

wq-3. M/M/S 待ち行列, アーランの即時式モデル

(待ち行列の数理)

<https://www.kkaneko.jp/cc/wq/index.html>

金子邦彦



アウトライン



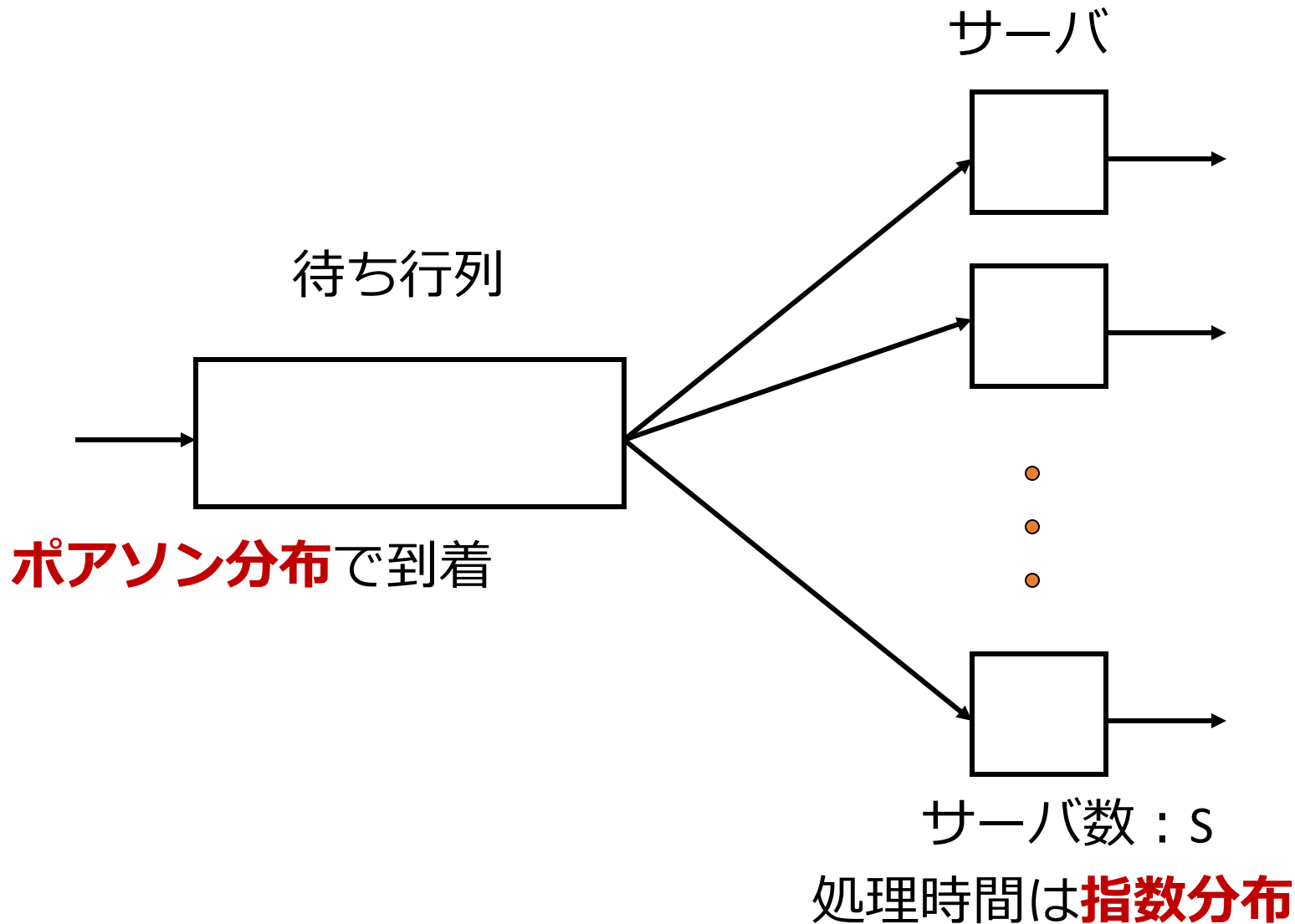
3-1 M/M/S 待ち行列

3-2 アーランの即時式モデル

3-1 M/M/S 待ち行列

- X: 到着過程
→ **ポアソン分布**のとき「M」と書く
- Y: 処理時間分布
→ **指数分布**のとき「M」と書く
- Z: サーバ数
個数を限定しないとき「S」と書く
- K: **待ち行列の長さ**の制限
制限しないときは何も書かない

M/M/S 待ち行列モデル



M/M/S 待ち行列モデルの解析



- 状態遷移
 - システム内のジョブ数を「状態」と見る
- 定常状態
 - 定常状態での、各「状態」の確率を求める
- システムの「処理率」を求める

状態 0 : システム内のジョブ数が 0

状態 1 : システム内のジョブ数が 1

⋮

状態 S: システム内のジョブ数が S

⋮

- ジョブが到着 :

状態 k から状態 $k + 1$ に遷移

- ジョブが完了 :

状態 k から状態 $k - 1$ に遷移

遷移確率に関する方程式



- 微小時間 Δt (限りなく 0 に近い) についての式
- 「ジョブの到着」と「ジョブの完了」は、「同時」には起きない
- 「ジョブの完了」の確率は、 Δt と μ の式
- 「ジョブの到着」の確率は、 Δt と λ の式

「ジョブの完了」に関する式



- $k \leq S$ ならば (k は状態番号)

- サーバに空きがある

$$k \mu \Delta t (1 - \mu \Delta t)^{k-1} \doteq k \mu \Delta t$$

- $k > S$ ならば (k は状態番号)

- サーバは全て忙しい

$$S \mu \Delta t (1 - \mu \Delta t)^{S-1} \doteq S \mu \Delta t$$

状態遷移



- ジョブが到着 :

状態 k から状態 $k + 1$ に遷移

$$\lambda \Delta t$$

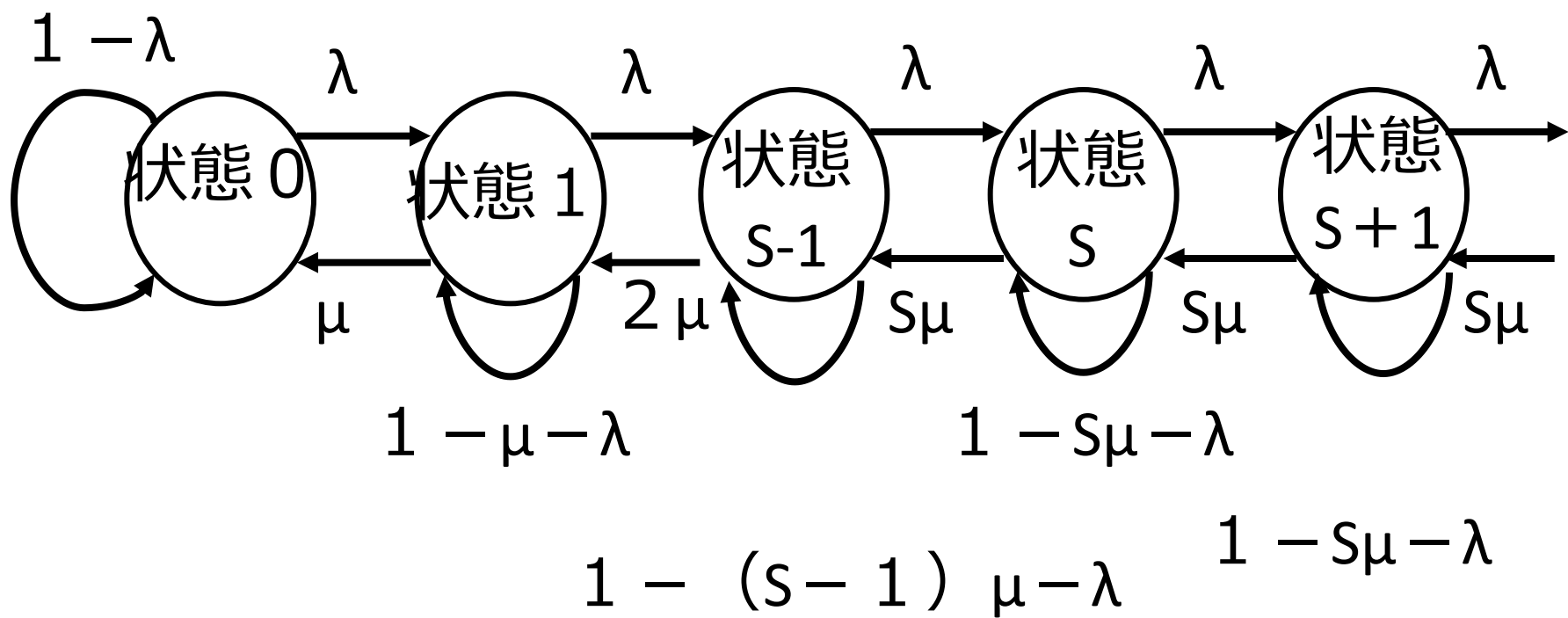
- ジョブが完了 :

状態 k から状態 $k - 1$ に遷移

$$k \leq S \text{ ならば : } k \mu \Delta t$$

$$k > S \text{ ならば : } S \mu \Delta t$$

状態遷移図



定常状態方程式



$0 < n < S$ では

$$(\lambda + n\mu) P_n = \lambda P_{n-1} + (n+1)\mu P_{n+1}$$

$S \leq n$ では

$$(\lambda + S\mu) P_n = \lambda P_{n-1} + S\mu P_{n+1}$$

$$\lambda P_0 = \mu P_1$$

定常確率をP0で表す



$0 < n < S$ では

$$P_n = \left(\frac{\lambda}{\mu} \right)^n \frac{P_0}{n!}$$

$S \leq n$ では

$$P_n = \left(\frac{\lambda}{\mu} \right)^n \frac{P_0}{S! S^{n-S}}$$

$$\rho = \lambda / \mu$$

- $\lambda \Delta t$: 「時間 $(t, t + \Delta t)$ に到着するジョブ数」の平均
- $\mu \Delta t$: 「サーバがジョブを処理中の間, Δt 内に完了する処理数」の平均

待ち合わせが生じる確率



- システム内のジョブ数が S 以上

- $P_{\text{queue}} = \sum_{n=S}^{\infty} P_n$
 $= \sum_{n=S}^{\infty} \left(\frac{\lambda}{\mu} \right)^n \frac{P_0}{S! S^{n-S}}$
 $= \sum_{n=S}^{\infty} P_0 \frac{(S\rho)^n}{S!} \frac{1}{S^{n-S}}$

3-2 アーランの即時式モデル

X: 到着過程

→ **ポアソン分布**のとき「M」と書く

Y: 処理時間分布

→ **指数分布**のとき「M」と書く

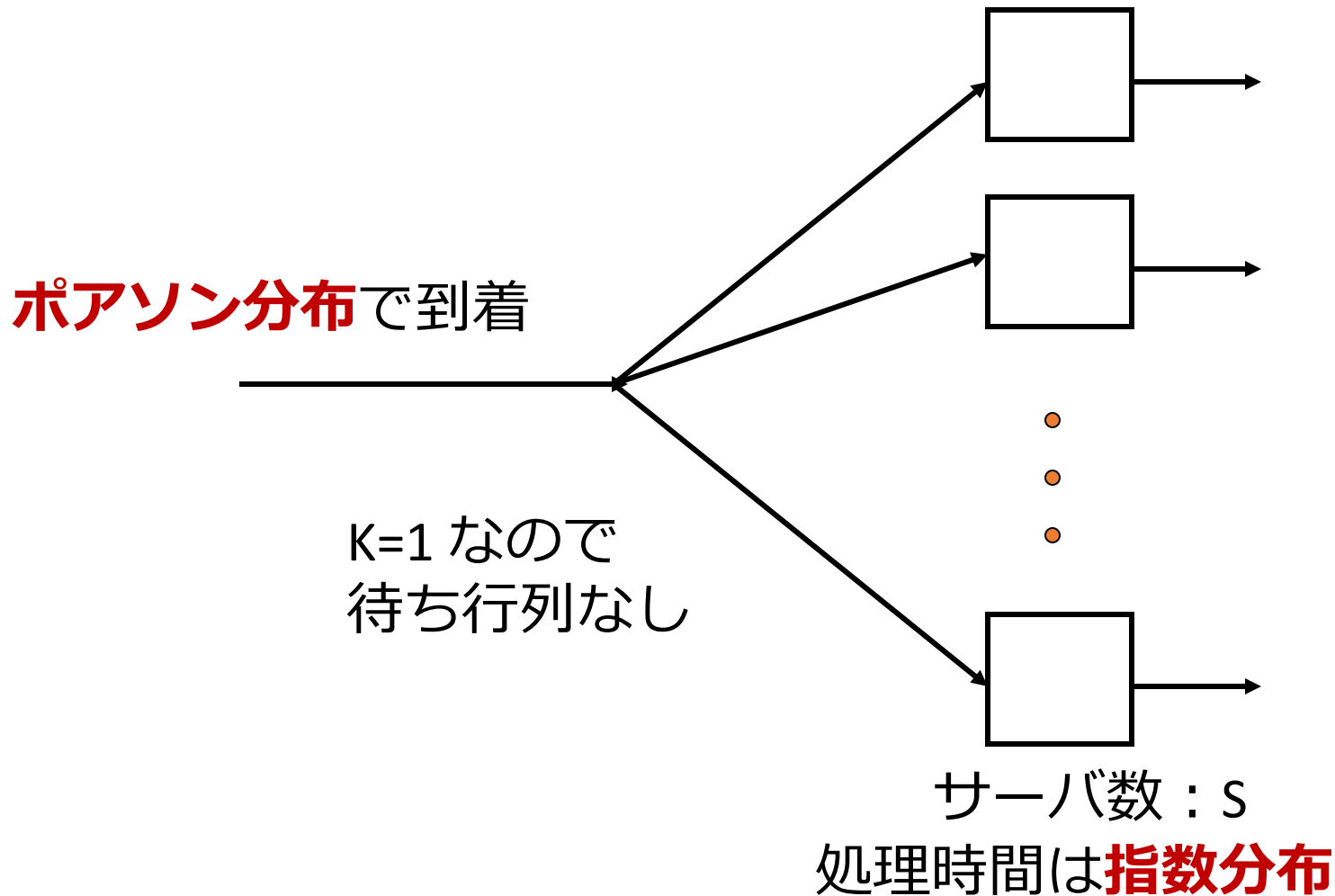
Z: サーバ数

個数を限定しないとき「S」と書く

K: **待ち行列の長さ**の制限

$K = 1$

M/M/S/1 のとき



アーランの即時式モデル



- 待ち行列モデル : $M/M/S/S$
- 入力回線数 : 無限
- 待ち行列長 : $L = 0$
- システム内の最大占有サーバ数 : $K = S$

「即時式」の意味



$K = 1$ なので、次の性質を持つ

- 待ち行列なし
- 待ち行列長は 0
- システム内の占有サーバ数が S （すべてのサーバが占有されている）のとき、到着したジョブは棄却される
- サーバの空きがあれば、直ちに処理される
→ **即時式**

• ジョブの到着

- 平均 λ のポアソン分布

→ ある時刻に客が到着してから時間 t 内に次の客が到着する確率はポアソン分布に従う： $1 - e^{-\lambda t}$

- 微少時間 Δt の間にジョブが到着する確率： $\lambda \Delta t$

• ジョブの処理時間

- 平均 $1 / \mu$ の指数分布

- 微少時間 Δt の間に処理が終わる確率： $\mu \Delta t$

システム処理能力： $\rho = \lambda / \mu$

状態 0 : 系内のジョブの数が 0

状態 1 : 系内のジョブの数が 1

⋮

状態 S: 系内のジョブの数が S

(状態はsまで)

- ジョブが到着：

- 1) $k < S$ のとき

- 状態 k から状態 $k + 1$ に遷移

- 2) $k = S$ のとき

- 状態 S のまま（新しい到着は棄却される）

- ジョブの回線の占有終わり：

- 状態 k から状態 $k - 1$ に遷移

- ジョブが到着 :

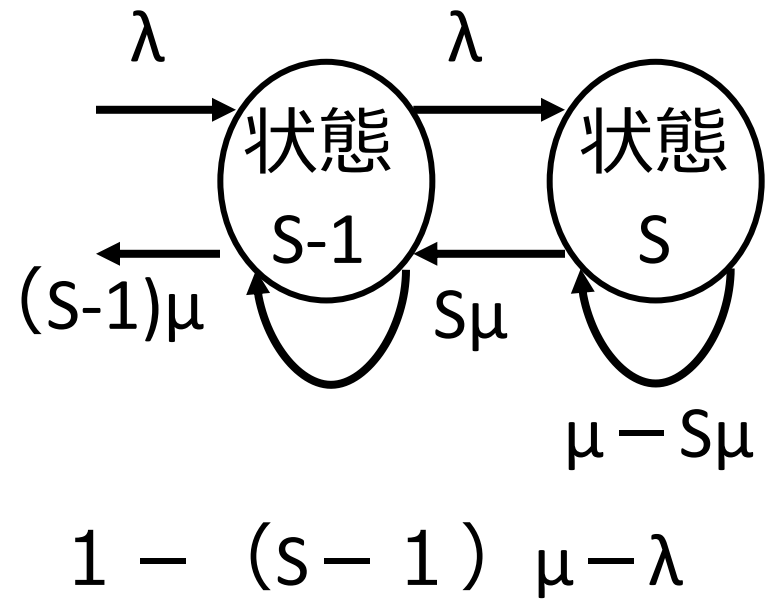
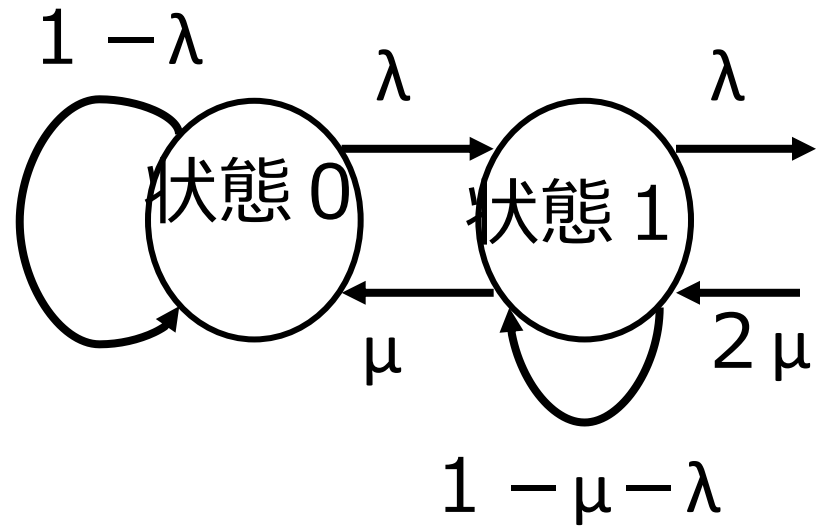
1) $k < S$ のとき

状態 k から状態 $k + 1$ に遷移 : $\lambda \Delta t$

- ジョブの回線の占有終わり :

状態 k から状態 $k - 1$ に遷移 : $k \mu \Delta t$

状態遷移図



定常状態についての方程式



$$(\lambda + n \mu) P_n = \lambda P_{n-1} + (n + 1) \mu P_{n+1}$$

$$\lambda P_0 = \mu P_1$$

$$\lambda P_{n-1} = S \mu P_n$$

- 定常状態で、系内に n 個のジョブ ($0 \leq n \leq s$) がある確率を P_n とすると

$$P_n(t) = \frac{\frac{\rho^n}{n!}}{1 + \rho + \frac{\rho^2}{2} + \dots + \frac{\rho^s}{s!}}$$

- 棄却率：
 - システム内の占有サーバ数が S になる確率： P_s
 - 上の式で、 $n=s$ として計算する