

# 전산통계

Chapter 2.3 확률분포의 생성

## 확률분포의 생성

---

- 조건 1) 생성된 값들의 분포의 형태가 원하는 확률분포를 따라야 한다.
- 조건 2) 분포의 형태가 생성개수 또는 생성의 시작위치에 불변해야 한다.
- 조건 3) 생성에 소요되는 시간이 빨라야 한다.
- 조건 4) 값들의 생성에서 이웃하는 값들 사이에 독립성을 만족해야 한다.
- 조건 5) 생성된 값들의 동일한 수열이 재구성될 수 있어야 한다.

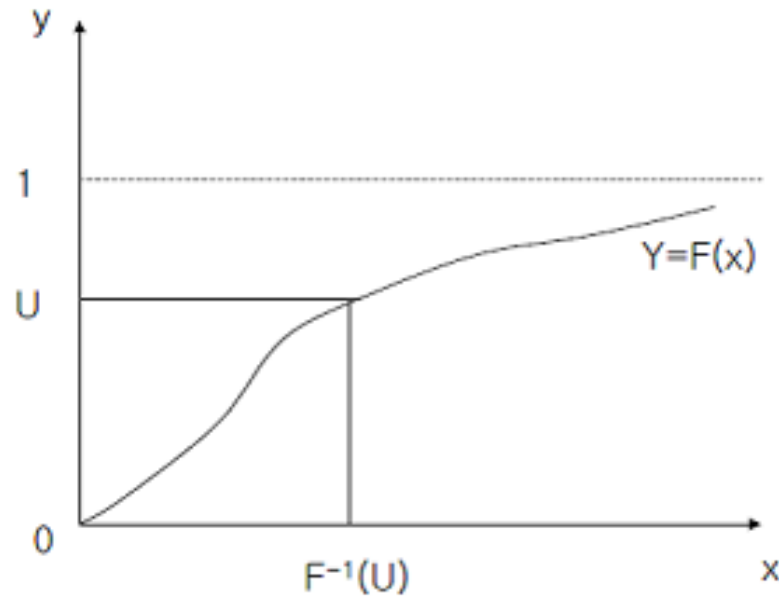
- 
- d--- : 각 분표별 확률값 계산
  - p--- : 누적확률값 계산
  - q--- : 분위수 계산
  - r--- : 분포를 따르는 값들을 생성

분포	함수	생성함수와 모수들
Beta	rbeta	rbeta(n, shape1, shape2)
Binomial	rbinom	rbinom(n, size, prob)
Cauchy	rcauchy	rcauchy(n, location, scale)
Chi-Square	rchisq	rchisq(n, df)
Exponential	rexp	rexp(n, rate)
F	rf	rf(n, df1, df2)
Gamma	rgamma	rgamma(n, shape, scale)
Geometric	rgeom	rgeom(n, prob)
Hypergeometric	rhyper	rhyper(n, m1, m2, k)
Logistic	rlogis	rlogis(n, location, scale)
Log Normal	rlnorm	rlnorm(n, meanlog, sdlog)
Negative Binomial	rnbinom	rnbinom(n, size, prob, mu)
Normal	rnorm	rnorm(n, mean, sd)
Poisson	rpois	rpois(n, lambda)
Student t	rt	rt(n, df)
Uniform	runif	runif(n, min, max)
Weibull	rweibull	rweibull(n, shape, scale)

## 역변환법

- 분포함수  $F(x)$ 의 역함수  $F^{-1}(u)$ 가 알려져 있는 경우에  $U$ 가 표준균일분포를 따르는 확률변수이면  $X = F^{-1}(U)$ 는  $F(x)$ 를 분포함수로 가진다.

$$- P(X \leq x) = P(F^{-1}(U) \leq x) = P(F(F^{-1}(U)) \leq F(x)) = P(U \leq F(x)) = F(x)$$



- 
- 단계 1 : 난수생성자로 부터 난수열을 생성한다.  $\{u_1, u_2, \dots, u_N\}$
  - 단계 2 : 생성한 난수들을 원하는 분포함수의 역함수  $F^{-1}(u)$ 에 대입하여 확률변수값을 생성한다.  $x_n = F^{-1}(u_n), n = 1, 2, \dots, N$
  - 단계 3 : 생성된 확률변수값들을 사용하여 통계적 실험을 한다.

---

- 지수분포

- $f(x) = \begin{cases} \lambda e^{-\lambda x}, 0 < x \\ 0, otherwise \end{cases}$

- $F(x) = 1 - e^{-\lambda x}$

- $F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u)$

---

- 와이블분포

- $$f(x) = \begin{cases} \frac{\beta}{\alpha^\beta} x^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta}, & 0 < x \\ 0, & otherwise \end{cases}$$

- $$F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta} = u$$

- $$x = F^{-1}(u) = \alpha^\beta \sqrt{-\ln(1-u)}$$



## 변수변환법

---

- $Y = g(X_1, X_2, \dots, X_n)$  : 생성하려고 하는 확률변수
  - 단계 1)  $X_1, X_2, \dots, X_n$ 의 값을 생성한다.
  - 단계 2)  $Y = g(X_1, X_2, \dots, X_n)$  를 계산한다.
  - 단계 3) 생성된  $Y$ 값을 통계적 실험에 사용한다.

## 변수변환법

---

- $Y = g(X_1, X_2, \dots, X_n)$  : 생성하려고 하는 확률변수
  - 단계 1)  $X_1, X_2, \dots, X_n$ 의 값을 생성한다.
  - 단계 2)  $Y = g(X_1, X_2, \dots, X_n)$  를 계산한다.
  - 단계 3) 생성된  $Y$ 값을 통계적 실험에 사용한다.

## 정규분포의 생성 (박스-뮐러 방법)

---

- $X, Y \sim N(0,1): i.i.d.$
- $R, \vartheta : (X, Y)$ 의 극좌표. 즉,  $R^2 = X^2 + Y^2, \tan \vartheta = \frac{Y}{X}$
- $f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$
- $d^2 = x^2 + y^2, \vartheta = \tan^{-1} \frac{Y}{X}$
- $f_{R^2, \vartheta}(d, \vartheta) = \frac{1}{2\pi} \frac{1}{2} e^{-d/2}, 0 < d < \infty, 0 < \theta < 2\pi$

- 
- 단계 1) 난수  $U_1$ 과  $U_2$ 를 생성
  - 단계 2)  $R^2 = -2\ln U_1, \vartheta = 2\pi U_2$
  - 단계 3)  $X = R\cos\vartheta = \sqrt{-2\ln U_1}\cos(2\pi U_2), Y = R\sin\vartheta = \sqrt{-2\ln U_1}\sin(2\pi U_2),$