More Probability Distributions for Decision-Making

Howard Hao-Chun Chuang (莊皓鈞)

Professor
College of Commerce
National Chengchi University

October, 2024
Taipei, Taiwan

Case Study: Operations at Conley Fisheries

- **A fisheries company operates one boat, assuming**

everyday it is able to capture full-boat load 3,500 lbs (磅/里拉) coldfish
operating expenses are $10,000 per day
sell all the fish at either Gloucester or Rockport

Revenue = PR * min(D, fish量) - Operating expenses

The daily demand at Gloucester is infinity & the price is fixed at 3.25/lb
The daily demand at Rockport is a RV *D* (Random Variables)
The price of Rockport is a RV *PR*
***PR* ~ Normal(μ_{PR}=3.65/lb, σ_{PR}=0.2/lb)**

Demand (lbs. of codfish)	Probability
0	0.02
1,000	0.03
2,000	0.05
3,000	0.08
4,000	0.33
5,000	0.29
6,000	0.20

What is the decision variable?
What are the parameters?
What is the daily earning w/o uncertainty?

在 Gloucester

Case Study: Operations at Conley Fisheries

隨機

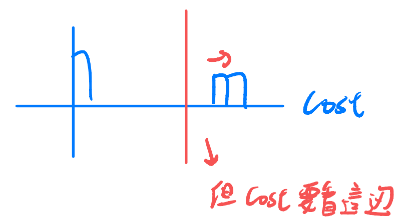
- Let *F* be the stochastic earnings from selling the fish at Rockport
  - What is the mathematical form of *F*?

    F = PR * Min (3500, D) - 10,000

  - What is the shape of probability distribution of *F*?
  - What is $P(F > \$ 1{,}375)$? What is $P(F < \$ 0)$?
  - What is the 95% confidence interval of *F*?
  - What is the conditional value at risk (CVarR) of *F*? Say at 5%?
    - Definition: The "expected shortfall at q% level" is the expected return on the portfolio in the worst q% of cases

      在最爛的5%的平均

      最壞 q% 的平均 =)



  - What else can be random? To name a few…
    - daily demand & daily price at Gloucester
    - the quantity of fish caught everyday
    - possible operating hours (and hence operating expenses) everyday
  - We need extra probability distributions to model those issues

$$F = PR * Min(D, 3500) - 10,000$$

Simultion

1、	PR_1	D_1	F_1
2、	PR_2	D_2	F_2
3、	PR_3	D_3	F_3
⋮	⋮	⋮	⋮
S	PR_S	D_S	F_S

CVaR

如果不幸落到最糟Q%的收益

倒數

倒數的Q%算Average

Case Study: Operations at Conley Fisheries

full yield 良率

- **A fisheries company operates one boat, assuming**

daily price at Gloucester is *random* and CORR(PR_{Rock}, PR_{Glou}) = 0

Clou Rock

PR_{Glou} ~ Normal (μ=3.5/lb, σ=0.5/lb)

PR_{Rock} ~ Normal(μ=3.65/lb, σ=0.2/lb)

價 Nor Nor

Demand triangu

一般 3,500 lbs

daily demand at Gloucester (D_{Glou}) is *random*

D_{Glou} ~ triangular (*min, max, mode*) (possible outcomes in (*min*, *max*))

set *min*=2000, *max*=6000, & *mode*=5000 (random.triangular in numpy)

隨機比例

everyday it is able to capture a *random fraction* (*FRAC*) of the full-boat load 3,500 lbs coldfish

FRAC ~ ??? (possible outcomes in 0～1 **)**

[0,1]

full capacity

uniform (a,b), triangular (-,-,-), Beta (a,b) distribution

What is the decision? What are the sources of uncertainty?

0～1

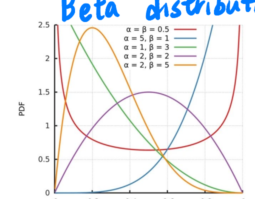
× 常態 不行 ⇒ 因負到正

√ 三角 0.6 0.75 1

if Manager

0.7 0.9 FRAC

Beta distribution



PR 下單

Purchase Request / FCST $\xrightarrow{\text{轉換率 } \frac{PO}{PR}}$ PO Purchase Order 實際下單

Case Study: Operations at Conley Fisheries

- ***G*: The stochastic earnings** from selling the fish at Gloucester

  What is the mathematical form of *G*?

  $G = PR_{Glou} * Min(3500 * FRAC, D_{Glou}) - 10{,}000$

  What is the shape of probability distribution of *G*?

  What is *P*(*G* > \$ 1,375)? What is *P*(*G* < \$ 0)?

  What is the 95% confidence interval of *G*?

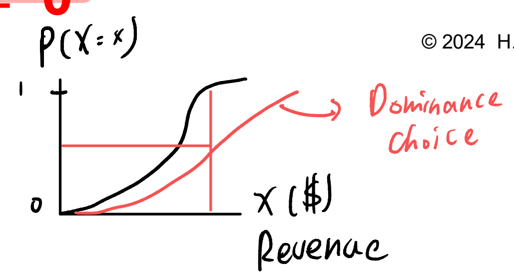
  What is the conditional value at risk (CVarR) of *G*? Say at 5%?

  Definition: The "expected shortfall at q% level" is the expected return on the portfolio in the worst q% of cases

- ***F*: The stochastic earnings from selling the fish at Rockport**

  Comparing *G* with *F*, where would you like to sell the fish? Why?

For HW: What if $CORR(PR_{Rock}, PR_{Glou}) \neq 0$



Modeling Stochastic Time-To-Event

- ***X* ~ exponential(*λ*)** 連續 **(closely related to Poisson(*λ*))** 離散

  possible outcomes *x* in [0, Infinity)

  scipy.stats.expon(*scale*) in Python

  λ* is a rate parameter (i.e., counts per time unit), *scale*=1/*λ

  Example: Modeling random processing time

  A call center agent on average spends 10 minutes per call. Assuming the time for a call (*T*) follows an exponential(*λ*) distribution.

  What is the distribution of the number of calls processed within 60 minutes?

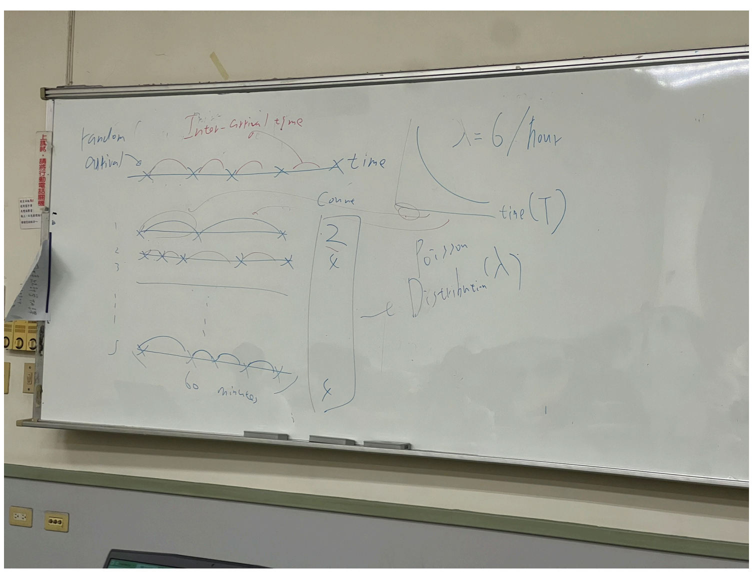
$\lambda \Rightarrow$ Rate

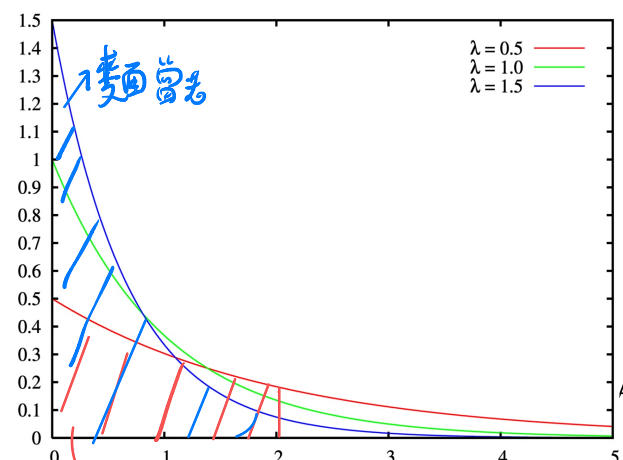
$\frac{1}{\lambda} \Rightarrow$ scale

平均每個時間要處理幾個人

$\lambda = 0.1$ call / min

$\lambda = 6$ calls / 60 mins





有較多Case 0~2 mins 的機率

↑ Utilization / Idle Time ↓ ⟷ Waiting Time ↑

Trade off

Case: Analysis of Queues (see Sadri et al 2023 too)

- Suppose a service is provided 8 hours per day

  Customers will NOT be accepted 30 minutes before the service is close

  Stochastic time

  - Arrival rate ~ exponential(6 customers per hour)
  - Service rate ~ exponential(3 customers per hour)
  - Time willing to wait: An exponential RV with mean 0.25 hours

  Idle time $\Rightarrow \frac{480 \times 2 - \text{use}}{480 \times 2}$

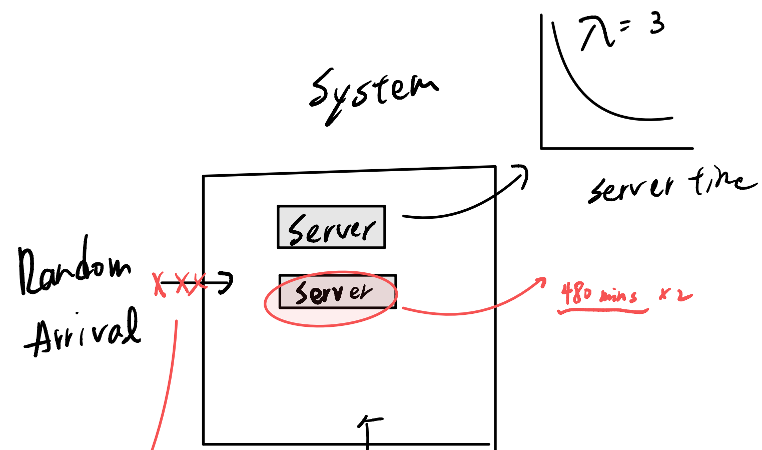
  Operational cost

  - Customer waiting cost: \$5 per hour
  - Idle server cost: \$10 per hour
  - Cost of losing a customer: \$30 per customer

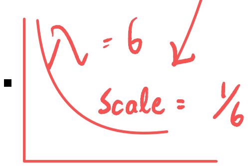
  What will be a **focal decision variable**?

  \# of servers → how much capacity

  What are the **key performance indicators**?



  The system has **Reneging** (等太久就不等) & NO **Balking** (看一看不進來).

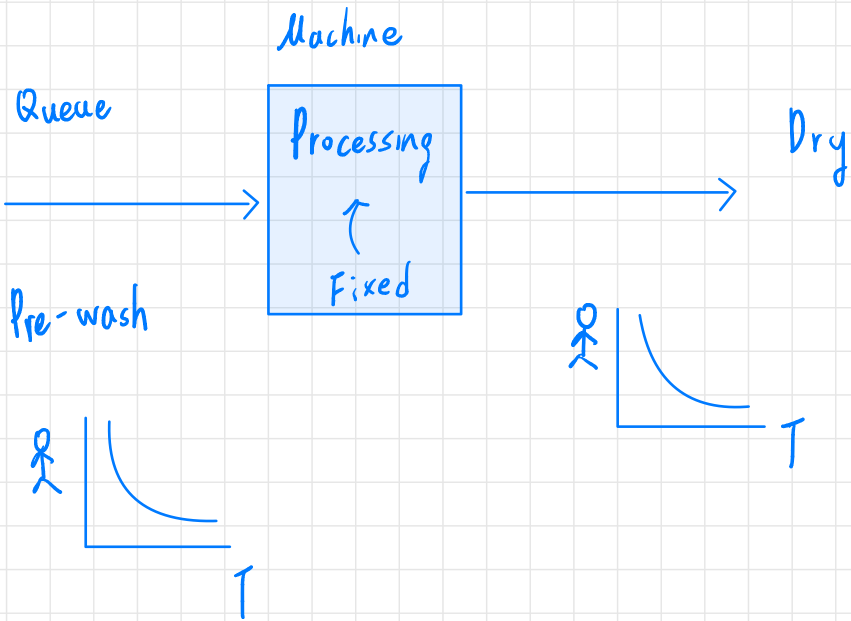


Enter the quese, but quit.

猶豫不決

抵達時間的機率分配

洗車:



為什麼很多真實世界的機率分配長得像子數函數？

飛機航班.

Modeling Stochastic Time-To-Event

沒有記憶性

- $X \sim$ **exponential(λ)** (closely related to Poisson(λ))

  possible outcomes x in [0, Infinity) 非負 Real number

  scipy.stats.expon(*scale*) in Python

  λ is a rate parameter (i.e., counts per time unit), *scale*=1/λ

  [illegible]好

  Example: Modeling random processing time

  A call center agent on average spends 10 minutes per call. Assuming the time for a call (T) follows an exponential(λ) distribution.

  What is the distribution of the number of calls processed within 60 minutes?

  沒記憶性

  塞車… 約會…

  Memoryless: $P(X>m) = P(X>n+m \mid X>n)$

  For instance, $P(T>5 \text{ minutes}) = P(T>15 \text{ mins} \mid T>10 \text{ mins})$

  Verify this property through simulation!

  我等了10mins 還要在等5mins

This is an important distribution for modeling time-to-XX. But it is also limited in terms of capturing stochastic time in the real world…

Case: Project Duration Simulation

- **A project has 8 tasks, each with uncertain time to complete**

  each activity time ~ ??? & independent of each other

  why not use the Normal(*mu, sigma*) distribution ?

Activity	Mean Time	Stdev. Time	Predecessors
A (pour foundation)	20	7	—
B (build support towers)	50*	10*	A
C (assemble cables)	60	12	—
D (install cables)	15	3	B, C
E (excavate connecting road)	65	30	—
F (pave connecting road)	35	15	E
G (install bridge roadbed)	30	5	D
H (pave bridge)	10	3	G

*Activity B's time could be reduced by a more expensive subcontractor.

Modeling Stochastic Time-To-Event

a 可考慮多個獨立事件

- ✱ X ~ gamma(a, $scale$)
  - possible outcomes x in (0, Infinity)
  - scipy.stats.gamma(a, $scale$) in Python
    - Be careful about the parameterization of gamma distribution
  - E[X] = a*$scale$; Var[X] = a*$scale^2$
    - Both a & $scale$ > 0

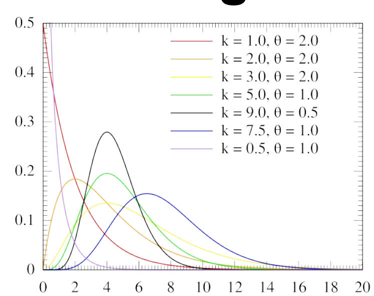
$a=1 \rightarrow \exp(\lambda)$

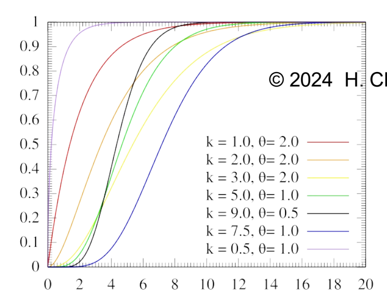
$Var(x) = \frac{1}{\sqrt{a}}$

- Let Y = the sum of k exponential(λ). Y ~ gamma(a=k, $scale$=1/λ)
- Example: Modeling random processing time
  - A call center agent on average spends 10 minutes per call. Assuming the time for a call (T) follows an exponential(λ) distribution.
  - What is the distribution of the time to process 5 calls?

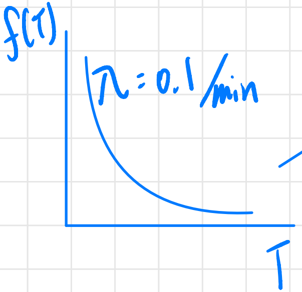
累積

- This is a flexible distribution for modeling time-to-XX and non-negative continuous random outcomes





What is the distribution of the time to process 5 calls?



$t_1 + t_2 + t_3 + t_4 + t_5 = X_1$

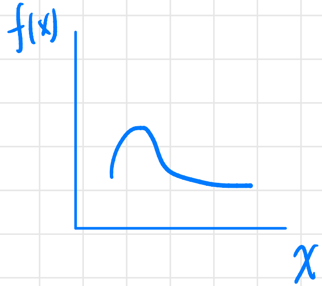
$\underline{t_1 + t_2 + t_3 + t_4 + t_5} = X_2$

$\vdots$

gamma 由子故来……

抽五次

$f(x) \quad \text{gamma}\left(\underline{a=5,}\ \text{scale} = 1/\lambda = 1/0.1\right)$



- k：形狀參數（shape parameter），決定分布的形狀。
- \theta：尺度參數（scale parameter），調整數據的範圍（拉伸或壓縮）。
- \Gamma(k)：Gamma 函數。
- 性質：
- 平均值 \mu = k \cdot \theta
- 方差 \sigma^2 = k \cdot \theta^2

- \alpha：形狀參數（shape parameter），與 k 相同，決定分布的形狀。
- \beta：速率參數（rate parameter），表示數據的頻率，與 \theta = 1 / \beta 成反比。
- \Gamma(\alpha)：Gamma 函數。
- 性質：
- 平均值 \mu = \frac{\alpha}{\beta}
- 方差 \sigma^2 = \frac{\alpha}{\beta^2}

Gamma 分配的適用性

- 連續非對稱數據： Gamma 分配適用於正值且 右偏分布 的連續型數據（與對數正態分布類似），例如購物金額。
- 靈活的形狀： Gamma 分配根據其形狀參數 \alpha 和尺度參數 \beta 可生成從右偏到接近對稱的多種分布形狀。
- 當 \alpha \to 1 ，Gamma 分布接近指數分布（極右偏）。
- 當 \alpha 增加，分布逐漸趨於對稱。
- 應用情境： Gamma 分配常用於建模正值的隨機變數，例如支出、時間持續量和金額類數據。

Case: Project Duration Simulation

- **A project has 8 tasks, each with uncertain time to complete**
  - **each activity time ~ gamma(*a*, *scale*) & independent of each other**
    - **any other distributions for activity time?**

Activity	Mean Time	Stdev. Time	Predecessors
A (pour foundation)	20	7	—
B (build support towers)	50*	10*	A
C (assemble cables)	60	12	—
D (install cables)	15	3	B, C
E (excavate connecting road)	65	30	—
F (pave connecting road)	35	15	E
G (install bridge roadbed)	30	5	D
H (pave bridge)	10	3	G

*Activity B's time could be reduced by a more expensive subcontractor.

$E(X)$ $\sqrt{Var(X)}$

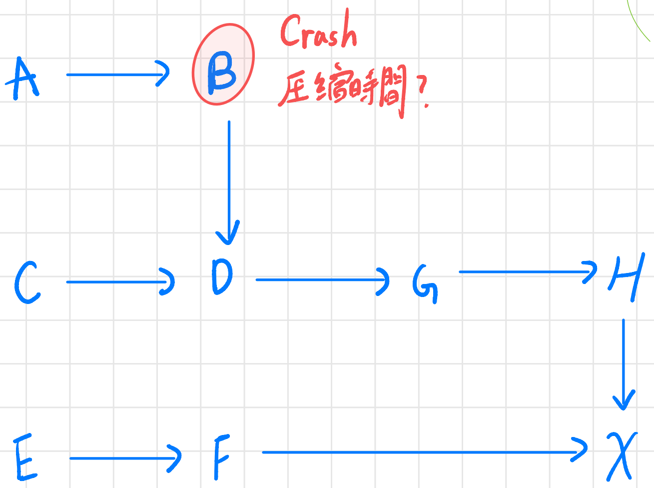
$\Rightarrow 20 = a * scale, \quad 49 = a * scale^2$

$scale = \frac{49}{20} \quad a = \frac{20}{scale}$

The goal is to finish the project within 130 days, and the delay penalty is \$100,000 per day. No bonus from finishing the project earlier.

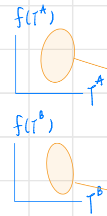
What are the questions we can answer through simulation modeling?

1. 視覺化

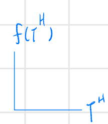


常態不可以，因為
① 負到正
② 現實大多是 gamma

2.

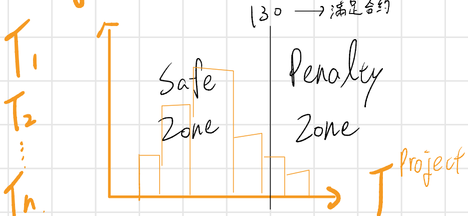


	Start	End
A	0.	
B	End A.	
C	0.	
D	Max (End A, End C)	
E	0.	
F	End E	
G	End D	
H	End G	



X: Max (End H, End F)

Project Duration ($T^{Project}$)



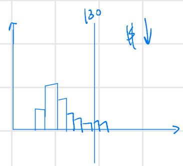
1. 如何抽樣（a, λ怎麼來）
2. 如果可以加速，要怎麼做都

1.

$E(x) = a * scale$, $Var = a * scale^2$

(Mean) (stdev^2)

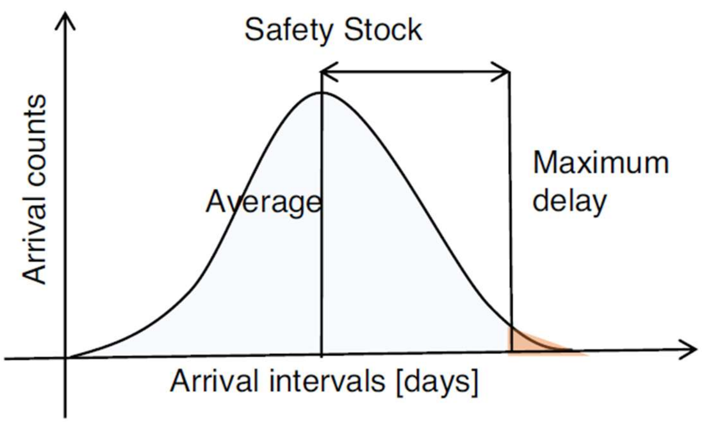
2.

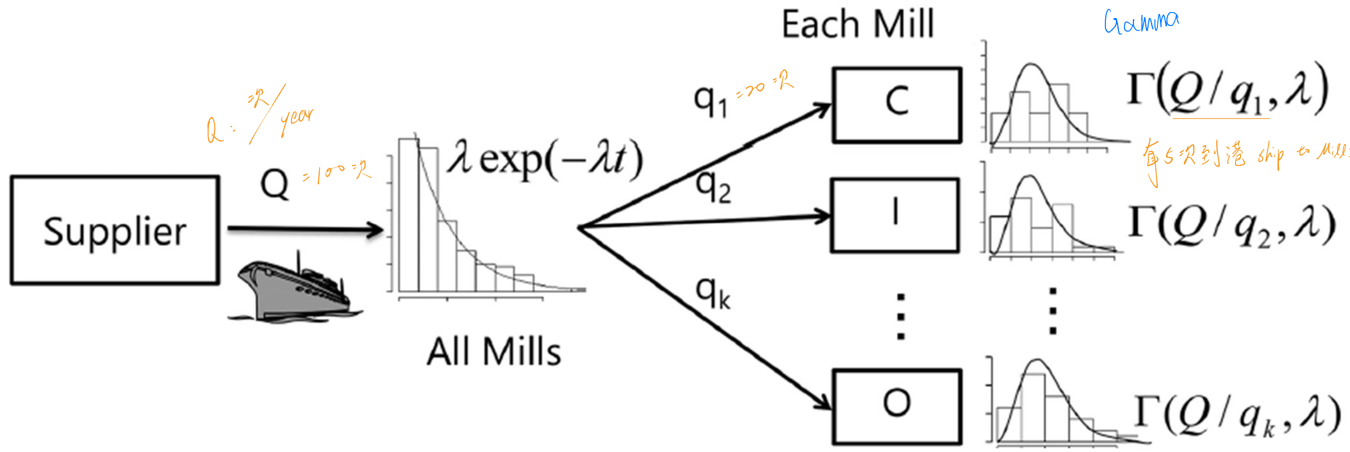


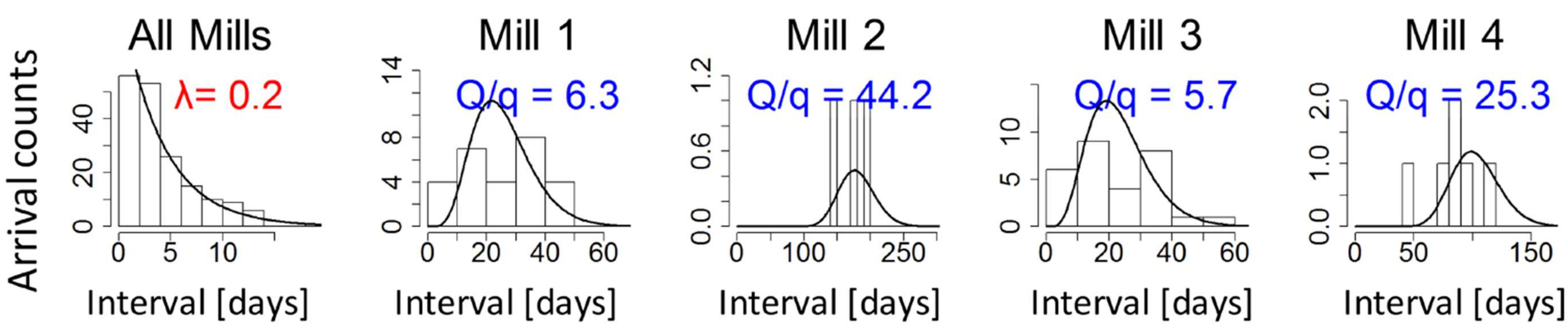
Project B'

Case: Nippon Steel Corporation

- **Simulating arrival intervals for safety stock decisions**







Discrete Probability Distributions

- For a **discrete random variable (RV)** X, we have learnt

  $P(X = x)$ is the probability distribution function   Countable

  $P(X <= x) = \sum_x P(X = x)$ is the cumulative distribution function (CDF)

  $X \sim$ **Discrete Uniform** (a, b) ∃ 骰子 

    possible outcomes $x=a, a+1,\ldots, b$

  $X \sim$ **Bernoulli**(p)

    possible outcomes $x=0, 1$ Coin tossing

  $X \sim$ **Binomial**(n, p) (sum of n Bernoulli(p) RVs)

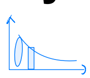
    possible outcomes $x=0, 1,\ldots, n$ n個 Coin tossing，總共幾個1.

  $X \sim$ **Geometric**(p) # of zeros till the first one，顧客追蹤.

    possible outcomes $x=0, 1,\ldots,$ Infinity

  $X \sim$ **Poisson**(λ) Count/times ⇒ $\underbrace{E(x)=Var(x)=\lambda}$ 缺点，but 通常 $Var(x) > E(x)$

    possible outcomes $x=0, 1,\ldots,$ Infinity

  $X \sim$ **Negative Binomial** (n, p) 跑得慢 

    possible outcomes $x=0, 1,\ldots,$ Infinity

Continuous Probability Distributions

- For a **continuous random variable (RV)** ***X***, we have learnt

  $F(x)=P(X <= x)=\int_{-\infty}^{x} f(x)dx$ is the CDF

  $f(x) = dF(x)/dx$ is the probability density function

  X **~ Uniform (*a*, *b*)**

  possible outcomes *x* in [*a*, *b*]

  X **~ Normal(*mu*, *sigma*)**

  possible outcomes *x* in [-Infinity, Infinity]

  X **~ Triangular (*min*, *mode*, *max*)**

  possible outcomes *x* in [*min*, *max*]

  X **~ Beta (*a*, *b*)**

  possible outcomes *x* in (0, 1)

  X **~ Exponential(*λ*)**

  possible outcomes *x* in [0, Infinity)

  X **~ Gamma(*a*, *scale*)**

  possible outcomes *x* in (0, Infinity)

1. Triangular 分配（ \text{min}, \text{mode}, \text{max} ）

- 適用情境：
- 模擬未知分布但能估計範圍和模式的數據，例如專案時間估算（最短、最可能和最長完成時間）。
- 原因：
- Triangular 分配是一種簡單且直觀的分布，適合在數據不足或難以確定分布形狀時使用。
- 提供了一個粗略的範圍並強調最可能值。

2. Beta 分配（ a, b ）

- 適用情境：
- 比例數據（如轉換率）或機率數據建模，數據範圍在（0, 1）。
- 原因：
- Beta 分配的靈活性（形狀由 a 和 b 決定）適合描述比例的右偏、左偏或對稱分布。

3. Exponential 分配（ \lambda ）

- 適用情境：
- 模擬隨機事件的等待時間，如顧客到達的間隔時間或機器故障時間。
- 原因：
- 具有無記憶性，適合建模單次事件的等待時間。

4. Gamma 分配（ a, \text{scale} ）

- 適用情境：
- 多次隨機事件的總等待時間，例如完成多個任務所需的總時間。
- 原因：
- Gamma 分配可累積多個指數分布，適合描述多次事件的累積時間。

2. Geometric 分配（ p ）

- 適用情境：
- 模擬第一次成功發生前的失敗次數（例如等待第一次到達的客戶失敗次數）。
- 原因：
- 記錄成功前的零次數，分布具有無記憶性。

3. Poisson 分配（ \lambda ）

- 適用情境：
- 模擬單位時間或空間內事件發生的次數（如每小時顧客到達數）。
- 原因：
- 適合描述隨機事件發生次數，平均值和方差均為 \lambda 。

4. Negative Binomial 分配（ n, p ）

- 適用情境：
- 模擬固定次數成功前的失敗次數（例如完成 5 次成交前的失敗訂單次數）。
- 原因：
- 是 Geometric 分配的延伸，適合描述累積成功的過程。