

마코프 조정 가우시안과정의 자기상관함수에 관한 연구

이혜연* · 장중순** · 신용백**

* 뉴욕주립대학교 산업공학과

** 아주대학교 산업정보시스템공학과

A Study on the Autocorrelation Function for Markov Modulated Gaussian Process

Hye-Yeon Lee* · Joong-Soon Jang** · Yong-Back Shin**

*State University of New York Dept. of Industrial Engineering

**Ajou University Dept. of Industrial and Information Systems Engineering

Most of process data control have been designed under the assumption that there are independence between observed data. However, it has been difficult to apply the traditional method to real-time data because they are autocorrelated, and they are not normally distributed. And the more, they have fluctuating means. Already the control method for these data was proposed by Markov Modulated Gaussian Process. Therefore, this study take into account MMGP's traits especially for the MMGP's autocorrelation.

Keywords : Markov Modulated Gaussian Process, Control Chart, Autocorrelated.

1. 서론

실시간 공정 데이터의 경우 실시간 관측치들은 일정한 자기상관을 갖는 경우가 많고, 공정변수들이 같을지라도 일정수준만큼의 공정평균의 변화는 있을수 있다. 기존 시계열 분석을 사용하여 공정평균의 변화를 감지하는 경우, 이러한 방법들이 가진 단점은 공정 평균을 단순히 상수로만 취급하여 변화를 감지하고자 했다는 것이다. 하지만, 공정 데이터의 평균 값 하나만을 고려하여 관리하는 시계열 관리의 경우에는 평균이 변화하는 공정을 관리하는데 한계가 있다고 할 수 있다. 마코프 조정과정은 데이터의 시간적 종속성을 잘 표현 할 수 있는 모델 중에 하나로 데이터를 표현하는 자기상관들이 마코프 조정과정에 의해 정확하게 표현되고 있다.

서주원[3]은 이러한 데이터를 마코프 가우시안 조정과정(Markov Modulated Gaussian Process)으로 정의하고 관

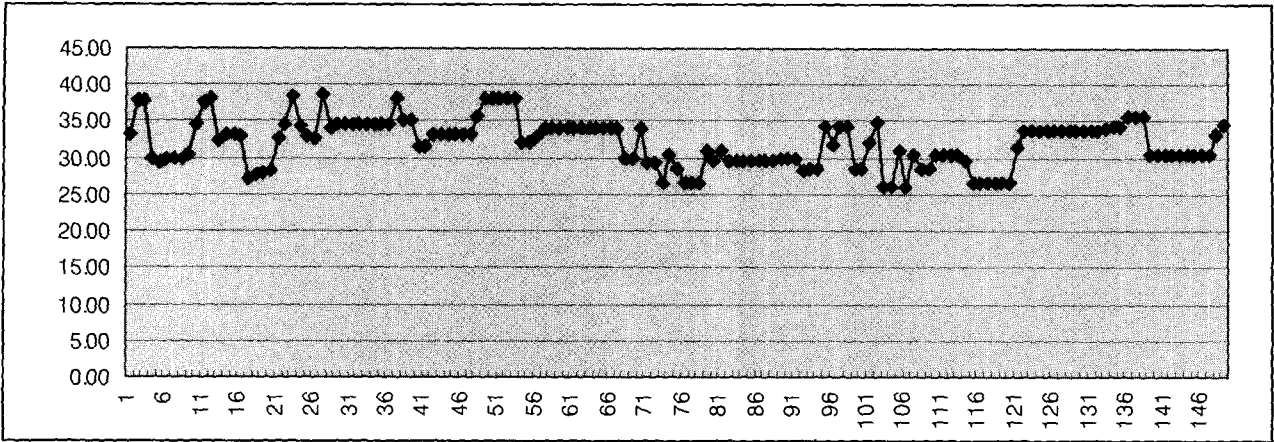
리방법을 제시하였지만, MMGP의 자기상관을 시뮬레이션을 통해 관찰하는데 그침으로써 자기 상관함수에 대한 연구가 필요하게 되었다.

본 연구에서는 특성치가 공정내에 다수의 평균을 가지며 이러한 각각의 평균값들을 μ 의 상태(state of μ)라 할 때, 공정이 μ 의 상태 사이를 일정시간마다 전이하는 상황을 고려하였고 이러한 모델의 자기상관함수를 정의한다. 각 상태에 전이되는 확률은 마코프 과정의 특성을 따르는 전이강도행렬(transition intensity matrix)을 갖는다.

2. MMGP의 성질

2.1 MMGP의 특성

MMGP는 평균이 유한개인 상태를 가진 공정에서, 일



<그림 2.1> 화력발전소에서 석탄교체에 따라 변화하는 SOX 배출량

정 시간에서 일정한 상태값 μ_i 를 중심으로 분산 σ^2 을 가지는 백색잡음(white noise)을 발생시키다가, 다시 일정 시점에 이르러서 마코프 과정의 확률적인 성질을 가지며 다른 μ_i 의 상태로 전이하는 형태를 가지는 시간적 추이의 확률과정으로 정의된다.

이를 수식으로 나타내어 보면, MMGP는 시간 t 에서의 특성치 X_t 가 식(1)

$$X_t = \mu_i + \varepsilon_t \dots\dots\dots (1)$$

를 따르는 모형인데, 여기서 $\varepsilon_t \sim N(0, \sigma^2)$ 인 백색잡음이며, μ_i 는 ε_t 와는 독립으로 전이강도행렬을 갖는 CTMP (Continuous Time Markov Process)이다. 이는 다시 말해 평균이 몇 가지의 일정수준에서 변화하는 과정이며, 각

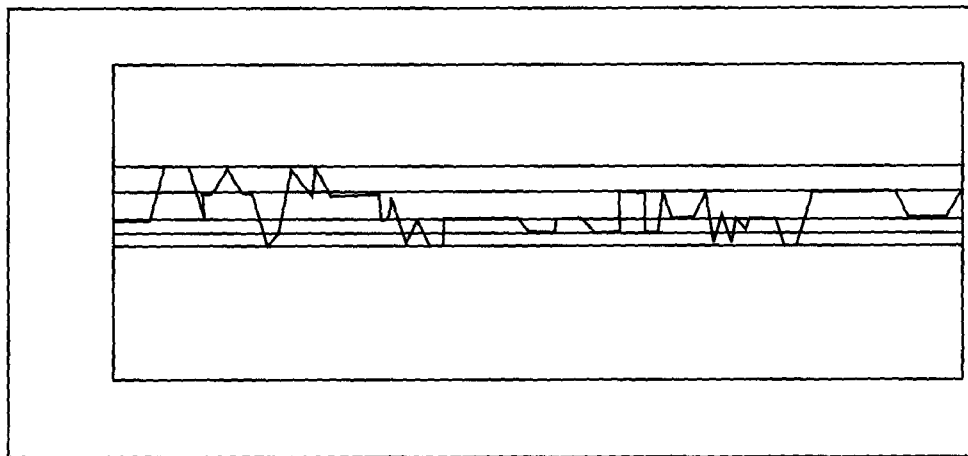
각의 특성치는 일정한 값 μ_i 에 백색잡음이 결합된 형태라고 볼 수 있다[3].

2.2 MMGP 데이터의 예

MMGP의 형태를 나타내는 공정의 요인은 자재 수급의 변화나 부품 품종의 변화 등과 같이 여러 가지의 작업에 영향을 주는 요소들의 변화로 볼 수 있다.

이러한 모델은 많은 사례에서 찾아볼 수 있는데, [그림 2.1]은 실제 화력발전소에서 석탄의 교체에 따라 배출되는 SOX 양의 변화를 시간에 따라 타점한 것이다.

<그림 2.2>는 <그림 2.1>이 상태별로 묘사된 것이다. 여기에서 보이는 바와 같이 각각의 데이터들은 일정수준의 μ 에서 움직이다가, 또 다른 μ 의 상태로 전이하는 것을 볼 수 있는데, 이는 발전에 투입되는 석탄의 종류



<그림 2.2> SOX 배출량에 대한 MMGP의 상태(state)별 묘사

가 교체됨에 따라 생기는 현상으로서, 전형적인 MMGP의 예라고 할 수 있다.

3. MMGP의 자기상관함수

특성치 X_t 는 상태 공간을 $S = \{1, \dots, M\}$ 로 하는 마코프 과정이다. MMGP는 X_t 의 전이강도 행렬 A 에 의해 특성화 된다.

MMGP의 특성치 X_t 에 대한 전이확률 행렬 $P(t)$ 는 마코프 과정의 성질에 의해 식(2)와 같이 정의된다.

$$F_{ij}(t) = P[X(t) = j | X(1) = i] \text{ for } i, j \in S \quad (2)$$

전이 강도 행렬 A 의 각 원소와 P_{ij} 는 식(3)의 관계를 가진다.

$$A = [\lambda_{ij}] = \begin{pmatrix} -\lambda_i & , & i=j \\ \lambda_i p_{ij} & , & i \neq j \end{pmatrix}$$

$$\sum_{i \neq j} \lambda_{ij} = \sum_{i \neq j} \lambda_i p_{ij} = \lambda_i, \quad (3)$$

$$\sum_j p_{ij}(t) = 1$$

안정상태의 $P(t) = \pi_i^T = \{\pi_1, \dots, \pi_M\}$ ($\pi^T > O^T$)이며, π_i^T 는 $\pi_i^T Q = 0^T$, $\pi_i^T e = 1$ 의 시스템 조건을 만족한다. 여기서 $e = (1, 1, \dots, 1)^T$ 이다.

MMGP에서 각 상태전이에 따른 지속시간 평균은 $m = \sum_{i=1}^M \lambda_i \pi_i$ 이며, 이의 n차 적률은 $m^n = \sum_{i=1}^M \lambda_i^n \pi_i$ 이다.

X_t 를 안정상태 마코프 프로세스라 하고, 상태 공간 $S = \{1, \dots, M\}$ 에서의 함수를 $f(n)$ 이라 정의하면, 이 상태공간에서의 $f(X_t)$ 에 대한 자기공분산함수 $C_f(\tau)$ 는 식(4)이다.

$$C_f(\tau) = E[f(X_t) f(X_{t+\tau})] - E^2[f(X_t)] \quad (4)$$

이 정의로부터 상태 지속시간에 관한 함수가 있는 MMGP 과정의 공분산 함수를 구하면,

$$C_\lambda(\tau) = E[\lambda(t) - m_t][\lambda(t+\tau) - m_t]$$

$$= E[\lambda(t)\lambda(t+\tau)] - E^2[m_t]$$

$$= \sum_{i \in S} \sum_{j \in S} \lambda_i \lambda_j P[\lambda(i)\lambda(j); i, j] - (\sum_{i \in S} \lambda_i \pi_i)^2$$

$$= \sum_{i \in S} \sum_{j \in S} \lambda_i \lambda_j P[\lambda_i; i] P[\lambda_j; j | \lambda_i; i] - (\sum_{i \in S} \lambda_i \pi_i)^2$$

즉, 식(5)와 같다.

$$C_\lambda(\tau) = \sum_{i \in S} \sum_{j \in S} \lambda_i \lambda_j \pi_i(t) P_{ij}(\tau) - (\sum_{i \in S} \lambda_i \pi_i(t))^2 \quad (5)$$

이 때,

$P_{ij}(\tau) = e^{A\tau}$ 이며, $\pi_i(t) = \pi_i(0)e^{(A \cdot t)}$ 의 관계가 성립된다.

한편, MMGP를 따르는 데이터의 분산을 v_λ 라 하면, 분산은 식(6)과 같다.

$$v_\lambda = \sum_{i \in S} (\lambda_i - m)^2 P[\lambda_i; i]$$

$$= \sum_{i \in S} (\lambda_i - m)^2 \pi_i \quad (6)$$

그러므로, 구하는 자기상관함수는 식(7)이 된다.

$$R_\lambda(\tau) = \frac{C_\lambda(\tau)}{v_\lambda} \quad (7)$$

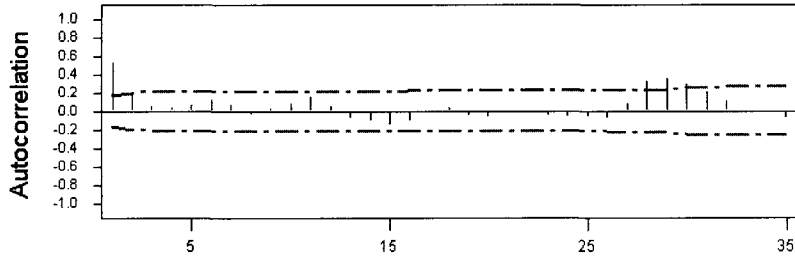
4. MMGP 모델의 적합성 비교

본 절에서는 실제 SOX 배출량 데이터에 제안하고 있는 MMGP 모델의 자기상관함수와 기존 시계열분석법의 자기상관함수를 비교한다. 각각의 분석은 Matlab 5.3 버전을 사용하여 프로그래밍 하였고, Minitab 12 버전을 이용하여 분석하였다. 사용된 데이터는 매일 확인한 일년 동안의 화력발전소 SOX 배출량 319개이다.

이 비교로 알 수 있는 것은 다음과 같다.

- 1) 시계열 모형으로 표현되는 SOX 데이터는 MMGP 모형으로도 적합하다고 할 수 있는가
- 2) 자기상관함수를 시계열 모형과 MMGP 모형으로 각각 분석한 경우 어느 모형으로의 식별결과가 자기상관도를 잘 나타내는가

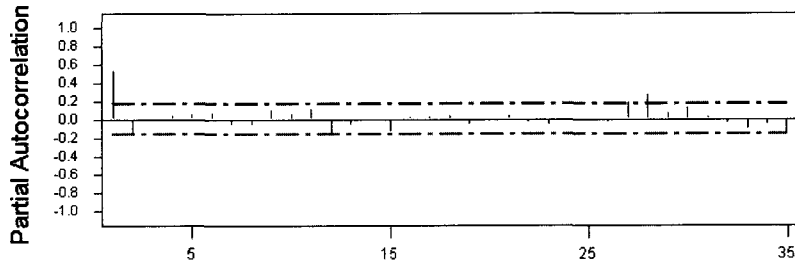
Autocorrelation Function for SOX



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.53	6.39	41.72	10	0.08	0.73	52.18	19	-0.03	-0.28	63.47	28	0.32	2.78	85.19
2	0.19	1.81	46.98	11	0.16	1.46	56.16	20	-0.04	-0.39	63.80	29	0.35	2.94	107.93
3	0.05	0.45	47.33	12	0.04	0.40	56.47	21	-0.01	-0.08	63.81	30	0.30	2.38	124.73
4	0.04	0.35	47.54	13	-0.07	-0.63	57.24	22	0.01	0.06	63.82	31	0.21	1.57	132.73
5	0.07	0.67	48.32	14	-0.10	-0.86	58.72	23	-0.04	-0.31	64.04	32	0.10	0.76	134.68
6	0.12	1.09	50.40	15	-0.14	-1.23	61.78	24	-0.06	-0.49	64.58	33	-0.00	-0.00	134.68
7	0.07	0.64	51.12	16	-0.09	-0.80	63.12	25	-0.04	-0.38	64.91	34	-0.01	-0.06	134.69
8	-0.02	-0.16	51.16	17	-0.01	-0.12	63.15	26	-0.06	-0.56	65.62	35	-0.07	-0.50	135.57
9	0.02	0.17	51.21	18	0.03	0.27	63.30	27	0.09	0.76	66.95				

<그림 4.1> SOX 데이터의 시계열 모형에 대한 ACF

Partial Autocorrelation Function for SOX



Lag	PAC	T	Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	0.53	6.39	10	0.05	0.58	19	-0.05	-0.63	28	0.26	3.14
2	-0.13	-1.58	11	0.11	1.28	20	-0.01	-0.10	29	0.07	0.86
3	0.01	0.08	12	-0.15	-1.82	21	0.04	0.49	30	0.13	1.54
4	0.04	0.52	13	-0.05	-0.62	22	-0.02	-0.24	31	0.03	0.31
5	0.05	0.61	14	0.00	0.01	23	-0.05	-0.65	32	-0.04	-0.42
6	0.07	0.85	15	-0.12	-1.46	24	0.01	0.13	33	-0.10	-1.16
7	-0.04	-0.54	16	0.02	0.25	25	0.01	0.13	34	-0.05	-0.61
8	-0.05	-0.65	17	0.02	0.26	26	-0.02	-0.22	35	-0.15	-1.84
9	0.09	1.08	18	0.04	0.47	27	0.18	2.20			

<그림 4.2> SOX 데이터의 시계열 모형에 대한 PACF

4.1 SOX 배출량 데이터의 시계열 모형에 대한 자기상관함수

4.1.1 SOX 배출량 데이터의 모형식별

이 데이터에 대한 자기상관함수(Autocorrelation Function)와 편자기상관함수(Partial Autocorrelation Function)를

살펴보면 <그림 4.1>, <그림 4.2>와 같다.

ACF와 PACF가 시차 1에서 유의하고 나머지 시차에서는 모두 신뢰한계 내에 존재하며 PACF는 지수적으로 감소하는 것을 알 수 있다. 이로부터 잠정적으로 식별된 모형은 ARIMA(0,0,1)이다. 아카이케 정보판단 기준(AIC)을 이용해 모형을 식별해 보아도 ARIMA(0,0,1)의 값이

<표 4.1> SOX 배출량의 전이강도행렬

	state1	state 2	state 3	state 4	state 5
state 1	-0.0333	0.0056	0.0222	0.0056	0
state 2	0.005	-0.0302	0.0101	0.0151	0
state 3	0.007	0.0093	-0.028	0.0093	0.0023
state 4	0.0051	0.0026	0.0103	-0.0257	0.0077
state 5	0	0	0.006	0.0179	-0.0239

최소값이 되었으므로 이 데이터는 ARIMA(0,0,1)로 판별한다. 한편, 이 때의 자기상관 함수는 time lag의 수를 35개로 하여 관찰하여보면 시차1에서만 0.53으로 유의한 값을 가진다.

4.2 SOX 데이터의 MMGP 모형에 대한 자기상관 함수

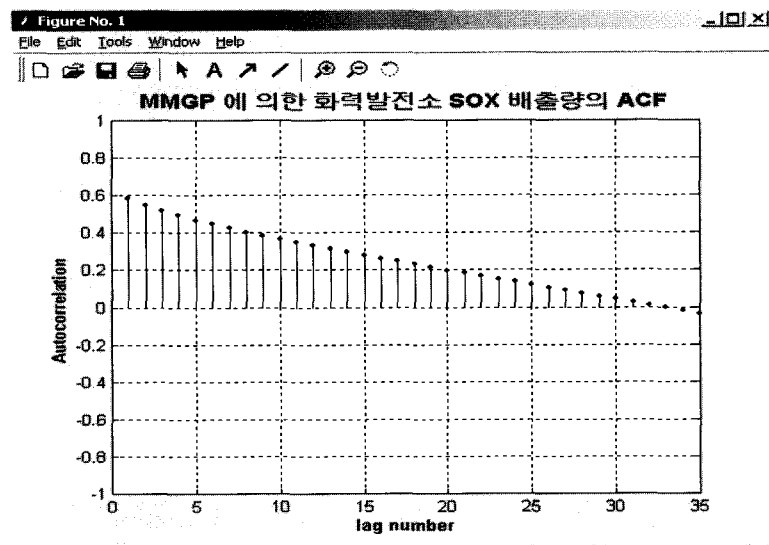
화력발전소의 sox 데이터는 MMGP 모형 식별단계에서 <그림 2.2>에서 보는바와 같이 5개의 μ 상태를 갖는

다. 각 상태간 전이강도 행렬 A 은 <표 4.1>과 같다.

이 때, $\pi_i(0)=(0.0232 \ 0.036 \ 0.813 \ 0.052 \ 0.076)$ 이며 안정 상태값은 $\pi_i(t)=(0.1293 \ 0.1404 \ 0.2925 \ 0.3047)$ 이다.

이제 time lag 의 수를 35개로 하여 자기상관함수를 살펴보면 <그림 4.3>과 같고, 그 값은 0.6 부근에서 점차 감소하다가 시차 33이상에서는 음의 값이 되는 것을 볼 수 있다. 그 수치는 <표 4.2>와 같다.

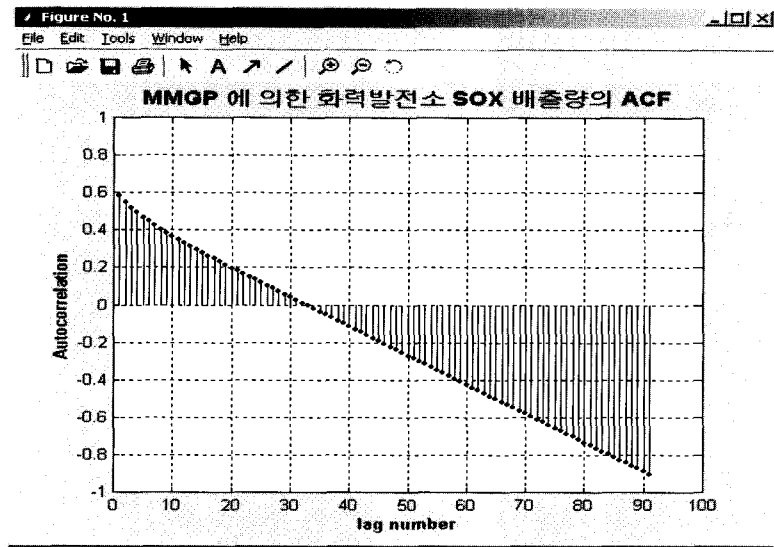
MMGP 모형으로 자기상관함수를 관찰하기 위해 구했던 MMGP 모형으로 분석한 안정상태의 $\pi_i(t)$ 는 t=15이



<그림 4.3> SOX 배출량의 ACF

<표 4.2> SOX 배출량의 ACF

Lag	Corr	Lag	Corr	Lag	Corr	Lag	Corr	Lag	Corr
1	0.5849	8	0.4040	15	0.2793	22	0.1667	29	0.0568
2	0.5511	9	0.3846	16	0.2628	23	0.1509	30	0.0412
3	0.5211	10	0.3658	17	0.2466	24	0.1352	31	0.0256
4	0.4939	11	0.3477	18	0.2304	25	0.1194	32	0.0100
5	0.4690	12	0.3301	19	0.2144	26	0.1038	33	-0.0056
6	0.4460	13	0.3129	20	0.1984	27	0.0881	34	-0.0212
7	0.4244	14	0.2960	21	0.1825	28	0.0725	35	-0.0367



<그림 4.4> time lag 에 따른 MMGP의 ACF

후로 안정값을 가진다. SOX데이터의 자기상관함수는 그 값이 선형으로 일정하게 감소한다. time lag을 늘려서 관찰 할 경우, time(t)를 100으로 설정하고 time lag을 250 까지 관찰해보면 <그림 4.4>를 얻을 수 있다.

5. 결 론

본 연구에서는 MMGP의 특성을 갖는 공정 데이터에 관한 자기상관함수를 제시하고 있다. 상태간 전이율 행렬과 각 상태에서의 발생확률을 통해 MMGP의 모습을 관찰하였다. 상태간 전이율 행렬은 마코프 프로세스가 갖는 기본적인 특성이기도 하다. 4절에서 볼 수 있듯이, 복수개의 평균치를 갖는 실제 화력발전소의 SOX 발생량의 데이터들은 시계열 데이터로도 분석가능하지만, MMGP 모형으로 분석하였을 경우 전자의 경우보다 높은 자기상관도를 표현할 수 있다.

실제 데이터 분석결과 복수개의 평균치가 나타나는 추이적 특성을 갖는 공정 데이터의 경우, 시계열 관리와 같은 기존의 데이터 관리 방법과는 구분해서 관리되어야 한다는 것을 알 수 있다. 이후에는 MMGP에서 모수들의 변화에 따른 자기상관함수의 변화에 대한 연구가 수행되어야 할 것이다.

참고문헌

[1] 안윤기, 확률과정입문, 자유아카데미, 1993.
 [2] 허 선, 확률과정론, 청문각 pp.181-296, 1997.
 [3] 서주원, "마코프 조정과정을 이용한 관리도의 설계"

석사학위논문, 아주대학교, 2000.

[4] Chen, D.T., Rieders, M., "Cyclic Markov Modulated Poisson Proces In Traffic Characterization", *Commun.S tatist-Stochastic Models*, 12(4), 585-610,1996.
 [5] Fischer, W., Meier-Hellstern, "The Markov-modulated Poisson process(MMPP) cook book", *Performance Eva luation*, 18, 1992, 149-171
 [6] Machihara F., "Covariance Structure Of Interrupted M arkov Modulated Poisson Process", *Journal of the Op erations Research*, Society of Japan, Vol. 36, No. 2, June 1993.
 [7] Keilson J., "Markov Chain Models-Rarity and Expone ntiality", Springer-Verlag, 1979.
 [8] J. F. Kingman, "On doubly stochastic Poisson proces es, *Proc. Camb. Phil. Soc.*, Vol. 60, pp923-930, 1964.
 [9] H. Heffes and D. M. Lucantoni, "A Markov modulate d characterization of packetized voice and data traffic and related statistical multiplexer performance", *IEEE J. Selected Areas Commun.*, Vol. SAC-4, No. 6, pp 8 56-868, 1986.
 [10] E. Cinlar, "Introduction to Stochastic Processes", Prent ice-Hall,(1975)