STOCHASTIC APPROXIMATION

(OUTLINE OF PRESENTATION)

MOTIVATION

HISTORICAL PAPERS

BASIC RESULTS:CONVERGENCE ASYMPTOTIC NORMALITY

CONNECTIONS WITH MORE GENRAL THEOREMS &CENTRAL LIMIT THEOREMS

ESTIMATION IDENTIFICATION

MOTIVATION

$f : R \to R,$ to solve $f(x) = 0, x_{n+1} = x_n - a_n f(x_n), x_{n=1} = x_n - a_n \dfrac{f(x_n)}{f^{'}(x_n)}$

Similar methods for solving $\min f(x), a_n$ is the stepsize.

In all these methods as they appear in standard texts and programms , the function f is assumed known, or at least that the values needed are known and the stepsize is chosen by some rule and in the limit it is usually constant. But in reality the values of f are known approximately and one can consider that the supplied value of f is equal to the true value plus noise or more generally that it is the outcome of a random experiment where the underlying random variable has hopefully as mean the true value and a very small variance.

ROBBINS-MONRO & KIEFER-WOLFOWITZ

This is the motivation of the Robbins –Monro (for solving f(x)=0) and of Kiefer-Wolfowitz (for minf(x)).

The results give conditions for convergence (some type of stochastic convergence concept) and also study rapidity of convergence and asymptotic normality. The choice of the sequence of the stepsize α_n and it's behaviour with as n increases is a special feature that we will motivate separately.

Let us consider that we want to solve the simple equation:f(x)=cx=0,

And we use the algorithm:

$$x_{n+1} = x_n - a_n f(x_n) = x_n - a_n c x_n$$

In order to converge to the solution x*=0,we impose the assumption $|1\text{-}a_n c\ | < 1-\varepsilon, \Rightarrow \varepsilon < a_n c < 2-\varepsilon,$ for some small positive epsilon.Let c=1,w.l.o.g.We have: $\varepsilon<\alpha_{\mathrm{n}}<2\text{-}\varepsilon$.We usually take α_{n} to be constant,and small.

But it is not necessary to be constant although it has to be small!What we really need is that it holds:

$$\lim \prod_{k=0}^{n}(1-a_k) = 0, \text{as } n \to +\infty \ \text{,since } x_{n+1} = \prod_{k=o}^{n}(1-a_k)x_o \to 0, \text{as n} \to \infty.$$

We thus see that the stepsize does not need to be constant.For example $a_n = 1/n$ is OK, whereas $1/n^2$ is not.Notice that

$$\sum_{\mathrm{k=o}}^{\mathrm{n}} a_k \to +\infty \qquad \text{IMPORTANT CONDITION} \quad (1)$$

imlies the above condition.This can also be justified as follows:The ODE $\frac{dx}{dt} = f(x)$ where x(t) $\to$ 0,as $t \to \infty$, approximated by $\frac{x_{n+1} - x_n}{a_n} = f(x_n)$, has $x(t_n) \approx x_{n+1}$, where $t_n = a_0 + a_1 + ... + a_n$ which clearly has to go to infinity!

This is a discussion based on the deterministic case!The stochastic considerations next.

Let us assume that in our basic formula we learn the value of f(x) with an error:

$x_{n+1} = x_n - a_n(f(x_n) + w_n) = x_n - a_n(cx_n + w_n)$, the w_n's are iid Gaussian,N(0,1).

Since now the x's are (gaussian) random variables we want them to go to zero in a stochastic sense.Let us use the m.s convergence,For the mean $E(x_n) = \mu_{n,}$ we have the previous analysis which needs the "important condition".

$\mathrm{E(x_{n+1})}^2 = E\{(1-a_n c)x_n - a_n w_n\}^2 = (1-a_n c)^2 E(x_n^2) + a_n E(w_n^2) - 2(1-a_n c)a_n E(x_n w_n) = (1-a_n c)^2 E(x_n^2) + a_n^2$

since the w_n is independent of the x_n, the product $E(x_n w_n)$ is zero.Let $E(x_n^2) = \sigma_n$.It holds:

$$\sigma_{n+1} = (1-a_n c)^2 \sigma_n + a_n^2$$

In order vto have that $\sigma_n \to 0,$ we need $a_n \to 0,$ and

$$\lim \sum_{\mathrm{k=0}}^{\mathrm{n}} a_n^2 < \infty \quad \text{,as } n \to \infty, \quad \text{IMPORTANT CONDITION} \quad (2)$$

The contribution of the w_0 ,$w_1, w_2, .., w_n$ to the variance of x_{n+1} is cumulative and equal approximately to:

$$a_0^2 E(w_0^2) + a_1^2 E(w_1^2) + \ldots + a_n^2 E(w_n^2) = a_0^2 + a_1^2 + a_2^2 + \ldots a_n^2 = \sum_{k=0}^{n} a_k^2$$

NOTICE THAT THE SEQUENCE $a_n = \frac{1}{\mathrm{n}}$ SATISFIES BOTH CONDITIONS 1&2

ROBBINS-MONRO

We want to solve the equation :

$M(x) = 0,$

where:

$$M(x) = E[y/x] = \int_{-\infty}^{+\infty} y\,dH(y/x), |y| \le C, 0 < C < +\infty$$

$$x_{n+1} = x_n - a_n y_n$$

I would like to use $x_{n+1} = x_n - a_n M(x_n)$ but M is not known exactly!The reason is that either $H(y/x)$ is not known or the integral which gives it is very hard to calculate. Instead I know y_n, ie given x_n , an experiment is performed with probability distribution $H(y/x_n)$ and an y_n is produced and given to me.

(We go now to material from:

1.the Robbins -Monro paper,2.Wasan's book and 3.Ljiung &Soderstrom for the Least Squares Application.)

KIEFER-WOLFOWITZ

$f : R \to R, \min f(x), x_{n+1} = x_n - a_n \nabla f(x_n)$

This is Steepest Descent. But If f cannot be calculated exactly we can use a finite difference approximation:

$$x_{n+1} = x_n - a_n \frac{f(x_n + c_n) - f(x_n - c_n)}{2c_n}$$

$f(x_n + c_n), f(x_n - c_n)$ are calculated with errors , respectively as y_n, y_{n-1},

ie y_n is a random variable with distribution: $F(y / x_n + c_n)$

and y_{n-1} is a random variable with distribution: $F(y / x_n - c_n$. We thus need to study the convergence of:

$$x_{n+1} = x_n - a_n \frac{y_n - y_{n-1}}{2c_n}$$

KIEFER-WOLFOWITZ PAPER :We need to solve :

$$\max \; M(X), M(x) = E[y/x] = \int_{-\infty}^{+\infty} y dH(y/x), Var(y/x) \le S < +\infty$$

$$Use: \quad z_{n+1} = z_n + a_n \frac{y_{2n} - y_{2n-1}}{c_n},$$

y_{2n-1} has prob.distribution $H(y / z_n - c_n),, y_{2n}$ has prob.distribution $H(y / z_n + c_n)$

(We go now to material from the Kiefer-Wolfowitz paper and Wasan's book.)

Material from:

1.H.Robbins&S.Monro, “A Stochastic Approximation Method”

Ann.Math.Stat.Vol.22(1951),pp.400-407.

2.J.Kiefer&J.Wolfowitz, “Stochastic Estimation of the Maximum of a RegressionFunction”Ibid,Vol.23,1952.pp462-466.

3.Ljung &Soderstrom”Theory&practice of Rec.Indentification”,MIT Press,1983,

Sections:2.1,2.2.1,2.4.1,2.4.4,2.4.5,2.4.6.

4.M.T.Wasan,”Stochastic Approximation” Cambridge Univ.Press,1969,pp.1-29,36-41,95-103.

HOMEWORK

Let $y = xw + x^2w^2 - 1,$ where w is a Gaussian random variable with $E(w) = 0$ and $E(w^2) = \sigma^2 = 1.$

1.$M_1(x) = E[y/x] = x^2 - 1$

Try to solve $M(x) = 0$,using the Robbins-Monro method. Experiment with several stepsizes and initial conditions.

2.$M_2(x) = E[y/x] = -x^2 + x^4 E(w^4) + 1$

Try to solve $\min M_2(x)$, by using the Kiefer-Wolfowitz method.Experiment with several stepsizes and initial conditions.

Notice that for both cases the conditions of the papers do not exactly hold,but this should not discourage you from using them.