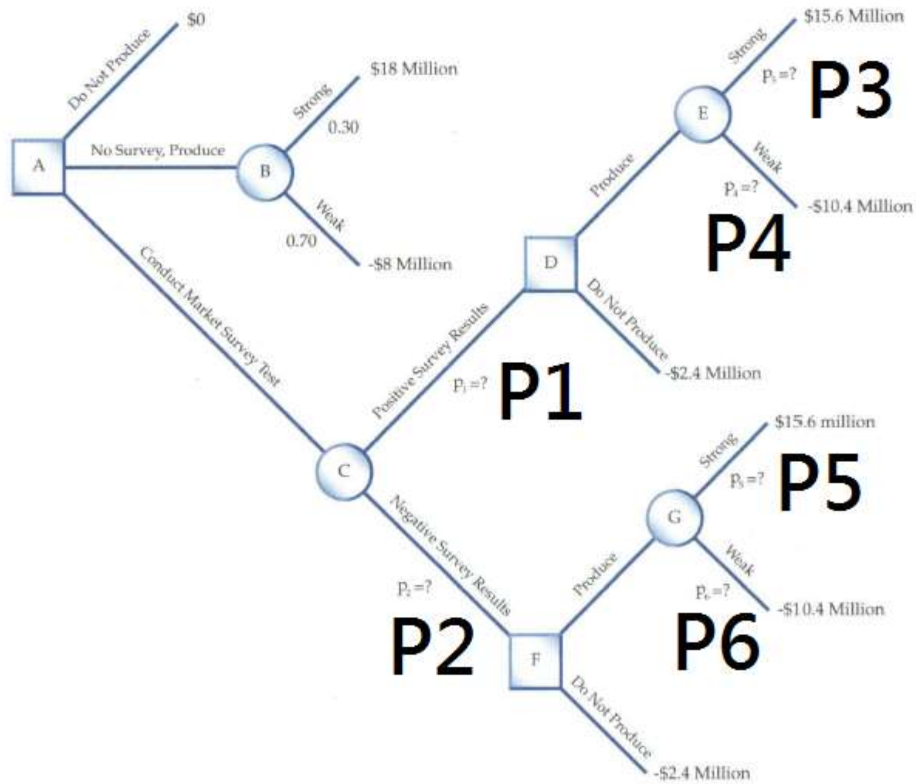
**September, 2024**
**Taipei, Taiwan**

# Example 2.3 Caroline Janes' Decision

- **The market will be strong(S) with *P*(S)=0.3 for +$18 million or weak(W) with *P*(W)=0.7 for -$8 million**
- **Market survey will cost $2.4 million. Let Q denote positive survey results and N for negative survey results.**
- **With *P*(Q|W)=0.1 & *P*(N|S)=0.2, what else can we infer right away?**
- **What are the decisions to be made?**

produce ?

survey ?

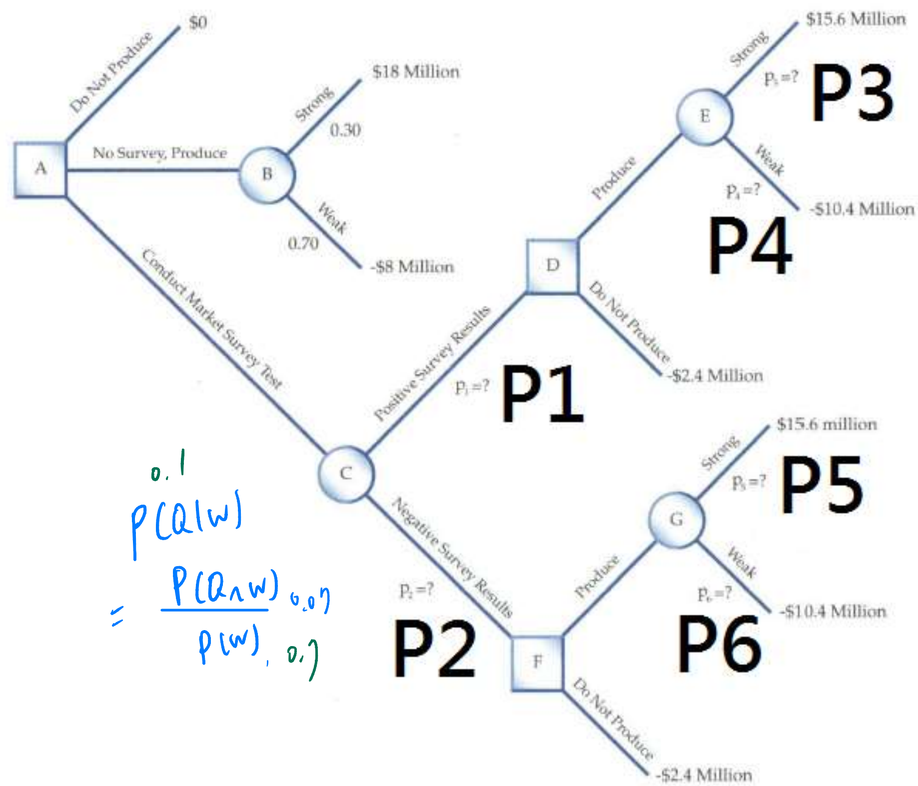


# Example 2.3 Caroline Janes' Decision

- **With $P$(Q|W)=0.1 & $P$(N|S)=0.2**

| | Strong (S) | Weak (W) | Total |
|---|---|---|---|
| Positive (Q) | $P$(Q&S) = 0.23 | 1. $P$(Q&W) = 0.07 | $P$(Q) = 0.31 |
| Negative (N) | $P$(N&S) = 0.07 | 2. $P$(N&W) = 0.63 | $P$(N) = 0.69 |
| Total | $P$(S) = 0.3 | $P$(W) = 0.7 | 1 |

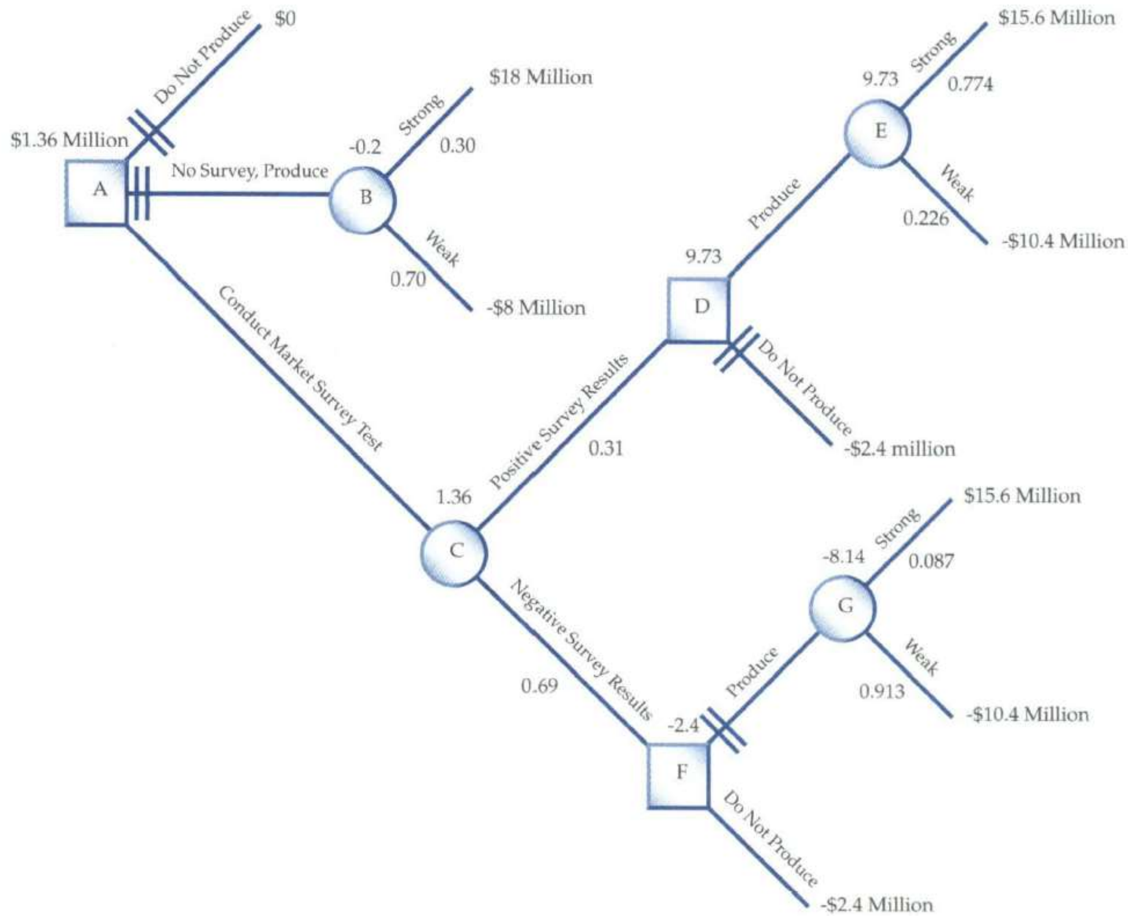
- $P_1 = P($ Q $)=$ 0.31
- $P_2 = P($ N $)=$ 0.67
- $P_3 = P($ S | Q $)=$ 0.24/0.31
- $P_4 = P($ W | Q $)=$ 0.07/0.31
- $P_5 = P($ S | N $)=$ 0.06/0.69
- $P_6 = P($ W | N $)=$ 0.63/0.69




0.1

$P(Q|W) = \frac{P(Q \cap W)}{P(W)} \; \frac{0.07}{0.7}$

# Example 2.3 Caroline Janes' Decision



# Foundation of Probability Theory

- **The Laws of Probability**

  **The first law**: The probability of any event should be between 0 to 1.

  **The second law**: If A & B are mutually exclusive, $P(\text{A or B})=P(\text{A})+P(\text{B})$

  **The third law**: $P(\text{A}|\text{B}) = P(\text{A and B})/P(\text{B})$

  **The fourth law**: If A & B are independent, $P(\text{A}|\text{B})=P(\text{A})$ so $P(\text{A \& B})=P(\text{A})\text{x}P(\text{B})$

- **Bayes' Theorem**: $P(\text{B}|\text{A}) = P(\text{A}|\text{B})P(\text{B})/P(\text{A})$

$$P(\text{Cause} \mid \text{Effect}) = \frac{\overset{\text{Sensor data}}{P(\text{Effect} \mid \text{Cause})} \cdot \overset{\text{Prior information}}{P(\text{Cause})}}{\underset{\text{Prior information}}{P(\text{Effect})}}$$

- **Law of Total Probability**

$$P(\text{A})=P(\text{A}|\text{B}_1)P(\text{B}_1) + P(\text{A}|\text{B}_2)P(\text{B}_2) + \ldots + P(\text{A}|\text{B}_n)P(\text{B}_n)$$

News — sport

News — International

Prior: 用过去資料推出來的

$$\begin{cases} P(sport) = \dfrac{\# \text{ or } \#}{n} \\ P(International) = \dfrac{\# \text{ or } \#}{n} \end{cases}$$

News: title Jordan

Posterior

$$\begin{cases} P(Sport \mid Jordan) = \dfrac{P(Sport) * P(Jordan \mid Sport)}{P(Jordan)} \\ P(International \mid Jordan) = \end{cases}$$

在 News 是 Sport 的情況下是 Jordan 的機率.

# Foundation of Probability Theory

- **Joint probability table**

| Dog barks | Burglar | Raccoon | Tally | P | Selected |
|---|---|---|---|---|---|
| false | false | false | 405 | 0.405 | ☐ |
| false | false | true | 225 | 0.225 | ☐ |
| false | true | false | 0 | 0.000 | ☐ |
| false | true | true | 0 | 0.000 | ☐ |
| true | false | false | 45 | 0.045 | ☐ |
| true | false | true | 225 | 0.225 | ☐ |
| true | true | false | 50 | 0.050 | ☐ |
| true | true | true | 50 | 0.050 | ☐ |
| | | | 1000 | 1.000 | 0.000 |

If ⇒ Dog = T.

$$P(B=T \mid D=T) = \frac{P(B=T \wedge D=T)}{P(D=T)} = \frac{100/1000}{370/1000}$$

2.75倍

$$P(R=T \mid D=T) = \frac{P(R=T \wedge D=T)}{P(D=T)} = \frac{275/1000}{370/1000} = \frac{275}{370}$$

Bayes factor ⇒ $\frac{275}{370} : \frac{100}{370} = 2.75$

# Foundation of Probability Theory

- **Bayes theorem**

probability a hypothesis is true given the evidence

probability a hypothesis is true (before any evidence is present)

probability of seeing the evidence if the hypothesis is true

probability of observing the evidence

$$P(H/E) = \frac{P(H)\,P(E/H)}{P(E)}$$

- **Bayes factor** $\dfrac{P(M_1 \mid D)}{P(M_2 \mid D)} = \dfrac{P(D \mid M_1)P(M_1)}{P(D \mid M_2)P(M_2)}$

**Ratio of posterior probs for two models/hypotheses**

**BF>100: Decisive evidence for $M_1$**

**BF>10: Strong evidence for $M_1$**

**10>BF>3: Moderate evidence for $M_1$**

**3>BF>1: Weak evidence for $M_1$**

# Foundation of Probability Theory

- **Conditional probability table**

| Dog barks | Burglar | Raccoon | Tally | P | Selected |
|---|---|---|---|---|---|
| false | false | false | 0 | 0.000 | ☐ |
| false | false | true | 0 | 0.000 | ☑ |
| false | true | false | 0 | 0.000 | ☐ |
| false | true | true | 0 | 0.000 | ☑ |
| true | false | false | 45 | 0.122 | ☐ |
| true | false | true | 225 | 0.608 | ☑ |
| true | true | false | 50 | 0.135 | ☐ |
| true | true | true | 50 | 0.135 | ☑ |
| ◉ T ○ F ○ ? | ○ T ○ F ◉ ? | ○ T ○ F ◉ ? | 370 | 1.000 | 0.743 |

# Simulation for Conditional Probabilities

- IC Modules

In ~70% of production runs, the electric current is regular so every IC module produced has an independent 10% chance of being defective. In the other 30% of production runs, when current is irregular, every IC module produced has an independent 40% chance of being defective.

We test 10 IC modules and aim to answer the following questions:

What is the probability of finding exactly 2 defective IC modules among the 10 tested?

What is the conditional probability that the current is exclusively regular given that 2 defective IC modules are found among the 10 tested?

What is *P*(at least 1 irregular current| *k* defectives among 10 tested), for *k*=0, 1, 2, 3?

# Random Variables

- Let *X* denote a discrete random variable (RV)

  *P*(*X* = *x*)=*p* is the probability distribution function (pdf) denoted by *f*(*x*)

  *P*(X <= *x*) is the cumulative distribution function (cdf) denoted by *F*(*x*)

  $\sum P(X = x) = 1$ for the defined range of *x*

  What is a probability distribution of *X* then?

  Expected value $\mu_X = E(X) = \sum_{i=1}^{n} p_i x_i$

  Variance $\sigma_X^2 = VAR(X) = \sum_{i=1}^{n} p_i \, (x_i - \mu_X)^2$

- For another RV *Y*=*aX*+*b*

  Expected value $\mu_Y = E(Y) = aE(X) + b$

  Variance $\sigma_Y^2 = VAR(Y) = a^2 VAR(X)$

- Wikipedia has great details for different probability distributions

# The Binomial Distribution (section 2.6)

- ***X*** **~ Binomial(*n*, *p*) (sum of *n* Bernoulli(*p*) RVs)**

  ***P*(*X*=*x*)/*f*(x): binom.pmf (scipy)**

  ***P*(*X* <= *x*)/*F*(x): binom.cdf**

  **Find the smallest *k* such that *P*(*X* <= *k*)>=*q*: binom.ppf** → $P(X \le k) \ge 0.98$

  predefined

  **Simulate *S* random realizations: binom.rvs**

  stats.binom.ppf(0.98, n=8, p=0.02) ⇒ $P(X \le k) \ge 0.98$

  ⇒ 成功次數最多是 k 次的情況下累積機率 ≥ 0.98

  **Example 2.12 in page 70 of Bertsimas & Freund (2004)**

  **50 mainframe computers were sold to 50 locations. On average, a client would send 5 service requests in the past year (250 working days). Each request would take a full day for a service engineer to process.**

  **With one engineer, what is the chance at least one client won't be served on any given day? What if two engineers are hired?**

  $1 - P(X=0) - P(X=1)$

  $= 1 - P(X \le 1) \to P(X > 1)$

  **How many engineers are need for a service level agreement (SLA) of 98%?**

  $P(X \le k) \ge 0.98$

  smallest

$X$: # of request in a day

$X \sim \text{Binomial}(n=50, p=\frac{5}{250})$

50個銅板，每一個翻起來的機率 p.

$X$ (possible outcome) = 0, 1, 2, 3, ..., 50.

Sum of n=50 Bernoulli (p).

# Simulation for Conditional Probabilities

- **IC Modules**

  电容

  **In ~70% of production runs, the electric current is regular so every IC module produced has an independent 10% chance of being defective. In the other 30% of production runs, when current is irregular, every IC module produced has an independent 40% chance of being defective.**

  **We test 10 IC modules and aim to answer the following questions:**

  **What is the probability of finding exactly 2 defective IC modules among the 10 tested?**

  电容都正常

  **What is the conditional probability that the current is exclusively regular given that 2 defective IC modules are found among the 10 tested?**

  **What is *P*(at least 1 irregular current| *k* defectives among 10 tested), for *k*=0, 1, 2, 3?**

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{\#(A \cap B)}{\cancel{S}}}{\frac{\#(B)}{\cancel{S}}}$$

电
- reg 0.7
- irreg 0.3

reg
- good 0.9
- bad 0.1

irreg
- good 0.6
- bad 0.4

1. P(2 defective IC) = ?
2. P(current all regular | 2 defective IC) = ?

$\Rightarrow \dfrac{P(\text{current all regular} \cap 2 \text{ defective IC})}{P(2 \text{ defective IC})}$

Simulation

run

1.

2.

3.

4.

⋮

defective or not

current regular or not

# of bad IC

10個

# Random Variables

- **Let *X* denote a discrete random variable (RV)**
  **$P(X = x)=p$ is the probability distribution function (pdf) denoted by $f(x)$**
  **$P(X <= x)$ is the cumulative distribution function (cdf) denoted by $F(x)$**
  **$\sum P(X = x) = 1$ for the defined range of $x$**
  **What is a probability distribution of *X* then?**

  **Expected value** $\mu_X = E(X) = \sum_{i=1}^{n} p_i x_i$
  **Variance** $\sigma_X^2 = VAR(X) = \sum_{i=1}^{n} p_i (x_i - \mu_X)^2$

- **For another RV $Y=aX+b$**
  **Expected value** $\mu_Y = E(Y) = aE(X) + b$
  **Variance** $\sigma_Y^2 = VAR(Y) = a^2 VAR(X)$

- **Wikipedia has great details for different probability distributions**

# The Binomial Distribution (section 2.6)

- $X \sim$ Binomial($n$, $p$) (sum of $n$ Bernoulli($p$) RVs)
  $P(X{=}x)/f(x)$: binom.pmf (scipy)
  $P(X <= x)/F(x)$: binom.cdf
  Find the smallest $k$ such that $P(X <= k) >= q$: binom.ppf
  Simulate $S$ random realizations: binom.rvs

  Example 2.12 in page 70 of Bertsimas & Freund (2004)

  50 mainframe computers were sold to 50 locations. On average, a client would send 5 service requests in the past year (250 working days). Each request would take a full day for a service engineer to process.
  With one engineer, what is the chance at least one client won't be served on any given day? What if two engineers are hired?

  How many engineers are need for a service level agreement (SLA) of 98%?

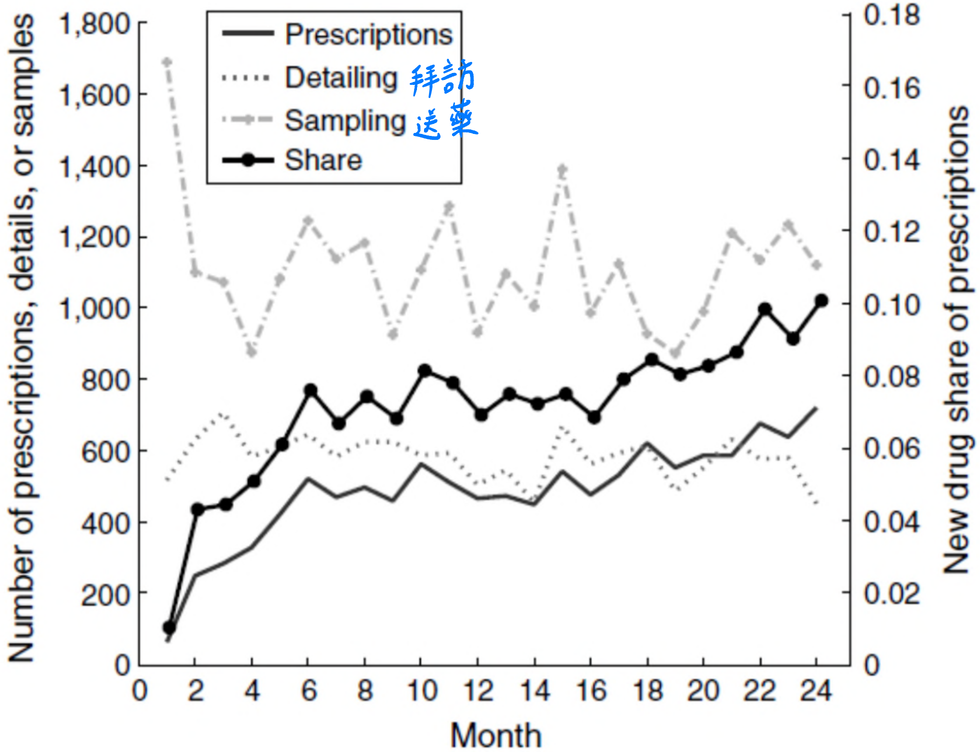
# Binomial Distribution in Action

$Y_{it}$ : # of prescriptions from client $i$ at month $t$

- **Data-driven detailing & sampling**

  發行**XX**用藥，***C***位醫生，***T***個月資料蒐集

  每個醫生每個月，面對***n***個此症狀病人，會開出***y***張用新藥處方的機率？




data

$$P(Y_{it}=y_{it}) = \binom{n_{it}}{y_{it}} (P_{it})^{y_{it}} (1-P_{it})^{n_{it}-y_{it}}$$

# Binomial Distribution in Action

代表都還alive

- Two coins of a customer

shown

客戶

- show 1 (prob $p$)
- no-show 0 (prob 1-$p$)

hidden

客人有來沒來

- alive A (prob 1-θ)
- dead D (prob θ)

$\leftarrow$ | $\rightarrow$ dead

| 1996 | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 0 | 0 |

| 1996 | 1997 | 1998 | 1999 | 2000 | 2001 |
|---|---|---|---|---|---|
| A | A | A | A | D | D |
| A | A | A | A | A | D |
| A | A | A | A | A | A |

data likelihood

$$f(100100 \mid p, \theta) = p(1-p)(1-p)p \underbrace{(1-\theta)^4\theta}_{P(\text{AAAADD})}$$
$$+ p(1-p)(1-p)p(1-p) \underbrace{(1-\theta)^5\theta}_{P(\text{AAAAAD})}$$
$$+ \underbrace{p(1-p)(1-p)p}_{P(Y_1=1, Y_2=0, Y_3=0, Y_4=1)}(1-p)(1-p) \underbrace{(1-\theta)^6}_{P(\text{AAAAAA})}$$

# Decision Tree for Sourcing

Outcome of a fair Die

- $X$ ~ discrete Uniform($a$, $b$)

  Can you think of any example(s) of a discrete Uniform($a$, $b$) RV?

  $a$: minimum outcome; $b$: maximum outcome

  possible outcomes of $X$ is [$a$, $a$+1,…,$b$]

  how to estimate $a$ & $b$ from data, if any?

- A **sourcing** problem

  two suppliers (see the right)

  unit price of 150€

  50%-50% chance demand ($D$)

  weak $D$ ~ Uniform(2000, 8000)

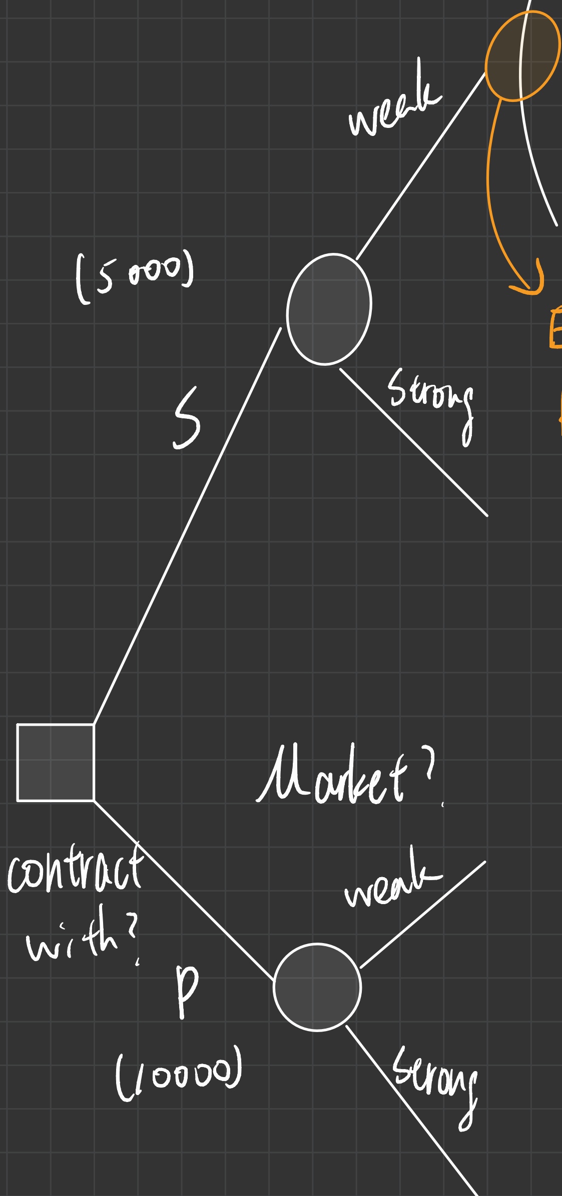
  strong $D$ ~ Uniform(6000, 14000)

  数量

| | Sweden (S) | Poland (P) |
|---|---|---|
| Order Quantity | 5,000 units | 10,000 units |
| Fixed Charge | 0€ | 50,000€ |
| Unit Cost | 120€ | 100€ |

how can we represent this in a decision tree?

which supplier should we choose from?




Demand

2000 : $\frac{1}{8000-2000+1}$ ↙ 机率

2001 :

⋮

7999 :

8000 :

EMV ?

EMV = $\cancel{\left(\frac{8000+2000}{2}\right)}$

$$\text{Profit}_{(S,\,weak)} = 150 * \text{Min}(5{,}000,\ D^{weak}) - 120 * 5{,}000$$

| Simulation | Profit |
|---|---|
| $D_1$ | $\$_1$ |
| $D_2$ | $\$_2$ |
| ⋮ | ⋮ |
| $D_s$ | $\$_s$ |

$\overline{\text{Profit}}$

(fresh food) 数量 life time (hour)

item $Q$ 0. 1. 2, ..... , 23

$Q=10$ $Q=9$

$\boxed{A}$ $Q_A$ $\underbrace{0.\ 1,\ 0.}_{\text{賣出的数量}}$

(if=1?)

B

C

⋮

$$D_{hour}\,(\text{Demand}) \sim \text{Binomial}\,(Q_{hour}, P)$$

$= f(x_1, x_2, \dots x_{xx})$

東西被買走的机率

# Model Formulation & Strategy Optimization

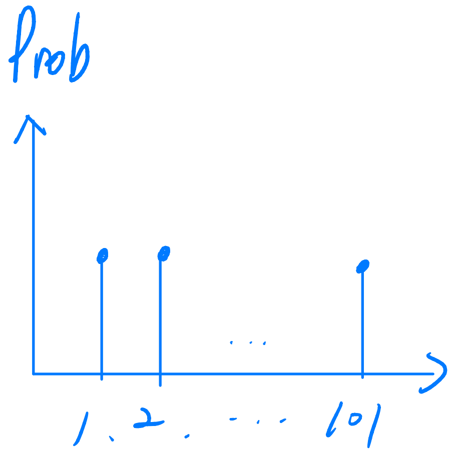
- **Hello Kitty Magnets**

# Model Formulation & Strategy Optimization

- **Hello Kitty Magnets**

  Suppose 7-11 launches a new set of magnets 101 pieces. You aim to buy the whole set, using two options

  - \$5 each from 7-11, not knowing the number a priori
  - \$25 each from dealers, knowing the number for sure



# Model Formulation & Strategy Optimization

- **Hello Kitty Magnets**

  **Suppose 7-11 launches a new set of magnets 101 pieces. You aim to buy the whole set, using two options**

  **\$5 each from 7-11, not knowing the number a priori** $5 \times 101 = 505$ 最低

  **\$25 each from dealers, knowing the number for sure**

  **What is a strategy with zero uncertainty? Cost?** $25 \times 101 = 2525$

  **any strategy to beat this and reduce cost?**

  **Decision model**

  # of cards from 7-11 (x), 101 - x from dealers

  **decision variable**

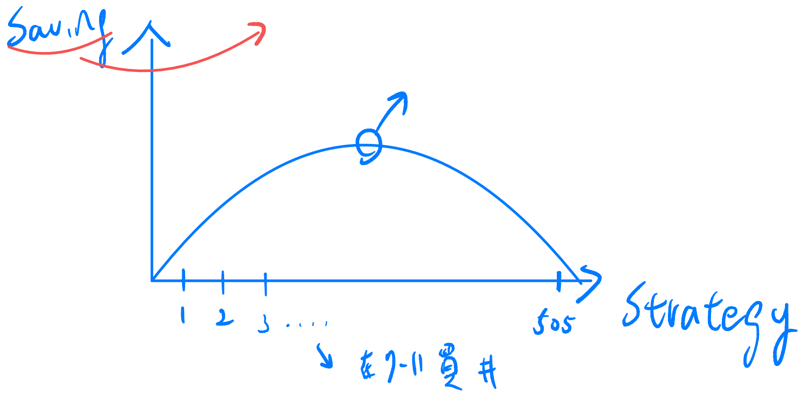
  **parameter**

  **objective**

  **uncertainty**

  **random variable**

  $\boxed{\text{Saving}} = 2525 - \text{cost} = 2525 - (5 * x + 25 * (101 - \underbrace{\text{Unique cards}}_{\text{缺的卡}}))$




  **What are the assumptions?**

  - 可以任意買 missing卡 with flat price
  - No exchange → buy x from 7-11, get unique xx, left x - xx (多出來重複的), 缺 101 - xx
  - No Resell
  - Constant price

  Assume 1:1 exchange ?

exchange ~ Binomial (n = Min (缺, left), $p^{exchange}$)

# Discrete RVs for Decision Modeling

1. Empirical
2. Binominal (n, p)
3. Uniform (a, b)

- **Three discrete distributions for infinite values**

  **$X$ ~ Poisson($\lambda$)**

  **a very, very, very important distribution for counting processes**

  **$E[X] = Var[X] = \lambda$**

  **$Y$~Poisson($\lambda_1$) & $Z$~Poisson ($\lambda_2$) => $W=Y+Z$~Poisson ($\lambda_1 + \lambda_2$)**

  **$Y$~Poisson($\lambda_1$) & $Z$~Poisson ($\lambda_2$) => $W=Y-Z$~Skellam**

  **check scipy.stats.skellam**

Discrete random count

ex:

Demand, Visits, Download.

離散隨機變數取特定值的機率

PMF: Probability Mass Function (PMF)

機率質量函數 $P(X=x)=f(x)$

Poisson:

$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$

$\lambda$: 單位時間 or 空間內事件的平均次數

# Discrete RVs for Decision Modeling

- **Three discrete distributions for infinite values**

  ***X* ~ Poisson(*λ*)**

  **a very, very, very important distribution for counting processes**

  **E[*X*] = Var[*X*] = *λ***

  **Example: To obtain a model for campus Internet security in NCCU, the number of cyber-attacks occurring each week was observed over a period of 1 year. It was found that,**

  **0 attacks occurred in each of 9 weeks**
  **1 attack occurred in each of 14 weeks**
  **2 attacks occurred in each of 13 weeks**
  **3 attacks occurred in each of 9 weeks**
  **4 attacks s occurred in each of 4 weeks**
  **5 attacks occurred in each of 2 weeks**
  **6 attacks occurred in each of 1 weeks**

$X$: # of attacks/week ~ Poisson($\lambda$)

$E(X)=\lambda$

mean: $\bar{X}=1.9038$

$P(X\le 6)\ge 0.99$

attacks

service level

status.poisson.ppf

ppf ⇒ Percent Point Function

= Inverse CDF

# Discrete RVs for Decision Modeling

- **Three discrete distributions for infinite values**

  **$X$ ~ Poisson($\lambda$)** ⇒ $\frac{\text{count}}{\text{time}}$  *:變異數和期望值相同

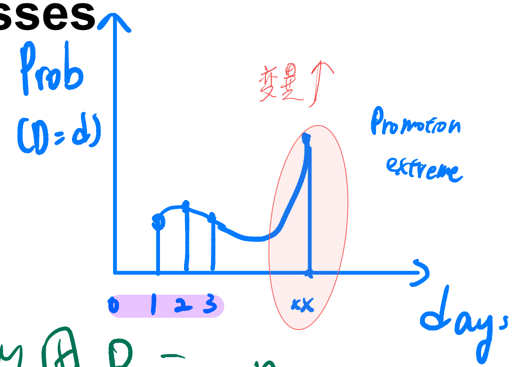
  a very, very, very important distribution for counting processes

  $E[X] = Var[X] = \lambda$ ⇒ 平均值

  1. $Y$~Poisson($\lambda_1$) & $Z$~Poisson ($\lambda_2$) => $W$=$Y$+$Z$~Poisson ($\lambda_1$ +$\lambda_2$)
  2. $Y$~Poisson($\lambda_1$) & $Z$~Poisson ($\lambda_2$) => $W$=$Y$-$Z$~Skellam

     check scipy.stats.skellam

不能處理 也就是尾 Varience↑的項




當 Var[x] 變大，和 E(x) 相距太遠就不可以用 Poisson

  **$X$ ~ negative binomial ($n$, $p$)**

  useful for modeling a count data distribution with $Var[X]/E[X]>>1$

  example: modeling retail sales

$p = \frac{\mu}{\sigma^2}$ ← $E[x]$ ← 4.899 ($\bar{x}$) ; ← $Var[x]$ ← 10.096 ($s^2$) 資料

$n = \frac{\mu^2}{\sigma^2 - \mu}$

  **$X$ ~ Geometric($p$)**

  related to Bernoulli ($p$) processes

  example: detecting empty shelves / missing members


$\lambda_2 = 20/\text{Day}$ $\qquad$ $\lambda_1 \Rightarrow 30/\text{Day}$

$$D = D_1 + D_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

---

100,000

→ 80k items → Prioritise

15k items

| | |
|---|---|
| 1. | $d_1$ |
| 2. | $d_2$ |
| 3. | $d_3$ |
| 4. | ⋮ |
| ⋮ | |
| 80k | $d_{80k}$ |

# 交友軟體

$\longrightarrow$

0000 00 1 00 1 ... 000000 1

↘ 要看幾0才能找到，

1.

1. 有來

0. 沒來

eg. $0.98 = p \rightarrow p \rightarrow 1$ 的 prob

$0.02 = q \rightarrow q \rightarrow 0$ 的 prob

0: 一天沒去: 0.02

00: 連續 2天沒來: $(0.02)^2$

0000: 連續 4天 沒來 = 0.0004

prob

類別
Category K

| | 1, 2, ... ... k |
| --- | --- |
| 會員 1. | |
| 2. | |
| 3. | |
| ⋮ | |

if $q \le$ tolearnc ← manager

(# of straight) < 0