일차원 비 선형방정식의 확률 근사법에 의한 해법

Solution of One-Dimensional Non-Linear Equations Using Probability Approximation

이성철1)

요약

본 논문은 일차원 Robbins-Monro 확률근사법에 사용된 계수 a_n을 진보시켜 새로운 알고리즘을 제안하고 실례를 들어 효율성을 입증 하였다.

Ⅰ.서론

선형 및 비선형 방정식 및 미분방정식의 근사근을 구하기 위한 해법은 미적분이 발견된 이래로 계속되어왔다. 선형방정식의 근사 해법에 있어서 확률근사법의 사용은 1951년 H. Robbins 교수와 당시의 대학원생 Monro에 의하여 처음시도 되었다. 흔히 Robinns-Monro방법으로 일컬어지는 이방법은 1차원 확률근사법에 의한 방정식 $M(x)=0$ ($M(x)$는 연속함수)의 근의 근사해법이다.

점 x에서 관찰치는 확률변수 $Y(x)=M(x)+Z(x)$의 실현치라고 하고 이것을 $y(x)$로 나타낸다. 여기서 $Z(x)$는 평균이 0 인 확률변수로서 그의 분산은 유한치 $\sigma^2<\infty$를 넘지 않는다 하자. $y(x)$는 x 에 의존함으로 그의 분포함수를 $H(y \mid x)$ 이란 조건부 분포함수로 나타낼 때 $M(x)=\int y dH(y \mid x)$ 로 쓰인다. 즉, $E[Y(x)]=M(x)$이다. 이 경우, $M(X)$를 x에 대한 y의 회귀함수(regression function)라 한다.

[정리1(Robbins-Monro)]

(1) $|M(x)| \le c$

(2) $\sigma^2(x)=\int [y-M(x)]^2 dH(y \mid x) \le \sigma^2 < \infty$

(3) $x<\theta$이고, $M(x)<\alpha$, $M(\theta)=\alpha$. $\theta<x$ 에서 $\alpha<M(x)$

(4) 양수 δ가 존재하여, 구간$|x-\theta|<\delta$ 에서 $M(x)$는 단조증가

1) 남서울대학교 교양학부 교수

(6) 양수열 $\{a_n\}$에 대하여 $\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty$

이상의 조건에서 근 θ의 임의의 제1근사값 x_1에 대하여 $x_{n+1} = x_n - a_n[y(x_n) - \alpha]$인 algorithm에 의하여 만들어지는 수열 $\{x_n\}$은 근 θ에 평균제곱 수렴한다.

즉, $\lim_{n \to \infty} E[(x_n - \theta)^2] = 0$.

1954년 Blum은 위의 정리를 다음과 같은 보다 완화된 조건하에서 확률 1의 수렴을 보증하는 정리로 발표하였다.

[정리 2(Blum)]

(1) $|M(x)| \le C + A|x| \quad (C, A > 0)$

(2) $\sigma^2(x) \le \sigma^2$

(3) $x < \theta$에서 $M(x) < \alpha$, $\theta < x$ 에서 $\alpha < M(x)$

(4) 임의의 $0 < \delta_1 < \delta_2$에 대하여

$inf_{\delta_1 \le |x-\theta| \le \delta_2} |M(x) - \alpha| > 0$ 의 조건하에 $\{x_n\}$에 대하여 $\Pr\{\lim_{\to\infty} x_n = \theta\} = 1$ 이 성립한다.

아울러 그는 1965년의 논문에서 상기정리를 다차원의 경우로 확장하였다.
이상의 정리에 의하여 w.p.1(with pnobability 1)의 수렴이 보증되는 Robbins-Monro algorithm $x_{n+1} = x_n - a_n(y(x_n) - \alpha)$는 $y(x_n)$이 확률변수 $Y(x_n)$의 실현치이므로, 1987년까지 이 Robbins-Monro 확률근사법이 소위 확률적인 잡음을 수반하는 경우라고는 생각되지 않았다.

$M(x) = \sin x - \frac{x}{2} = 0$ 의 수치해법으로서, 또 방정식 $x^2 - x - 2 = 0$, $x^3 = n$, $x^6 - x - 1 = 0$ 등의 근을 구하기 위한 효율적인 수치해법으로서 또한 복잡하게 보인다. 그러나 gamma분포의 미지 parameter α 를 구하기 위한 최우방정식(maximum likelihood equation)

$$M(\alpha) = \ln\alpha - \frac{d}{d\alpha}\ln[\Gamma(\alpha)] - \ln[\frac{1}{n}\sum_{i=1}^{n} x_i] + \frac{1}{n}\sum_{i=1}^{n}\ln x_i \tag{1}$$

에서 $M(\alpha) = 0$ 의 수치해법으로서 쓰일 것이라고는 누구도 생각 못하였다.

2.본론

그동안 통계적인 관측잡음이 개입되지 않는다고 생각되는 방정식 $M(x) = 0$의 수치해법 에는 주로 Newton-Raphsòn 근사법이 쓰여졌다. 그것은 방정식을 $f(x) = 0$로 쓸 때, x_0을 초기값으로 하고

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (n = 0, 1, 2. \ldots, n, \ldots) \tag{2}$$

으로 나타내진다.

가령 $f(x_n)>0$, $f'(x_n)<0$ 인 경우는, x_n은 $x_0<x_1<x_2<\ldots<\theta$ 근 θ에 단조수렴(monotone convergence)하고, 이 경우 소위 2차수렴 즉 $|x_{n+1}-\theta|\le k|x_n-\theta|^2$ 이 성립함으로 수렴속도는 대단히 빠르다. 그러나 이 Newton 법은 $f'(x_n)$의 계산이 필요함으로서 그만큼의 연산회수가 증가한다. 뿐만 아니라 계산도중 임의의 x_n 에서 $f'(x_n)$의 값이 0이 되었을 경우 더 이상의 연산을 수행할 수 없다. 또 $f(x)$가 알려진 경우에는 직접 미분을 하든지 아니면 충분히 작은 양수 ε 에 대하여 수치적으로

$$f'(x_n)\cong\frac{f(x_n+\varepsilon)-f(x_n-\varepsilon)}{2\varepsilon} \tag{3}$$

를 구하여 Newton 근사법을 사용한다. 그러나 $x_0<x_1<\ldots<\theta<\ldots<x_3<x_2$ 과 같이 θ의 상하에 옮겨가면서 수렴하는 경우의 수학적 취급은 곤란하며, 또 초기값을 어떻게 선정하느냐도 문제이다. 그래서 결정계의 방정식 $f(x)=0$ $(M(x)=0)$ 의 수치적 해법에도 RM 확률근사법을 쓰는 것을 생각하게 된다. 관측잡음이 부가되는데도 사용하는 그 이유는 결정계라해도 수치해법인 이상수치를 모두 유한자리의 수만으로 근사적으로 계산하고, 소수이하 d자리째의 수는 그 이하의 수를 rounding off계산하고 있는 것이다. 그 때문에, 절단오차(truncation errors)와 마무리 오차(raunding off errors)가 축차 축적된다. 그러므로 $M(x)$의 값이라해도, 참값이 아니고, $M(x_n)+Y(x_n)$의 꼴로 나타내지는 값을 써서 $M(x)$의 값을 계산하고 있는 것이다.
여기서 $Y(x_n)$은 random인 오차이다.

그러므로
$M^*(x_n)=M(x_n)+Y(x_n)$이라고 쓰면 R-M법의

$$\begin{aligned} x_{n+1} &= x_n-a_n\{M^*(x_n)-0\} \\ &= x_n-a_nM^*(x_n) \end{aligned}$$

즉,

$$x_{n+1}=x_n-a_nM(x_n) \tag{4}$$

더욱 이경우 계산상의 오차이외에 Noise는 들어가지 않으므로 자리수를 소수점이하 충분히 크게 잡아서 계산한다면(예컨대 $10^{-d}; d=8,\ d=10$), $\sigma^2(Y(x_n))\le\sigma^2$ 은 $M(x_n)$에 대하여 무시될 정도작다.

$f(x)=0$의 근을 RM-법으로 구하려는 경우, $f(x)$에 대한 조건은 $f'(x)$가 존재하고, 또한 연속이어야한다.

다음 계수 a_n의 선정이 매우 중요하다.

그 선정법(특히 a_n의 크기의 order)을 잘못하면 실패한다. 또 a_n을 적절히 선정하면 Newton의 근사법과 같이 매회 $f'(x)$의 계산을 필요로 하지 않으나, RM -법이므로 오히려 적은연산회수와 적은 반복회수로 효율적으로 근 θ의 값을 충분한 精度로 구할수 있다. 과거에는 $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$ 을 만족시키는 $\{a_n\}$으로서

$$a_n = \frac{c'}{n} \quad (c'\text{은 적절한 크기의 상수}) \tag{5}$$

이 쓰여졌으나, 이것으로는 a_n에 의한 수정의 gain이 바로 떨어져서, 반복회수를 아무리 크게 해도 θ의 좋은 근사값은 구해지지 않는다. 일반적으로 $x_2 - x_1$, $x_3 - x_2$, · · · , $x_{n+1} - x_n$ 로 이어지는 차($\triangle_n = x_{n+1} - x_n$)의 부호 $sgn(\triangle_n)$이 변하지 않으면 a_n을 계속 사용한다. 그렇게해도 $w.p.1$의 수렴이 수량적으로 증명된다.[1] 그러나 이것 역시 불충분하다.

이제 우리는 확률적 잡음을 수반하는 경우 $f(x)=0$의 근을 효율적으로 구하기 위하여 (5)식을 계량하여 다음과 같은 a_n을 제안하고자 한다.

$$a_n = \frac{c'}{c'' + n} \quad (c',\ c''\text{은 2개의 양의 상수}) \tag{6}$$

이것은 물론 만일 c'과 c''이 충분히 큰 상수이면 n이 바뀌어도 a_n은 거의 같은 값을 취해 나아간다. 그러므로 (6)식이 (5)식 보다도 더욱 일반적이고, 아울러 $\sum_{n=1}^{\infty} a_n = \infty$, $\sum_{n=1}^{\infty} a_n^2 < \infty$ 이란 조건은 물론 만족되어, 수정의 gain은 크고, RM-법은 효율적이다. 또 $f(x)$가 근 θ의 근방에서 단조증가이면 a_n 앞의 부호를 음으로, 단조감소이면 양으로 해야하는데, 이와같은 구간안에서 제1근사값 x_1을 취할때 a_n의 값을($w.p.1$로 수렴할 근 θ에) 초기 몇번 반복단계만으로 효율을 높이기 위하여는 근 θ의 되도록 가까운 근사값 추정치 $\bar{\theta}$ 의 근방에서의 $f'(\bar{\theta})$의 역수를 수치적으로 구하여 이 값을 $\{a_n\}$으로 사용하는 것이 유효하다.

(예제1) $\sin x - \frac{x}{2} = 0$의 근 θ를 구하여 보자.

풀이.

$$M(x) = \sin x - \frac{x}{2} \quad \text{로놓으면}$$

$$M\left(\frac{\pi}{2}\right) = 0.215,\ \ M(1.7) = 0.142$$

$M(1.8)=0.074,\ M(1.9)=-0.004$

$M(1.89)=0.004$

$d(x)=\frac{d}{dx}M(x)=\cos x-\frac{1}{2}$

$\frac{1}{d(1.89)}=-1.229 \fallingdotseq -1.23$

$a_n=1.23\frac{100}{100+n}=\frac{123}{100+n}$ 을 써서 다음결과를 얻는다.

n	x_n	$M(x_n)$
1	1.9	-0.004
2	1.89549417	8.259×10^{-8}
3	1.89549427	1.02×10^{-9}
4	1.89549427	2.239×10^{-11}
5	1.89549427	7.02×10^{-13}
6	1.89549427	2.853×10^{-14}
7	1.89549427	1.443×10^{-15}
8	1.89549427	0

$x_1=1.9,\qquad x_{n+1}=x_n+a_nM(x_n)$

(예제2) $x_1=16.2$, $x_2=20.4$, $x_3=26.2$일 때 Gamma 분포의 미지 parameter α 의 최우추정법

$$M(\alpha)=\ln\alpha-\frac{d}{d\alpha}\ln[\Gamma(\alpha)]-\ln[\frac{1}{n}\sum_{i=1}^{n}x_i]+\frac{1}{n}\sum_{i=1}^{n}\ln x_i=0$$ 의 해를 구하여 보자.

풀이.

$M(20)=0.006\qquad M(25)=8.924\times10^{-4}$

$M(26)=1.131\times10^{-4}\qquad M(26.2)=-3.558\times10^{-5}$

$d(\alpha)=\frac{d}{d\alpha}M(\alpha)\qquad \frac{1}{d(26.2)}=-1.356\times10^{3}$

$a_j=-1356\frac{100}{100+j}=-\frac{135600}{100+j}$

$\alpha_1=26.2\qquad \alpha_{j+1}=\alpha_j-a_jM(\alpha_j)$

j	a_j	$M(a_j)$
1	26.2	-3.558×10^{-5}
2	26.152228	-2.781×10^{-7}
3	26.151858	-4.372×10^{-9}
4	26.151852	-1.105×10^{-10}
5	26.151852	-4.216×10^{-12}
6	26.151852	1.75×10^{-13}

(예제3) 연립방정식 $M(x, y) = x^3 + 2y^2 - 1 = 0$　(1)

$K(x, y) = 5y^3 + x^2 - 2xy - 4 = 0$　(2) 의 해를 구하여 보자.

풀이. (1)로부터

$$y(x) = \frac{1}{\sqrt{2}}\sqrt{1 - x^3}$$

를 (2)에 대입하면

$$f(x) = 5\{y(x)\}^3 + x^2 - 2xy(x) - 4 = 0$$

이다. 1차원 R-M 방법을 사용하면

$d(x) = \frac{d}{dx} f(x)$　$\frac{1}{d(-0.5)} = -0.206$

$x_1 = 0$　$x_{j+1} = x_j + 0.2f(x_j)$　$j = 1, \cdot\cdot\cdot, 10$

$x_{10} = -0.6493337$　$\frac{1}{d(-0.65)} = -0.139$

$x_1 = -0.5$　$x_{n+1} = x_n + 0.2f(x_n)$　$n = 1, \cdot\cdot\cdot, 10$

$x_5 = -0.6494416$　$y(x_5) = 0.798087$　$f(x_5) = -2.143 \times 10^7$

$x_{10} = -0.64941597$　$y(x_{10}) = 0.798086$　$f(x_{10}) = 0.$

3. 결론

상기 예에서 입증되었듯이 제안된 계수 $a_n = \frac{c'}{c'' + n}$ (c', c''은 2개의 양의 상수)의 사용은 적은 반복횟수에도 정도높은 근사값을 구할 수 있어 효율성 면에서 진일보하였다고 할 수 있다.

참고문헌

[1] H. Kesten, "Accelerated Stachastic Approxination" Annals of Mathematical Statistics 29, 1958

[2] Richard L. Burden, Numerical Analysis, Thomson Information Publishing Group, 1993

[3] Robert J. Serfling, Approximation Theorem of Mathematical Statistics, John Wiley & sons 1980

[4] Shoichiro Nakamura, Applied Numerical Methods With Software, Prentice - Hall International Editions, 1991

[5] S. SWierczkowski, Stochastic Stability of Differential Equation Sijthoff & Noordhoff 1980