1997年度日本オペレーションズ・リサーチ学会
秋季研究発表会

非定常ポアソン過程における隣接時間区間のデータを流用した平均事象発生数の擬似推定法

01506240 (株)日立製作所 生産技術研究所 ＊船木謙一 FUNAKI Kenichi
(株)日立製作所 生産技術研究所 的場秀彰 MATOBA Hideaki

1．緒言

非定常ポアソン過程(NHPP)は、店への客の到着数や台風の発生数など多くの時系列確率過程のモデルとして用いられている。NHPP の性質は、時刻tの関数として表される平均値関数Λ(t)によって一意に決定される。したがって、シミュレーションモデルなどにおいて NHPP を用いるためには、対象事象の発生データを時系列に並べた標本データ系列を観測し、Λ(t)の形を推定することが必要となる。Λ(t)を表す関数式の形(パラメータ構成)が分かっている場合には、尤度関数を導き、観測した標本データ系列に対して最大尤度を与えるパラメータ値を求めれば良い[1]。標本データ系列の表現方法には、事象の発生時刻を観測して時点列として表す方法と単位時間区間当たりの事象発生数を観測して件数列として表す方法があるが、どちらの場合にも尤度関数は簡単に導くことができ、推定は容易である。しかし、実際にはΛ(t)の形を予め特定することが困難で、尤度関数も導けないことも多い。そのような場合には、時系列上を適当な時間区間に区切り、複数の標本データ系列を観測して各時間区間における事象発生数の区間平均値を求める方法がある[2]。この方法は、標本データ系列が多数得られる場合には有効であるが、標本データ系列が少ない場合には推定精度が悪くなるという問題がある。

本発表では、Λ(t)の形が分からず、かつ標本データ系列が一つまたは少数しか観測できない場合に、上記の事象発生数の区間平均値をとる方法において、隣接時間区間のデータを流用した擬似推定値を用いて推定精度を向上する方法を提案する。

2．隣接時間区間のデータを流用した推定とその有効性

2．1 用語、記号の定義

まず、時系列上のある時間区間中の平均事象発生件数を推定する場合を考える。推定の対象となる時間区間を推定区間、推定区間に隣接する時間区間を併せた区間を参照区間と呼ぶことにし(図1)、以下のように記号を定義する。

N：観測した標本データ系列数
Ha：推定区間中の真の平均事象発生件数
Hb：参照区間中の真の平均事象発生件数
τa、τb：推定区間、参照区間の長さ
Xa：推定区間中の観測事象発生件数の標本平均
Xb：参照区間中の観測事象発生件数の標本平均
Ya：擬似推定値(≡Xb・τa／τb)

但し、観測事象発生件数の標本平均とは、各標本データ系列の当該時間区間中に観測した事象発生件数を合計して、標本データ系列数Nで割った値。

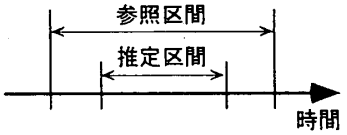

図1 推定区間と参照区間

2．2 擬似推定値の有効性

単純に標本データからの区間平均値を用いるならば、Haの推定値はXaである。しかし、ここではXaの代わりに擬似推定値Yaを用いた方が統計的に精度が良くなる場合があることを示す。精度とは、真の平均値Haの周辺の一定領域に入る確率が高い方が良いという尺度で考える。

まず単純平均値Xaについて、任意の正定数εaに対して、チェビシェフの不等式より、

$$Prob\{|Xa-Ha|\geqq \varepsilon a\}\leqq \frac{Ha}{N\cdot \varepsilon a^2} \qquad 式1$$

が成り立つ。これを変形して

$$Prob\{|Xa-Ha|<\varepsilon a\}\geqq 1-\frac{Ha}{N\cdot \varepsilon a^2}=\alpha \qquad 式2$$

とおくと

$$\varepsilon a=\sqrt{\frac{Ha}{(1-\alpha)N}} \qquad 式3$$

であるから、式2に戻して、XaがHaの周辺で確率α以上で入る領域Dxは

$$Dx=\{Xa\,|\,Xa\in(Ha-\sqrt{\frac{Ha}{(1-\alpha)N}},\ Ha+\sqrt{\frac{Ha}{(1-\alpha)N}})\} \qquad 式4$$

と表される。次に、擬似推定値Yaについても同様にチェビシェフの不等式から、

$$Prob\{|Ya-Ha|\geqq \varepsilon a\} \leqq \frac{1}{\varepsilon a^2}\left[\left\{Ha-\frac{Hb\cdot \tau a}{\tau b}\right\}^2+\frac{Hb\cdot \tau a^2}{N\cdot \tau b^2}\right] \qquad 式5$$

であり、これを変形して

$$Prob\{|Ya-Ha|<\varepsilon a\} \geqq 1-\frac{1}{\varepsilon a^2}\left[\left\{Ha-\frac{Hb\cdot \tau a}{\tau b}\right\}^2+\frac{Hb\cdot \tau a^2}{N\cdot \tau b^2}\right]=\alpha \qquad 式6$$

とおくと、

$$\varepsilon a=\sqrt{\frac{\left\{Ha-\frac{Hb\cdot \tau a}{\tau b}\right\}^2+\frac{Hb\cdot \tau a^2}{N\cdot \tau b^2}}{1-\alpha}} \qquad 式7$$

であるから、ＹａがHaの周辺で確率α以上で入る領域Ｄｙは

$$Dy=\{Ya \mid Ya\in(Ha-\sqrt{\frac{\{Ha-\frac{Hb\cdot\tau a}{\tau b}\}^2+\frac{Hb\cdot\tau a^2}{N\cdot\tau b^2}}{1-\alpha}},\ Ha+\sqrt{\frac{\{Ha-\frac{Hb\cdot\tau a}{\tau b}\}^2+\frac{Hb\cdot\tau a^2}{N\cdot\tau b^2}}{1-\alpha}})\}$$ 式８

と表される。Ｄｙ⊆Ｄｘであれば推定値としてＹａを用いた方が真値Haにより近い値を得る確率が高いといえるので、式４と式８より、

$$\sqrt{\frac{\{Ha-\frac{Hb\cdot\tau a}{\tau b}\}^2+\frac{Hb\cdot\tau a^2}{N\cdot\tau b^2}}{1-\alpha}}\leqq\sqrt{\frac{Ha}{(1-\alpha)N}}$$ 式９

を満たすことが、Ｙａによる推定の方がＸａによる推定よりも精度が良くなるための十分条件である。すなわち、ある推定区間と参照区間に対して、式９を満たす関係があるとき、単純区間平均値Ｘａの代わりにＹａを推定値とした方が良いといえる。

しかし、真の値Ha、Hbは事前に知りえないので、式９の条件を吟味する際には、代わりにそれぞれＸａおよびＸｂを用いて計算することを提案する。

３．推定アルゴリズムと実験結果

３．１　推定アルゴリズム

本発表によるNHPPの推定では、時系列上を推定する時間区間(推定区間)に区切って、各推定区間ごとに平均事象発生件数を逐次推定していく。推定アルゴリズムを次のようにする。

For 全ての推定区間
　当該推定区間の単純区間平均値算出
　推定値←単純区間平均値
　Do While 参照区間長さが限界値以下
　　参照区間設定(逐次広げる)
　　If 式９を満たす *Then*
　　　推定値←擬似推定値
　　　Exit Do
　　End If
　Loop
Next

上記において、各推定区間に対する参照区間の設定には、参照区間長さの限界値を決めておき、推定区間の両側から限界値に達するまで逐次単位時間ずつ広げていく方法をとる。そして、各参照区間について式９の判定を行い、式９を満たしたら、そのときの参照区間による擬似推定値を採用する。

３．２　実験結果

提案する推定法の有効性を見るため、次の$\Lambda(t)$を持つNHPPを実験的に発生させ、その発生結果を標本データ系列と見立てて上記アルゴリズムを用いて推定した(図２、表１)。

$$\Lambda(t)=6t^5-45t^4+130t^3-180t^2+210t+\frac{15}{2}-\frac{15}{2}\cos 6t+\frac{45}{8}\sin 8t$$

この$\Lambda(t)$の形は不規則で、標本データから事前にその形を予想することが困難な場合をうまく表している。標本データには、0.05時間ごとの事象発生件数を１系列観測したものを用いた。また、推定区間を各0.05時間ごとにとり、参照区間長さの限界値は0.15、0.25、0.35時間の３通りを考えた。図２は、参照区間長さの限界値が0.15時間のときの単純区間平均値と提案した方法による推定値とを各推定区間ごとにプロットした例である。提案した方法による推定では、隣接時間区間の変動が式９を満たす範囲ならば、単純区間平均値よりも滑らかな値をとるように修正していることが分かる。これは、隣接時間区間のデータを似ていると判断して、これらの区間で平均化しているからである。表１は、各推定区間の真値との相対２乗誤差の全区間の平均値を、単純区間平均値を用いて推定した場合と上記アルゴリズムを用いて推定した場合とを比較したものである。但し、実験は５０回繰り返し、表中の値はその平均を示している。この結果から、上記アルゴリズムによる推定値の方が全体として精度を上げていること、および参照区間長さの限界値を大きくとれば精度が上がることが分かり、提案した方法の有効性が実験的に証明された。

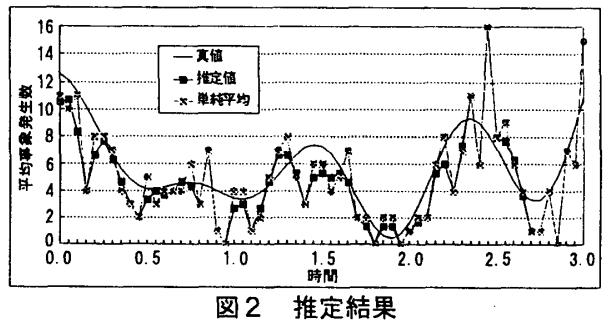

図２　推定結果

表１　各方法における誤差の比較

単純区間平均	提案した方法		
	限界値＝0.15	0.25	0.35
0.22518	0.17909	0.17026	0.16711

４．結言

本発表では、$\Lambda(t)$の形が分からず、かつ標本データ系列が一つまたは少数しか観測できない場合でも隣接時間区間のデータを流用して平均事象発生数の推定精度を向上する方法を提案した。また、その有効性を実験的に検証し、確かめることができた。

参考文献

[1] I.Bar-David,"Communication under the Poisson regime,"*IEEE Trans. Information Theory*, vol.IT-15, no.1, pp.31-37 (1969)
[2] A.M.Law,et al,*Simulation modeling & analysis*, pp.406-408, McGraw-Hill (1991)