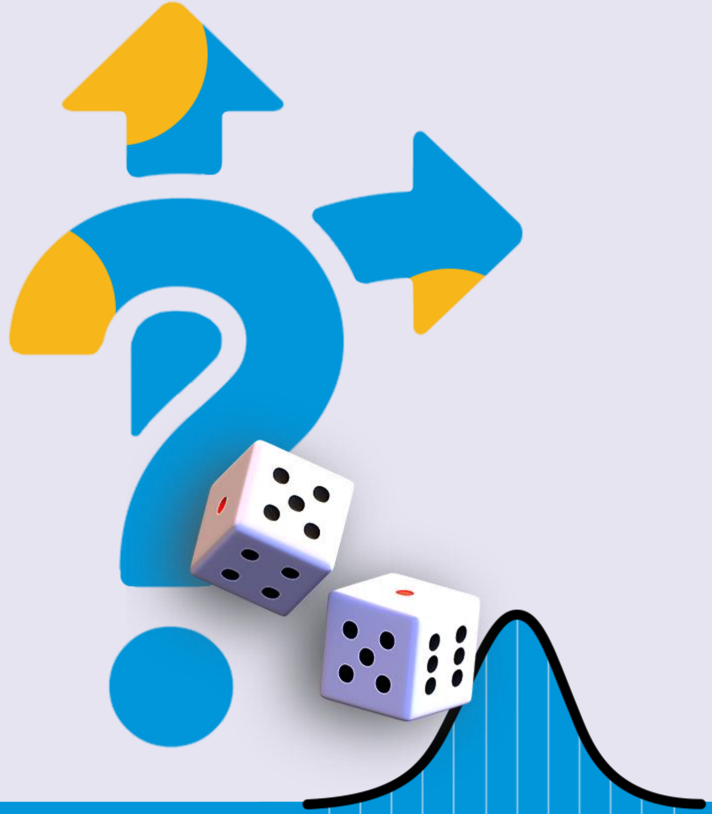
베이즈데이터분석 / 이재용 교수

08강

랜덤 숫자 발생



목차

목차

역함수방법

- F는 역함수가 존재하는 누적분포함수라 하자.
- $u \sim U(0,1)$이라면, $F^{-1}(u) \sim F$이다.

예: 지수분포

- $x \sim Exp(1) \Leftrightarrow x \overset{d}{=} -\log(1 - u), u \sim U(0,1)$
  $\Leftrightarrow x \overset{d}{=} -\log u,\ u \sim U(0,1)$

- $x \sim Exp(\lambda), \lambda > 0 \Leftrightarrow x = z|\lambda, z \sim Exp(1)$

예: 코시 확률변수의 생성

- $x \sim Ca(0,1)$일 때, x의 밀도함수는

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, x \in \mathbb{R}$$

이고, 누적분포함수는

$$F(x) = \frac{1}{\pi}arctan(x) + \frac{1}{2}, x \in \mathbb{R}$$

이다. 누적분포함수의 역함수는

$$F^{-1}(u) = tan\left(\pi(u - \frac{1}{2})\right), u \in (0,1)$$

이다.

예: 코시 확률변수의 생성

- $u \sim U(0,1)$일 때,

$$\tan(\pi(u - \frac{1}{2})) \sim Ca(0,1)$$

- $z \sim Ca(0,1)$일 때, $x = \mu + \sigma z, z \sim Ca(\mu, \sigma)$

참고

1. $arctan(x)$는 $tan(x)$의 역함수이다.

   즉, $y = arctan(x) \Leftrightarrow x = tan(y)$가 성립한다.

2. $\int \frac{1}{1+x^2} dx = arctan(x).$

예: 코시 확률변수의 생성

목차

합격–불합격 방법(acceptance-rejection method)

목적과 상황

밀도함수 f 에서 난수를 발생시키고자 한다. f 에서 직접 난수를 발생시키는 것은 어렵지만, 아래의 조건을 만족하는 밀도함수 g에서 난수를 발생시키는 것은 쉽다고 하자.

$$f(x) \leq Mg(x), x : f(x) > 0.$$

여기서 $M > 0$은 상수이다.

f 를 목표 밀도 함수(target density),

g를 제안 밀도 함수(proposal density),

M을 봉투 상수(envelope constant)라 한다.

합격-불합격 방법(acceptance-rejection method)

알고리듬

(단계 1) $x \sim g$와 $u \sim U(0,1)$를 발생시킨다.

(단계 2) $u \le f(x)/Mg(x)$이면, $y = x$이라 하고,
그렇지 않으면 단계 1로 돌아간다.

합격–불합격 방법(acceptance-rejection method)

참고

- f 와 g 의 적분값이 1이 아니어도 알고리듬은 성립한다.
- x 가 주어져 있을 때의 x 가 합격이 될 비율

$$\alpha(x) := \frac{f(x)}{Mg(x)}$$

를 합격 비율($acceptance\ ratio$)

- g 에서 생성한 x 가 f 의 확률 변수로 합격이 될 확률

$$\mathbb{P}\left(u \le \frac{f(x)}{Mg(x)}\right)$$

를 합격 확률($acceptance\ probability$)이라 한다.

합격 확률이 너무 작지 않아야 사용할 수 있다.

예. 삼각분포에서 확률변수의 생성

- 다음의 확률밀도함수에서 확률변수를 생성하는 문제를 고려해 보자.

$$f(x) = \begin{cases} x, & 0 \le x \le 1, \\ 2 - x, & 1 \le x \le 2 \end{cases}$$

$$g(x) = 1, 0 \le x \le 2$$

라 정의하면

$$\sup_{x \in [0,2]} \frac{f(x)}{g(x)} = 1$$

임을 알 수 있다.

예. 삼각분포에서 확률변수의 생성

단계 1. $x \sim U(0,2), u \sim U(0,1)$을 생성한다.

단계 2. 합격비율 $f(x)$를 계산한다.

단계 3. $u \leq f(x)$이면, $y = x$를 반환하고,
그렇지 않으면 단계 1로 돌아간다.

예. 정규 확률변수의 생성

- (목적)

$$y \sim c_1 f(x) = c_1 e^{-y^2/2}, c_1 = \frac{1}{\sqrt{2\pi}}$$

- (제안 밀도함수)

$$x \sim c_2 g(x) = c_1 \frac{1}{1+x^2}, c_2 = \frac{1}{\pi}$$

$$\sup_x \frac{f(x)}{g(x)} \leq M, M = 2/\sqrt{e}.$$

- (합격확률)

$$\mathbb{E}\alpha(x) \approx 0.66$$

예. 정규 확률변수의 생성

알고리듬

단계 1. $M = \frac{2}{\sqrt{e}}$이라 놓는다.

단계 2. $x \sim Cauchy(0, 1)$와 $u \sim Unif\,(0, 1)$를 생성한다.

단계 3. 합격 비율

$$\alpha(x) = \frac{(1 + x^2)e^{-x^2/2}}{M}$$

를 계산한다.

단계 4. $u \leq \alpha(x)$이면, $y = x$를 반환하고,
아니면 (2)로 돌아가서 다시 시작한다.

제안 분포의 비교

- f : 목적 밀도함수. 밀도함수의 상수배 가능.
- $g_i, i = 1, 2$: 제안 밀도함수. 밀도함수이어야.
- (합격 확률)

$$P\left(u \le \frac{f(x)}{M_i g_i(x)}\right)$$

$$= EP\left(u \le \frac{f(x)}{M_i g_i(x)} \middle| x\right)$$

$$= E\frac{f(x)}{M_i g_i(x)} = \int \frac{f(x)}{M_i g_i(x)} g_i(x) dx = \frac{1}{M_i}\int f(x) dx.$$

- M_i가 작은 g_i가 큰 합격 확률을 갖는다.

다음시간

09강

깁스추출법