

# ADF 모의실험 및 비교

김동완

고려대학교 정책대학원  
데이터 통계학과

2016년 3월 5일

# Overview

## 1 개요

- 문제점 분석
- 실험 목적
- 비교 대상 방법론

## 2 모의 실험

- 모의 데이터
- 회귀 계수 추정: 최대 우도
- 모의 실험 결과:  $w_0$
- 모의 실험 결과:  $w_1$
- 모의 실험 결과: 비교

## 문제점 분석

- 지난달 구현에서의 문제점은  $\theta_t$  의 예측분포를 update 해야하는데,  $\theta_{t-1}$  의 사후 분포를 update하여 생긴 문제로 보임. 이를 보완하기 위해 정규분포의 특성에 따라  $\sigma_{t,i}^2$ 를 update하는 단계에서  $\theta_t$ 의 분산을 더해줌.  $\theta_t$ 의 분산 값으로는 iteration이 증가함에 따라 불확실성이 감소함을 반영하기 위해  $\theta_{t-1}$ 의 사후 평균의 절대값을 iteration 수로 나눈 값을 사용함.

$$\mu_{t,i} = \mu_{t|t-1,i} + a_i \delta_m \quad (1)$$

$$\sigma_{t,i}^2 = \sigma_{t-1,i}^2 + a_i^2 \delta_v + |\mu_{t,i}|/t \quad (2)$$

$$a_i \triangleq \frac{x_{t,i} \sigma_{t|t-1,i}^2}{\sum_j x_{t,j}^2 \sigma_{t|t-1,i}^2} \quad (3)$$

# 실험 목적

- ADF(Assumed Density Filtering)를 GLM에 적용하는 것은 온라인 경사 하강법(Online Gradient Descent)과 많은 부분이 유사하며, 매 반복에서 선형식  $\eta$ 와 각 회귀 계수의 분포를 Step size에 반영하는 것이 차이점.
- 모의 실험을 통해 로지스틱 회귀에 있어서 '온라인 경사 하강법'과 'ADF 방법'을 비교 하고자 함.

# 비교 대상 방법론

- 경사 하강법(Gradient Descent): 1차 최적화 알고리즘 중 하나로 함수의 극솟값(local minimum) 및 값수값이 극솟값을 갖게 하는 모수를 점진적으로 찾는 방법
- 배치 경사 하강법(Batch Gradient Descent): 매 반복(iteration)마다 전체 데이터를 사용하여 극솟값을 찾는 방법
- 온라인 경사 하강법(Online Gradient Descent): 매 반복에서 하나의 샘플을 사용하여 극솟값을 찾는 방법

# 모의 데이터

- 아래 조건으로 단순 로지스틱 회귀모형을 위한 데이터를 3000건 생성.

$$\text{logit}(E[Y_i]) = w_{0i} + w_{1i}x_{1i}, \quad i = 1, \dots, n$$

$$w_{0i} \stackrel{i.i.d}{\sim} N(3, 1), \quad i = 1, \dots, n$$

$$w_{1i} \stackrel{i.i.d}{\sim} N(10, 1), \quad i = 1, \dots, n$$

$$y_i \sim \text{Bern}(p_i = \text{logit}^{-1}(w_{0i} + w_{1i}x_{1i})), \quad i = 1, \dots, n$$

- 경사하강법(GD)에서 사용할 step size  $\alpha$ 값은  $w_0$ 에는 0.08,  $w_1$ 에는 1.1을 사용.

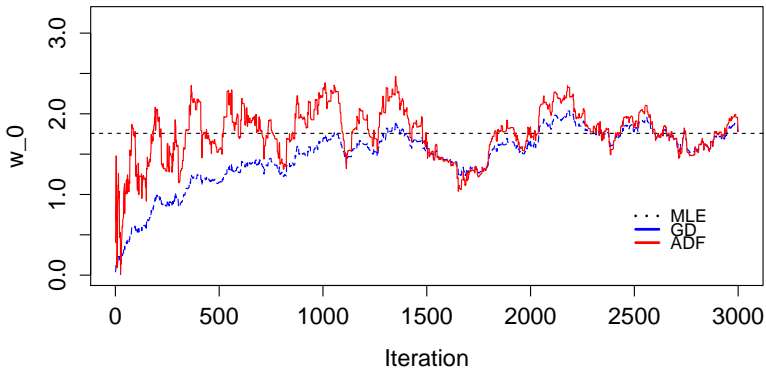
## 회귀 계수 추정: 최대 우도

- 전체 데이터를 이용하여 회귀 계수( $w_{0i}$ ,  $w_{1i}$ )에 대한 최대 우도 추정치를 구하면 아래와 같다.

회귀 계수	추정치	P-value
$w_0(\text{intercept})$	2.6620	$<2e-16$
$w_1$	8.6870	$<2e-16$

## 모의 실험 결과: $w_0$

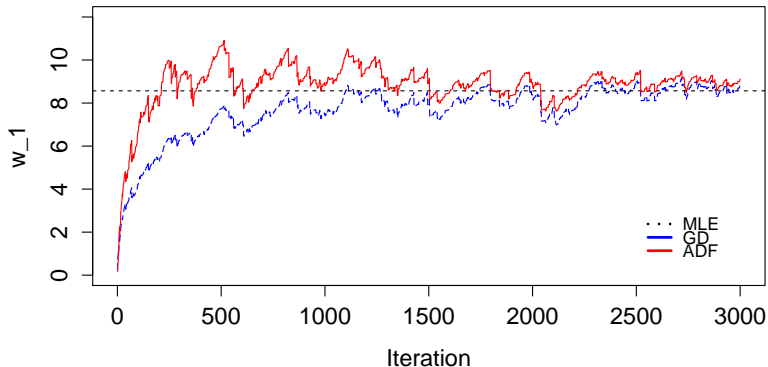
- iteration이 증가함에 따른  $w_0$ (intercept) 값의 변화 비교  
(온라인 경사하강법(GD) vs ADF)





## 모의 실험 결과: $w_1$

- iteration0이 증가함에 따른  $w_1$ (intercept) 값의 변화 비교  
(온라인 경사하강법(GD) vs ADF)



## 모의 실험 결과: 비교(1)

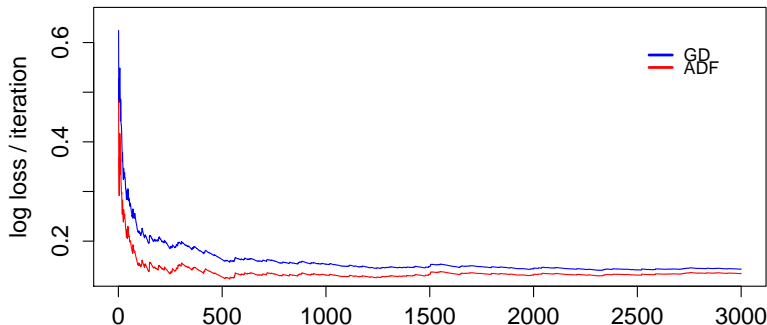
- 두 방법 모두 iteration이 증가함에 따라 MLE 추정치로 근사함을 알 수 있고, ADF의 경우 초기 각 계수의 큰 분산의 영향으로 그 범위 폭이 GD에 비해 큼을 알 수 있다.

Iteration	GD	ADF
200	0.2071293	0.1498564
500	0.1632435	0.1278876
1500	0.1493539	0.1335120
3000	0.1441112	0.1352678

표:  $\frac{\text{Logloss}}{\text{iteration}}$

## 모의 실험 결과: 비교(2)

- 두가지 방법론에서 iteration 당 log-loss를 비교해 보면 아래와 같고, ADF가 약간 더 좋은 결과를 보임을 알 수 있다. 그러나 iteration이 많아 지면 그 차이가 점차 줄어들게 된다.



# References



오만숙, (2012)

R몬테칼로와 함께하는 베이지안 통계추론

자유아카데미 Ch17, p115 - 116.



Murphy, K.P.(2012)

Machine learning

*The MIT Press* Ch18.5.3, p652 - 655.



Zoeter, O. (2007)

Bayesian Generalized Linear Models in a Terabyte World.

# The End