確率過程の基礎

1　確率過程

例 1. (確率過程の例). 天候，地震の発生，遺伝，株価，為替ルート，⋯

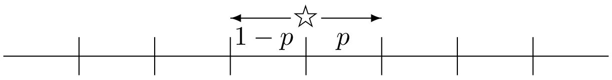
用語

★. 確率過程＝「時点 $\mapsto$ 観測値」
離散型：$n \mapsto X_n$
連続型：$t \mapsto X_t$
確率変数の集合 $\{X_n\}, \{X_t\}$ のことを**確率過程**という．特に $X \in \mathbb{R}$ のとき $\{X_n\}, \{X_t\}$ のことを**標本関数**または**サンプルパス**という．

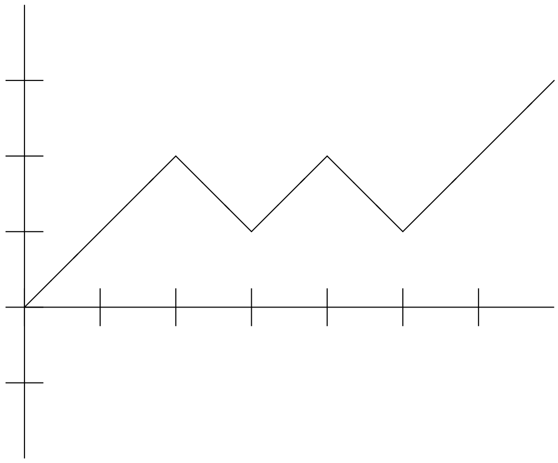
注意 1. $\{X_i\}$ は互いに独立で同一な分布に従うとは限らない．

2　１次元ランダムウォーク

$0<p<1$ なるパラメータを固定する．各時点 $0, 1, 2, \cdots$ で，ある粒子☆が確率 p で右に移動，確率 $1-p$ で左に移動しているとする．



時点 0 において粒子☆が原点にいるとする．このとき時間の推移と共に変化する粒子☆の位置を表したものが次の図である．



このような運動は次のように表すことができる．まず $X_1, X_2, \cdots, X_n$ を

$$P(X_i = 1) = p, \quad P(X_i = -1) = 1 - p$$

であるような確率変数列とする．このとき時点 n における粒子☆の位置は

$$S_n = X_1 + X_2 + \cdots + X_n$$

と表現できる．

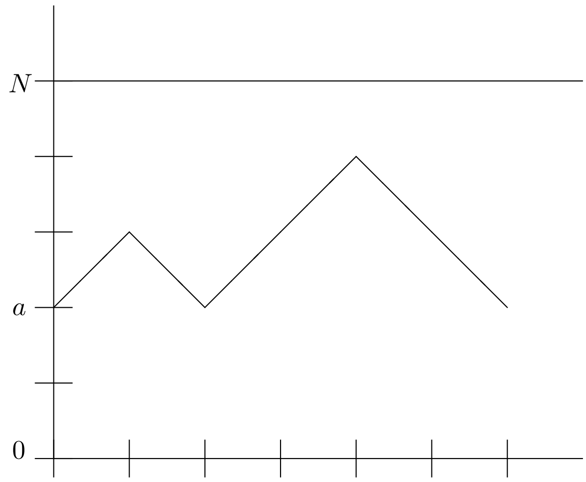
3　ギャンブラーの破産問題

次のようなゲームを考える．

ルール

◎プレイヤーは A, B
◎最初に A は a 枚，B は $N-a$ 枚のチップを持つ．
◎コインを投げて表が出たら A は B からチップを１枚もらい，裏が出たら A は B にチップを１枚渡す．
◎これを繰り返し，どちらかのもち手のチップがなくなって「破産」したらゲームは終了．

A のチップ数を表したものが次の図である．



もし 0 に到達すれば A は破産したことになる．逆に N に到達すれば，B が破産したことになる．0 と N は**吸収壁**と呼ばれる．

コインの表が出る確率を p，裏が出る確率を $q(=1-p)$ とする．このとき次の問題を考える．

問題

A の破産する確率を求めよ．

A が破産する確率は，最初のチップ数に依存するので $r(a)$ で破産確率を表すことにする．事象 R, H を次で定める．

$$R = \{A \text{ はいつか破産する}\}$$
$$H = \{\text{第 1 回目にコインの表が出る}\}$$

このとき条件付確率の定理から次のことが言える：

$$P(R) = P(R|H)P(H) + P(R|H^c)P(H^c)$$

第 1 回目にコインの表が出れば A のチップ数は $a+1$ 枚となり，裏が出れば $a-1$ 枚となるので，

$$P(R|H) = r(a+1) \to \text{初期条件を } a+1 \text{ にした}$$
$$P(R|H^c) = r(a-1) \to \text{初期条件を } a-1 \text{ にした}$$

となる．従って $a = 1, 2, \cdots, N-1$ に対して，差分方程式

$$r(a) = pr(a+1) + qr(a-1) \qquad (\diamondsuit)$$

が成り立つ．境界条件[1]

$$r(0) = 1, \quad r(N) = 0$$

を考慮して，差分方程式 $(\diamondsuit)$ を解くと，破産確率は

$$r(a) = \begin{cases} \frac{(q/p)^N - (q/p)^a}{(q/p)^N - 1} & (p \neq q) \\ 1 - a/N & (p = q) \end{cases} \qquad (\heartsuit)$$

となる．

ここで次のような疑問を考えてみる．

疑問

このゲームは必ず勝負が付くのだろうか？

このことを調べるため B の破産する確率を $w(a)$ とする．これは $(\heartsuit)$ において $p \leftrightarrow q$, $a \leftrightarrow N-a$ と入れ替えたものなので

$$r(a) = \begin{cases} \frac{(p/q)^N - (p/q)^{N-a}}{(q/p)^N - 1} & (p \neq q) \\ a/N & (p = q) \end{cases}$$

[1]最初に A のチップ数が 0 枚なら既に破産していて，チップ数が N ならば B が破産しているからこうなる．

となるので確かに

$$r(a) + w(a) = 1$$

となってこの賭けは必ず勝負が付くことがわかる．

$p = q$ のとき，この掛けは**公平**であるという．賭けが公平でも勝負が付くことを上は示している．ここで A の最終的な利益または損失 G の期待値を考える．ここに G とは最終的な手持ちから最初の手持ち a を引いたものである．

$$\begin{aligned} E(G) &= (1 - r(a)) \cdot (N - a) + r(a) \cdot (-a) \\ &= N(1 - r(a)) - a \end{aligned}$$

よって，公平な賭けのとき $(\heartsuit)$ より，$E(G) = 0$ だから最初の所持金に関係なく公平な賭けであることがわかる．

疑問

このゲームが終了するまでに，平均してどのくらいの時間 (回数) がかかるのだろう？

A または B のどちらかが破産してゲームが終了するまでに，コインを投げた回数を F とする．F の期待値のことを**期待終了時間**という．これももちろん最初の所持金 a に依存する．

$$E(F) = E(F|H)P(H) + E(F|H^c)P(H^c)$$

が成立するので $e(a) = E(F)$ とおけば

$$E(F|H) = 1 + e(a+1), \quad E(F|H^c) = 1 + e(a-1)$$

であるから，$a = 1, 2, \cdots, N-1$ に対して，差分方程式

$$e(a) = \{1 + e(a+1)\}p + \{1 + e(a-1)\}q$$

が成立する．ここで境界条件

$$e(0) = e(N) = 0$$

より差分方程式を解くと，期待終了時間として

$$e(a) = \begin{cases} \left(N \dfrac{(q/p)^a - 1}{(q/p)^N - 1} - a\right) / (p - q) & (p \neq q) \\ a(N - a) & (p = q) \end{cases}$$

が導ける．

4 ブラウン運動

この節のテーマ

★．ランダムウォークにおいて時間の刻み幅 Δt, 移動する幅 Δr を 0 に近づけるとどうなるだろうか？(ランダムウォークの連続化)

ランダムウォーク S_n の期待値と分散は

$$
\begin{aligned}
E(S_n) &= nE(X_i) = n\{1\cdot p + (-1)\cdot q\} = n(p-q)\\
V(S_n) &= nV(X_i) = n\{p+q-(p-q)^2\} = 4npq
\end{aligned}
\qquad (\spadesuit)
$$

である．１回に移動する幅を Δr，それにかかる時間を Δt とする．このとき時間 t の間に粒子は $t/\Delta t$ 回動くので $(\spadesuit)$ において

$$n \to t/\Delta t, \quad 幅 \to \Delta r$$

とおきかえれば，時間 t までの粒子の総変位の期待値と分散は

$$(p-q)t(\Delta r/\Delta t), \quad 4pqt(\Delta r)^2/\Delta t$$

となる．ここで期待値，分散が ∞ にならないように $\Delta t, \Delta r \to 0$ とすることを考える．天下り的に次を仮定する．

仮定

$$\frac{(\Delta r)^2}{\Delta t} = 2D, \quad p = \frac{1}{2} + \frac{C\Delta r}{2D}, \quad q = \frac{1}{2} - \frac{C\Delta r}{2D}$$

ここで D は**拡散係数**，C は**ずれ**と呼ばれる定数である．このとき，期待値と分散は次のようになる．

$$(p-q)t\cdot\frac{\Delta r}{\Delta t} = t\left(\frac{C}{D}\right)\cdot\frac{(\Delta r)^2}{\Delta t}$$

$$
\begin{aligned}
4pqt\cdot\frac{(\Delta r)^2}{\Delta t} &= 4\left(\frac{1}{2}-\frac{C\Delta r}{2D}\right)\left(\frac{1}{2}+\frac{C\Delta r}{2D}\right)t\cdot\frac{(\Delta r)^2}{\Delta t}\\
&= t\cdot\frac{(\Delta r)^2}{\Delta t} - \left(\frac{C\Delta r}{D}\right)^2 t\cdot\frac{(\Delta r)^2}{\Delta t}.
\end{aligned}
$$

ここで $\Delta r, \Delta t \to 0$ とすると，時点 t までの総変位の期待値と分散の極限値は $2Ct, 2Dt$ となる．ここで次の中心極限定理を思い出しておく．

中心極限定理

$X_1, X_2, \cdots, X_n$ が互いに独立で，同一分布に従うとき，$S_n = X_1 + \cdots + X_n$ とおくと，$(S_n - n\mu)/\sqrt{n}\sigma$ は $n \to \infty$ のとき正規分布に従う．

よって，時点 t における粒子の位置は，正規分布 $N(2Ct, 2Dt)$ に従う．これを**ブラウン運動**といい，B_t と表す．

ところで時点 s から t $(0<s<t)$ の間の**増分** $B_t - B_s$ は，同様に正規分布 $N(2C(t-s), 2D(t-s))$ に従う．従って，B_t と B_s の共分散は

$$
\begin{aligned}
\mathrm{Cov}(B_t, B_s) &= \frac{1}{2}\{V(B_t) + V(B_s) - V(B_t - B_s)\}\\
&= D\{t + s - (t-s)\} = 2Ds \quad (s<t)
\end{aligned}
$$

さらに $t_1 < t_2 \leq t_3 < t_4$ として，異時点間の増分 $B_{t_2} - B_{t_1}$ と $B_{t_4} - B_{t_3}$ の共分散を考える．

$$
\begin{aligned}
&\mathrm{Cov}(B_{t_2} - B_{t_1}, B_{t_4} - B_{t_3})\\
=&\mathrm{Cov}(B_{t_2}, B_{t_4}) - \mathrm{Cov}(B_{t_2}, B_{t_3})\\
&- \mathrm{Cov}(B_{t_1}, B_{t_4}) + \mathrm{Cov}(B_{t_1}, B_{t_3})\\
=&2D(t_2 - t_2 - t_1 + t_1) = 0
\end{aligned}
$$

となり無相関となる．$B_{t_2} - B_{t_1}$ と $B_{t_4} - B_{t_3}$ は正規分布に従うので，これは独立性を意味する．これよりブラウン運動は**独立増分**の**正規過程**とも呼ばれる．

5 マルコフ連鎖

ランダムウォークとマルコフ性

◎ランダムウォーク：次の行動に過去の影響を受けない

◎マルコフ性：次の行動に過去の影響を受ける

この節では，$n \mapsto X_n : \mathbb{Z}_+ \to \{1, 2, \cdots, N\}$ という形の確率過程を考える．ここで $1, 2, \cdots, N$ は**状態**と呼ばれる．

定義 2. 確率過程 $\{X_n\}_{n\in\mathbb{N}}$ が $\forall n, \forall j_0, \cdots, j_{n-1}, k \in \{1, 2, \cdots, N\}$ に対し

$$
\begin{aligned}
&P(X_{n+1} = k | X_0 = j_0, \cdots, X_{n-1} = j_{n-1}, X_n = j)\\
&= P(X_{n+1} = k | X_n = j)
\end{aligned}
$$

←次の状態が今の状態にしか依存しない

を満たすとき $\{X_n\}$ のことを**マルコフ連鎖**という．

定義 3. 確率過程 $\{X_n\}$ が有限個の過去の状態にも依存するとき $\{X_n\}$ を**多重マルコフ過程**という．また条件付分布が時点 n にも依存しないとき**斉次マルコフ過程**という．

以下，X_n を斉次マルコフ連鎖とする．

定義 4. 状態が 1 期の間に j から k に移る確率

$$p_{jk} = P(X_{n+1} = k | X_n = j)$$

を**推移確率**といい，それらを並べた行列

$$P = (p_{jk})$$

を**推移確率行列**という．

命題 5. $\sum_{k=1}^{N} p_{jk} = 1$.

検討課題

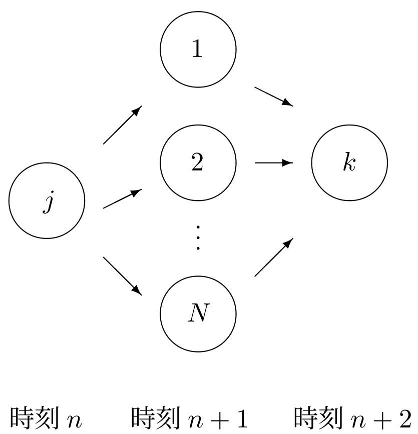
マルコフ連鎖 X_n において，2 期後，3 期後，$\cdots$，n 期後の状態はどうなっているだろうか？

まず 2 期後の推移確率や推移確率行列を考える．

$$p_{jk}^{(2)} = P(X_{n+2} = k | X_n = j)$$

とおく．

参考図



上の図より

$$p_{jk}^{(2)} = \sum_{i=1}^{N} p_{ji} p_{ik}$$

である．これは $P^2 = PP$ の (j,k) 成分なので 2 期後の推移確率行列は P^2 である．同様に n 期後の推移確率行列は P^n となる．$\forall m, n \in \mathbb{N}$ に対し

$$P^{m+n} = P^m P^n$$

であるので次が成り立つ．

定理 6. (**チャップマン・コルモゴロフ方程式**).

$$p_{jk}^{(m+n)} = \sum_{i=1}^{N} p_{ji}^{(m)} p_{ik}^{(n)}$$

次に時点 n で，状態 j である確率を求めよう．

定義 7. 時点 n で状態 j である確率を

$$p_n(j) = P(X_n = j) \quad j = 1, 2, \cdots, N$$

とおく．このときベクトル

$$\boldsymbol{p}_n = (p_n(1), p_n(2), \cdots, p_n(N))$$

を**状態確率分布**という．特に$\boldsymbol{p}_0$ のことを**初期分布**という．

ここで

$$\begin{aligned} p_n(j) &= \sum_{i=1}^{N} P(X_{n-1} = i) P(X_n = j | X_{n-1} = i) \\ &= \sum_{i=1}^{N} p_{n-1}(i) p_{ij} \end{aligned}$$

に注意すると

$$\boldsymbol{p}_n = \boldsymbol{p}_{n-1} P \qquad (\clubsuit)$$

が成立する．これを繰り返すことにより

$$\boldsymbol{p}_n = \boldsymbol{p}_0 P^n$$

を得る．

疑問

$n \to \infty$ のとき$\boldsymbol{p}_n$ はどうなるのだろうか？

定義 8. もし，初期条件に関係なく$\boldsymbol{\pi} = \lim_{n\to\infty} \boldsymbol{p}_n$ が存在し$\boldsymbol{\pi} = (\pi_1, \pi_2, \cdots, \pi_N)$ が確率分布になっているとき，これを**定常分布**または**不変分布**という．

もし定常分布が存在すれば ($\clubsuit$) において $n \to \infty$ として

$$\boldsymbol{\pi} = \boldsymbol{\pi} P$$

を満たす$\boldsymbol{\pi}$を求めればよい．

6　ポアソン過程

ポアソン過程

★．ポアソン過程：$t \mapsto N_t$．但し N_t は時間 $[0,t]$ にある事象が起きた回数.

N_t にいくつかの仮定をおいて，N_t の状態の確率分布を導こう.

仮定

◎ $P(N_0=0)=1$
◎ $N_s \le N_t\,(s \le t)$：区間が長いほど回数は多い
◎ $N_s-N_0,\ N_t-N_s\,(0<s\le t)$ は独立：区間 $[0,s]$ と区間 $[s,t]$ で起きる回数は互いに影響しない
◎推移確率について

$$\begin{aligned} &P(N_{t+h}=n+1|N_t=n)=\lambda h+o(h) \\ &P(N_{t+h}=n|N_t=n)=1-\lambda h+o(h) \qquad (\star) \end{aligned}$$

：微小な時間 h の間に起こる回数はほとんど 0 または 1 である

定理 9. 上の仮定の下で $k \in \mathbb{N}$ に対し

$$P(N_t=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t}$$

が成り立つ.

証明. $k \in \mathbb{N}$ に対し

$$\begin{aligned} &P(N_{t+h}=k) \\ &=P(N_{t+h}=k|N_t=k-1)P(N_t=k-1) \\ &\quad +P(N_{t+h}=k|N_t=k)P(N_t=k)+o(h) \\ &=\{\lambda h+o(h)\}P(N_t=k-1) \\ &\quad +\{1-\lambda h+o(h)\}P(N_t=k)+o(h) \\ &=\lambda hP(N_t=k-1)+(1-\lambda h)P(N_t=k)+o(h) \end{aligned} \tag{1}$$

が成立する．ここで

$$p_k(t)=P(N_t=k)$$

とおくと (1) より

$$p_k(t+h)-p_k(t)=\lambda h\{p_{k-1}(t)-p_k(t)\}+o(h) \tag{2}$$

となる．ここで (2) の両辺を h で割って $h \to 0$ とすると，$o(h)/h \to 0$ から

$$p'_k(t)=\lambda\{p_{k-1}(t)-p_k(t)\}\,(k=1,2,\cdots) \tag{3}$$

を得る．$k=0$ の場合も同様に

$$p'_0(t)=-\lambda p_0(t) \tag{4}$$

が導ける．これは $p_k(t),\,(k=0,1,2,\cdots)$ に対する微分差分方程式である．$N_0=0$ だから，境界条件

$$p_k(0)=\begin{cases} 1 & (k=0) \\ 0 & (k\ne 0) \end{cases}$$

の下で，帰納法により示せる.

□

7　出生死滅過程

出生死滅過程

★．ポアソン過程は常に増加するような確率過程であった．増加・減少する確率過程 (出生死滅過程) も同様に考えることができる.

L_t：時点 t における細菌の総数

定義 10. 次の仮定を満たす確率過程を**出生死滅過程**という.

仮定

各小区間 $(t,t+h]$ において 1 つの個体が生じる確率が $\lambda_k h+o(h)$，死滅する確率が $\mu_k h+o(h)$ となる：

$$\begin{aligned} &P(L_{t+h}=k+1|L_t=k)=\lambda_k h+o(h) \\ &P(L_{t+h}=k|L_t=k)=1-\lambda_k h-\mu_k h+o(h) \\ &P(L_{t+h}=k-1|L_t=k)=\mu_k h+o(h) \\ &P(L_{t+h}\le k-2\,or\,\ge k+2|L_t=k)=o(h) \end{aligned}$$

ポアソン過程のときと同様に

$$\begin{cases} p'_k(t) & \\ =-(\lambda_k+\mu_k)p_k(t)+\lambda_{k-1}p_{k-1}(t)+\mu_{k+1}p_{k+1}(t) & (k\ge 1) \\ p'_0(t)=-\lambda_0p_0(t)+\mu_1p_1(t) & (k=0) \end{cases}$$

が導ける.

定義 11. 出生死滅過程において $\lambda_k=k\lambda, \mu_k=k\mu$ としたものを**フェラー・アレイ過程**という.

定義 12. 同様に $\mu_k=0$ としたものを**純出生過程**または**ユール過程**という.

定義 13. 同様に $\lambda_k = k\lambda + \mu$ としたものを**ケンドール過程**という.

ケンドール過程はフェラー・アレイ過程で外部からの**移民**がいると考えている.

出生死滅過程はポアソン過程と異なり，一般的に解を求めることはできない．フェラー・アレイ過程の場合は次が成り立つ.

定理 14. フェラー・アレイ過程の下で次が成り立つ：

$$\begin{aligned} p_0(t) &= \frac{\lambda\{1-\exp((\lambda-\mu)t)\}}{\mu--\lambda\exp((\lambda-\mu)t)} \quad (\mu \neq \lambda) \\ &= \frac{\lambda t}{1+\lambda t} \quad (\mu = \lambda) \end{aligned}$$

$k \geq 1$ に対して

$$\begin{aligned} p_k(t) &= \frac{\lambda^{k-1}(\lambda-\mu)^2 e^{(\lambda-\mu)t}(1-e^{(\lambda-\mu)t})^{k+1}}{(\mu-\lambda e^{(\lambda-\mu)t})^{k+1}} \quad (\mu \neq \lambda) \\ &= \frac{(\lambda t)^{k-1}}{(1+\lambda t)^{k+1}} \quad (\mu = \lambda) \end{aligned}$$

但し，境界条件を

$$p_1(0) = 1,\, p_k(0) = 0 \quad (k \neq 1)$$

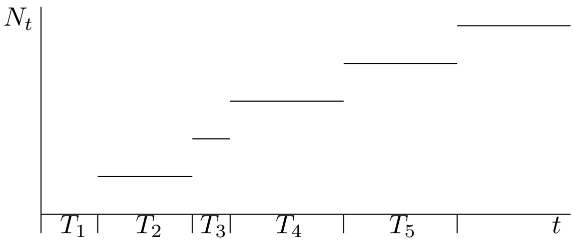
とする.

8　待ち行列

待ち行列分布

出来事が起きる時間間隔 (待ち時間) はどのような分布に従っているか？

ポアソン過程で最初に事象が起きるまでの時間間隔を T_1，事象が $(i-1)$ 回起きてから i 回起きるまでの時間間隔を T_i $(i = 2, 3, \cdots)$ とする.



このとき事象 $\{T_1 \leq t\}$ は $\{N_t \geq 1\}$ に等しいので，T_1 の分布関数を $F_1(t)$ とすると

$$F_1(t) = P(T_1 \leq t) = P(N_t \geq 1) = \sum_{i=1}^{\infty} \frac{(\lambda t)^i}{i!} e^{-\lambda t} = 1 - e^{-\lambda t}$$

密度関数は

$$f_1(t) = \lambda e^{-\lambda t}\,(t > 0)$$

となる．これは T_1 の分布が指数分布であることを示している．次に

$$S_n = T_1 + T_2 + \cdots + T_n$$

の分布を考える．S_n は n 回目の事象が起きるまでに要する時間である．ポアソン過程は，定常で，各増分は互いに独立であるから，T_i は互いに独立に指数分布 $E(\lambda)$ に従っている．事象 $\{S_n \leq t\}$ は事象 $\{N_t \geq n\}$ に等しいから，S_n の確率分布を $F_n(t)$ とすると

$$F_n(t) = P(S_n \leq t) = P(N_t \geq n) = \sum_{i=1}^{\infty} \frac{(\lambda t)^i}{i!} e^{-\lambda t}$$

密度関数 $f_n(t)$ は，$F_n(t)$ の右辺を項別積分して

$$\begin{aligned} f_n(t) &= \sum_{i=n}^{\infty} \lambda \left\{ \frac{(\lambda t)^{i-1} e^{-\lambda t}}{(i-1)!} - \frac{(\lambda t)^i e^{-\lambda t}}{i!} \right\} \\ &= \lambda \sum_{i=n-1}^{\infty} \frac{(\lambda t)^i e^{-\lambda t}}{i!} - \lambda \sum_{i=n}^{\infty} \frac{(\lambda t)^i e^{-\lambda t}}{i!} \\ &= \frac{\lambda(\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!} \end{aligned}$$

となる.

定義 15. 上の確率分布を**相** n **のアーラン分布**という.

相 n のアーラン分布はチケットを購入するために並んだ n 番目の人が，発売開始から購入するまでの**待ち時間**の分布である.

さらにチケットの発売所の例では，発売開始後も順番待ちの人は増加する一方，チケットを受け取った人は立ち去るので一種の出生死滅過程と見ることができる．これを**待ち行列**という.

今，人々の到着がパラメータ λ のポアソン分布に従い，係員の処理時間がパラメータ λ の指数分布に従うとすれば，行列に並んでいる人の数は $\lambda_k = \lambda,\, \mu_k = \mu$ の出生死滅過程になる．発売開始までに i 人が並んでいたとすると，初期状態を

$$\begin{aligned} p_k(0) &= 1 \quad (k = i) \\ &= 0 \quad (k \neq i) \end{aligned}$$

とすればいい．また係員が s 人いる場合は

$$\begin{aligned} \mu_k &= k\mu \quad (0 \leq k < s) \\ &= s\mu \quad (k \geq s) \end{aligned}$$

とすればよい.