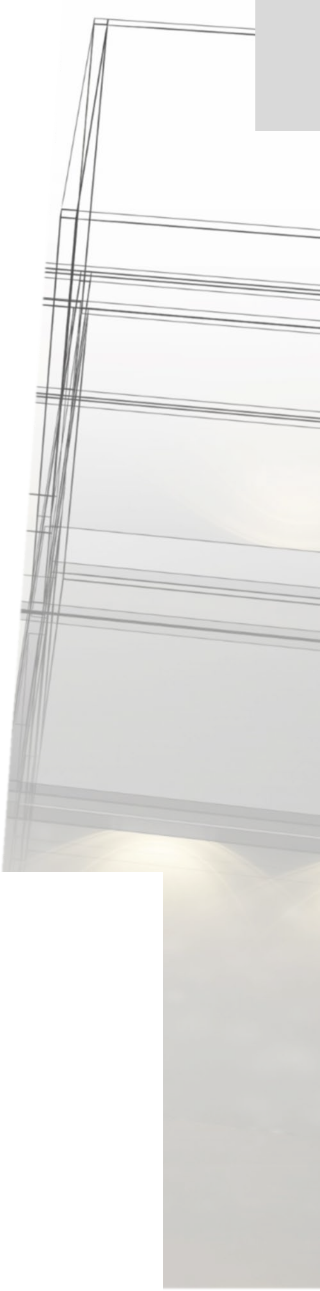
15강

난수생성과 몬테카를로 시뮬레이션

통계·데이터과학과
이긍희 교수

학습목표

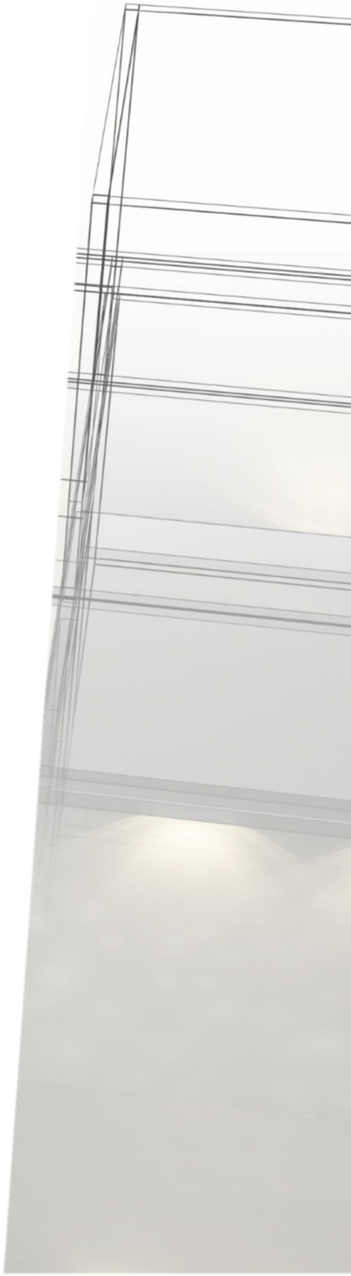
1. 연속형 확률변수의 난수를 생성할 수 있다.
2. 이산형 확률변수의 난수를 생성할 수 있다.
3. 몬테카를로 적분을 이해할 수 있다.



01

15강 난수생성과
몬테카를로 시뮬레이션

몬테카를로 시뮬레이션 의 개요



시뮬레이션

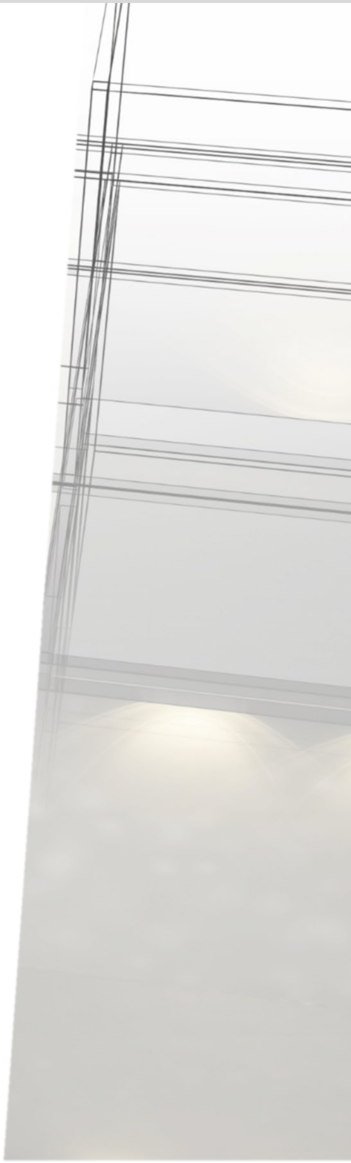
- 컴퓨터 환경에서 현실과 유사한 가상 환경을 수학적 모형으로 만들고, 조건 변화하면서 가상 환경의 변화 살펴보는 것

몬테카를로 시뮬레이션

- 확률분포로부터 난수(random number) 생성
  → 복잡한 문제의 해 근사적으로 구하는 방법
- 기본 원리 : 대수의 법칙

몬테카를로 시뮬레이션

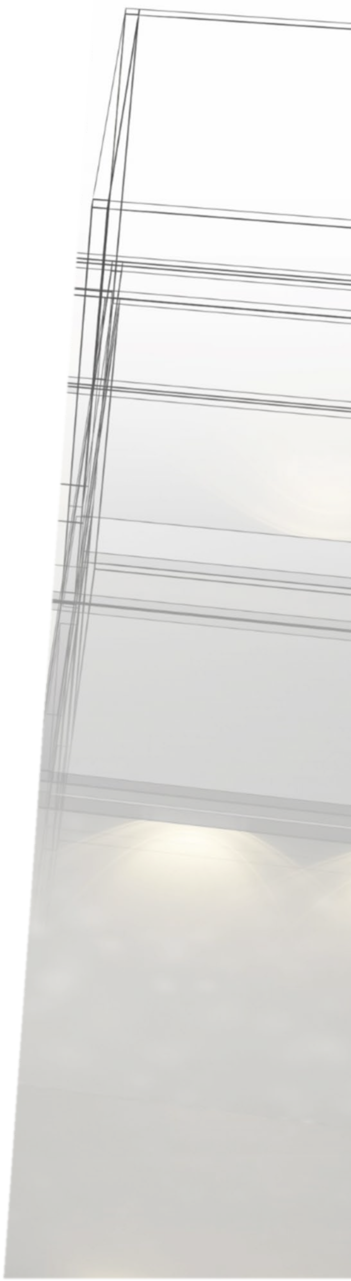
- 1940대 후반 E. Fermi와 S. Ulam의 연구로 시작
- 활용 분야 : 확률분포 생성, 최적화 문제, 수학적 적분 등



02

15강 난수생성과 몬테카를로 시뮬레이션

연속형 확률분포의 난수 생성



연속형 균등분포

◆ 연속형 균등분포의 난수 생성

- R의 runif 함수 : 메르센 트위스터 (Mersenne Twister) 알고리즘 이용
- 난수에는 씨앗(seed)값이 존재

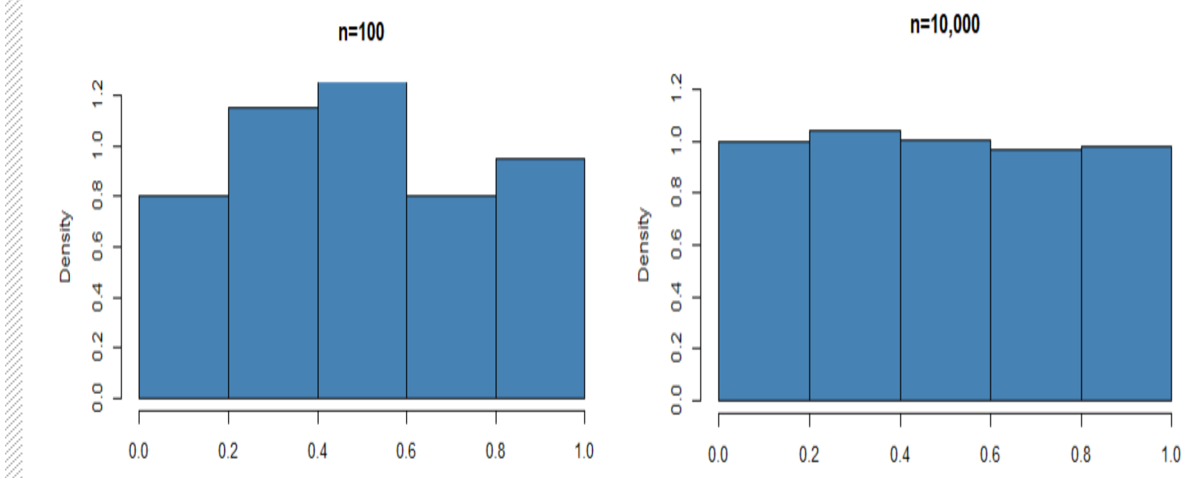
연속형 균등분포

◆ 연속형 균등분포에서 100개와 10,000개 난수 생성과 히스토그램 작성

```
> set.seed(1234567)
> hist(runif(100), xlab=" ", main="n=100", freq =FALSE, ylim=c(0, 1.2),
      breaks = c(0,0.2,0.4,0.6,0.8,1.0), col="steelblue")
> hist(runif(10000), xlab=" ", main="n=10,000", freq =FALSE, ylim=c(0,
      1.2), breaks = c(0,0.2,0.4,0.6,0.8,1.0), col="steelblue")
```

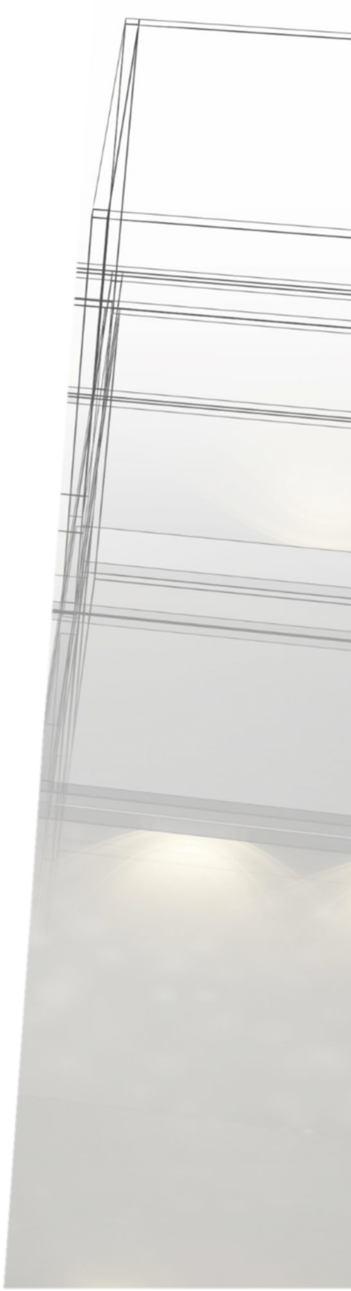
연속형 균등분포

- 연속형 균등분포에서 100개와 10,000개 난수 생성과 히스토그램 작성



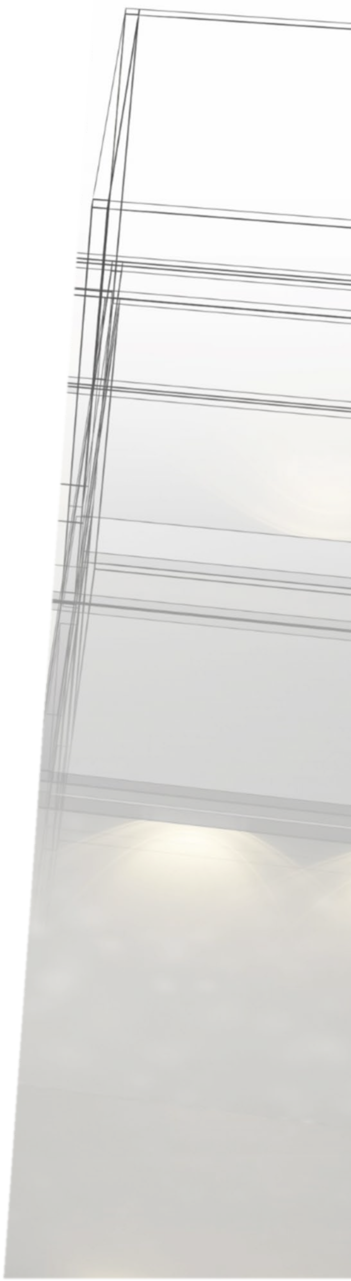
누적분포함수

- $F_X^{-1}(y) = inf\{x|F_X(x) > y\}$: $F_X(x)$ 의 역함수
- $F_X(x) \sim U(0,1)$



역변환 방법

- $U \sim U(0,1) \Rightarrow X = F_X^{-1}(U) \sim F_X(x)$



역변환 방법

◆ 연속형 확률변수 $X \sim F_X(x)$의 난수 생성

(1) $U(0,1)$에서 난수 u_i

(2) X 의 누적분포함수 역함수 F_X^{-1} 구함

(3) X의 난수 생성 : $F_X^{-1}(u_i)$

역변환 방법의 예

예

연속형 균등분포 이용하여 지수분포의 난수 생성

역변환 방법의 예

예

연속형 균등분포 이용하여 지수분포의 난수 생성

역변환 방법의 예

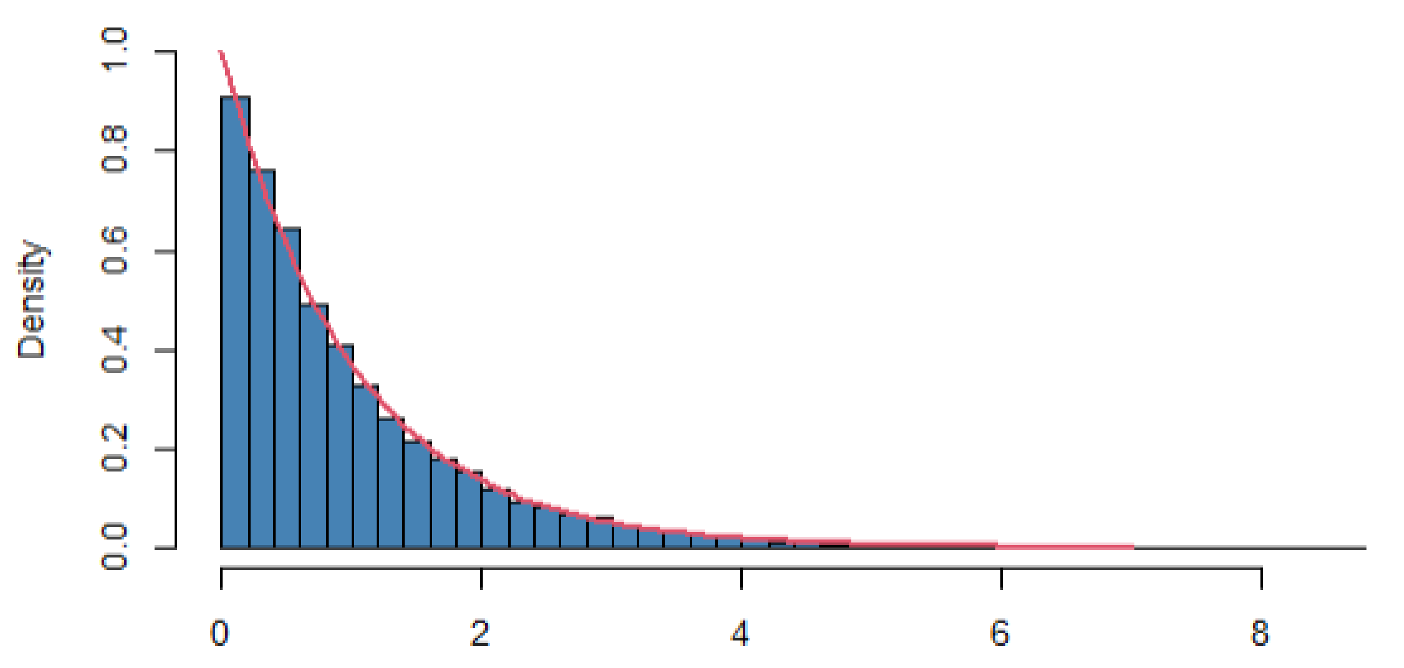
◆ 연속형 균등분포 이용하여 지수분포의 난수 생성

```
> set.seed(1234567)
> n = 10000
> u1 = runif(n)
> e1 = -log(1-u1)
>
> hist(e1, freq=FALSE, breaks=40, ylim=c(0, 0.8),
       xlab="", main="", col="steelblue")
> curve(dexp(x,1), xlim = c(0, 7), add=TRUE, col=2, lwd=2)
```

역변환 방법의 예

예

연속형 균등분포 이용한 지수분포의 난수 생성



기각법

◆ 확률밀도함수 $f_X(x)$ 의 난수 X 생성

-모든 x, $\frac{f_X(x)}{g(x)} \le c$ 만족하는 확률밀도함수 : $g(x)$

① 확률밀도함수 $g(x)$ → 난수 Y 생성

② 연속형 균등분포 $U(0,1)$ 난수 U 생성

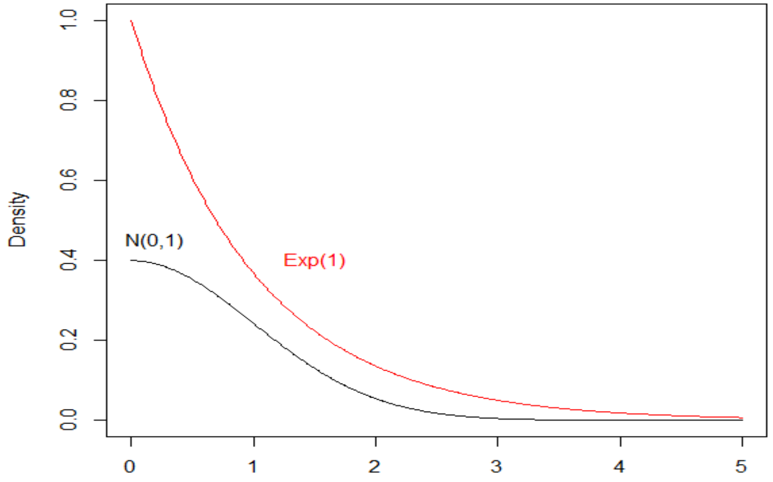
③ $U \le \frac{f_X(x)}{cg(x)}$ 이면 $X = Y$, 그 외 경우 다시 ①

기각법의 예

예

기각법으로 표준정규분포의 난수 생성.

비교대상 확률분포 : 지수분포



기각법의 예

예

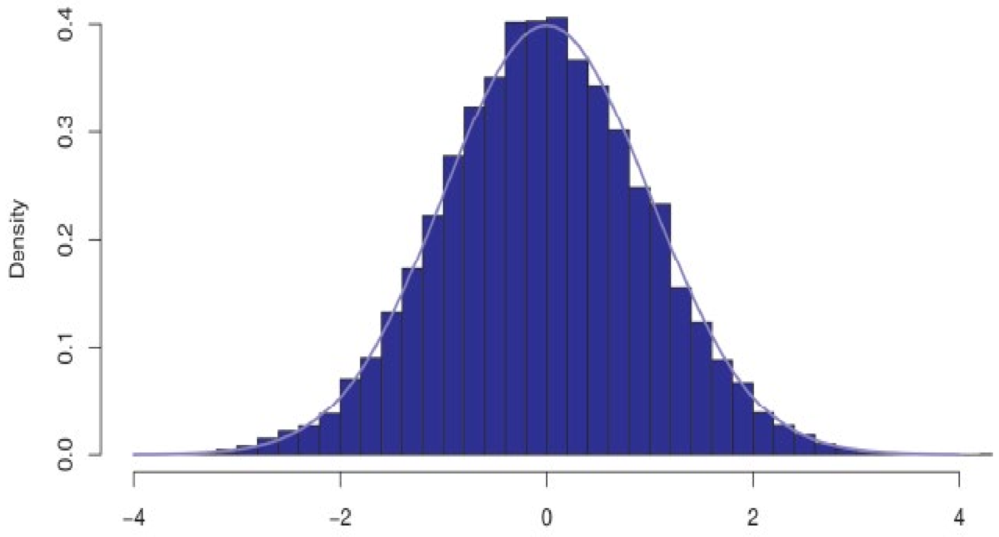
기각법으로 표준정규분포의 난수 생성.

비교대상 확률분포 : 지수분포

기각법의 예

- 기각법으로 표준정규분포의 난수 생성.

```
> set.seed(1234567)
> n = 10000
> x1 = rep(NA, n)
> ii = 1
> while (ii <= n) {
    y1 = rexp(1)
    ratio = exp(-(y1-1)^2/2)
    u1 = runif(1)
    u2 = runif(1)
    if(u1 < ratio){
       x1[ii] = ifelse(u2 > 0.5, y1, -y1)
       ii = ii + 1
    }
 }
```



정규분포 난수 생성

- 박스와 뮬러(1958) : 두 개의 균등분포 확률변수로부터 정규분포 확률변수 도출방법 제안

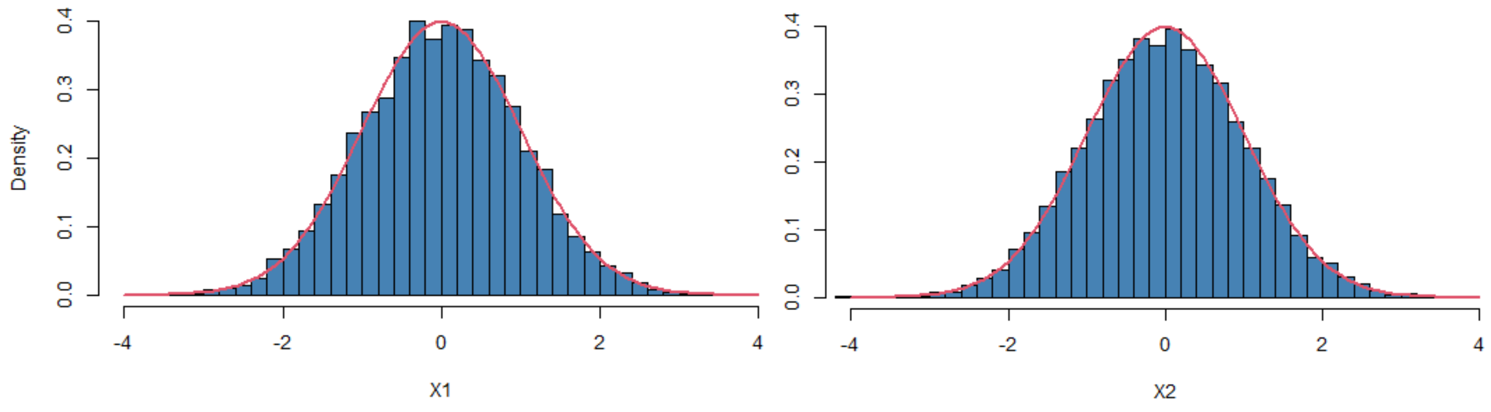
정규분포 난수 생성

◆ 박스와 뮬러의 방법

```
> set.seed(1234567)
> u1 = runif(10000)
> u2 = runif(10000)
> x1 = sqrt(-2*log(u1))*cos(2*pi*u2)
> x2 = sqrt(-2*log(u1))*sin(2*pi*u2)
> hist(x1, freq=FALSE, xlim=c(-4,4), breaks=40, ylim=c(0,0.4),
       xlab=" ", main=" ", col="steelblue")
> curve(dnorm(x), add=TRUE, col=2, lwd=2)
> hist(x2, freq=FALSE, xlim=c(-4,4), breaks=40, ylim=c(0,0.4),
       xlab=" ", main=" ", col="steelblue")
> curve(dnorm(x), add=TRUE, col=2, lwd=2)
```

정규분포 난수 생성

◆ 박스와 뮬러의 방법



03

15강 난수생성과
몬테카를로 시뮬레이션

이산형 확률분포의
난수 생성

베르누이 분포

◆ 베르누이 분포 $Ber(p)$ 를 따르는 확률변수

① 연속형 균등분포 $U(0,1)$ 난수 U 생성

② $U \leq p$ 이면 $X = 1$, 아니면 $X = 0$

이항 분포

◆ 이항 분포 $B(k,p)$ 를 따르는 확률변수

① 연속형 균등분포 $U(0,1)$ 난수 $U_1, U_2, \cdots, U_k$ 생성

② $U_i \leq p$ 이면 $X_i = 1$, 아니면 $X_i = 0$

③ $Y = \sum_{i=1}^{k} X_i$

이항분포의 예

예

이항분포 $B(10, 0.25)$와 이항분포 $B(100, 0.25)$를 따르는 난수를 생성하여라.

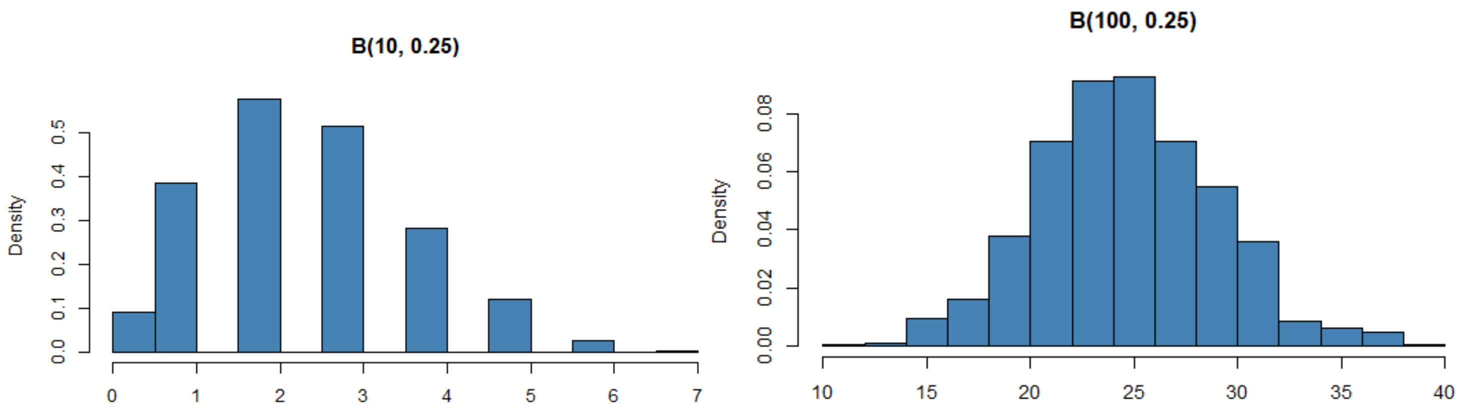
이항분포의 예

◆ 이항분포 난수의 생성

```
> set.seed(1234567)
> # 이항분포 B(10, 0.25) 생성
> bn1 = rep(NA, 1000)
> for(i in 1:1000) {bn1[i] = sum(runif(10)<0.25)}
> hist(bn1, freq=FALSE, xlab=" ", main="B(10, 0.25)", col="steelblue")
> # 이항분포 B(100, 0.25) 생성
> bn2 = rep(NA, 1000)
> for(i in 1:1000) {bn2[i] = sum(runif(100)<0.25)}
> hist(bn2, freq=FALSE, xlab=" ", main="B(100, 0.25)", col="steelblue")
```

이항분포의 예

◆ 이항분포 난수의 히스토그램



포아송 분포

- 연속형 균등분포 $U(0,1)$로부터 난수 U_i 생성
- $\prod_{i=1}^{Y+1} U_i < e^{-\lambda}$를 만족하는 최소 정숫값 Y

포아송분포의 예

예

평균이 3과 10인 포아송 분포 따르는 난수를 생성하여라.

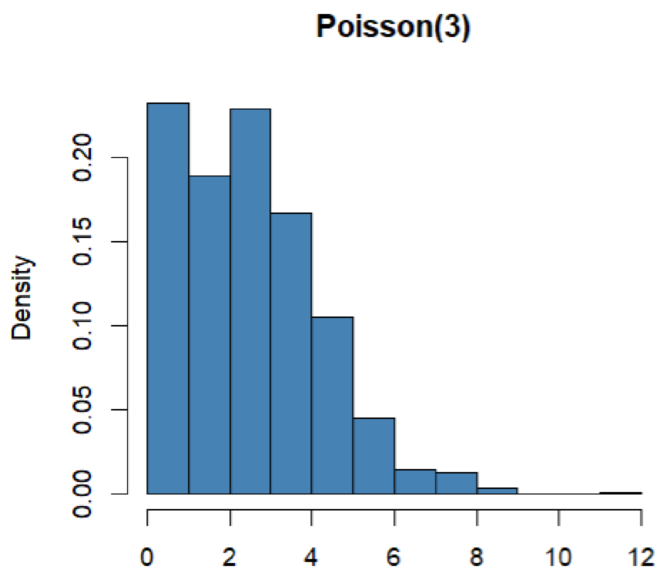
포아송분포의 예

- 포아송 분포 난수의 생성

```
> set.seed(1234567)
> poisson = function(lambda) {
  el = exp(-lambda)
  Y = 0;  up = 1
  while(up >= el) {
    Y = Y + 1
    up = up * runif(1)
   }
  return(Y-1)
 }
> n = 10000
> lambda1 = 3; lambda2 = 10
> Y1 = rep(NA, n); Y2 = rep(NA, n)
> for (i in 1:n){Y1[i] = poisson(lambda1)}
> for (i in 1:n){Y2[i] = poisson(lambda2)}
```

포아송 분포의 예

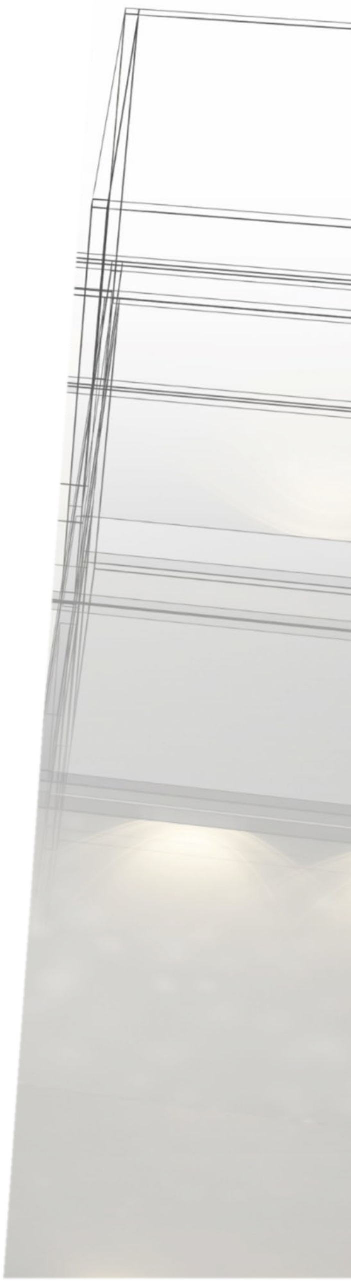
◆ 포아송 분포 난수의 히스토그램



04

15강 난수생성과 몬테카를로 시뮬레이션

몬테카를로 적분



몬테카를로 적분의 개념

- 연속형 확률변수 X의 확률밀도함수 : $f(x)$
- $h(X)$의 기댓값 : $E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x) dx$
- 몬테카를로 적분 : 대수의 법칙 이용

몬테카를로 적분의 예

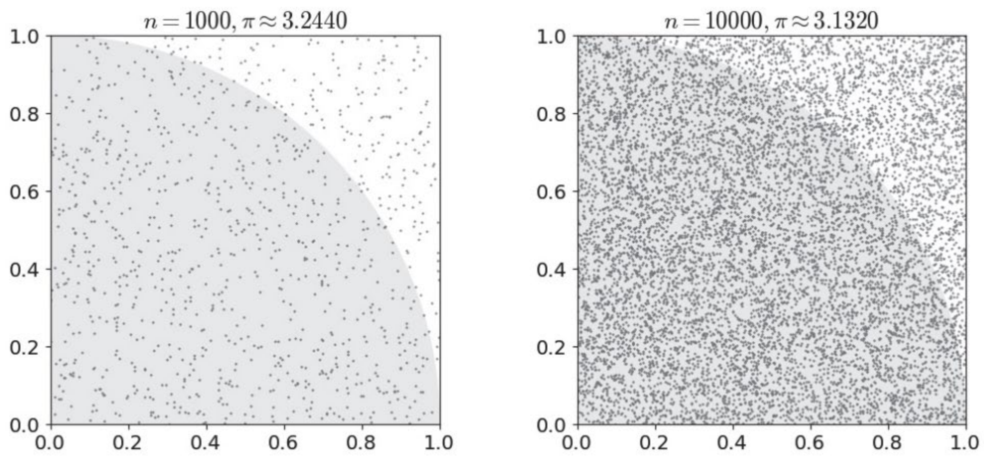
예

각변의 길이 2m인 정사각형 안에 반지름 1m인 원이 있는 표적. 정사각형 안에 임의로 화살을 쏠 때 화살이 원안에 들어갈 확률은?

몬테카를로 적분의 예

예

각변의 길이 2m인 정사각형 안에 반지름 1m인 원이 있는 표적. 이 정사각형 안에 임의로 화살을 쏠 때 화살이 원안에 들어갈 확률은?



몬테카를로 적분의 예

- 원주율 구하기

```
> set.seed(1234567)
> pical = function(n){
    u1 = runif(n)
    u2 = runif(n)
    x1 = rep(0,n)
    x1[u1^2 + u2^2 <= 1] = 1
    pi1 = mean(x1)*4
    return(pi1)
 }
```

```
> pical(100)
[1] 3.28
> pical(10000)
[1] 3.1668
> pical(1000000)
[1] 3.139024
```

몬테카를로 적분의 예

예

함수 $y = x^3$, $y = x\sqrt{1 - x^2}$ 의 0에서 1까지 하단 면적을 균등분포의 난수를 이용하여 구하여라.

몬테카를로 적분의 예

◆ 적분의 예

```
> par(mfrow=c(1, 2))
> set.seed(1234567)
> rint = function(f, gg, n){
    sam = matrix(runif(2*n), ncol=2)
    qq = gg(sam[,1], sam[,2])
    plot(sam[!qq, 1], sam[!qq, 2], col='steelblue', pch=1,
         xlim=c(0, 1), ylim=c(0, 1), xlab=" ", ylab=" ")
   points(sam[qq, 1], sam[qq, 2], col='gray', pch=1)
   curve(f, 0,1, n=100, col='black', add=TRUE, lwd=2)
   return(length(qq[qq]) / n)
  }
> f1 = function(x) x^3
> g1 = function(x, y) y <= x^3
> f2 = function(x) sqrt(1-x^2)*x
> g2 = function(x, y) y <= sqrt(1-x^2)*x
```

몬테카를로 적분의 예

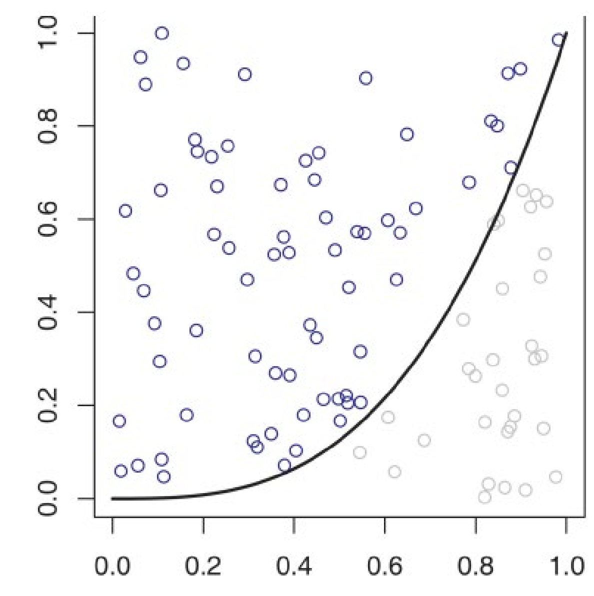
- 적분의 예

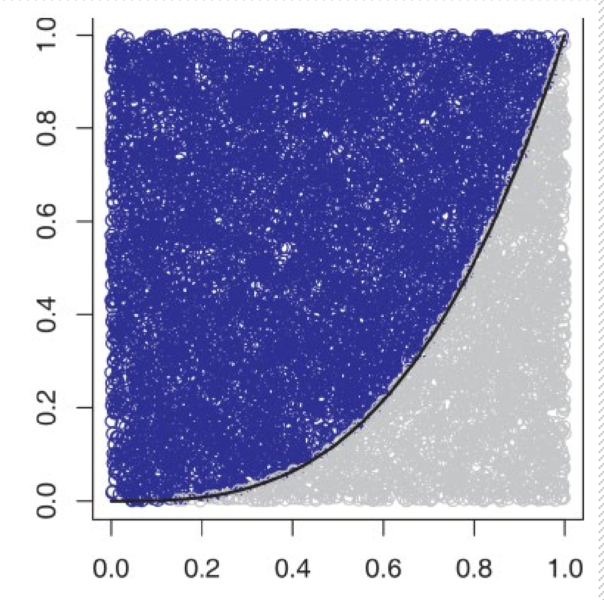
```
> integrate(f1, 0, 1)
0.25 with absolute error < 2.8e-15
> rint(f1, g1, 100)
[1] 0.28
> rint(f1, g1, 10000)
[1] 0.249
> integrate(f2,0,1)
0.3333334 with absolute error < 0.00012
> rint(f2, g2, 100)
[1] 0.32
> rint(f2, g2, 10000)
[1] 0.335
```

몬테카를로 적분의 예

예

함수 $y = x^3$의 0에서 1까지 하단 면적을 균등분포의 난수를 이용하여 구하여라.

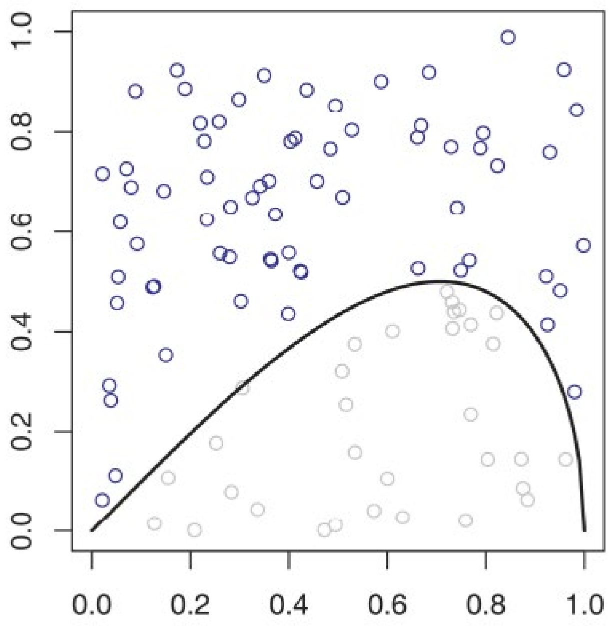


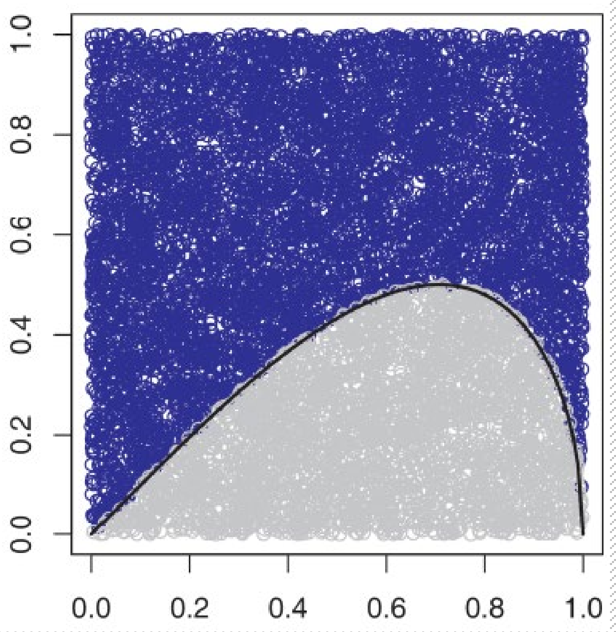


몬테카를로 적분의 활용 예

예

함수 $y = x\sqrt{1-x^2}$ 의 0에서 1까지 하단 면적을 균등분포의 난수를 이용하여 구하여라.





몬테카를로 적분의 예

예

평균이 0이고 분산이 1인 정규분포를 따르는 확률변수의 기댓값과 분산을 난수를 이용하여 구하여라.

몬테카를로 적분의 예

- 정규분포 확률변수의 기댓값과 분산

```
> set.seed(1234567)
> x1 = rnorm(100)
> x2 = rnorm(10000)
>
> f1 = function(x) x*dnorm(x)
> mu = integrate(f1, -Inf, Inf)
> mu
0 with absolute error < 0
> mean(x1)
[1] -0.0771234
> mean(x2)
[1] -0.009002444
```

```
> f2 = function(x) (x-mu$value)^2*dnorm(x)
> sigma2 = integrate(f2, -Inf, Inf)
> sigma2
1 with absolute error < 1.2e-07
> var(x1)
[1] 0.9188922
> var(x2)
[1] 0.9895978
```

학습정리

- 몬테카를로 시뮬레이션은 확률분포로부터 난수를 수 없이 반복, 생성하여 복잡한 문제 해를 근사적으로 구하고, 이를 통해 확률적 문제를 해결하는 방법이다.

- 확률분포의 난수는 누적확률분포의 역함수와 기각법 등을 통해 생성될 수 있다.

정리하기

학습정리

- 몬테카를로 적분은 난수를 통해 구한 확률변수 함수의 표본 평균을 통해 확률변수 함수의 적분값을 구하는 방법이다.

이번 학기
수고하셨습니다.