Pitman 壺過程を用いた 2ch.net の定常投稿過程のモデル化と検証

SP13150 宮本　優也　　量子物理学講座

1　はじめに

この論文の目的は人間の行動やその行動が周囲に与える影響をモデル化し、2ch.net という電子掲示板の投稿の時系列データを用いて検証することにある。ここで、2ch.net とは、スレッドフロート型掲示板と呼ばれる形態である。秒単位で投稿が行われるため、膨大なデータの取得が可能であり、ヒト集団挙動のモデル化と検証には格好の場を与えてくれる。本研究では、2ch.net の投稿過程を Pitman 壺過程を用いてモデル化する。用いたデータは、2009 年の 6ヶ月および 10ヶ月の期間に収集した「デジタルカメラ」「ビジネスニュース+」「ニュース速報+」の 3 つのカテゴリーの投稿ログを用いた。投稿数はそれぞれ 40 万,100 万,2000 万である。

2　Pitman 壺過程

自然数 n を k 個の自然数に分割することを考える。l 番目の自然数を c_l とすると、$n=\sum c_l$ が成立する $c_l=j$ となる自然数の個数を m_j とすると、$k=\sum_{j=1}^{n} m_j, \sum_{j=1}^{k} m_j j=n$ が成立する。Pitman 分布はこの分割 $(k,m_1,m_2,...,m_k)$ に対する次の確率分布として定義される。

$$P(k,m_1,m_2,...,m_k)=\frac{n!\theta^{[k:\alpha]}}{\theta^{[k]}}\prod_{j=1}^{k}\left(\frac{(1-\alpha)^{[j-1]}}{j!}\right)^{m_j}\frac{1}{m_j}$$

ここで $0\leq\alpha<1$ と $\theta\ (>-\alpha)$ はパラメーターで $x^{[j:\alpha]}=x(x+\alpha)\cdots(x+(j-1)\alpha), x^{[j]}=x[j:1]$, である。

この Pitman 分布をポリア壺過程としてモデル化することができる。番号のついた玉 $B_1,B_2,...$ を、この順序に従い、番号のついた壺 $U_1,U_2,...$ に逐次、ランダムに入れる。先ず、玉 B_1 は壺 U_1 に入れる。次いで、$n(=1,2,...)$ 個の玉 $B_1,...,B_n$ を壺に入れたとし、玉は k 個の壺 $U_1,...,U_k$ の中に入っているものとする。なお、$j=1,...,k$ に対し、壺 U_j には $c_j(>0)$ 個の玉が入っているとする。次の玉 B_{n+1} は、新しい壺 U_{k+1} へ、確率 $\frac{\theta+k\alpha}{\theta+n}$ で入るか, または, 既に玉が入っている壺の何れか U_j へ確率 $\frac{c_j-\alpha}{\theta+n}$　$(j=1,...k)$ で入る。このモデルの非定常分布として Pitman 分布が導出される [1]。また、一定個数 n 個の玉を見て $n+1$ 番目を決める。有限記憶のポリア壺過程においても、同様に Pitman 分布が導出される [2]。今回は有限記憶のポリア壺過程に従って解析する。(図 1)

3　解析と考察

壺をスレッド、投稿を玉の追加と考えて、2ch.net の投稿の時系列データを解析する。有限記憶のポリア壺過程のモデルのモデルパラメータ（θ、α）を推定する。

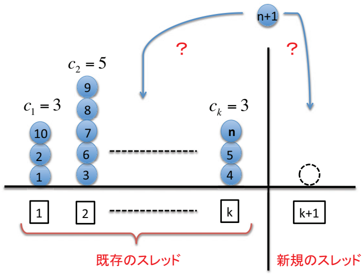
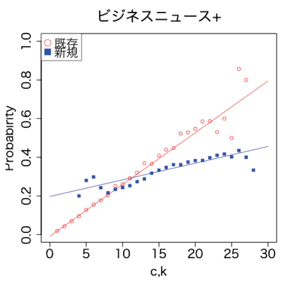


図 1: 有限記憶の Pitman 壺過程

n 個の投稿の中に存在するスレッドの数 k、それぞれのスレッドへの投稿数 $c_j(j=1,...,k)$ を記録することでスレッド数と投稿数の分布を得る。次に、$n+1$ 番目の投稿が既存のスレッドならそのスレッドの投稿数 c_j、新規のスレッドならその時のスレッド数 k を記録することで投稿に対する確率法則を得る。最尤推定法を用いてモデルパラメーター (θ、α) を推定し（図 2)、スレッド数 K の分布と比較した。(図 3)



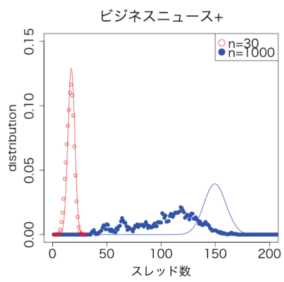


図 2: 投稿の確率法則 (n=30)　図 3: スレッド数 k の分布

図 3 を見ると n が小さい場合、Pitman モデルによく従っていると言える。Pitman モデルはスレッドの実力差はないとするモデルなので、n が大きくスレッドの実力（スレッドの伸び）が見えると Pitman モデルに従わなくなると考える。今後は n が大きい場合に従うモデルについて記述する予定である。

[1] 大和元, 渋谷政照, ピットマン確率分割と関連する話題, 統計数理研究所,2003.

[2]M.Hisakado and S. Mori, Pitman sampling formula in Equilibrium and Non-epuilibrium processes, preprint.