

Fundamentals of Discrete Probability

(Bertsimas & Freund 2004, Chapter 2)

Howard Hao-Chun Chuang (莊皓鈞)

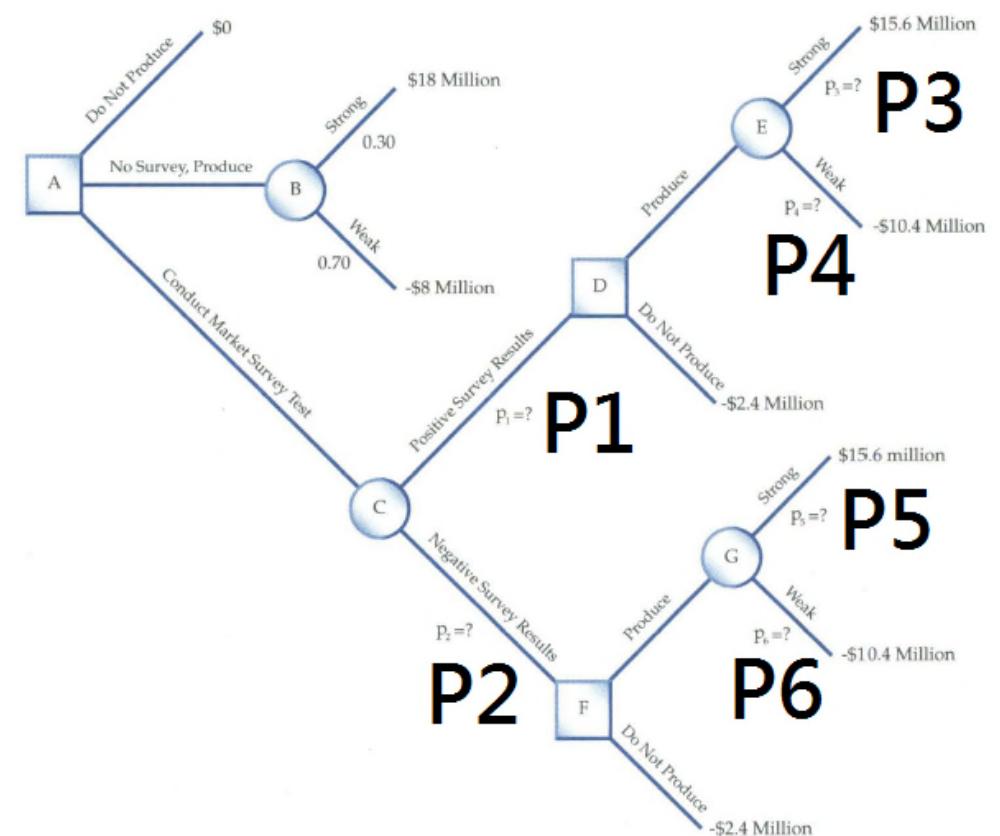
**Professor
College of Commerce
National Chengchi University**

**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?

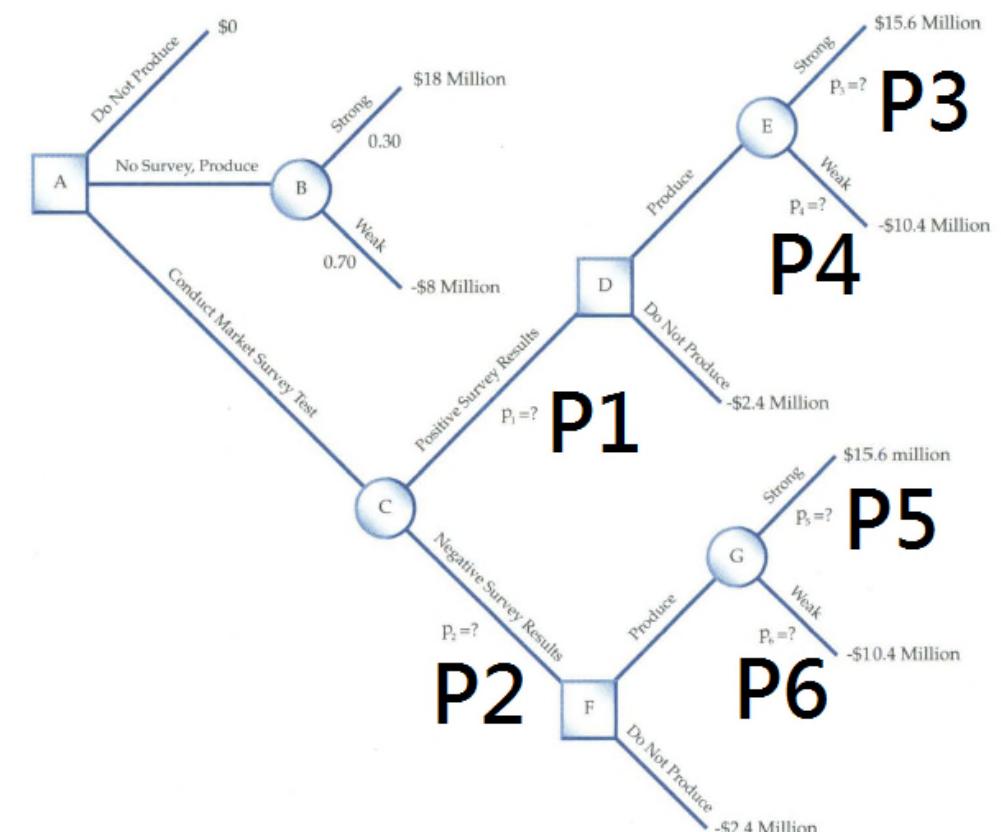


Example 2.3 Caroline Janes' Decision

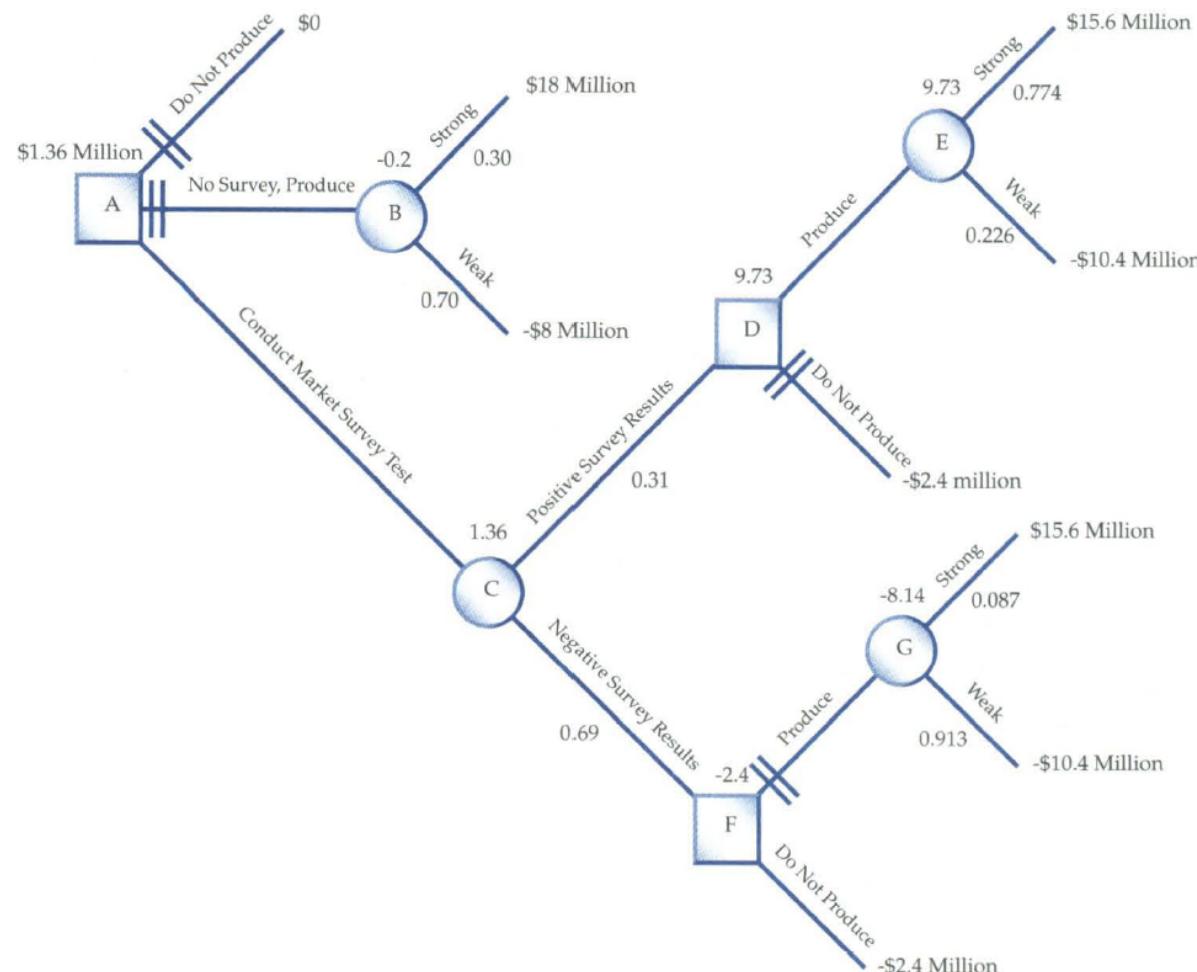
- With $P(Q|W)=0.1$ & $P(N|S)=0.2$

| | Strong (S) | Weak (W) | Total |
|--------------|---------------|---------------|----------|
| Positive (Q) | $P(Q \& S) =$ | $P(Q \& W) =$ | $P(Q) =$ |
| Negative (N) | $P(N \& S) =$ | $P(N \& W) =$ | $P(N) =$ |
| Total | $P(S) = 0.3$ | $P(W) = 0.7$ | 1 |

- $P_1 = P() =$
- $P_2 = P() =$
- $P_3 = P() =$
- $P_4 = P() =$
- $P_5 = P() =$
- $P_6 = P() =$



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, $P(A \text{ or } B) = P(A) + P(B)$

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

The fourth law: If A & B are independent, $P(A|B) = P(A)$ so $P(A \& B) = P(A) \times P(B)$

- **Bayes' Theorem:** $P(B|A) = P(A|B)P(B)/P(A)$

$$P(\text{Cause} | \text{Effect}) = \frac{P(\text{Effect} | \text{Cause}) \cdot P(\text{Cause})}{P(\text{Effect})}$$

Sensor data
Prior information
Prior information

- **Law of Total Probability**

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n)$$



Foundation of Probability Theory

- Joint probability table

| Dog barks | Burglar | Raccoon | Tally | P | Selected |
|-------------|---------|---------|-------|-------|--------------------------|
| false | false | false | 405 | 0.405 | <input type="checkbox"/> |
| false | false | true | 225 | 0.225 | <input type="checkbox"/> |
| false | true | false | 0 | 0.000 | <input type="checkbox"/> |
| false | true | true | 0 | 0.000 | <input type="checkbox"/> |
| true | false | false | 45 | 0.045 | <input type="checkbox"/> |
| true | false | true | 225 | 0.225 | <input type="checkbox"/> |
| true | true | false | 50 | 0.050 | <input type="checkbox"/> |
| true | true | true | 50 | 0.050 | <input type="checkbox"/> |
| Grand Total | | | | 1000 | 1.000 |
| | | | | | 0.000 |



Foundation of Probability Theory

- **Bayes theorem**

probability a hypothesis is true
given the evidence

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

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

- **Bayes factor** $\frac{P(M_1 | D)}{P(M_2 | D)} = \frac{P(D | M_1)P(M_1)}{P(D | M_2)P(M_2)}$

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 | <input type="checkbox"/> |
| false | false | true | 0 | 0.000 | <input checked="" type="checkbox"/> |
| false | true | false | 0 | 0.000 | <input type="checkbox"/> |
| false | true | true | 0 | 0.000 | <input checked="" type="checkbox"/> |
| true | false | false | 45 | 0.122 | <input type="checkbox"/> |
| true | false | true | 225 | 0.608 | <input checked="" type="checkbox"/> |
| true | true | false | 50 | 0.135 | <input type="checkbox"/> |
| true | true | true | 50 | 0.135 | <input checked="" type="checkbox"/> |
| RT GF RT | | | 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.

电容

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 $P(\text{at least 1 irregular current} | k \text{ defectives among 10 tested})$, for $k=0, 1, 2, 3$?



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

電
 reg 0.7
 irreg 0.3

1. $P(\geq \text{defective IC}) = ?$

2. $P(\text{current all regular} | \geq \text{defective IC}) = ?$

$$\frac{P(\text{current all regular} \wedge \geq \text{defective IC})}{P(\geq \text{defective IC})}$$

defective or not

reg
 good 0.9
 bad 0.1

Simulation

run.

current regular or not

of bad IC.

1. 00, 00 ... 00 —

2. 00

10.個

3.

4.

:

irreg
 good 0.6
 bad 0.4

Random Variables

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

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

$P(X \leq x)$ 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)

- $X \sim \text{Binomial}(n, p)$ (sum of n Bernoulli(p) RVs)

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

$P(X \leq x)/F(x)$: `binom.cdf`

Find the smallest k such that $P(X \leq k) \geq q$: `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%?

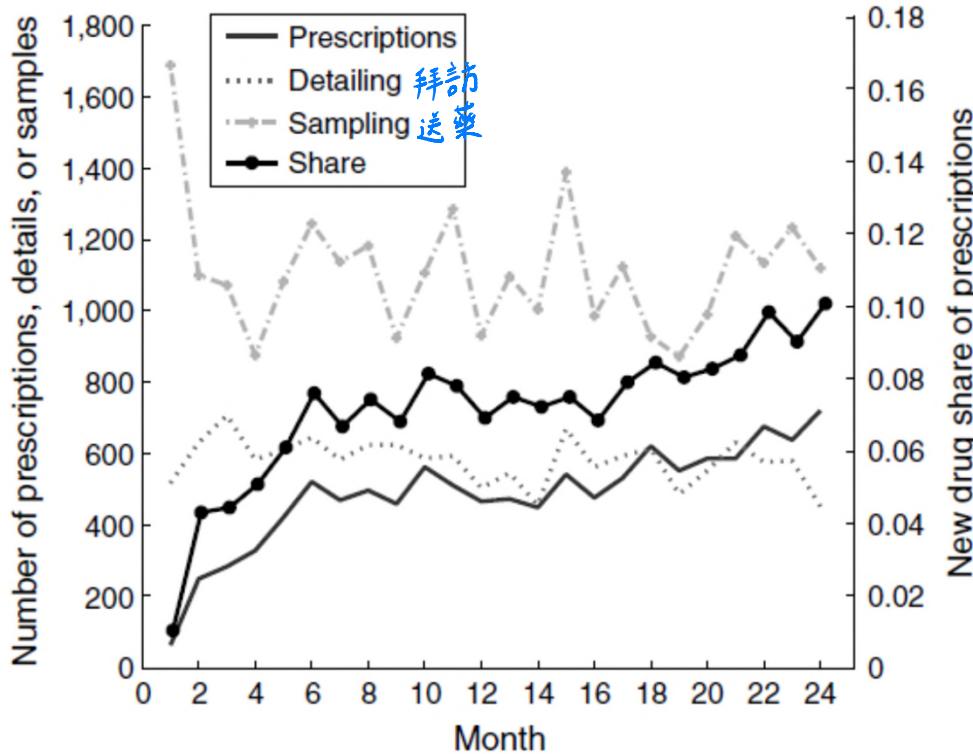


Binomial Distribution in Action

- 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？



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

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

data



Binomial Distribution in Action

- Two coins of a customer

✓ shown

show 1 (prob p)

no-show 0 (prob 1- p)

hidden

alive A (prob 1-θ)

dead D (prob θ)

data likelihood

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

代表都還alive

→ dead

| er | 1996 | 1997 | 1998 | 1999 | ← | → 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 | |



Decision Tree for Sourcing

- $X \sim \text{discrete Uniform}(a, b)$

Outcome of a fair Die

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 $[a, a+1, \dots, b]$

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 $D \sim \text{Uniform}(2000, 8000)$

strong $D \sim \text{Uniform}(6000, 14000)$

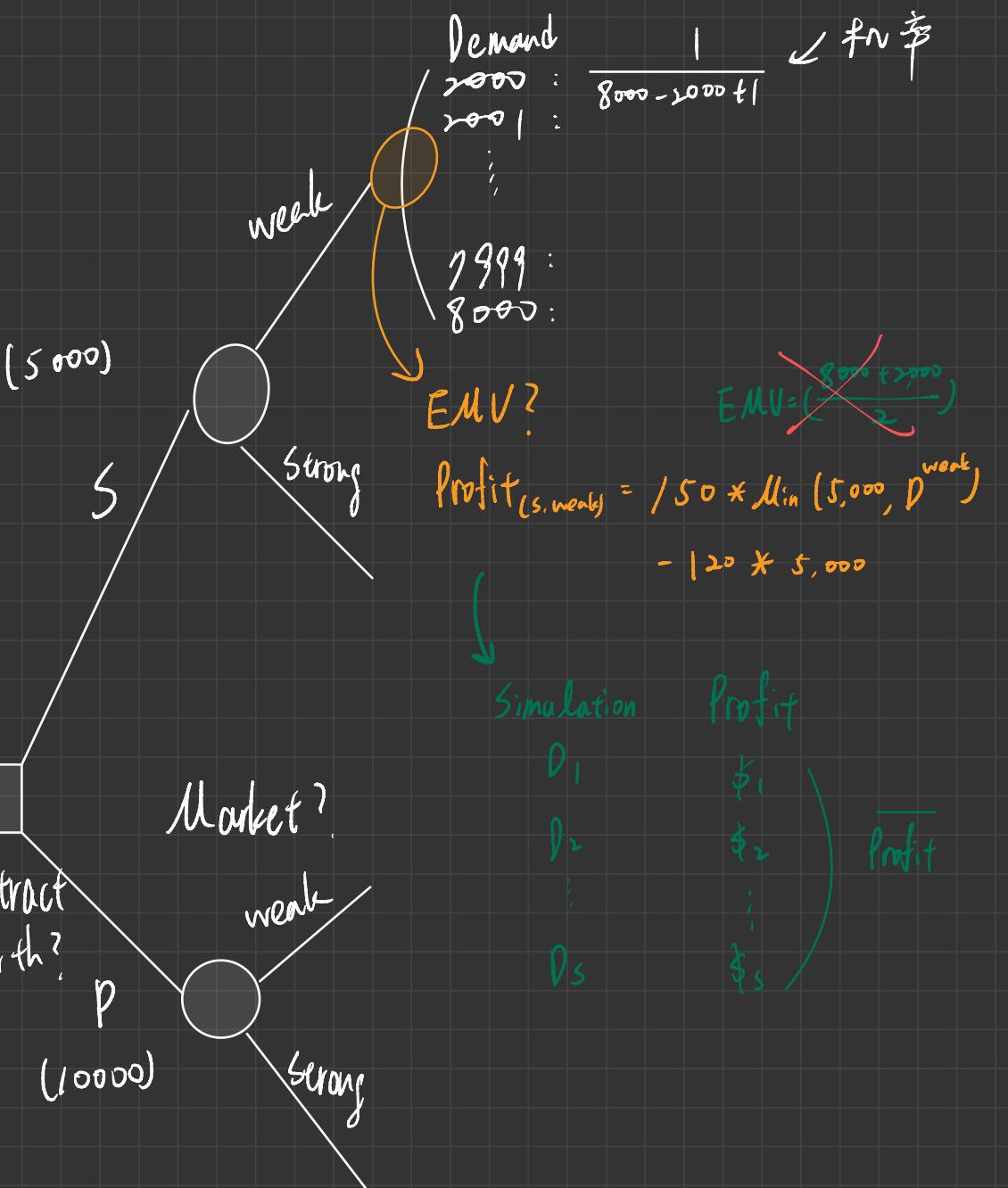
| | <u>Sweden (S)</u> | <u>Poland (P)</u> |
|----------------|-------------------|-------------------|
| Order Quantity | 5,000 units | 10,000 units |
| Fixed Charge | 0€ | 50,000€ |
| Unit Cost | 120€ | 100€ |

数量

how can we represent this in a decision tree?

which supplier should we choose from?





(fresh food), 現量 life time (hour)

item Q (D) 1. 2. , 23

A Q_A $\frac{0.1}{(f=1)}$ 0. 1. 0.

賣出的現量

B

C

:

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

$$= f(x_1, x_2, \dots, x_m)$$

東西被買走的機率

Model Formulation & Strategy Optimization

- ## • Hello Kitty Magnets



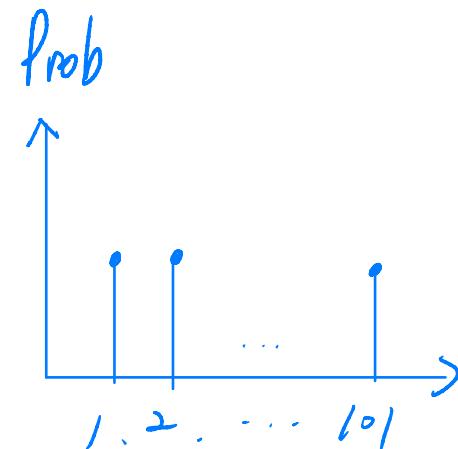
Model Formulation & Strategy Optimization

- **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

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

$$5 \times 101 = 505 \text{ 最低}$$

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

any strategy to beat this and reduce cost?

Decision model

decision variable

parameter

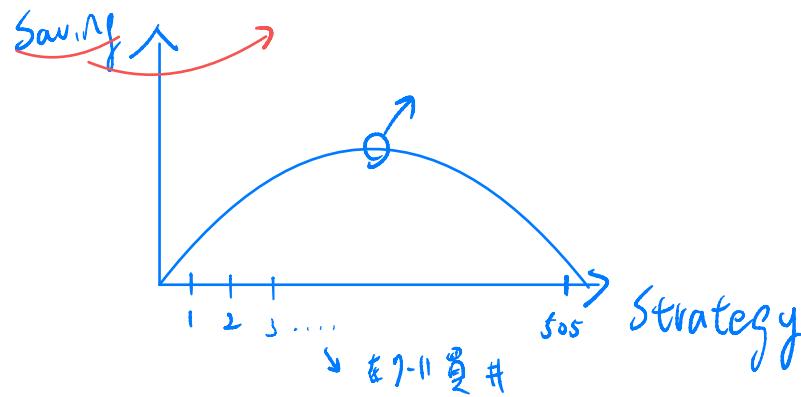
objective

uncertainty

random variable

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

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



What are the **assumptions**?

可以任意買 Missing with flat price

No exchange → buy x from 7-11,
get unique xx , left $x - xx$, 例如 $101 - xx$
Assume 1:1 exchange ↴

Discrete RVs for Decision Modeling

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

- Three discrete distributions for **infinite values**

$X \sim \text{Poisson}(\lambda)$

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

$$E[X] = \text{Var}[X] = \lambda$$

$Y \sim \text{Poisson}(\lambda_1)$ & $Z \sim \text{Poisson}(\lambda_2) \Rightarrow W = Y + Z \sim \text{Poisson}(\lambda_1 + \lambda_2)$

$Y \sim \text{Poisson}(\lambda_1)$ & $Z \sim \text{Poisson}(\lambda_2) \Rightarrow W = Y - Z \sim \text{Skellam}$

check `scipy.stats.skellam`

Discrete random count

ex:

Demand, Visits, Download.



離散隨機變數取特定值的機率

*: Probability Mass Function (PMF)] Poisson:

機率質量函數 $P(X=x) = f(x)$

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

λ : 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**

$X \sim \text{Poisson}(\lambda)$

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

$$E[X] = \text{Var}[X] = \lambda$$

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

$X: \# \text{ of attacks/week} \sim \text{Poisson}(\lambda)$

$$E(X) = \lambda$$

$$\text{mean} = \bar{\lambda} = 1.9038$$

$$P(X \leq 6) \geq 0.99$$

attacks

service level

status.poisson.ppf
ppf \Rightarrow Percent Point Function
 $=$ Inverse CDF



Discrete RVs for Decision Modeling

- Three discrete distributions for infinite values

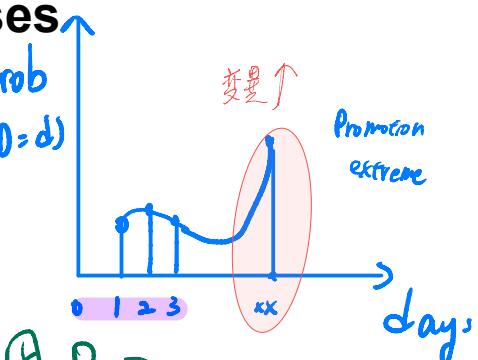
$X \sim \text{Poisson}(\lambda) \rightarrow \frac{\text{count}}{\text{time}}$ *变量数和期望值相同

a very, very, very important distribution for counting processes

$E[X] = \text{Var}[X] = \lambda$ 平均值

1. $Y \sim \text{Poisson}(\lambda_1)$ & $Z \sim \text{Poisson}(\lambda_2) \Rightarrow W = Y + Z \sim \text{Poisson}(\lambda_1 + \lambda_2)$ ($D=d$)

2. $Y \sim \text{Poisson}(\lambda_1)$ & $Z \sim \text{Poisson}(\lambda_2) \Rightarrow W = Y - Z \sim \text{Skellam}$
check `scipy.stats.skellam`



$X \sim \text{negative binomial}(n, p)$

useful for modeling a count data distribution with $\text{Var}[X]/E[X] \gg 1$

example: modeling retail sales

$$p = \frac{\mu}{\sigma^2} \leftarrow \frac{E(x)}{\text{Var}(x)} \leftarrow \frac{4.899}{10.076} \approx 0.4899$$

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

related to Bernoulli (p) processes

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

example: detecting empty shelves / missing members



D_2

[9-11]

[Post]

D_1 (Demand)

[9-11]



Road



$$\lambda_2 = 2\% \text{ day}$$

$$\lambda_1 = 3\% \text{ day}$$

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

100,000

80k items

1.

$\frac{d_1}{d_2}$

Prioritise

2.

$\frac{d_2}{d_3}$

15k items

3.

\vdots

4.

\vdots

80k

d80k

文長軟林

0000 00 | 00 1 ... 00000 2

題別
Category

大

要直線才能找到，

1, 2, k

1. 有米
0. 沒米

會
直
3.
3.

$$\text{eg. } 0.98 = p \rightarrow p > 1 \text{ 的 prob}$$

:

$$0.02 = q \rightarrow q > 0 \text{ 的 prob}$$

：

$$0: \text{一天沒去} = 0.02 \\ 00: \text{連續 2 天沒去} = (0.02)^2 \\ 0000: \text{連續 4 天沒去} = 0.0004$$

q^1

if $\# \text{ of straight} < \text{manages}$
 $l \leq \underline{\text{tolerance}}$

