

국내 우편통계의 현황과 배달 우편물량에 대한 추정

김성주¹⁾

요약

이 논문에서는 국내 우편통계의 현황에 대하여 간략히 언급하고 있으며 접수 우편물량에 대한 기술통계와 배달 우편물량에 대한 추정을 다루고 있다. 성장곡선모형을 이용하여 접수 우편물량에 대한 예측을 논의하였고 성장곡선모형 중에서 삼차곡선이 매우 잘 적합시키는 것을 확인할 수 있었다. 또한 체신청별 배달 우편물량을 체신청별 접수 우편물량과 마코프 연쇄에서 나타나는 전이행렬을 이용하여 추정하였으며 이는 표본이론의 比推定의 기본 개념과 일맥 상통함을 발견하였다.

1. 서론

정보통신부 우정국에서는 3종류의 우편통계 조사를 실시하고 있다. 첫째는 접수 우편물 조사로서 매달 5일, 15일, 25일 3일간 전국의 모든 우체국에서 접수한 우편물량을 집계하는 것이다. 이 접수 우편물 조사에서 3일 동안 접수한 우편물량의 평균값과 해당 달의 일수를 곱하여 월별 통계를 구하고 이를 누적하여 연간 접수 우편물량을 구하는 것이다. 둘째는 배달 우편물 조사로서 이는 집배원들이 매일 배달한 우편물량을 집계하여 연간 배달 우편물량을 구하는 것이다. 셋째는 3년마다 실시하는 우편센서스로서 일년 중 하루를 택하여 우편이용실태를 조사하며 이에 는 우편이용에 관한 설문조사, 통상우편물 중량별 통계조사, 소포우편물 중량별 통계조사 및 접수 우편물 이동실태조사의 네 부분으로 구성된다.

정보통신부 우정국 담당자에 의하면 접수 우편물 통계는 정확도가 매우 높을 것이라 한다. 그 이유는 접수 우편물량의 70%~80%는 자동 소인기에 의해 소인되면서 집계되므로 착오가 있을 수 없고 나머지 20%~30%가 수동으로 소인되지만 그것도 우표 팔린 금액을 알기 때문에 거의 정확히 추정될 수 있다고 한다. 그러나 배달 우편물 조사는 오직 집배원들의 기록에만 의존하기 때문에 신뢰하기 어렵다고 한다. 더욱이 접수 우편물량이 많다고 해서 해당 우체국이나 체신청에 득이 될 것은 없지만 배달 우편물량이 많을 경우에는 집배원수를 증원해야 하고 그에 따른 예산을 증액해야 하는 매우 민감한 문제라고 한다.

연간 총 접수 우편물량과 배달 우편물량을 비교하면 배달 우편물량이 접수 우편물량에 비해 월등히 많다고 한다. 여러 이유로 인하여 반송된 우편물이 재차 배달될 수도 있기 때문에 배달 우편물량이 접수 우편물량 보다 어느 정도 많을 것으로 예상되지만 현재 상황은 그 차이가 심각하다고 한다. 따라서 정보통신부 우정국에서는 적어도 국내 8개 체신청 단위라도 배달 우편물량을 검증할 수 있는 방안을 모색해 보고자 한다.

고동희, 최중범(1992)에 의하면 집배원이 배달하는 우편물량은 하루 평균 1,895통이고 중량은 74kg이며 주행거리는 16km라고 한다. 여러 이유로 인하여 반송되는 우편물은 평균

1) (110-745) 서울시 종로구 명륜동, 성균관대학교 경영학부, 교수

26통으로 전체 배달 우편물량의 1.4%라고 한다. 한편 집배원의 옥내 준비작업, 옥외 작업 그리고 배달 후 정리 작업등 하루 평균 12시간을 근무하는 것으로 조사되었다. 열악한 작업 조건하에서 과중한 업무를 수행하고 있는 집배원들에게 매일 같이 2,000통에 가까운 우편물을 집계하라고 요구하는 것은 무리라고 생각된다. 따라서 이 논문에서는 배달 우편물량을 집배원에 의존하지 않고 추정할 수 있는 대안을 제시하고자 한다.

이 논문은 다음과 같이 구성되어 있다. 연간 접수 우편물량에 대하여 제2절에서는 기술통계를 다루고 있으며 제3절에서는 성장곡선모형의 적합을 다루고 있다. 체신청별 연간 배달 우편물량의 추정에 대하여 제4절에서는 일반적인 문제를 다루고 있으며 제5절에서는 실제 사례를 다루고 있다. 제6절에서는 결론 및 건의사항을 간략히 기술하고 있다.

2. 연간 접수 우편물량에 대한 기술통계

이 논문에서 이용할 데이터는 표 2.1과 표 2.2에 나타나 있다. 이는 체신부 우정국에서 월 3회 실시한 접수 우편물 조사에서 발췌한 것이다. 표 2.1은 80년부터 94년까지 연도별 접수 우편물량을 나타내고 있으며 이를 그림 2.1에 산점도로 나타내었다. 그림 2.1에 의하면 우리나라의 연간 접수 우편물량은 85년 이후 급격히 증가하고 있다는 것을 알 수 있다. 그림 2.1에 나타난 곡선은 성장곡선모형 중에서 삼차모형을 적합시킨 것으로서 제3절에서 다시 언급하고자 한다.

이제 연간 접수 우편물량 중에서 각 체신청이 차지하는 백분율이 연도에 따라 어떻게 변화하는지를 살펴보기로 하자. 이를 위하여 표 2.2에서 (a)는 91년부터 94년까지 연도별 체신청별 접수 우편물량을 나타내고 있고 (b)는 (a)를 백분율로 나타낸 것이다. 표 2.2(b)에 의하면 서울청의 백분율은 수도권 인구 증가로 인하여 완만한 상승세를 나타내고 있으나 나머지 7개 청의 백분율은 상당히 안정된 형태를 띠고 있음을 알 수 있다. 이 논문에서는 접수 우편물량에 대해서만 다루고 있으나 이를 보통통상, 특수통상, 소포 등으로 세분화한 결과는 김성주(1996)에서 발견할 수 있고 동일한 결론을 얻을 수 있다.

여기서 우리는 매우 중요한 사실을 발견할 수 있다. 즉, 각 체신청의 접수 우편물량은 매년 증가하고 있지만, 접수 우편물량 중에서 각 체신청이 차지하는 백분율은 연도에 무관하게 안정적이라는 점이다. 이는 大數의 法則과 연관지어 설명할 수 있다. 예컨대 91년도 총 접수 우편물을 모집단이라 하고 그 중 서울청에 접수된 모비율(population proportion)을 θ_{91} 이라고 표시하자. 그러면 월 3회 실시한 접수 우편물 조사에 의해 θ_{91} 은 표 2.2(b)에서 91년도 서울청의 접수 비율인 58.3%로 추정된다. 만약 이 모집단에서 단순확률표본을 얻을 수 있다면 그 때의 표본비율은 대수의 법칙에 의해 모비율 θ_{91} 에 대한 일치 추정량이 될 것이다. 비록 월 3회 실시한 접수 우편물 조사는 단순확률표본에 의한 것은 아니지만 여기서 얻어진 표본 비율인 58.3%는 모비율에 매우 가까운 것으로 예상된다. 마찬가지로 92년도, 93년도, 94년도 각각에 대하여 서울청에 접수된 모비율도 정의될 수 있으며 이 값은 θ_{91} 에 근사할 것으로 예상된다. 왜냐하면 짧은 기간 동안에는 모비율의 변화는 매우 작을 것으로 예상되기 때문이다.

표 2.1: 연도별 접수 우편물량(단위는 백만 통)

연도	80년	81년	82년	83년	84년	85년	86년	87년	88년	89년	90년	91년	92년	93년	94년
접수량	1055	1072	1028	1103	1172	1268	1350	1582	1825	2092	2324	2579	2807	2960	3129

표 2.2: 연도별 체신청별 접수 우편물

(a) 접수 우편물량(단위는 백만 통)

체신청	91년	92년	93년	94년
서울청	1,503	1,645	1,761	1,930
부산청	306	337	359	345
충청청	217	232	236	249
전남청	166	171	178	171
전북청	82	88	86	89
경북청	221	240	241	245
강원청	67	75	79	80
제주청	17	19	20	20
합계	2,579	2,807	2,960	3,129

(b) 백분율

체신청	91년	92년	93년	94년
서울청	58.3%	58.6%	59.5%	61.7%
부산청	11.9%	12.0%	12.1%	11.0%
충청청	8.4%	8.3%	8.0%	8.0%
전남청	6.4%	6.1%	6.0%	5.5%
전북청	3.2%	3.1%	2.9%	2.8%
경북청	8.6%	8.5%	8.1%	7.8%
강원청	2.6%	2.7%	2.7%	2.6%
제주청	0.6%	0.7%	0.7%	0.6%
합계	100%	100%	100%	100%

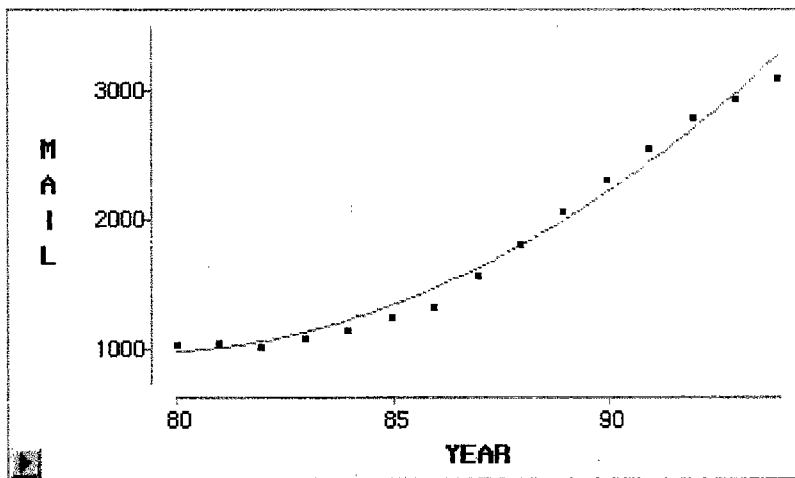


그림 2.1: 연도별 접수 우편물량의 산점도와 추정된 삼차 곡선

3. 연간 접수 우편물량에 대한 성장곡선모형

여기서는 표 2.1에 있는 연도별 접수 우편물량에 대하여 성장곡선모형(growth curve model)을 적합시켜 보고자 한다. 성장곡선모형이란 t 시점에서의 접수 우편물량을 y_t 라 표시할 때 y_t 가 t 의 함수로서 표시될 수 있다고 가정하는 것이며 함수의 형태는 흔히 아래와 같이 7가지 형태로 대별된다.

- (1) 직선 모형: $y_t = a + bt$
- (2) 포물선 모형: $y_t = a + bt + ct^2$
- (3) 삼차 모형: $y_t = a + bt + ct^2 + dt^3$
- (4) 지수 모형: $\log y_t = a + bt$
- (5) 수정된 지수 모형: $y_t = a + bc^t$
- (6) 고펜퍼츠 모형(Gompertz Model): $\log y_t = a + bc^t$
- (7) 펄 모형(Pearl Model): $y_t = (a + bc^t)^{-1}$

표 2.1에 있는 데이터에 대하여 위의 7가지 모형을 적합시켜 본 결과 모형의 단순성은 물론이고 적합성에 있어서도 삼차 모형이 매우 우수한 것으로 나타났다. SAS를 이용하여 추정한 삼차 곡선이 식 (3.1)에 나타나 있으며 그래프는 그림 2.1에 나타내었다. 그림 2.1에 의하면 삼차 곡선이 매우 잘 적합시키는 것을 알 수 있으며 수정 결정계수는 0.996이다.

$$y_t = 863,878 - 29,212t + 328t^2 - 1.2t^3, \quad t = 80, \dots, 94 \quad (3.1)$$

식 (3.1)은 80년부터 94년까지 데이터를 이용하여 추정한 곡선인데 우리의 궁극적인 관심은 이 추정 곡선이 95년 이후의 값을 어느 정도 정확하게 예측할 수 있겠느냐는 것이다. 성장곡선모형이 주로 짧은 기간의 예측 목적으로 이용된다는 점을 감안하여 95년 이후 3년 동안 접수 우편물량을 예측하여 보고 이를 실제 값과 비교해 보고자 표 3.1을 작성하였다. 표 3.1에서 실제 값은 정보통신부 우정국(1997)에서 발췌한 것으로 표 2.1의 값과 동일한 방식으로 얻어진 것이다. 표 3.1에 의하면 1년 후 예측의 상대오차는 4.0%로서 상당히 만족스러운 결과를 얻었다. 그러나 예측 기간이 길어지면 상대오차가 증가하는 것을 알 수 있다.

표 3.1: 식 (3.1)에 의한 예측 값과 실제 값(단위는 백만 통) 및 상대오차

연도	95년	96년	97년
예측값	3,291	3,408	3,480
실제값	3,427	3,778	3,918
상대오차	4.0%	9.8%	11.2%

4. 체신청별 연간 배달 우편물량의 추정

여기서는 8개 체신청별 연간 배달 우편물량의 추정에 대하여 논의하고자 한다. 설명의 편의상 연도를 고정시키고 8개 체신청(서울청, 부산청, 충청청, 전남청, 전북청, 경북청, 강

원청, 제주청) 각각에 대하여 순서대로 1부터 8까지 자연수를 대응시키자. 또한 i 번째 체신청에 접수된 우편물 중에서 j 번째 체신청으로 배달된 우편물의 표본 비율을 p_{ij} 라 표시하자. 그러면 8개 체신청간의 표본 비율 p_{ij} 로 이루어진 8×8 행렬 P 를 얻게 되며 한 예가 표 5.1에 나타나 있다. 이 행렬 P 는 매우 특수한 형태를 띠고 있다. 즉, 모든 원소들이 음이 아니며 각 행의 원소의 합은 언제나 1이 된다. 따라서 이 행렬 P 는 마코프 연쇄(Markov Chain)에서 전이행렬(transition matrix)의 표본값으로 이해될 수 있으며 이 논문에서는 간략히 전이행렬이라 지칭하고자 한다.

전이행렬 P 는 3년마다 실시하는 우편센서스에서 접수 우편물 이동실태조사로부터 구할 수 있으며 이 값은 연도에 따라 크게 달라지지 않을 것으로 예상된다. 그 이유는 제2절에서 설명한 것을 다음과 같이 수정하면 될 것이다. 예를 들어 어느 특정 연도에 서울청에 접수된 우편물을 모집단이라 하고 그 중 서울청에 배달된 모비율을 θ_{11} 이라 표시하자. 그러면 모비율 θ_{11} 은 표본 비율 p_{11} 으로 추정되며 대수의 법칙에 의해 p_{11} 은 θ_{11} 에 근사할 것으로 예상된다. 또한 기간이 짧을 경우 θ_{11} 은 연도에 따라 크게 달라지지 않을 것으로 예상된다. 이러한 주장을 일반화하면 p_{ij} 은 θ_{ij} 에 근사하고 기간이 짧을 경우 θ_{ij} 는 연도에 따라 크게 달라지지 않을 것으로 예상되기 때문에 전이행렬 P 는 연도에 따라 크게 달라지지 않을 것으로 예상된다. 이러한 주장은 여러 차례 실시된 접수 우편물 이동 실태조사를 통하여 확인할 수 있다.

이제 i 번째 체신청의 접수 우편물량을 s_i 라 하고 배달 우편물량을 r_i 라고 표시하자. 그러면 체신청별 배달 우편물량은 식 (4.1)과 같이 추정된다.

$$[r_1 \cdots r_8] = [s_1 \cdots s_8] \begin{bmatrix} p_{11} & \cdots & p_{18} \\ \vdots & \dots & \vdots \\ p_{81} & \cdots & p_{88} \end{bmatrix} \quad (4.1)$$

우리는 식 (4.1)에서 매우 중요한 사실을 발견할 수 있다. 즉, 식 (4.1)에 의하면 체신청별 배달 우편물량은 체신청별 접수 우편물량과 전이행렬의 곱으로 추정되는데 이는 표본이론에서 말하는 比推定(ratio estimate)의 기본 개념과 일맥상통한다는 것이다. 박홍래(1989)의 388쪽에 의하면 비추정이란 Y 를 추정할 때 Y 와 밀접한 관계가 있는 보조 변수 X 를 이용하자는 것이다. 예컨대 쌀의 총 생산량 Y 를 추정하고자 한다면 벼 재배 면적 X 와 단위당 생산량 R 을 이용하여 식 (4.2)와 같이 추정하자는 것이다.

$$\hat{Y} = X\hat{R} \quad (4.2)$$

식 (4.1)과 식(4.2)를 비교하면, 식(4.1)에 있는 체신청별 배달 우편물량, 체신청별 접수 우편물량, 전이행렬은 각각 식 (4.2)에서 \hat{Y} , X , \hat{R} 의 역할을 하고 있음을 알 수 있다. 또 다른 공통점은 식 (4.2)에서 X 는 정확히 알 수 있어야 하는데 식 (4.1)에서도 체신청별 접수 우편물량은 비교적 정확히 알 수 있다는 점이다.

만약 전년도 체신청별 배달 우편물량을 추정하고자 한다면 체신청별 접수 우편물량은 알려진 값이라고 할 수 있으므로 식 (4.1)에 의해 체신청별 배달 우편물량을 직접 추정할 수 있다. 그러나 내년도 체신청별 배달 우편물량을 추정하고자 한다면 체신청별 접수 우편

물량을 모르기 때문에 이 값부터 추정해야 할 것이다. 이 논문에서는 체신청별 접수 우편물량을 성장곡선 모형에 의한 전국의 접수 우편물량에 대한 추정 값과 체신청별 접수 우편물량의 비율의 곱으로 추정하고자 한다. 즉, 성장곡선 모형에 의한 전국의 접수 우편물량에 대한 추정 값을 T 라 표시하고 i 번째 체신청에 접수된 우편물량의 비율을 q_i 라고 표시하자. 그러면 체신청별 접수 우편물량은 식 (4.3)과 같이 추정된다. 식 (4.3)에서 i 번째 체신청에 접수된 우편물량의 비율 q_i 는 표 2.2(b)에 나타나 있으며 이 값은 제2절에서 설명한 바와 같이 연도에 따라 크게 달라지지 않는 매우 안정적인 값이다.

$$[s_1 \cdots s_8] = T [q_1 \cdots q_8] \quad (4.3)$$

따라서 체신청별 접수 우편물량을 모를 경우에는 체신청별 배달 우편물량은 식 (4.4)와 같이 추정된다.

$$[r_1 \cdots r_8] = T [q_1 \cdots q_8] \begin{bmatrix} p_{11} & \cdots & p_{18} \\ \vdots & \cdots & \vdots \\ p_{81} & \cdots & p_{88} \end{bmatrix} \quad (4.4)$$

본 연구에서 비율을 이용하여 추정하고자 하는 이유는 비율을 이용할 경우 몇 가지 매우 중요한 장점이 있기 때문이다. 첫째, 비율에 관하여는 앞에서 언급한 大數의 法則이 성립한다는 점이다. 둘째, 비율은 그 값이 0과 1사이로 한정된다는 점이다. 셋째, 비율에 관한 추정은 통계학뿐만 아니라 산업공학 분야에서도 매우 중요한 과제로서 이에 관하여는 방대한 기존의 연구결과를 이용할 수 있다는 점이다.

5. 체신청별 배달 우편물량의 추정 사례

제5절에서는 실제 데이터를 이용하여 앞에서 언급한 체신청별 배달 우편물량을 추정해 보고자 한다. 이를 위하여 전이행렬 P 를 고동희(1994)의 169쪽에 있는 "체신청별 우편물 발송 및 도착 현황"을 바탕으로 하여 구했으며 이를 표 5.1에 나타내었다.

첫째, 체신청별 접수 우편물량을 아는 경우 체신청별 배달 우편물량을 추정해 보자. 예컨대 1994년도 체신청별 배달 우편물량을 식 (4.1)을 이용하여 추정해 보자. 이 경우는 체신청별 접수 우편물량이 표 2.2(a)에 주어져 있고 전이행렬은 표 5.1에 나타나 있기 때문에 이를 식 (4.1)에 대입하면 체신청별 배달 우편물량은 다음과 같이 추정된다.

$$\begin{aligned} [r_1 \cdots r_8] &= [1, 930 \cdots 20] P \\ &= [1, 757 \ 372 \ 297 \ 171 \ 97 \ 321 \ 86 \ 28] \end{aligned}$$

즉, 서울청의 배달 우편물량은 1,757(단위는 백만 통), 부산청의 배달 우편물량은 372, 제주청의 배달 우편물량은 28로 추정된다.

둘째, 체신청별 접수 우편물량을 모르는 경우 체신청별 배달 우편물량을 추정해 보자. 예컨대 1994년까지 데이터를 바탕으로 1995년도 체신청별 배달 우편물량을 식(4.4)를 이

표 5.1: 8개 체신청간 전이행렬 P

	배 달								
	서울청	부산청	충청청	전남청	전북청	경북청	강원청	제주청	
서울청	0.832	0.054	0.041	0.018	0.011	0.028	0.012	0.004	
부산청	0.098	0.711	0.018	0.009	0.004	0.151	0.007	0.002	
집	충청청	0.138	0.019	0.795	0.010	0.010	0.015	0.011	0.002
	전남청	0.172	0.029	0.025	0.740	0.015	0.010	0.006	0.003
	전북청	0.177	0.009	0.023	0.017	0.762	0.005	0.006	0.001
수	경북청	0.084	0.035	0.018	0.005	0.003	0.846	0.008	0.001
	강원청	0.205	0.035	0.034	0.016	0.009	0.025	0.675	0.001
	제주청	0.053	0.018	0.007	0.011	0.002	0.004	0.003	0.902

용하여 추정해 보자. 식 (3.1)에 의해 1995년도 전국의 접수 우편물량은 3,291(단위는 백만 통)로 추정되고 표 2.2(b)에 있는 94년도 체신청별 접수 비율을 이용하면 체신청별 배달 우편물량은 다음과 같이 추정된다.

$$\begin{aligned}
 [r_1 \cdots r_8] &= 3,291 [0.617 \cdots 0.006] P \\
 &= [1,849 \ 391 \ 313 \ 181 \ 101 \ 337 \ 91 \ 28]
 \end{aligned}$$

6. 결론 및 건의사항

3년마다 실시하는 우편센서스는 엄밀히 말해 센서스라고 할 수 없다. 왜냐하면 센서스라고 하면 통계청에서 실시하는 인구센서스와 같이 그 결과가 참 값이라는 것을 전제로 하는데 94년 우정국에서 실시한 센서스는 일년 중 어느 하루를 택해 전국 8781개 배달 구역 중에서 약 2%에 해당하는 205개 배달 구역을 조사한 결과를 바탕으로 일년동안의 총 물량을 추정하기 때문이다. 이러한 이유로 인하여 94년 센서스 결과 보고서인 고동희(1994)에 의하면 국내 우편물은 줄고 국제 우편물량은 늘고 있다는 받아들이기 매우 어려운 결론을 얻었으며 이는 접수 우편물량이 급증하고 있는 것을 보여주는 그림 (2.1)과 상충된다. 따라서 우편센서스에서 실시하는 4가지 통계조사를 비록 소규모로 실시하는 한이 있더라도 적어도 분기마다 한번씩 실시할 것을 건의한다. 이렇게 얻어진 접수 우편물 이동실태조사를 이용하면, 집배원을 통한 배달 우편물 통계를 지양하고 본 연구결과인 식 (4.1)과 식 (4.4)에 의해 배달 우편물량을 추정할 수 있을 것이다.

현재 특송업체(DHL, Fedral Express등)에서는 우편물마다 발송인과 수취인의 주소를 비롯한 여러 정보를 전산으로 입력하여 처리하고 있다. 이렇게 한다면 우편통계는 더 이상 논의의 대상이 될 필요가 없을 것이다. 국내 우편물의 경우 이렇게 까지는 못하더라도 미국의 우체국에서 실시하고 있는 바와 같이 우편물마다 수취인 우편번호를 바코드로 입력

하는 방안을 강구해 볼 수 있다. 이렇게 할 경우 우편물을 자동으로 분류할 수 있을 뿐만 아니라 우편통계업무도 대폭 개선될 수 있을 것이다.

현재 우편업무 담당자들은 과도한 우편업무를 처리하면서 아울러 우편통계업무를 취급하는 관계로 그들에게서 통계조사의 연속성, 전문성, 효율성을 기대하기 어렵다고 생각된다. 따라서 접수물량과 배달물량을 포함한 모든 우편통계에 관하여 우편통계 업무만을 전담하는 인원을 배정하는 것이 절실하다고 생각한다.

끝으로 이 논문의 심사과정에서 여러 가지 건설적인 제안을 해주신 익명의 심사위원 두 분께 감사드린다

참고문헌

- [1] 고동희 (1994). <우편서비스의 품질 및 이용특성 정보시스템의 설계와 개발>, 우편센서스 결과보고, 통신개발연구원.
- [2] 고동희, 최중범 (1992). <대도시 우편물 배달의 효율성 제고 방안>, 통신개발연구원.
- [3] 김성주 (1996). <우편통계업무 개선방안>, 정보통신학술연구지원국.
- [4] 박홍래 (1989). <통계조사론>, 영지문화사, 서울.
- [5] 정보통신부 우정국 (1998). <우편통계편람>.
- [6] 체신부 우정국 (1994). <우편통계편람>.

[1998년 9월 접수, 1999년 7월 최종수정]

Korean Mail Statistics and Estimation of the Amount of Delivered Mail

Seong-Ju Kim¹⁾

ABSTRACT

This paper concerns Korean mail statistics and their description. We discuss the forecasting of the annual amount of incoming mail using the growth curve models. It is observed that the cubic model works very well. Also we discuss the estimation of the annual amount of delivered mail by a local district. It is observed that the result is expressed in terms of the so called transition matrix in Markov chain. Also it is noticed that the result is related to the ratio estimation in sampling theory.

1) Professor, School of Business Administration, Sungkyunkwan University, Seoul 110-745, Korea