

[수치해석] Simulation

Modeling

| Unmodeled world (현실 세계) → Modeled world

- 현실 세계가 굉장히 복잡하기 때문에 어떤 규칙이 있는지 알 수 없다.
- 이를 표현하고 이해하기 위해 우리가 이해하기 쉬운 형태로 변형한 것.

Fitting

| Observed data를 통해 우리가 Modeling한 Model의 Parameter를 찾는 과정

- $\{X_n, Y_n\} \rightarrow y = f(w, x) + Noise$
- 최적의 w 를 찾아내는 과정

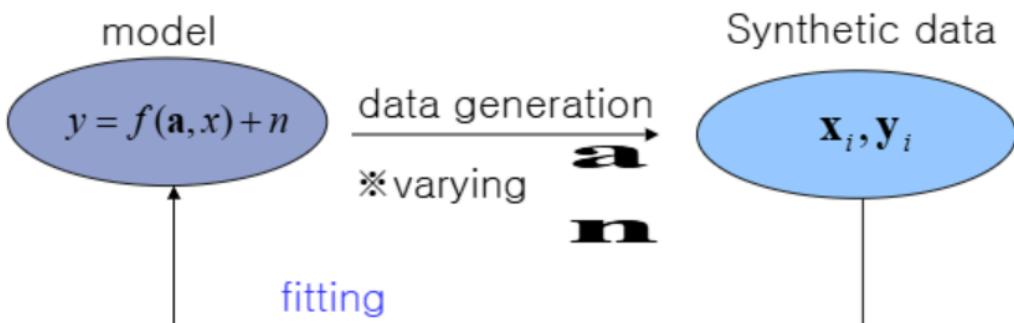
모델링은 모델의 복잡도가 높을수록 좋은 성능을 보이지만, 그만큼 큰 비용이 요구된다.

현실을 완벽하게 모델링하는 것은 어렵다.

Simulation

| 현실 세계의 시스템을 수학적·컴퓨터 모델로 모방하여 실험하는 과정

Simulation



1. 현실 세계를 모방한 Model을 이용하여 **Synthetic data**를 생성한다.
2. Synthetic data를 이용하여 **Model을 Fitting**한다.

일반적으로 **Data generation**에서 a는 Uniform distribution에서, n은 Gaussian distribution에서 Sampling한다.

- a를 고정하고 x, n을 sampling해서 y를 얻고, (x, y) 쌍으로 묶어 Data를 생성한다.
- 이 Synthetic data 자체가 **현실처럼 동작하는 가상의 실험 데이터를 생성하는 것이다.**

Data generation에서 Random number를 다루는 것은 굉장히 중요하다.

- 실세계와 비슷하게 모방하기 위해선 **다양한 분포에서 샘플링을 할 필요가 있다.**

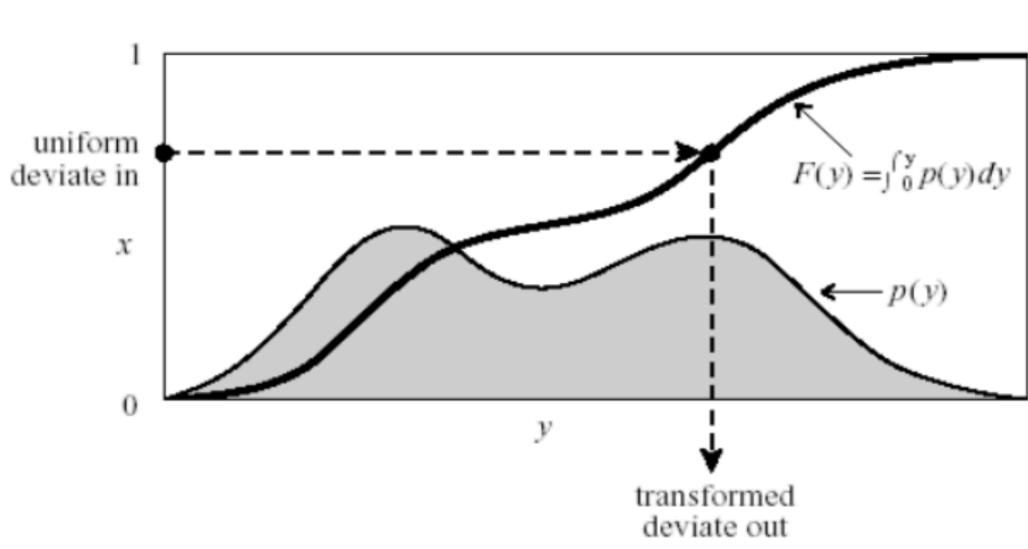
Random number를 생성하는 일반적인 과정

1. 어떤 분포를 사용할 지 정한다.
2. Uniform distribution을 얻는다.
 - 컴퓨터는 Uniform distribution만 쉽게 얻을 수 있다.
 - 컴퓨터는 원하는 분포를 바로 얻을 수 없어, Uniform distribution에서 변환하는 방식을 이용해야 하기 때문이다.
3. **Transformation method, Rejection method 등을 이용하여 (2)의 Uniform distribution을 변환해 원하는 분포를 얻는다.**

강의에선 두 가지 방법을 소개한다.

1. Transformation method

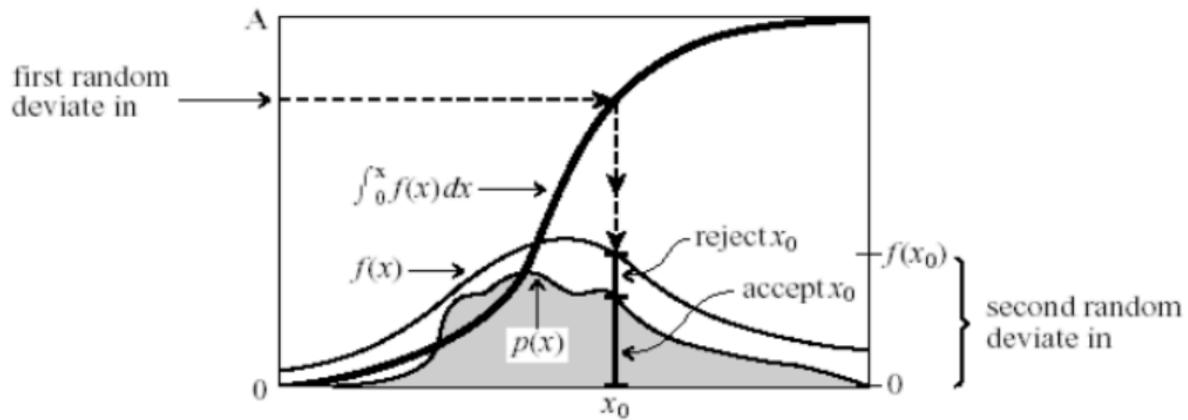
$p(x)$ 는 우리가 원하는 분포이다.



1. Uniform distribution에서 샘플링하여 x_0 를 얻는다.
2. $F(y)$ 는 $p(x)$ 의 CDF인데, $F(y) = x_0$ 가 되는 y' 를 얻는다.
3. y' 를 사용하면 된다.

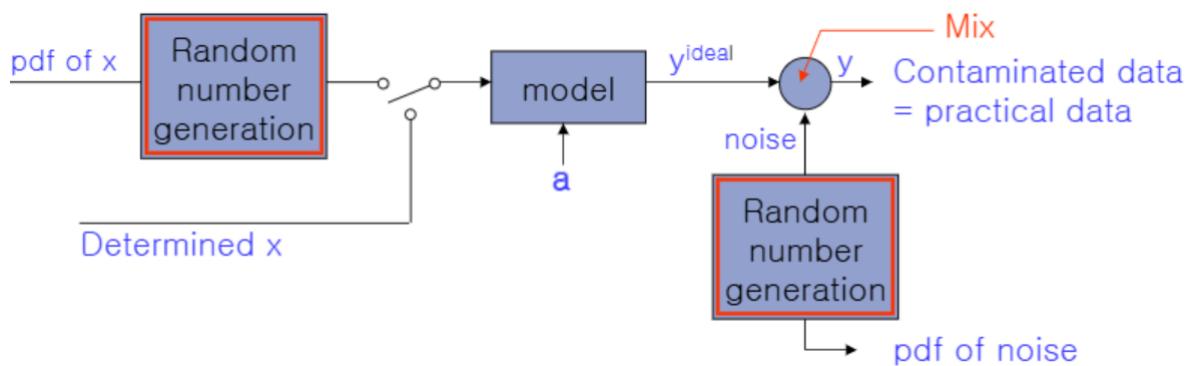
Rejection method

$p(x)$ 는 실제 분포, $f(x)$ 는 우리가 이미 모델링할 수 있는 분포이다.



1. 먼저 x 를 Uniform distribution에서 샘플링하여 x_0 를 얻는다.
2. $[0, f(x_0)]$ 구간의 한 점을 Random하게 얻는다.
3. (3)에서 얻은 값이 회색 영역 ($p(x)$) 내에 존재한다면 x_0 를 사용한다.
4. (3)에서 얻은 값이 회색 영역 내에 존재하지 않는다면 (1) ~ (3) 동작을 반복한다.

Synthetic data generation detail



1. X : 특정 PDF에서 샘플링된 랜덤한 값일 수도 있고 미리 정해진 값일 수도 있다.
2. X 가 a 라는 Parameter를 사용하는 모델을 통과하면 y_{ideal} 을 얻을 수 있다.
 - y_{ideal} : Noise가 없다는 의미에서 ideal

3. Random하게 생성된 Noise가 생성된다.

- 현실 데이터가 특정 Distribution만 따르듯이 깔끔하지 않기 때문에 추가한다.

4. y_{ideal} 에 Noise를 추가하여 **Practical data** y 를 얻는다.