

기말평가(온라인과제물) 과제명

2022학년도 2학기

대상학과	통계데이터과학과	교과목명	베이즈데이터분석
대상학년	4학년	평가유형	과제물

[과제명]

- (10점) 밀도함수 $f(x) = \sin^2(x) + \cos^2(2x)$, $x \in [0, 2\pi]$ 를 따르는 랜덤 숫자를 합격-불합격 방법으로 생성하고자 한다. 다음의 질문에 답하시오.
 - 밀도함수의 개형을 R을 이용해 그리시오.
 - $f(x)$ 의 최대값을 구하시오.
 - $U[0, 2\pi]$ 를 제안밀도함수로 하여 합격-불합격 방법 알고리듬을 서술하시오.
 - 위에서 서술한 알고리듬으로 5000개의 랜덤표본을 R을 이용해 $f(x)$ 에서 추출하고, 히스토그램을 그리시오. 표본평균과 표본 표준편차를 구하시오.
- (10점) $\theta \sim N(0, 1)I(\theta \geq 3)$ 에서 분할추출법을 이용하여 표본을 추출하고자 한다. 다음에 답하시오.
 - 책의 예 9-4를 참고하여 분할추출 알고리듬을 서술하시오.
 - 5000개의 표본을 추출하고 히스토그램을 그리시오. 표본평균과 표본 표준편차를 구하시오.
- (10점) 10명의 대한민국 남자들의 머리둘레와 발길이가 다음과 같이 주어졌다.

머리둘레 57.8, 58.6, 60.8, 57.6, 58.1, 60.4, 56.7, 60.0, 56.7, 54.3

발길이 27.2, 26.8, 29.9, 28.0, 27.8, 26.7, 30.8, 27.1, 26.5, 27.9

위의 데이터가

$$x_i = (x_{i1}, x_{i2}) \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right), i = 1, 2, \dots, 10$$

를 따른다고 하자. 각 모수의 사전분포가

$$\begin{aligned}\mu_1, \mu_2 &\sim U(R) \\ \sigma_1, \sigma_2 &\sim U(R_+) \\ \rho &\sim U(-1, 1)\end{aligned}$$

를 따른다고 할 때, 스텐을 이용하여 사후표본을 추출하고자 한다. 다음의 질문에 답하시오.

- (a) 스텐을 이용하여 번인 5000개를 포함하여 총 15,000개의 사후표본을 추출하시오.
- (b) 모수들의 사후표본의 밀도함수 그림, 시계열 그림, 자기상관계수 그림을 그리고 마르코프 체인이 수렴했는지 판단하시오. 수렴하지 않았다고 판단하면 수렴했다고 판단할 때까지 사후표본의 크기를 늘리시오.
- (c) 모수들의 사후평균, 사후표준편차, 95% 신용구간을 구하시오.

4.(20점) 다음은 10명의 남성들의 2라운드 골프 기록이다.

1라운드 92, 91, 85, 93, 87, 82, 100, 108, 82, 87

2라운드 93, 86, 88, 83, 79, 77, 108, 110, 85, 88

위의 데이터에 다음의 모형을 적합하고자 한다.

$$y_i = \alpha + \beta x_i + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, \dots, 10$$

$$\pi(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2} d\alpha d\beta d\sigma^2$$

$$\alpha, \beta \in R, \quad \sigma^2 > 0$$

x_i 와 y_i 는 i 번째 남성의 1라운드와 2라운드 기록을 나타낸다. 다음의 질문에 답하시오.

- (a) 위 데이터의 산점도를 그리시오.
- (b) 위 모형을 적합하기 위한 스템과 R 코드를 작성하고, 사후표본을 구하시오.
- (c) 모수들의 사후표본의 밀도함수 그림, 시계열 그림, 자기상관계수 그림을 그리고 마르코프 체인이 수렴했는지 판단하시오. 수렴하지 않았다고 판단하면 수렴했다고 판단할 때까지 사후표본의 크기를 늘리시오.
- (d) $(\alpha, \beta, \sigma^2)$ 의 베이즈 추정량을 구하고, 기록의 예측식을 써라.
- (e) $(\alpha, \beta, \sigma^2)$ 의 95% 신용구간을 구하라.
- (f) 1라운드 기록이 84인 남성이 있었다고 한다. 이 남성의 2라운드 기록을 예측해보시오. 이 남성의 2라운드 기록의 베이즈 추정량을 구하고, 95% 예측구간을 구하시오. 확률변수의 신용구간은 예측구간이라 한다.

※ 스텐이 작동이 되지 않는 학생들은 위의 3, 4번 대신 다음의 문제를 수행하시오.

기존 3, 4번 문제에 대하여 아래와 같은 문제로 대체하여 과제 제출하여도 불이익을 받지 않으니 과제 제출에 참고 바랍니다.

3. 다음의 이변량 정규분포

$$\pi = N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right), \mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 0.8$$

를 불변분포로 갖는 깁스추출 알고리듬을 R로 구현하고자 한다.

- (a) 알고리듬을 서술하시오.
- (b) 알고리듬을 R로 구현하시오.
- (c) m = 5000개의 표본을 추출하시고, 요약통계량들과 히스토그램을 그리시오.
- (d) 시계열 그림을 그리시오.
- (e) 자기상관계수 그림을 그리시오.

4. 다음은 10명 남성들의 골프 기록이다.

92, 91, 85, 93, 87, 82, 100, 108, 82, 87

위의 데이터에 다음의 모형을 적합하고자 한다.

$$x_1, x_2, \dots, x_n | \theta, \sigma \sim N(\theta, \sigma^2)$$

$$\pi(\theta, \sigma) \propto \frac{1}{\sigma} d\theta d\sigma$$

(θ, σ)의 사후분포에서 사후표본을 깁스추출법이나 메트로폴리스-헤이스팅스 알고리듬을 이용하여 R로 구현하고자 한다.

- (a) 알고리듬을 서술하시오.
- (b) 알고리듬을 R로 구현하시오.
- (c) m = 5000개의 표본을 추출하시고, 요약통계량들과 히스토그램을 그리시오.
- (d) 시계열 그림을 그리시오.
- (e) 자기상관계수 그림을 그리시오.

[과제 작성 시 지시사항]

★ 과제물 제출 시 기말시험(과제물) 표지를 과제물 파일로 제출하시고, 마크다운 파일과 컴파일된 html 파일을 zip파일로 압축하여 보조첨부파일로 제출하시오.

**※ 마크다운 파일과 컴파일된 html 파일에 작성자 학번과 성명이 기재될 수 있게 작성
하시오.**

<기말 과제물 표지 서식 다운로드 방법>

1. 메인홈페이지-학생서식-표지 및 기타서식-기말시험(과제물) 표지-한글/워드
2. 메인홈페이지-공지사항-시험/성적 공지사항-2022학년도 2학기 기말시험(과제물) 시행공고- 불임파일 참조

[참고문헌]

- 이재용, 이기재. (2022) 베이즈데이터 분석. 한국방송통신대학교출판문화원.