2003 Proceedings of the Spring Conference, Korean Statistical Society

이변량 Laplace 분포와 응용

홍성식[1], 홍종선[2]

Abstract

주변분포가 Laplace 분포인 세 가지 형태의 이변량 Laplace 분포를 연구한다. 각각의 이변량 Laplace 분포의 확률밀도함수와 누적분포함수를 유도하고, 분포의 그래프를 그려봄으로써 분포의 형태를 알아본다. 조건부 적률을 정리하여 조건부 첨도와 조건부 왜도를 구하고 분포의 성질을 파악한다. 상관계수를 구하여 다른 이변량 분포의 상관계수와 비교해 보았다. 그리고 정의된 분포함수를 응용하여 이변량 Laplace 분포를 따르는 난수벡터를 발생하는 알고리즘을 제안하였으며, 생성된 난수벡터의 표본으로부터 구한 표본평균과 중앙값의 분산-공분산 행렬식을 구하고 이변량 정규분포에 대응하는 행렬식과 비교 토론하였다.

Keywords: 난수벡터, 상관계수, 이변량 Laplace, 조건부첨도, 조건부왜도.

1. 서 론

지수분포(exponential distribution)는 어떤 사상이 정해진 시간으로부터 다음 번 사상이 발생하기까지의 시간에 관한 확률분포함수이며, 실제 많이 응용되고 있는 확률분포 중의 하나이다. 또한, 극단값(extreme value)의 이론에 대한 시발점으로 사용될 수 있다. Laplace 분포 또는 이중지수분포(double exponential distribution)의 경우에 모수의 값에 따라 이봉분포(bimodal distribution)의 형태를 나타내기도 하는데 이는 두 개의 정규(normal)분포의 선형결합(linear combination)분포로 생각해 볼 수 있다. 이변량(bivariate) 분포에 대한 대부분의 연구는 정규분포인 경우에 Bravais 와 Karl Pearson (19C 초) 시대 이래로 집중적으로 진행되어 왔는데, 이변량 정규분포에서는 다음과 같은 잘 알려진 성질들이 있다.

1) 등확률밀도(equal probability density)의 곡선(curve)은 동심 타원(concentric ellipses)이다.
2) $E[Y|x]$, $E[X|y]$은 직선이고, 기대값에서 교차된다.
3) 한 변수의 값이 증가함에 따라 다른 한 변수의 조건부 기대값은 무한히 증가 또는 감소한다.
4) $-1 \le \rho \le 1$

주변(marginal)분포가 그에 상당하는 이변량 분포를 결정하는 것은 아니다. Frechet (1951)은 주변분포가 주어졌을 때 이 주변분포를 갖는 무한히 많은 이변량분포가 존재한다는 것을 증명하였는데, 이 결과는 이변량 분포의 존재에 관한 것이지 그들의 구조 및 형태에 관한 것은 아니었다. Gumbel (1960)은 위의 네 가지 성질을 모두 만족하지 않는, 주변분포가 정규분포인

1) (135-070) 서울시 종로구 명륜동 3가 성균관대학교 경제학부 통계학전공
2) (135-070) 서울시 종로구 명륜동 3가 성균관대학교 경제학부 통계학전공, 교수

이변량 분포가 존재함을 보였다. 그리고 주변분포가 지수분포인 특별한 형태의 이변량 분포를 연구했는데, 여기에서도 역시 위의 네 가지 성질들이 만족되지 않았다. 특히, 상관관계는 협소한 범위 내에서 변화한다는 것을 발견하였다.

2. 이변량 Laplace 분포의 정의와 정리

본 연구에서는 주변분포가 지수분포인, 세 가지 형태의 이변량 지수분포에 대한 Gumble (1960)의 연구 내용을 확장하여 주변분포가 Laplace 분포인 이변량 Laplace 분포의 형태와 성질을 연구한다.

첫 번째 형태의 이변량 Laplace 분포

Gumbel (1960)이 연구한 주변분포가 지수분포인 이변량 지수분포함수를 확장하여 첫 번째 이변량 Laplace 확률밀도함수와 누적분포함수를 정의한다. Gumbel (1960)이 연구한 첫 번째 형태의 이변량 지수분포에서의 상관계수는 $-0.40 \le \rho \le 0$ 이나 첫 번째 형태의 이변량 Laplace 분포에서의 상관계수는 0임을 구하였다.

두 번째 형태의 이변량 Laplace 분포

Morgenstern(1956)이 연구한 주변분포가 지수분포인 이변량 지수분포분포함수를 확장하여 두 번째 이변량 Laplace 확률밀도함수와 누적분포함수를 정의한다. 두 번째 형태의 이변량 지수분포에서의 상관계수 값의 범위는 $-0.25 \le \rho \le 0.25$이나 두 번째 형태의 분포로부터 얻어진 이변량 Laplace분포에서의 상관계수는 $-0.328 \le \rho \le 0.28125$ 이다. 이 값의 구간은 이변량 정규분포보다는 협소한 구간을 형성했지만, 이변량 지수분포보다는 약간 넓은 구간을 형성했다.

세 번째 형태의 이변량 Laplace 분포

Gumbel (1960)의 세 번째 형태의 이변량 지수분포를 이용하여 세 번째 형태의 이변량 Laplace 분포의 확률밀도함수를 정의하였으며 누적분포함수를 구하기 위해서는 적분계산과정이 복잡하기 때문에 누적분포함수와 분포의 성질에 관한 그 외의에 대하여는 향후 연구 과제로 남겨두었다.

각각의 형태의 이변량 Laplace 분포에 대하여 조건부 적률을 정리하여 홀수차 적률과 짝수차 적률에 유도하였다. 첫 번째 형태의 이변량 Laplace 분포에서 조건부 기대값은 홀수차수일 경우에는 항상 0으로 일정하므로 조건부 왜도는 0으로 그래프는 좌우대칭이다. 두 번째 형태의 조건부 첨도는 3으로서 정규분포와 비교해 볼 때, 정규분포보다 중심 부분이 뾰족하게 솟아있는 형태를 나타내었다. 왜도는 $Y=0$일 때는 0을 중심으로 좌우 대칭이고 α 값이 증가할수록 증가한다.

3. 이변량 Laplace 분포에서의 난수벡터 생성

첫 번째와 두 번째 형태의 Laplace 분포의 누적분포함수의 역함수를 이용하여 두 분포를 따르는 확률난수벡터를 발생시킬 수 있다. 다음은 난수벡터를 발생하는 알고리즘이다.

<Algorithm>

Step 1. $u_1 \sim U(0,1)$ 을 생성한다.

① $0 < u_1 < F(0,0)$ 이면, $x<0$, $y<0$ 생성할 예정이다.

② $F(0,0) < u_1 < 0.5$ 이면, $x<0$, $y \ge 0$ 생성할 예정이다.

③ $0.5 < u_1 < 1-F(0,0)$ 이면, $x \ge 0$, $y<0$ 생성할 예정이다.

④ $1-F(0,0) < u_1 < 1$ 이면, $x \ge 0$, $y \ge 0$ 생성할 예정이다.

Step 2. $0<u_1<0.5$일 때, $u_2 \sim U(0,0.5)$를 발생시키면 X의 주변누적분포함수의 역함수 $F_X^{-1}(u_2)$를 이용하여 음수인 x 발생시킬 수 있고, $0.5<u_1<1$일 때, $u_2 \sim U(0.5,1)$를 발생시키면 X의 주변누적분포함수에서 비음수인 x 발생시킬 수 있다.

Step 3. Step 1과 Step 2에서 얻어진 u_1과 x를 조건부 누적분포함수의 역함수 $F_{Y|x}^{-1}(u_1)$를 이용하여 y를 발생한다.

Step 4. 생성된 난수 y 값의 범위가 Step 1에서 기대한 바와 같이 적절하다면 Step 1로 돌아가서 설정된 표본크기의 난수벡터를 얻을 때까지 반복한다. 만약 생성된 난수 y가 적절하지 않은 범위에서 발생되는 경우에는, Step 2로 돌아가 새로운 x 를 발생한 후 조건에 맞는 난수 y를 다시 발생한다.

4. 응용-이변량 Laplace 분포의 표본평균과 중앙값

본 연구에서는 앞서 제시된 첫 번째 형태의 이변량 Laplace 분포와 두 번째 형태의 분포에 의한 이변량 난수 벡터와 이변량 정규분포를 따르는 각각의 난수 벡터를 발생하여 발생된 표본의 표본평균과 표본중앙값에서 분산-공분산 행렬을 비교하고자 한다.

실험은 각각 표본크기 50인 표본을 1,000회 반복 발생하여 표본평균과 표본중앙값의 효율을 비교한다. 이 때 첫 번째와 두 번째 Laplace 분포에서는 θ와 α 를 각각 0에서 1까지와 -1에서 1까지 각각 0.2 씩 증가시키며 실험하고, 이변량 정규분포의 실험에서도 마찬가지로 상관계수의 변화에 따라 결과를 살펴본다. 일변량일 때의 표본평균과 표본중앙값의 분산비율이 정규분포에서 1보다 작고 Laplace 분포에서는 1보다 크다는 사실을 확인할 수 있다. 이변량 난수벡터의 표본평균과 중앙값의 분산을 비교하기 위하여 각각의 분산공분산행렬 $\Sigma_{\overline{X}\,\overline{Y}}$, 와 $\Sigma_{M_X M_Y}$의 행렬식(determinant)의 비를 비교한다. 결과는 단변량일 때와 유사하게 이변량 정규분포에서는 1보다 작고 이변량 Laplace분포에서는 1보다 큰 값을 얻을 수 있다. 첫 번째 형태의 이변량 Laplace분포의 분산공분산행렬 $\Sigma_{\overline{X}\,\overline{Y}}$, 와 $\Sigma_{M_X M_Y}$의 행렬식의 비율은 두 번째 형태의 분포보다 크다는 것을 발견하였다.

<표 1> 이변량 정규분포에서의 표본평균과 표본중앙값의 행렬

상관계수	$V(\overline{X})$	$V(M_X)$	$\frac{V(\overline{X})}{V(M_X)}$	$Cov(\overline{X}, \overline{Y})$	$Cov(M_X, M_Y)$	$\frac{Det(\Sigma_{\overline{X}\,\overline{Y}})}{Det(\Sigma_{M_X M_Y})}$
-0.6	0.0199	0.0296	0.6716	-0.0121	-0.0128	0.3511
-0.2	0.0203	0.0303	0.6693	-0.0028	-0.0026	0.4178
0.2	0.0202	0.0317	0.6376	0.0047	0.0044	0.3857
0.6	0.0206	0.0312	0.6606	0.0113	0.0134	0.3649

<표 2> 첫 번째 형태의 이변량 Laplace에서 표본평균과 표본중앙값의 행렬

θ	$V(\bar{X})$	$V(M_X)$	$\frac{V(\bar{X})}{V(M_X)}$	$Cov(\overline{X}, \overline{Y})$	$Cov(M_X, M_Y)$	$\frac{Det(\Sigma_{\overline{XY}})}{Det(\Sigma_{M_X M_Y})}$
0.0	0.0191	0.0087	2.1881	0.0001	0.0002	4.2139
0.2	0.0154	0.0081	1.8929	-0.0006	-0.0003	3.6100
0.4	0.0151	0.0077	1.9655	-0.0005	-0.0004	3.3323
0.6	0.0133	0.0074	1.7948	-0.0003	0.0003	3.6479
0.8	0.0137	0.0074	1.8491	-0.0006	-0.0005	3.2745
1.0	0.0132	0.0073	1.8132	0.0004	0.0000	3.4505

<표 3> 두 번째 형태의 이변량 Laplace에서의 표본평균과 표본중앙값의 행렬

α	$V(\bar{X})$	$V(M_X)$	$\frac{V(\bar{X})}{V(M_X)}$	$Cov(\overline{X}, \overline{Y})$	$Cov(M_X, M_Y)$	$\frac{Det(\Sigma_{\overline{XY}})}{Det(\Sigma_{M_X M_Y})}$
-1.0	0.0417	0.0362	1.3958	-0.0139	-0.0079	1.8362
-0.6	0.0415	0.0310	1.4027	-0.0083	-0.0049	2.0152
-0.2	0.0358	0.0259	1.4754	-0.0045	-0.0008	2.0198
0.2	0.0343	0.0244	1.4047	-0.0011	0.0013	2.0391
0.6	0.0388	0.0286	1.3005	0.0064	0.0038	2.2621
1.0	0.0423	0.0314	1.3265	0.0114	0.0060	2.1414

참 고 문 헌

1. Frechet, M. (1951),"Sur les tableau de correlation dont les marges sont donnees", *Annales de l'Universite de Lyon,* Ser.Ⅲ, 14, 53-77.
2. Gumbel, E. J. (1960), "Bivariate Exponential Distributions", *Journal of the American Statistical Association,* Vol. 55, pp.698-707.
3. Morgenstern, D. (1956). "Einfache Beispiele zweidimensionaler Verteilungen", *Mitteiungsblatt Fur mathematicsche Statistik,* Vol. 8, 234-235.
4. Norman, L. Johnson, and Samuel Kotz (1970), *Continuous Univariate Distribution-2, Boston,* Houghton Mifflin.
5. Norman, L. Johnson, and Samuel Kotz (1970), *Continuous Multivariate Distributions, Boston,* Houghton Mifflin.