

t분포의 극단 꼬리부분으로부터의 효율적인 난수생성¹⁾

오 만 숙²⁾, 김 나 영³⁾

요 약

제한된 t분포로부터의 난수 생성은 특히 베이지안 분석에서 제한이 있는 모수의 사후밀도함수를 추정하기 위하여 몬테카를로 적분을 하는 경우 등에 필요하다. 그런데 제한영역이 t분포의 극단 꼬리부분으로 주어졌을 때 기존의 난수생성기법의 적용은 매우 비효율적이 될 수 있다. 본 논문에서는 극단꼬리부분으로 제한된 t분포로부터의 난수생성에 쉽게 사용할 수 있는 효율적인 난수생성 알고리즘을 제시하고 기존의 기법들과 시뮬레이션을 통하여 효율을 비교하였다.

1. 서론

제한된 정규분포나 t분포로부터의 효율적인 난수 생성은 많은 통계학자들의 관심을 끌고 있는 문제이다. 이는 특히 베이지안 분석에서 제한이 있는 모수의 사후밀도함수 추정을 위한 몬테카를로 기법에서 필수적으로 요구되는데 중요한 예로 정규선형모형에 대한 베이지안 분석에서 계수에 선형제약이 있는 경우를 들 수 있다. 또한 고전적인 통계에서도 제한이 있는 우도함수의 적분이 필요한 경우 등에서 요구되고 있다.

이와 관련된 연구로 Moran(1984)과 Evans와 Swartz(1986)는 정규분포에서 orthant probability의 계산에 대한 연구를 하였고 McFadden(1989), Hajivassiliou와 McFadden(1990)는 고전적인 통계에서의 적용예와 여러 난수생성기법을 제시하였다. 또한 고성능 컴퓨터의 고장률과 같은 매우 작은 확률계산에 몬테카를로 주표본기법(importance sampling)을 적용한 연구로 Chen, Lu, Sadowsky(1993), Mihram(1988), Sadowsky와 Bucklew(1990), Stadler와 Roy(1993)가 있다. 그리고 Geweke(1989, 1991, 1993)는 베이지안 추론에서의 적용예와 난수생성기법을 제안한 바 있다. 특히 Geweke(1991)는 일반적인 선형제한이 있는 경우의 난수생성을 위해 최근에 유용한 몬테카를로 기법으로 관심을 모으고 있는 김스표본기법 (Gelfand와 Smith(1990))을 적용하였다. 그런데 김스표본기법의 적용에서는 제한된 일차원 정규나 t분포로부터의 난수생성이 필수적으로 요구된다. 일반적으로, 제한된 일차원 정규나 t분포로부터의 난수생성은 역누적분포함수기법(inverse cumulative distribution function technique,

1 이 논문은 1994 이화여자대학교 교내연구비에 의하여 지원된 연구임.

2 서울시 서대문구 대현동 이화여대 통계학과 조교수.

3 서울시 성북구 안암동 고려대학교 통계학과 박사과정.

Devroye(1986))으로 쉽게 해결할 수 있지만 제한영역이 극단꼬리부분인 경우에는 역누적분포함수 기법이 불가능한 경우가 발생할 수 있다. 이의 해결을 위해 Geweke(1991)는 정규분포의 경우 지수밀도함수(Exponential density function)를 봉투함수(envelope function)로 사용하는 효율적인 기각기법(accept-reject method)을 제안하였다. 그리고 t 분포의 경우에는 t 밀도함수가 정규밀도함수의 척도모수에 대한 적분으로 표시될 수 있음을 이용하여 일차원 t 난수의 생성을 위하여 깃스표본기법을 적용할 것을 제안하였다. 그러나 이러한 깃스기법의 적용에는 깃스기법의 수렴시점 포착이 어렵고 또한 효율이 떨어지는 단점이 있다. 특히 다차원으로의 확장에서 깃스기법이 사용되기 때문에 이중적인(nested) 깃스기법의 사용은 효율면에서 바람직하지 못하다.

본 논문에서는 일차원 t 분포의 극단 꼬리 부분(extreme tail area)으로부터의 난수생성에 초점을 맞추어 기존의 기법에 비하여 효율적인 알고리즘을 제시하고자 한다. 즉, 변수 T 가

$$T \sim t_a(0, 1) I(T > a) \quad (1)$$

분포를 따를 때 T 의 난수를 생성하는 경우로, 여기에서 $t_a(0, 1)$ 은 자유도 α 를 갖는 표준 t 분포를 나타내며 I 는 지시함수(indicator function)이고 a 는 임의의 양수이다. 구체적으로, Geweke(1991)가 정규분포의 경우에 적용한 것 처럼 $T > a$ 구간에서 t 분포의 밀도함수와 유사한 모양을 가지면서 난수생성이 쉬운 봉투함수를 찾아 이를 이용하는 기각기법을 제시하고자 한다. 일차원에서 효율적인 알고리즘이 개발되면 제한된 다차원 t 분포로부터의 난수생성으로의 확장은 깃스표본기법을 이용하여 매우 손쉽게 이루어질 수 있을 것이다.

우선 2절에서는 보편적으로 사용되는 기존의 난수생성 기법에 관하여 간단히 기술하고 3절에서는 새로운 난수생성 알고리즘을 제시한 뒤 4절에서는 모의실험을 통하여 효율을 비교하였다. 마지막 5절에서는 제언을 간단히 기술하였다.

2. 기존의 난수생성기법

2.1 단순기각기법

제한된 t 분포로부터 난수를 생성하는 가장 단순하고 보편적인 방법은 일단 제한조건을 무시한 채 t 분포의 전체 영역으로부터 난수를 생성한 뒤 제한영역에 속하는 난수만을 취하고 나머지는 버리는 방법이다. 이는 매우 단순한 기각기법으로 사용이 쉬우나 제한영역이 전체영역에 비하여 작을수록 거의 모든 난수가 기각되어 버리기 때문에 낭비가 심한 매우 비효율적인 방법으로, 특히 a 가 큰 경우에는 심각한 낭비를 초래한다. 예를 들면, 표 4.1에서 보는 바와 같이

$\alpha=1, a=5$ 인 경우 90% 이상의 난수가 기각되어 버리고 10% 이하의 난수만 채택되는 것이다.

2.2 역누적함수 기법

제한된 영역에서의 난수 생성을 위하여 사용되는 또 하나의 보편적인 방법은 역누적분포함수 기법이다. (Devroye(1986)) 난수 T 가 $t_\alpha(0, 1) I(T > a)$ 분포를 따르면 아래 식을 만족하는 난수 U 는 균일분포(0,1)를 따른다.

$$U = (G_\alpha(T) - G_\alpha(a)) / (1 - G_\alpha(a))$$

위 식에서 G_α 는 자유도 α 를 갖는 표준 t 분포의 누적분포함수이다. 따라서 먼저 균일분포(0,1)로부터 난수 U 를 생성시킨 뒤

$$T = G_\alpha^{-1}(U(1 - G_\alpha(a)) + G_\alpha(a)) \quad (2)$$

의 관계식으로부터 목적하는 난수 T 를 얻는 방법이다. 이 방법은 낭비하는 난수가 없는 장점이 있으나 누적분포함수 G_α 와 이의 역함수를 계산하여야 하는 문제가 있다. 자유도가 1인 경우에는 $G_1^{-1}(t)$ 를 수리적으로 구할 수 있으나 그 외의 경우에는 누적분포함수 G_α 가 수리적으로 구해지지 않는다. 따라서 G_α 의 근사식을 사용해야 하는데 이 과정에서 비용이 많이 소모되는 단점이 있다. 이는 표 4.1에서 확인할 수 있는데 각각의 α 와 a 에 대해 10,000개의 난수생성에 2분 이상의 시간이 소모됨을 볼 수 있다.

2.3 깁스표본기법

깁스표본기법을 사용하는 방법은 다음과 같다. 변수 Z 가 표준정규분포 $N(0, 1)$ 를 따르고 변수 X 가 자유도 α 를 갖는 카이제곱분포, χ_α^2 , 를 따르면 $T = Z/\sqrt{X/\alpha}$ 는 $t_\alpha(0, 1)$ 을 따른다. 이를 $T > a$ 의 제한조건이 있는 경우에 적용하면 Z 와 X 의 조건부 확률분포가 각각

$$X | Z \sim \chi_\alpha^2 I(X < \alpha(Z/a)^2) \quad (3)$$

$$Z | X \sim N(0, 1) I(Z > a\sqrt{X/\alpha}) \quad (4)$$

로 주어진다. 따라서 (3)과 (4)를 사용하여 Z 와 X 의 난수를 교대로 생성시키는 과정을 반복하면 $T = Z/\sqrt{X/a}$ 는 $t_a(0, 1)I(T > a)$ 분포로 수렴한다. 따라서 수렴시점 이후에 $Z/\sqrt{X/a}$ 를 T 의 난수로 간주하면 된다. (Gelfand와 Smith(1990)) 이 방법은 사용이 간편하나 깃스기법이 일반적으로 가지는 단점인, 수렴시점까지의 난수를 버리기 때문에 효율이 떨어지고 수렴시점을 포착하기 어려운 문제점이 있다.

3. 효율적인 난수생성을 위한 알고리즘

3.1 기각기법

t 분포의 극단 꼬리부분으로부터의 효율적인 난수생성을 위해 본 논문에서는 Geweke(1991)가 정규분포의 경우에 사용했던 것과 같이 기각기법을 제안하고자 한다.

함수 $g(t)$ 에 비례하는 확률밀도함수를 갖는 분포로부터 난수를 생성하기 위한 기각기법 알고리즘은 다음과 같이 요약될 수 있다.

단계 1 : 확률밀도함수 $f(t)$ 를 선택하고 $g(t)/f(t) \leq K < \infty$ 인 상수 K 를 찾는다. (보통 $g(t)/f(t)$ 의 최대값을 K 로 사용한다.)

단계 2 : $f(t)$ 를 밀도함수로 갖는 분포로부터 난수 T 를 생성한다.

단계 3 : 균일분포(0,1)로부터 난수 U 를 생성한다.

단계 4 : 만약 $U \leq \frac{g(T)}{Kf(T)}$ 이면 T 를 채택하고 아니면 단계 2로 돌아간다.

단계 1에서 확률밀도함수 $f(t)$ 를 봉투함수라 부른다. 위의 알고리즘에서 난수의 채택률은

$$\begin{aligned} P_A &= P\left(U \leq \frac{g(T)}{Kf(T)}\right) = E\left(I\left(U \leq \frac{g(T)}{Kf(T)}\right)\right) = E^T E^{U|T}\left(I\left(U \leq \frac{g(T)}{Kf(T)}\right)\right) \\ &= E^T\left(\frac{g(T)}{Kf(T)}\right) = \int \frac{g(t)}{Kf(t)} f(t) dt = \frac{\int g(t) dt}{K} \end{aligned}$$

로 주어짐을 알 수 있다. 따라서 기각기법의 효율은 $f(t)$ 를 밀도함수로 갖는 분포로부터의 난수생성(간단히 $f(t)$ 로부터의 난수생성)이 용이하고 K 값이 작을수록 좋아진다. 그런데 K 는

$g(t)/f(t)$ 의 최대값이므로 $f(t)$ 에 의존하며 $f(t)$ 가 $g(t)$ 와 모양이 유사할수록 작아진다. 그러므로 난수생성이 용이하고 $g(t)$ 와 모양이 유사한 좋은 봉투함수 $f(t)$ 의 선택이 기각기법의 핵심이라 하겠다. (Rubinstein(1989))

3.2 봉투함수

t 분포의 극단꼬리 부분으로부터 난수를 생성하기 위한 기각기법에서 좋은 봉투함수의 선택을 위하여 다음과 같은 형태의 확률밀도함수를 생각하여 보자.

$$f_b^*(t) = \frac{\sqrt{a}}{b} \left(1 + \frac{t}{\sqrt{ab}}\right)^{-(1+a)} I(t \geq 0). \quad (5)$$

이 밀도함수에 대응하는 누적분포함수는

$$F_b^*(t) = 1 - \left(1 + \frac{t}{\sqrt{ab}}\right)^{-a} I(t \geq 0)$$

이다. 만약 위 밀도함수에 대응하는 분포를 따르는 변수 T 에 $T > a$ 의 제한조건이 첨가되면 T 의 밀도함수는

$$f_b(t) = \frac{f_b^*(t)}{1 - F_b^*(a)} I(t > a) = \frac{\sqrt{a}/b \left(1 + \frac{t}{\sqrt{ab}}\right)^{-(1+a)}}{\left(1 + \frac{a}{\sqrt{ab}}\right)^{-a}} I(t > a) \quad (6)$$

가 되며 누적분포함수는

$$F_b(t) = \frac{F_b^*(t) - F_b^*(a)}{1 - F_b^*(a)}$$

이다. 그런데 $F_b^*(t)$ 가 간단히 주어지므로 난수 T 의 생성은 역누적분포함수 기법을 사용하여 매우 손쉽게 이루어질 수 있다. 따라서 식 (6) 형태의 밀도함수는 좋은 봉투함수의 첫번째 조건을 만족시킨다.

다음, $t_a(0, 1) I(T > a)$ 분포의 밀도함수는

$$g(t) = \left(1 + \frac{t^2}{a}\right)^{-(\alpha+1)/2} I(t > a) \quad (7)$$

에 비례한다. 식 (6)의 $f_b(t)$ 와 식 (7)의 $g(t)$ 를 비교하면 $t \rightarrow \infty$ 일 때 0에 수렴하는 속도가 둘 다 $t^{-(\alpha+1)}$ 로 같다. 따라서 $f_b(t)$ 에서 상수 b 를 적절히 선택하면 $t > a$ 구간에서 $f_b(t)$ 를 $g(t)$ 와 유사하게 만들 수 있을 것이다. 달리 말하면, $g(t)/f_b(t)$ 의 최대값 K_b 를 최소로 만드는 b 값을 찾으면 그 때의 b 값을 갖는 $f_b(t)$ 가 식 (6)으로 주어진 형태의 봉투함수 중 가장 적절한 것이 될 것이다. 이를 위하여 $t > a$ 인 구간에서

$$w_b(t) \equiv g(t)/f_b(t) = \frac{b}{\sqrt{a}} \left(1 + \frac{a}{\sqrt{ab}}\right)^{-\alpha} \left(1 + \frac{t^2}{a}\right)^{-\frac{1}{2}(1+\alpha)} \left(1 + \frac{t}{\sqrt{ab}}\right)^{(1+\alpha)} I(t > a)$$

를 최소화하는 t^* 를 구하면

$$t^* = \begin{cases} \frac{\sqrt{a}}{b}, & \sqrt{a}/b \leq a \\ a, & \sqrt{a}/b > a \end{cases}$$

이다. 따라서 $w_b(t)$ 의 최대값 K_b 는

$$K_b = \begin{cases} \frac{b}{\sqrt{a}} \left(1 + \frac{a}{\sqrt{ab}}\right)^{-\alpha} \left(1 + \frac{1}{b^2}\right)^{\frac{1}{2}(1+\alpha)}, & \sqrt{a}/b \leq a \\ \frac{b}{\sqrt{a}} \left(1 + \frac{a}{\sqrt{ab}}\right) \left(1 + \frac{a^2}{a}\right)^{-\frac{1}{2}(1+\alpha)}, & \sqrt{a}/b > a \end{cases} \quad (8)$$

로 주어진다. 다음, 위의 K_b 를 b 에 대하여 최소화시키면 b 값이 아래의 b^* 와 같을 때 K_b 값이 최소화됨을 알 수 있다.

$$\begin{aligned} b^* &= \min \left\{ \frac{\sqrt{a}}{a}, \sqrt{\alpha + \frac{a^2(1+\alpha)^2}{4\alpha}} - \frac{a(1+\alpha)}{2\sqrt{a}} \right\} \\ &= \sqrt{\alpha + \frac{a^2(1+\alpha)^2}{4\alpha}} - \frac{a(1+\alpha)}{2\sqrt{a}}. \end{aligned}$$

참고로, 위 식을 보면 자유도 α 가 고정되어 있을 때는 하한경계치 a 가 증가할수록 b^* 가 작아지고 또한 하한경계치 a 가 고정되어 있을 때는 자유도 α 가 커지면 b^* 는 커짐을 알 수 있다.

위의 결과로부터 $f_{b^*}(t)$ 가 식 (6)의 형태를 갖는 봉투함수 중 가장 채택률을 높이는, 즉, 가

장 효율을 좋게하는 봉투함수이다. 또한 $\sqrt{a/b^*} > a$ 이므로 식 (8)로부터 $g(t)/f_{b^*}(t)$ 의 최
대값은

$$K_{b^*} = \frac{b^*}{\sqrt{a}} \left(1 + \frac{a}{\sqrt{ab^*}}\right) \left(1 + \frac{a^2}{a}\right)^{-\frac{1}{2}(1+a)}$$

이다.

몇 가지 경우에 $g(t)$ 와 $f_{b^*}(t)$ 를 그림으로 그려 본 결과가 그림 3.1에 나타나 있다. 단,
적절한 시각적 비교를 위하여 두 함수 모두 $t \in a$ 구간에서 적분값이 1 이 되도록 만드는 비례
상수를 곱하여 그림에 나타내었다. 그림에서 보면 $g(t)$ 와 $f_{b^*}(t)$ 는 거의 일치함을 알 수 있
다.

3.3 난수생성 알고리즘

이상의 결과를 종합하여 $t(0,1)I(T > a)$ 분포를 따르는 난수 T 를 생성하는 알고리즘을
정리하면 다음과 같다.

단계 1 : 주어진 α, a 에 대하여 b^* 와 K_{b^*} 를 계산한다.

단계 2 : 균일분포(0,1)로부터 난수 U_1 를 생성시킨 다음

$$T = F_{b^*}^{-1}(U_1(1 - F_{b^*}(a)) + F_{b^*}(a))$$

로 놓는다.

단계 3 : 균일분포(0,1)로부터 난수 U_2 를 생성시킨다.

단계 4 : 만약

$$U_2 \leq \frac{g(T)}{K(b^*)f_{b^*}(T)}$$

이면 T 를 채택하고 그렇지 않으면 단계 2로 돌아간다.

위 알고리즘에서 단계 2는 $f_{b^*}(t)$ 로부터 역누적분포함수 기법을 이용하여 난수를 생성하는 과
정이다.

4. 모의실험을 통한 효율 비교

2장에서 소개된 기존의 난수생성 기법과 3장에서 제안한 새로운 알고리즘의 효율을 모의실험을 통하여 비교하여 보기로 한다. 이를 위하여 각 알고리즘을 적용하여 t 분포의 자유도인 α 가 각각 1, 5, 10, 20인 경우와 제한영역의 한계인 a 가 3, 5, 7, 11인 경우에 대하여 10,000개의 난수를 생성하는데 드는 계산시간을 비교하여 결과를 표4.1에 정리하였다. 표에서 N 은 10,000개의 난수 생성을 위하여 기각되는 난수까지 포함하여 실제로 생성되는 총 난수의 수를 나타낸다. 모의실험은 퍼스컴 486-DX2에서 SAS(DOS용) 패키지를 사용하여 수행하였다. 그리고 계산시간은 SAS 명령어 TIME을 사용하여 초기화단계를 제외한 순수한 난수생성시간을 측정하였다.

단순기각기법에서는 제한이 없는 t 분포를 따르는 변수 T 의 난수생성을 위하여 RANNOR로부터 정규난수 Z 를 생성하고 RANGAM을 이용하여 자유도 α 인 카이제곱난수 X 를 생성하여 $T = Z/\sqrt{X/\alpha}$ 를 취하는 방식을 택하였다. 이 경우 예를 들면 $\alpha=1$, $a=3$ 일 때 10,000개의 난수생성을 위하여 총 97,719개의 난수가 생성되어 그 중 87,719개는 기각되고 10,000개만 a 보다 큰 난수로 채택되는데 이에 소모되는 총시간은 1분 34초이다. 시행마다 생성되는 난수가 달라지기 때문에 총 난수의 수나 계산시간은 다소 변동이 있을 수 있다. 참고로, 제한이 없는 t 분포로부터의 난수생성에 Kinderman, Monahan, Ramage(1977)의 알고리즘을 사용하는 경우에는 제한된 구간에 속하는 10,000개의 난수생성에 드는 시간이 $\alpha=1$ 인 경우 다소 덜 소요되었고 그 외의 경우에는 다소 더 소요되었으나 그리 큰 차이는 나타나지 않았다. 제한이 없는 t 분포로부터의 난수생성에 어떤 효율적인 알고리즘을 적용하더라도 결국은 거의 모든 난수가 기각되어 버리기 때문에 전체적으로 단순기각기법은 매우 비효율적이며 특히 α 와 a 가 커짐에 따라 거의 사용이 어렵게 됨을 알 수 있다.

역누적분포함수 기법을 사용하는 경우에 t 분포의 누적분포함수와 역누적분포함수로는 SAS 함수 PROB T 와 TINV를 각각 사용하였다. 이 경우에는 기각되는 난수가 없으므로 $N=10,000$ 이다. 표를 살펴보면 α 와 a 값이 변하여도 시간은 거의 일정함을 알 수 있다. 그러나 누적분포함수와 역누적분포함수를 계산하기 때문에 대략 2분-4분 정도로 제안된 기법에 비교하면 효율이 떨어짐을 알 수 있다.

깁스기법의 경우에는 식 (3)과 (4)에 주어진 제한이 있는 정규분포와 카이제곱분포로부터의 난수생성에 역누적분포함수 기법을 사용하였다. 이에 사용된 SAS함수로는 PROBNOR, PROBIT, PROBCHI, CINV이다. 그런데 $a=11$ 에서는 표준정규분포의 분포함수 $\Phi(11)$ 이 거의 1과 같아 역누적분포함수 기법의 사용이 불가능하였다. 그리고 (3)과 (4)의 분포로부터 난수를 반복적으로 생성해 가면서 깁스 알고리즘의 수렴성을 조사한 결과 각각의 α 와 a 에 대하여 약 100번 후에 수렴이 일어나는 것으로 보인다. 깁스기법에서 나타나는 난수들간의 심각한 상호관련성(serial correlation)을 피하기 위하여 수렴시점 100번 후에 매 5번째의 난수를 취하는 경우

를 표에 나타내었다. 만약 상호관련성이 상관 없이 없다면 수렴 후 연속적으로 10,000개의 난수를 취하면 되므로 표에 나타난 계산시간의 약 1/5이 소요될 것이다. 그러나 독립적인 난수들을 원한다면 보통 난수를 취하는 구간을 더 늘려야 하므로 계산시간이 위 표에 나타난 것 보다 훨씬 더 길어질 것이다. 일반적으로 알려진 바와 같이 깃스기법은 효율면에서 우수하지 못하여 역누적분포함수기법이나 제안된 기법에 비하여 시간이 많이 소모됨을 볼 수 있다.

끝으로, 본 논문에서 제안된 기각기법에서는 균일난수 생성을 위하여 SAS 난수생성함수 RANUNI를 사용하였다. 그림 3.1에서 보는 바와 같이 $f(t)$ 와 $g(t)$ 가 거의 일치하기 때문에 제안된 알고리즘의 단계 4에서 기각되는 난수는 매우 적고 계산시간이 α 나 a 값에 상관없이 모두 대략 10초 정도로 다른 기법들보다 월등히 뛰어난 것을 알 수 있다.

결론적으로, 제안된 기법은 봉투함수로부터의 난수생성이 매우 간편하고 빠르며 기각되는 난수가 거의 없기 때문에, 컴퓨터에 익숙하지 않은 비전문인들도 쉽게 사용할 수 있으며 또한 효율이 매우 뛰어난 알고리즘이라 할 수 있다.

5. 제언

임의의 구간으로 제한된 일차원 t 분포의 경우, 즉, $-\infty < a_1 < a_2 < \infty$ 에 대해

$$T \sim t_{\alpha}(0, 1) I(a_1 < T < a_2) \quad (9)$$

인 난수 T 의 생성을 요하는 경우에는 구간 (a_1, a_2) 을 꼬리부분과 0을 중심으로 한 구간으로 나누어 0을 중심으로 한 구간에서는 정규분포를 봉투함수로 사용하는 기각기법을 사용하고 꼬리부분에 속하는 구간에서는 본 논문에서 제안한 봉투함수를 사용하는 기각기법을 사용하여 난수를 생성시킬 수 있을 것이다.

다차원 공간에서 선형제약이 있는 경우는 일반적으로

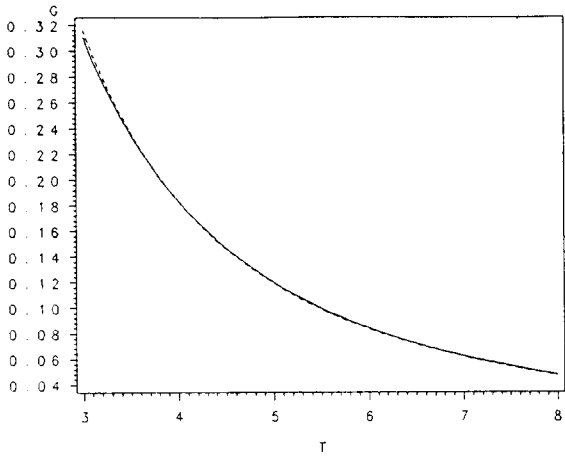
$$T = (t_1, t_2, \dots, t_p) \sim T_{\alpha}(0, \Sigma) I(a_1 < T < a_2)$$

로 나타낼 수 있다. 위 식에서 $T_{\alpha}(0, \Sigma)$ 는 자유도 α , 위치모수 0, 척도모수 Σ 인 다변량 t 분포를 나타내며 a_1 과 a_2 는 p 차원 벡터이다. 위의 분포를 따르는 벡터변수 T 의 난수를 생성시키고자 할 때는 깃스표본기법을 이용하여 각 원소변수 t_i 의 난수를 그 조건부 확률분포로부터 차례로 생성하면 되므로 다차원 난수의 생성이 결국은 (9)와 같은 일차원 난수의 생성으로 귀결된다. 따라서 3장의 알고리즘은 일반적인 선형제약이 있는 다차원 t 분포로부터의

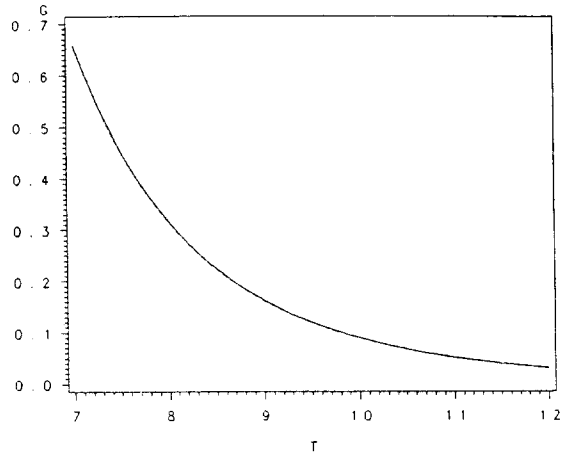
난수생성에 효율적으로 사용될 수 있을 것이다.

참 고 문 헌

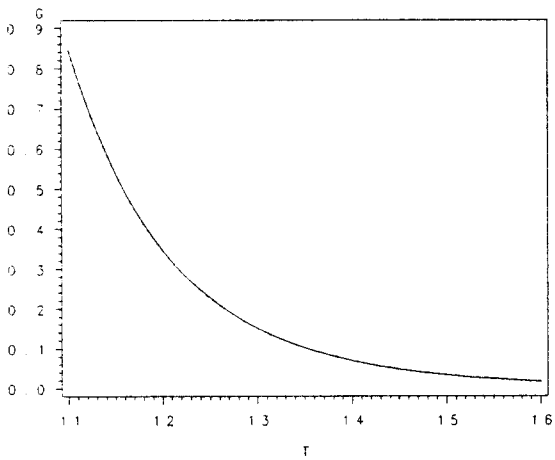
- [1] Chen, J. C., Lu, D. Q., Sadowsky, J. S. and Yao, K. (1993). On Importance Sampling in Digital Communications-Part 1: Fundamentals, *IEEE Journal on Selected Areas in Communications*, Vol. 11, 289-298.
- [2] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York.
- [3] Evans, M. and Swartz, T. (1986). Monte Carlo Computation of Some Multivariate Normal Probabilities, *Technical Report No. 7*, Dept. of Stat., Univ. of Toronto.
- [4] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities, *The Journal of the American Statistical Association*, Vol. 85, 398-409.
- [5] Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica*, Vol. 57, 1317-1339.
- [6] Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints, *Computing Science and Statistics (Proceedings of the 23rd symposium on the Interface)*
- [7] Geweke, J. (1993). Bayesian Inference for Linear Models subject to Linear Inequality Constraints, *Technical Report*, Dept. of Economics, Univ. of Minesota.
- [8] Hajivassiliou, V. A. and McFadden, D. L. (1990). The Method of Simulated Scores for the Estimation of LDV Models with an Application to External Debt Crises, *Colwes Foundation working paper*, Yale Univeristy.
- [9] McFadden, D. L. (1989). A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration, *Econometrica*, Vol. 57, 995-1026.
- [10] Mihram, G. A. (1988). Simulation-Statistical Foundations and Methodology, *Academic Press*, New York and London.
- [11] Moran, P. A. P. (1984). The Monte Carlo Evaluation of Orthant Probabilities for Multivariate Normal Distributions, *Australian Journal of Statistics*, Vol. 26, 39-44.
- [12] Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*, Wiley, NewYork.
- [13] Sadowsky, J. W. and Bucklew, J. A. (1990). On Large Deviations Theory and Asymptotically Efficient Monte Carlo Estimation, *IEEE Transactions on Information Theory*, Vol. 36, 579-588.
- [14] Stadler, J. S. and Roy, S. (1993). Adaptive Importance Sampling, *IEEE Journal on Selected Areas in Communications*, Vol. 11, 309-316.



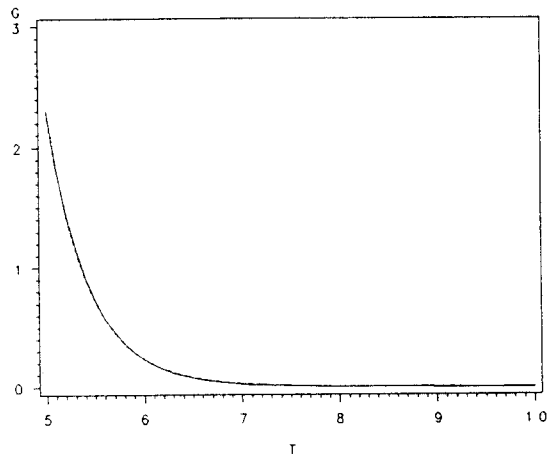
a) $\alpha=1, a=3$



b) $\alpha=5, a=7$



c) $\alpha=10, a=11$



d) $\alpha=20, a=5$

그림 3.1 $g(t)$ (—)와 $f(t)$ (- - -)

표 4.1 모의실험 결과

자유도	a	단순기각기법		역누적합수법		깁스표본기법		기각기법	
		N	시간	N	시간	N	시간	N	시간
1	3	97,719	1분34초	10,000	1분57초	50,100	6분31초	10,084	10초
	5	159,372	2분34초	10,000	1분54초	50,100	6분37초	10,026	10초
	7	221,162	3분33초	10,000	1분52초	50,100	6분34초	10,019	10초
	11	346,006	5분34초	10,000	1분51초	50,100	6분00초	10,004	10초
5	3	650,184	9분30초	10,000	3분06초	50,100	6분56초	10,261	10초
	5	.	>1시간	10,000	2분25초	50,100	6분59초	10,111	10초
	7	.	>1시간	10,000	2분21초	50,100	7분19초	10,072	9초
	11	.	>1시간	10,000	2분21초	*	불가능	10,031	9초
10	3	1,488,221	21분28초	10,000	3분46초	50,100	7분22초	10,331	9초
	5	.	>1시간	10,000	3분32초	50,100	7분38초	10,124	10초
	7	.	>1시간	10,000	2분54초	50,100	8분11초	10,069	9초
	11	.	>1시간	10,000	2분51초	*	불가능	10,041	9초
20	3	2,814,172	40분19초	10,000	3분47초	50,100	8분11초	10,386	10초
	5	.	>1시간	10,000	4분39초	50,100	8분09초	10,160	9초
	7	.	>1시간	10,000	4분02초	50,100	9분15초	10,091	9초
	11	.	>1시간	10,000	3분51초	*	불가능	10,033	9초

Efficient Random Number Generation from Extreme Tail Areas of a t-distribution⁴⁾

Man-Suk Oh⁵⁾, Kim Na Young⁶⁾

Abstract

It is often needed to generate random numbers from truncated t-distributions to carry out Bayesian inferences, especially in Monte Carlo integration for estimation of posterior densities of constrained parameters. However, when the restricted area is an extreme tail area with a small probability most existing random generation methods are not efficient. In this paper, we propose an efficient acceptance-rejection method to generate random numbers from extreme tail areas of a t-distribution. Using some simulation results, we compare the proposed algorithm with other popular methods.

4 This research was supported by the 1994 Ewha Womans University Research Fund.

5 Assistant Professor, Dept. of Statistics, Ewha Womans Univ., Sodaemoon Gu, Seoul, Korea.

6 Ph.D. student, Dept. of Statistics, Korea Univ., Sungbook Gu, Seoul, Korea.