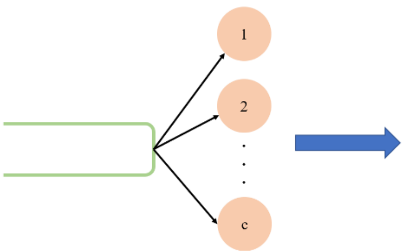
Exercise: General Formulation with G/G/C, G/G/1, G/G/c/k

對於一個排隊系統而言，我們最想要知道的是他在穩定狀態時候的系統效能。所謂的穩定狀態，其實就是一個長時間觀察取平均的概念，數學上也就是所謂的計算期望值。當一個系統到達穩定狀態，也就反映出了一個系統的效能，例如可以知道一個系統大概會面臨多少使用者，型對應的我們可以配置多少服務器。

但是在實際情況中，計算穩定狀態是十分困難的事情。舉例來說，計算一個隊伍的人一周內大概要吃多少食物，這是很難計算的，例如有的人今天運動量大進食比較多，有的人今天生病了毫無食欲，因此難以評估一個準確的需求。評估準確需求是一個排隊系統希望能夠計算出來的，這是排隊理論的一個重要的目標。

G / G / C Queue



現有一個排隊系統，Arrival Rate 服從任意分佈，service 的時間服從任意分佈，系統內的 server 數量為 c 個，系統大小無限。

結論：

i. Traffic Intensity (負載): $\rho = \frac{\lambda}{c\mu} \qquad (\rho < 1)$

ii. $W = W_q + \frac{1}{\mu}$ （在隊列的等待時間加上服務時間）

iii. Utilization (Traffic Intensity) of any server U

$$U = \Pr\{\text{a server is busy}\} = \rho = \frac{\lambda}{c\mu}$$

有 1 台服務器忙碌，換句話說就是系統在忙碌，也就是 1-Pr{系統不忙碌} ，這部

分完全可以用負載來等價。舉例負載為 0.6，可以想象成有 60%的時間系統是忙碌的。

iv. $L = L_q + \mathrm{E}[\text{number of users in server}]$

$$= L_q + cU$$

$$= L_q + c * \frac{\lambda}{c\mu}$$ (例有 100 台服務器，0.6 負載表示有 60 台在服務，即有 60 人)

E.g. 現有一個排隊系統，Arrival Rate 服從任意分佈，$\lambda = 1/10$, service 的時間服從任意分佈，$1/\mu = 5$, $\mu = 1/5$, 系統內的 server 數量為 c=3 個。

i. Traffic Intensity (負載): $\rho = \frac{\lambda}{c\mu} = (\frac{1}{10}) / (\frac{1}{5} * 3) = \frac{1}{6}$

ii. Utilization (Traffic Intensity) of any server U

$$\mathrm{U} = \Pr\{\text{a server is busy}\} = \rho = \frac{\lambda}{c\mu} = \frac{1}{6}$$

✧ 現系統的負載率為 1/6，即一段時間內 (假設 60 分鐘)，有 1/6 的時間處於忙碌狀態，而有 5/6 的時間處於空閑狀態。現在每分鐘進入系統看一次，看系統是否空閑，自然就是有 50 次觀察到的結果是空閑，有 10 次忙碌。

$$\mathrm{U} = \Pr\{\text{a server is busy}\}$$

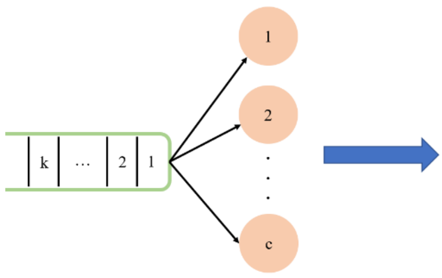
$$= \frac{\text{觀察到忙碌的次數}}{\text{觀察的總次數}} = \frac{1}{6}$$

G / G / 1 Queue

現有一個排隊系統，Arrival Rate 服從任意分佈，service 的時間服從任意分佈，系統內的 server 數量為 1 個，系統大小無限。

$$L = L_q + \frac{\lambda}{\mu} = L_q + (1 - P_0)$$ （P_0指的是系統空閑的機率）

G / G / c / k Queue



現有一個排隊系統，Arrival Rate 服從任意分佈，service 的時間服從任意分佈，系統內的 server 數量為 c 個，系統大小為 k。

結論：

i. Pr{ blocking } $\triangleq P_B = P_k$

P_k 指的是系統有 k 個人的機率。當系統有 k 個人的時候，說明系統就塞不下其他人了，此時有人進入到系統就會被 block。那麼對一個人來說，被 block 的機率就是系統處於 k 個人的時候的機率。

ii. $\boldsymbol{\lambda_{eff} = \lambda\,(1 - P_B)}$ ，稱爲有效到達率。

iii. Utilization (Traffic Intensity) of any server U

$$\mathrm{U} = \Pr\{\text{a server is busy}\} = \rho = \frac{\boldsymbol{\lambda\,(1 - P_B)}}{c\mu} = \frac{\boldsymbol{\lambda_{eff}}}{c\mu}$$

Summary

✧ **事實上，不管是 λ，還是 μ，在上面的講述中都是一個期望值。在統計的意義中，期望值是均值的概念，需要經過一段時間的觀察統計求得。但實際系統中，期望值是很難知道的，雖然可以觀察一段時間，但是觀察的這段時間的所有統計量，對於接下來的時間沒有任何意義。**

✧ **因此，會對排隊系統做出假設，假設說 λ 或 μ 服從某些分佈，通過這些分佈的性質來一步步計算出系統的穩定狀態，從而對系統進行效能分析。因此接下來的內容，λ 或 μ 會被賦予機率分佈。**