전산통계

Chapter 2.6 거절법

정규분포군의 생성

- 단계 1) 난수 U_1과 U_2를 생성
- 단계 2) $R^2 = -2lnU_1, \vartheta = 2\pi U_2$
- 단계 3) $\mathrm{X} = \mathrm{R}\cos\vartheta = \sqrt{-2lnU_1}\cos(2\pi U_2)$, $\mathrm{Y} = \mathrm{R}\sin\vartheta = \sqrt{-2lnU_1}\sin(2\pi U_2)$,
- 로그-정규분포의 생성 : 로그-정규분포는 로그변환을 하면 정규분포에 따르는 확률변수의 분포. 즉, 확률변수 X의 분포가 정규분포이면 $Y = e^X$의 분포가 로그-정규분포.

지수분포군의 생성

- 와이블 분포 (Weibull distribution)
  - 제품의 신뢰수명분포
  - $X \sim \exp(\lambda)$일 때, $Y = X^{1/\beta}$로 놓으면, Y는 와이블 분포
  - β는 형태모수 (shape parameter), λ는 척도모수 (scale parameter)
  - $\beta = 1$인 경우, 지수분포
  - $\beta = 2$인 경우, 라이레히 분포

- 얼랑 분포 (Erlang distribution)
  - 포아송 과정에서 n번째 사건이 발생할 때까지 걸리는 시간의 분포
  - 사건발생 간격시간은 지수분포이므로 얼랑분포는 지수분포 확률변수들의 합의 분포. 즉, $\{N(t), t > 0\}$를 모수가 λ인 포아송과정이라 하고, 서로 독립인 발생간격시간들은 $X_1, X_2, \cdots,\ X_n \sim \exp(\lambda)$이며, $S_n = X_1 + X_2 + \cdots + X_n$로 놓으면 얼랑분포를 따른다.

거절법

- 적당한 분포로부터 확률변수 값을 생성하고 그 중에서 특정 조건에 맞는 경우만 채택하고 맞지 않는 경우는 거절하는 방법

- 확률변수 X의 확률밀도함수가 $f_X(x) = Cg(x)h(x)$, 여기서 $C \geq 1,\ 0 < g(x) \leq 1$, $h(x)$는 어느 확률밀도함수이다. $U \sim U(0,1)$이고, 확률변수 Y는 $h(y)$를 확률밀도함수로 가지며 서로 독립이면, $f_Y(x|U \leq g(Y)) = f_X(x)$를 만족한다.
- $f_X(x) = Cg(x)h(x) = \phi(x)h(x)$

- 기본 알고리즘
  - 1단계 : 난수생성자를 사용하여 난수 U를 생성한다.
  - 2단계 : Y값의 생성단계 : $h(y)$값에서 Y값을 생성한다.
  - 3단계 : $U \leq g(Y)$이면, Y값을 X값으로 사용한다. 아니면, 다시 1단계로 돌아간다.

- 효율성
  - 채택이 될 확률이 높을수록 생성의 속도가 빨라지므로 좋은 알고리즘
  - $C \geq 1$이고 채택이 될 확률은 $P\big(U \leq g(Y)\big) = 1/C$이므로 효율적인 거절법 알고리즘의 조건은 i) C가 작을수록 알고리즘의 효율성이 좋아진다. ii) $h(y)$으로부터 Y값의 생성이 쉬워야 한다.

- 일반화
  - 확률변수 X의 값을 생성
  - 확률변수 X가 구간 $[a, b]$에서 정의되어 있고, 확률밀도함수에 대해서 $f_X(x) \leq M$이 만족되는 값 M이 존재한다.
  - $\phi(x) = M$이라 놓으면 $f_X(x) \leq M$이므로 Y가 $U(a, b)$에 따르면 확률밀도함수는 $h(x) = \frac{1}{b-a}$이 되며, $C = M(b-a) \geq 1, g(x) = \frac{f_X(x)}{M},\ a \leq x \leq b$

- 균일분포
  - 1단계 : 난수생성자를 사용하여 난수 u_1와 u_2를 생성
  - 2단계 : Y값의 생성단계 : $\mathrm{y} = a + (b - a) * u_2$
  - 3단계 : $u_2 \leq g(y)$이면, y값을 x값으로 사용

합성법

- 합성법 (composition method)은 생성하고자 원하는 분포를 여러 분포의 혼합분포 (mixture distribution)로 생각
- $f(x|\theta)$는 모수 θ에 의하여 분포가 결정되는 확률밀도함수
- 모수 θ의 확률밀도함수가 $h(\theta)$라면, 확률변수 X의 확률밀도함수는

$$f_X(x) = \int f(x|\theta)h(\theta)d\theta \,, f_X(x) = \sum_i f(x|\theta = \theta_i)h(\theta_i)$$

- 알고리즘

  - 1단계 : $h(\theta)$로부터 θ값을 생성한다.
  - 2단계 : 생성된 θ값에 의한 $f(x|\theta)$로부터 x값을 생성한다.