전산통계

Chapter 3.3 표본평균법

표본평균법 (sample mean method)

- 모수 추정에 있어 불편성과 최소 분산을 갖는 추정량을 선호
- Hit or Miss 방법보다 분산이 작으며, 기댓값 개념을 이용하는 방법
- $I = \int_a^b g(x)dx\,, \int_a^b \frac{g(x)}{f(x)} f(x)dx = E\left[\frac{g(x)}{f(x)}\right]$, $f(x)$는 $[a, b]$에서 정의된 확률변수 X의 확률밀도함수
- 적당한 $f(x)$를 선택하고, $f(x)$를 따르는 N개의 표본을 생성하여 $\frac{g(x)}{f(x)}$을 계산하고 이들의 평균을 구하여 I의 추정값으로 사용한다. 즉, $\hat{I} = \frac{1}{N}\sum_{i=1}^{N} \frac{g(x_i)}{f(x_i)}$

표준균일분포를 이용한 알고리즘

- 1단계 : 반복 단계 $i = 1,\ 2,\ \cdots,\ N$
  - 난수 u_i를 생성
  - $x_i = a + u_i(b-a)$를 계산
  - $g(x_i)$를 계산
- 2단계 : 종료 단계 $\hat{I} = (b-a) \times \frac{1}{N}\sum_{i=1}^{N} g(x_i)$

- $X \sim U(a,\ b)$이므로, $I = (b-a)E[g(X)]$
- $\hat{I} = (b-a) \times \frac{1}{N}\sum_{i=1}^{N} g(x_i)$, x_i는 $U(a,b)$에서 생성된 표본
- $\hat{I}$은 I의 불편추정량

  - $E\left[\hat{I}\right] = (b-a)\frac{1}{N}\sum_{i=1}^{N} E[g(X_i)] = (b-a)\frac{1}{N}\sum_{i=1}^{N}\int_a^b \frac{g(x_i)}{b-a}dx_i = I$

- $Var\left(\hat{I}\right) = \frac{(b-a)^2}{N^2}\sum_{i=1}^{N} Var[g(X_i)] = \frac{(b-a)^2}{N} Var[g(X_1)] = \frac{(b-a)^2}{N}\left[\int_a^b \frac{g(x)^2}{b-a}dx - \frac{I^2}{(b-a)^2}\right] =$

  $\frac{1}{N}\left[(b-a)\int_a^b g(x)^2 dx - I^2\right]$

- Hit or Miss 추정량보다 표본평균법의 추정량의 분산이 작다.

- Ex) $\int_0^1 4\sqrt{1-x^2}\,dx$

주표본기법 (importance sampling)

- 표본평균법의 일반화된 형태
- $I = \int_a^b g(x)dx$, $\int_a^b \frac{g(x)}{f(x)} f(x)dx = E\left[\frac{g(x)}{f(x)}\right]$, $f(x)$는 $[a, b]$에서 정의된 확률변수 X의 확률밀도함수
- 적당한 $f(x)$를 선택하고, $f(x)$를 따르는 N개의 표본을 생성하여 $\frac{g(x)}{f(x)}$을 계산하고 이들의 평균을 구하여 I의 추정값으로 사용한다. 즉, $\hat{I} = \frac{1}{N}\sum_{i=1}^{N} \frac{g(x_i)}{f(x_i)}$

- 1단계 : 반복 단계 $i = 1,\ 2,\ \cdots,\ N$
  - 난수 u_i를 생성
  - $x_i = a + u_i(b-a)$를 계산
  - $\frac{g(x_i)}{f(x_i)}$를 계산
- 2단계 : 종료 단계 $\hat{I} = \hat{I} = \frac{1}{N}\sum_{i=1}^{N}\frac{g(x_i)}{f(x_i)}$

- $\hat{I}$은 I의 불편추정량

  - $E[\hat{I}] = I$

- $Var(\hat{I}) = \frac{1}{N}\left[\int_a^b \frac{g(x)^2}{f(x)} dx - I^2\right]$

- $Var(\hat{I})$을 가능한 작아지도록 $f(x)$를 선택한다. $f(x)$를 주요함수 (importance function)라고 한다.
- 코쉬-슈바르츠 (Cauchy-Schwarz) 부등식
  - $\left[\int_a^b |g(x)|dx\right]^2 = \left[\int_a^b \frac{|g(x)|}{\sqrt{f(x)}}\sqrt{f(x)}dx\right]^2 \le \int_a^b \frac{g(x)^2}{f(x)}dx \int_a^b f(x)dx = \int_a^b \frac{g(x)^2}{f(x)}dx$
  - 분산의 최소값 = $\frac{1}{N}\left[\left\{\int_a^b |g(x)|dx\right\}^2 - I^2\right]$, 여기서 $f(x) = \frac{|g(x)|}{\int_a^b |g(x)|dx}$
- $f(x) \propto |g(x)|$인 형태이므로 $|g(x)|$에 비례하는 확률밀도함수를 선택하면 분산이 작아질 것으로 예상

- Ex) $\int_0^1 \sqrt{1-x^2}\,dx$
  - $g(x) = \sqrt{1-x^2} = g(0) + g'(0)x + \frac{1}{2}g''(0)x^2 + \cdots = 1 - \frac{1}{2}x^2 + \cdots$
  - $f(x) = \frac{6}{5}(1 - \frac{1}{2}x^2)$