Outline

1. The Robbins-Monro algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

2. The Kiefer-Wolfowitz algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

3. Acceleration by averaging
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.
Outline

1. The Robbins-Monro algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

2. The Kiefer-Wolfowitz algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

3. Acceleration by averaging
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.
Stochastic approximation.

Herbert Robbins
Let $f$ be an **unknown function** from $\mathbb{R}^d$ to $\mathbb{R}^d$.

**Goal**

For a **given vector** $\alpha$ of $\mathbb{R}^d$, find a vector $x^*$ which satisfies

$$f(x^*) = \alpha.$$ 

We will assume in all the sequel that for all $n \geq 1$, we can compute $X_1, \ldots, X_n$ of $\mathbb{R}^d$ and we can find $Y_{n+1}$ of $\mathbb{R}^d$ such that

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = f(X_n)$$

where $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. 
Stochastic approximation for $d = 1$.

Goal

→ Find the value $x^*$ with very few knowledge on $f$. 

The Robbins-Monro algorithm

Introduction.
The Robbins-Monro algorithm

Introduction.

Stochastic approximation.

Basic Idea

If you are able to say that \( f(X_n) > \alpha \), then increase the value of \( X_n \).
The Robbins-Monro algorithm

Introduction.

Stochastic approximation.

Basic Idea

If you are able to say that \( f(X_n) < \alpha \), then decrease the value of \( X_n \).
Outline

1. The Robbins-Monro algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

2. The Kiefer-Wolfowitz algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

3. Acceleration by averaging
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.
Let \((\gamma_n)\) be a sequence of positive real numbers decreasing to zero

\[
\sum_{n=1}^{\infty} \gamma_n = +\infty \quad \text{and} \quad \sum_{n=1}^{\infty} \gamma_n^2 < +\infty.
\]

For the sake of simplicity, we shall make use of

\[
\gamma_n = \frac{1}{n}.
\]

The Robbins-Monro algorithm

\[
X_{n+1} = X_n + \gamma_n \left( Y_{n+1} - \alpha \right)
\]

where the initial state \(X_0\) is a square integrable random vector of \(\mathbb{R}^d\) which can be arbitrarily chosen.
Almost sure convergence.

Let $g$ be the positive function defined on $\mathbb{R}^d$ by

$$g(X_n) = \mathbb{E}[\|Y_{n+1}\|^2|\mathcal{F}_n].$$

**Theorem (Robbins-Monro)**

Assume that the function $f$ is continuous from $\mathbb{R}^d$ to $\mathbb{R}^d$ such that $f(x^*) = \alpha$, and for all $x$ different from $x^*$,

$$\langle x - x^*, f(x) - \alpha \rangle < 0.$$

Assume that for $K > 0$ and for all $x \in \mathbb{R}^d$,

$$g(x) \leq K(1 + \|x\|^2).$$

Then, we have the almost sure convergence

$$\lim_{n \to \infty} X_n = x^* \quad \text{a.s.}$$
Proof of the almost sure convergence.

Proof.

First of all, denote

\[ V_n = \|X_n - x^*\|^2. \]

For all \( n \geq 0 \), we clearly have

\[
V_{n+1} = \|X_{n+1} - x^*\|^2,
\]

\[
= \|X_n + \gamma_n(Y_{n+1} - \alpha) - x^*\|^2,
\]

\[
= \|X_n - x^*\|^2 + 2\gamma_n \langle X_n - x^*, Y_{n+1} - \alpha \rangle + \gamma_n^2 \|Y_{n+1} - \alpha\|^2,
\]

which leads to

\[
V_{n+1} = V_n + \gamma_n^2 \|Y_{n+1} - \alpha\|^2 + 2\gamma_n \langle X_n - x^*, f(X_n) + \varepsilon_{n+1} - \alpha \rangle
\]

where \( \varepsilon_{n+1} = Y_{n+1} - f(X_n) \).
Proof.

Since $\mathbb{E}[Y_{n+1} | \mathcal{F}_n] = f(X_n)$, $\mathbb{E}[\varepsilon_{n+1} | \mathcal{F}_n] = 0$. It means that $(\varepsilon_n)$ is a martingale difference sequence. Consequently,

$$\mathbb{E}[V_{n+1} | \mathcal{F}_n] = V_n + \gamma_n^2 \mathbb{E}[\|Y_{n+1} - \alpha\|^2 | \mathcal{F}_n] - B_n$$

where $(B_n)$ is the positive sequence given by

$$B_n = -2\gamma_n \langle X_n - x^*, f(X_n) - \alpha \rangle$$

Moreover,

$$\mathbb{E}[\|Y_{n+1}\|^2 | \mathcal{F}_n] \leq K(1 + \|X_n\|^2) \leq L(1 + V_n)$$

where $L = 2K(1 + \|x^*\|^2)$. 

Proof.

Therefore, we obtain that

\[ \mathbb{E}[V_{n+1} | \mathcal{F}_n] \leq V_n(1 + a_n) + A_n - B_n \]

where \( a_n = 2L\gamma_n^2 \) and \( A_n = 2(L + \|\alpha\|^2)\gamma_n^2 \). The assumption

\[ \sum_{n=1}^{\infty} \gamma_n^2 < +\infty \]

clearly implies that

\[ \sum_{n=1}^{\infty} a_n < +\infty \quad \text{and} \quad \sum_{n=1}^{\infty} A_n < +\infty \quad \text{a.s.} \]
Proof.

Hence, it follows from Robbins-Siegmund theorem that \((V_n)\) converges a.s. to a random variable \(V_\infty\) and

\[
\sum_{n=1}^{\infty} B_n < +\infty \quad \text{a.s.}
\]

It remains to prove that \(V_\infty = 0\). Assume by contradiction that \(V_\infty > 0\). Then, we can find two finite constants \(0 < a < b\) such that, for \(n\) large enough, \(a \leq \|X_n - x^*\| \leq b\). Denote by \(\Delta\) the annulus of \(\mathbb{R}^d\),

\[
\Delta = \left\{ x \in \mathbb{R}^d \text{ such that } a \leq \|x - x^*\| \leq b \right\}.
\]

Let \(F\) be the continuous negative function defined, for all \(x \in \mathbb{R}^d\), by

\[
F(x) = \langle x - x^*, f(x) - \alpha \rangle.
\]
Proof.

One can find $c > 0$ such that, for all $x \in \Delta$,

$$F(x) \leq -c.$$

However, for $n$ large enough $X_n \in \Delta$, which implies that $F(X_n) \leq -c$. Consequently, for $n$ large enough,

$$B_n = -2\gamma_n F(X_n) \geq 2c\gamma_n.$$

Finally, the assumption

$$\sum_{n=1}^{\infty} \gamma_n = +\infty \quad \Rightarrow \quad \sum_{n=1}^{\infty} B_n = +\infty$$

leading to a contradiction. It means that $V_\infty = 0$ so $X_n \rightarrow x^*$ a.s.
Outline

1. The Robbins-Monro algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

2. The Kiefer-Wolfowitz algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

3. Acceleration by averaging
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.
Asymptotic normality.

The asymptotic normality requires more assumption on the function $f$. We now assume that $f$ is twice differentiable. It follows from Taylor’s formula that

$$f(x) = \alpha + H(x - x^*) + O(||x - x^*||^2)$$

where $H$ is the Jacobian matrix of $f$ at $x^*$. We also assume that $H$ is an Hurwitz matrix. It means that the real parts of all the eigenvalues of $H$ are negative. Let $\lambda_{max}(H)$ be the eigenvalue of $H$ with the largest real part and denote

$$\rho = -\text{Re}(\lambda_{max}(H)).$$

In dimension $d = 1$, we have $f(x^*) = \alpha$, $H = f'(x^*)$ and $\rho = -f'(x^*)$. 
Asymptotic normality, continued.

Let \((\varepsilon_n)\) be the **martingale difference sequence** given by

\[ \varepsilon_{n+1} = Y_{n+1} - \mathbb{E}[Y_{n+1} | \mathcal{F}_n] = Y_{n+1} - f(X_n). \]

**Theorem (Robbins-Monro, continued)**

Assume that the function \(f\) is twice differentiable from \(\mathbb{R}^d\) to \(\mathbb{R}^d\) such that \(f(x^*) = \alpha\). Suppose that \(f\) and \(g\) satisfy the same assumptions as in Robbins-Monro Theorem. Moreover, assume that

\[
\lim_{n \to \infty} \mathbb{E}[\varepsilon_{n+1} \varepsilon_{n+1}^T | \mathcal{F}_n] = \Gamma \quad \text{a.s.}
\]

where \(\Gamma\) is a symmetric definite positive matrix and that \((\varepsilon_n)\) has a conditional moment of order \(> 2\). If \(\rho > 1/2\), we have the asymptotic normality

\[
\sqrt{n}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma).
\]
Asymptotic normality for $d = 1$.

The limiting covariance matrix $\Sigma$ is the unique solution of the **Lyapunov equation**

$$
(H + \frac{1}{2} I_d) \Sigma + \Sigma (H^T + \frac{1}{2} I_d) = -\Gamma.
$$

It is quite complicated to evaluate $\Sigma$. However, in the special case $d = 1$ and $\Gamma = \sigma^2$, we have $\rho = -H = -f'(x^*)$,

$$
\Sigma = \frac{\sigma^2}{2\rho - 1}.
$$

Consequently, as soon as $\rho > 1/2$, we have

$$
\sqrt{n}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{\sigma^2}{2\rho - 1}).
$$
Outline

1 The Robbins-Monro algorithm
   • Introduction.
   • Almost sure convergence.
   • Asymptotic normality.

2 The Kiefer-Wolfowitz algorithm
   • Introduction.
   • Almost sure convergence.
   • Asymptotic normality.

3 Acceleration by averaging
   • Introduction.
   • Almost sure convergence.
   • Asymptotic normality.
Stochastic approximation.
Let $f$ be an **unknown differentiable function** from $\mathbb{R}^d$ to $\mathbb{R}$.

**Goal**

→ Find a vector $x^*$ of $\mathbb{R}^d$ which satisfies

$$\nabla f(x^*) = 0.$$ 

We will make use of the **directional derivative** of $f$ at $x \in \mathbb{R}^d$ along the vector $y \in \mathbb{R}^d$, given by

$$\langle \nabla f(x), y \rangle = \lim_{t \to 0} \frac{f(x + ty) - f(x - ty)}{2t}. $$
Stochastic approximation.

We will assume in all the sequel that for all \( n \geq 1 \), we can compute \( X_1, \ldots, X_n \) of \( \mathbb{R}^d \) and we can find \( Y_{n+1} \) and \( Z_{n+1} \) of \( \mathbb{R}^d \) such that

\[
\mathbb{E}[Y_{n+1} | \mathcal{F}_n] = \begin{pmatrix} f(X_n + c_n e_1) \\ \vdots \\ f(X_n + c_n e_d) \end{pmatrix} = \Phi(X_n)
\]

and

\[
\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = \begin{pmatrix} f(X_n - c_n e_1) \\ \vdots \\ f(X_n - c_n e_d) \end{pmatrix} = \Psi(X_n)
\]

where \((e_1, \ldots, e_d)\) is the canonical basis of \( \mathbb{R}^d \), \((c_n)\) is a sequence of positive real numbers decreasing to zero, and \( \mathcal{F}_n = \sigma(X_1, \ldots, X_n) \). In dimension \( d = 1 \), \( \mathbb{E}[Y_{n+1} | \mathcal{F}_n] = f(X_n + c_n) \), \( \mathbb{E}[Z_{n+1} | \mathcal{F}_n] = f(X_n - c_n) \).
Outline

1. The Robbins-Monro algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

2. The Kiefer-Wolfowitz algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

3. Acceleration by averaging
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.
The Kiefer-Wolfowitz algorithm

Let \((\gamma_n)\) be a sequence of positive real numbers decreasing to zero

\[
\sum_{n=1}^{\infty} \gamma_n = +\infty, \quad \sum_{n=1}^{\infty} \left(\frac{\gamma_n}{c_n}\right)^2 < +\infty, \quad \sum_{n=1}^{\infty} \gamma_n c_n < +\infty.
\]

For the sake of simplicity, we can choose \(0 < c < 1/2\),

\[
\gamma_n = \frac{1}{n} \quad \text{and} \quad c_n = \frac{1}{n^c}.
\]

The Kiefer-Wolfowitz algorithm

\[
X_{n+1} = X_n + \frac{\gamma_n}{2c_n} \left( Y_{n+1} - Z_{n+1} \right)
\]

where the initial state \(X_0\) is a square integrable random vector of \(\mathbb{R}^d\) which can be arbitrarily chosen.
Let $g$ and $h$ be the two positive functions defined on $\mathbb{R}^d$ by

$$g(X_n) = \mathbb{E}[\| Y_{n+1} \|^2 | \mathcal{F}_n] \quad \text{and} \quad h(X_n) = \mathbb{E}[\| Z_{n+1} \|^2 | \mathcal{F}_n].$$

**Theorem (Kiefer-Wolfowitz)**

Assume that the function $f$ is twice continuously differentiable from $\mathbb{R}^d$ to $\mathbb{R}$ such that $\nabla f(x^*) = 0$, and for all $x$ different from $x^*$,

$$\langle x - x^*, \nabla f(x) \rangle < 0.$$ 

Assume that for $L > 0$ and for all $x \in \mathbb{R}^d$,

$$\| \nabla^2 f(x) \| \leq L(1 + \| x \|).$$
Theorem (Kiefer-Wolfowitz, continued)

Moreover, assume that for $K_g > 0$, $K_h > 0$ and for all $x \in \mathbb{R}^d$,

\[ g(x) \leq K_g(1 + \|x\|^2) \quad \text{and} \quad h(x) \leq K_h(1 + \|x\|^2). \]

Then, we have the almost sure convergence

\[ \lim_{n \to \infty} X_n = x^* \quad \text{a.s.} \]
The Kiefer-Wolfowitz algorithm

1. Introduction.
2. Almost sure convergence.
3. Asymptotic normality.

The Robbins-Monro algorithm

1. Introduction.
2. Almost sure convergence.
3. Asymptotic normality.

Acceleration by averaging

1. Introduction.
2. Almost sure convergence.
3. Asymptotic normality.
The asymptotic normality requires more assumption on the function $f$. We now assume that $f \in C^3(\mathbb{R}^d)$ with $\nabla f(x^*) = 0$. As $\nabla f \in C^2(\mathbb{R}^d)$, it follows from Taylor’s formula that

$$\nabla f(x) = H(x - x^*) + O(||x - x^*||^2)$$

where $H = \nabla^2 f(x^*)$ is the Hessian matrix of $f$ at $x^*$. We also assume that $H$ is a negative definite matrix. It means that all the eigenvalues of $H$ are negative. Denote

$$\rho = -\lambda_{max}(H).$$

In dimension $d = 1$, we have $f'(x^*) = 0$, $H = f''(x^*)$ and $\rho = -f''(x^*)$. 
Asymptotic normality, continued

Let \((\varepsilon_n)\) and \((\xi_n)\) be the **martingale difference sequences** given by

\[
\begin{align*}
\varepsilon_{n+1} &= Y_{n+1} - \mathbb{E}[Y_{n+1}|\mathcal{F}_n] = Y_{n+1} - \Phi(X_n), \\
\xi_{n+1} &= Z_{n+1} - \mathbb{E}[Z_{n+1}|\mathcal{F}_n] = Z_{n+1} - \Psi(X_n).
\end{align*}
\]

**Theorem (Kiefer-Wolfowitz, continued)**

Assume that the function \(f \in C^3(\mathbb{R}^d)\) such that \(\nabla f(x^*) = 0\). Suppose that \(f\) and \(g\) satisfy the same assumptions as in Kiefer-Wolfowitz Theorem. Moreover, assume that

\[
\lim_{n \to \infty} \mathbb{E}[\varepsilon_{n+1}\varepsilon^T_{n+1}|\mathcal{F}_n] = \Gamma_g \quad \text{a.s.}
\]

\[
\lim_{n \to \infty} \mathbb{E}[\xi_{n+1}\xi^T_{n+1}|\mathcal{F}_n] = \Gamma_h \quad \text{a.s.}
\]

where \(\Gamma_g\) and \(\Gamma_h\) are symmetric definite positive matrices and that \((\varepsilon_n)\) and \((\xi_n)\) have conditional moments of order \(> 2\).
Asymptotic normality, continued

Theorem (Kiefer-Wolfowitz, continued)

If $\rho > 2c$ where $1/6 < c < 1/2$, we have the asymptotic normality

$$
\sqrt{nc_n^2}(X_n - x^*) \overset{\mathcal{L}}{\to} \mathcal{N}(0, \Sigma).
$$

In the special case $\rho > 2c$ with $c = 1/6$, we also have

$$
n^{1/3}(X_n - x^*) \overset{\mathcal{L}}{\to} \mathcal{N}(m, \Sigma)
$$

where the mean $m$ can be explicitly calculated.
The Kiefer-Wolfowitz algorithm

Asymptotic normality.

Asymptotic normality for $d = 1$.

The limiting covariance matrix $\Sigma$ is the unique solution of the **Lyapunov equation**

$$
\left( H + \left( \frac{1}{2} - c \right) l_d \right) \Sigma + \Sigma \left( H + \left( \frac{1}{2} - c \right) l_d \right) = -\frac{1}{4} \Gamma.
$$

It is quite complicated to evaluate $\Sigma$. However, in the special case $d = 1$ and $\Gamma = \sigma^2$, we have

$$
\Sigma = \frac{\sigma^2}{8(\rho + c - 1/2)}.
$$

Consequently, as soon as $\rho > 2c$ where $1/6 < c < 1/2$, we have

$$
\sqrt{nc_n^2(X_n - x^*)} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{8(\rho + c - 1/2)}\right).
$$
Outline

1. The Robbins-Monro algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

2. The Kiefer-Wolfowitz algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

3. Acceleration by averaging
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.
Stochastic averaging.

David Ruppert

Boris Polyak
Stochastic averaging.

The idea of using averaging to accelerate stochastic algorithms in due to Poliak and Ruppert. It consists in introducing a Cesaro mean over the iterations of the original stochastic algorithm

\[
\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k.
\]

Goal

→ Improve the convergence properties of the stochastic algorithm by minimizing its asymptotic variance.
→ Substantially weaken the conditions of the previous asymptotic results on the stochastic algorithm.
Consider the slow down Robbins-Monro algorithm given by

\[ X_{n+1} = X_n + \gamma_n \left( Y_{n+1} - \alpha \right) \]

where the initial state \( X_0 \) is a square integrable random vector of \( \mathbb{R}^d \) which can be arbitrarily chosen and the step

\[ \gamma_n = \frac{1}{n^c} \quad \text{with} \quad \frac{1}{2} < c < 1. \]

At time \( n \geq 1 \), compute de Cesaro mean

\[ \overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k. \]
A second-order recursive equation.

We have

\[(n + 1)X_{n+1} = \sum_{k=1}^{n} X_k + X_{n+1} = nX_n + X_{n+1},\]

which implies that

\[X_{n+1} = X_n + \frac{1}{n+1} \left( X_{n+1} - X_n \right).\]

However,

\[X_{n+1} = X_n + \gamma_n \left( Y_{n+1} - \alpha \right).\]
A second-order recursive equation.

Consequently, as \( X_n = n \bar{X}_n - (n - 1) \bar{X}_{n-1} \) we obtain that

\[
\bar{X}_{n+1} = \bar{X}_n + \frac{1}{n+1} \left( X_n - \bar{X}_n + \gamma_n \left( Y_{n+1} - \alpha \right) \right),
\]

\[
= \bar{X}_n + \frac{1}{n+1} \left( (n-1) \left( \bar{X}_n - \bar{X}_{n-1} \right) + \gamma_n \left( Y_{n+1} - \alpha \right) \right),
\]

\[
= \bar{X}_n + \left( \frac{n-1}{n+1} \right) \bar{X}_n - \left( \frac{n-1}{n+1} \right) \bar{X}_{n-1} + \frac{\gamma_n}{n+1} \left( Y_{n+1} - \alpha \right),
\]

leading to the second-order recursive equation

\[
\bar{X}_{n+1} = \left( \frac{2n}{n+1} \right) \bar{X}_n - \left( \frac{n-1}{n+1} \right) \bar{X}_{n-1} + \frac{\gamma_n}{n+1} \left( Y_{n+1} - \alpha \right).
\]
A second-order recursive equation.

Consequently, as $X_n = nX_n - (n-1)X_{n-1}$ we obtain that

$$X_{n+1} = X_n + \frac{1}{n+1} \left( X_n - X_n + \gamma_n(Y_{n+1} - \alpha) \right),$$

$$= X_n + \frac{1}{n+1} \left( (n-1) \left( X_n - X_{n-1} \right) + \gamma_n(Y_{n+1} - \alpha) \right),$$

$$= X_n + \left( \frac{n-1}{n+1} \right)X_n - \left( \frac{n-1}{n+1} \right)X_{n-1} + \frac{\gamma_n}{n+1}(Y_{n+1} - \alpha),$$

leading to the second-order recursive equation

$$X_{n+1} = \left( \frac{2n}{n+1} \right)X_n - \left( \frac{n-1}{n+1} \right)X_{n-1} + \frac{\gamma_n}{n+1}(Y_{n+1} - \alpha).$$
A second-order recursive equation.

Consequently, as $X_n = n\bar{X}_n - (n - 1)\bar{X}_{n-1}$ we obtain that

\[
\bar{X}_{n+1} = \bar{X}_n + \frac{1}{n+1} \left( X_n - \bar{X}_n + \gamma_n \left( Y_{n+1} - \alpha \right) \right),
\]

\[
= \bar{X}_n + \frac{1}{n+1} \left( (n-1) \left( \bar{X}_n - \bar{X}_{n-1} \right) + \gamma_n \left( Y_{n+1} - \alpha \right) \right),
\]

\[
= \bar{X}_n + \left( \frac{n-1}{n+1} \right) \bar{X}_n - \left( \frac{n-1}{n+1} \right) \bar{X}_{n-1} + \frac{\gamma_n}{n+1} \left( Y_{n+1} - \alpha \right),
\]

leading to the second-order recursive equation

\[
\bar{X}_{n+1} = \left( \frac{2n}{n+1} \right) \bar{X}_n - \left( \frac{n-1}{n+1} \right) \bar{X}_{n-1} + \frac{\gamma_n}{n+1} \left( Y_{n+1} - \alpha \right).
\]
Outline

1. The Robbins-Monro algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

2. The Kiefer-Wolfowitz algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

3. Acceleration by averaging
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.
Theorem (Robbins-Monro averaging)

Assume that the function $f$ is continuous from $\mathbb{R}^d$ to $\mathbb{R}^d$ such that $f(x^*) = \alpha$, and for all $x$ different from $x^*$,

$$\langle x - x^*, f(x) - \alpha \rangle < 0.$$

Assume that for $K > 0$ and for all $x \in \mathbb{R}^d$,

$$g(x) \leq K(1 + ||x||^2).$$

Then, we have the almost sure convergence

$$\lim_{n \to \infty} \overline{X}_n = x^* \quad \text{a.s.}$$
Outline

1. The Robbins-Monro algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

2. The Kiefer-Wolfowitz algorithm
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.

3. Acceleration by averaging
   - Introduction.
   - Almost sure convergence.
   - Asymptotic normality.
Assume that the function $f$ is twice differentiable from $\mathbb{R}^d$ to $\mathbb{R}^d$ such that $f(x^*) = \alpha$. Suppose that $f$ and $g$ satisfy the same assumptions as in Robbins-Monro Theorem. Moreover, assume that

$$\lim_{n \to \infty} \mathbb{E}[\varepsilon_{n+1}\varepsilon_{n+1}^T | \mathcal{F}_n] = \Gamma \quad \text{a.s.}$$

where $\Gamma$ is a symmetric definite positive matrix and that $(\varepsilon_n)$ has a conditional moment of order $> 2$. Then, we have the asymptotic normality

$$\sqrt{n}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

where the asymptotic matrix $\Sigma$ is given by

$$\Sigma = H^{-1}\Gamma(H^{-1})^T.$$
Asymptotic normality for $d = 1$.

It is not necessary to assume that

$$\rho = -\text{Re}(\lambda_{\text{max}}(H)) > \frac{1}{2}.$$ 

In the special case $d = 1$ and $\Gamma = \sigma^2$, we have $\rho = -H = -f'(x^*)$, which means that

$$\Sigma = \frac{\sigma^2}{(f'(x^*))^2}.$$ 

Consequently, the asymptotic normality reduces to

$$\sqrt{n}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{(f'(x^*))^2}\right).$$
Acceleration by averaging  
Asymptotic normality.