Probability and Random Processes

Stochastic Convergence

- Motivation
- Convergence of function sequences (deterministic)
- Convergence of random variables
- Some useful results for proving convergence
- Motivation continued

Motivation: Asymptotic Analysis

Making exact statistical inference often impossible. Remedy: *asymptotic analysis*.

Example: Suppose

$$x_n = A + w_n;\; n = 0, 1, \ldots, N-1$$

where $E\{w_n\} = 0$, but the distribution is unknown. An estimate of A is the sample mean $\hat{A} = \overline{x}$. Statistical properties? We have

$$\hat{A} = A + \frac{1}{N}\sum_{n=0}^{N-1} w_n$$

How does the noise average behave for large N? Convergence is a *random event*. We need a new theory!

Sequences of Functions

Formally, random variable x_n is a function $X_n(\omega)$, where $\omega \in \Omega$ defines the realization (outcome). Consider a sequence of such functions:

$$f_0(x), f_1(x), \ldots$$

- **Pointwise convergence:** $f_n(x) \to f(x)\ \forall x$. (except subset of measure zero $\to$ **Almost Everywhere**; a.e.)
- **Norm convergence:** $\|f_n - f\| \to 0$. Special cases:
  - **Mean-square convergence:**

$$\|f\| = \|f\|_2 = \sqrt{\int_x |f(x)|^2 dx}$$

- **Uniform convergence:**

$$\|f\| = \|f\|_\infty = \max_x |f(x)|$$

- **Convergence in measure:** $d(f_n, f) \to 0$, with "distance" function $d(g, h)$. Common example

$$d_\epsilon(g, h) = \int_{x:\ |g(x)-h(x)|>\epsilon} dx$$

If $f_n(x)$ converges or not depends on the chosen definition!

Convergence of Random Variables

Consider sequence x_n, corresponding to $X_n(\omega)$.

Pointwise convergence $\leftrightarrow$ convergence for all $\omega \leftrightarrow$ deterministic convergence - of little interest! Useful definitions:

- $x_n \to x$ **almost surely** (w.p.1) if $P(x_n \to x) = 1$. This means $X_n(\omega) \to X(\omega)$ a.e.
- $x_n \to x$ **in** $r\text{th}$ **mean** if $E\{|x_n|^r\} < \infty$ and $E\{|x_n - x|^r\} \to 0$. Special cases: $r = 1$ l.i.m.; $r = 2$ mean-square convergence
- $x_n \to x$ **in probability** if $P(|x_n - x| > \epsilon) \to 0 \; \forall \epsilon > 0$. Related to convergence in measure!
- $x_n \to x$ **in distribution** if $P(X_n \leq x) \to P(X \leq x)$ for all x where $P(X \leq x)$ is continuous. Also *weak* or *law* convergence.

Implications

$$\left.\begin{array}{ll} x_n \to x & \text{w.p.1} \\ x_n \to x & \text{in } r\text{th mean} \end{array}\right\} \Rightarrow x_n \to x \text{ in prob.} \Rightarrow \ x_n \to x \text{ in distr.}$$

If $r > s \geq 1$, then $x_n \to x$ in rth mean $\Rightarrow$ $x_n \to x$ in sth mean

Other implications generally false!

Interpretations

- w.p.1 conv. says how *realizations* behave
- mean-square conv. says how variance behaves
- conv. in prob. how "most realizations behave most of the time"
- conv. in distr. useful for statistical tests and confidence interval

Example:

$$x_n = \begin{cases} 1 & \text{w.p. } 1/n \\ 0 & \text{w.p. } 1-1/n \end{cases}$$

Then $x_n \to 0$ in probability, but not w.p.1.

Useful Results

- **Dominated Convergence:** Assume $x_n < z \, \forall \, n$, $E[z] < \infty$. Then $x_n \to x$ in probability implies $E\{|x_n - x|\} \to 0$
- **Generalized Markov** Let h be a non-negative fcn and $a \geq 0$. Then

$$P(h(x) \geq a) \leq \frac{E\{h(x)\}}{a}$$

- **Hölder:** Let $p, q > 1$ and $1/p + 1/q = 1$. Then

$$E|xy| \leq (E|x^p|)^{1/p} + (E|x^q|)^{1/q}$$

- **Minkovski:** Let $p \geq 1$. Then

$$(E|x+y|^p)^{1/p} \leq (E|x^p|)^{1/p} + (E|y^p|)^{1/p}$$

- **Borel-Cantelli:** Let A_n be an infinite sequence of events and define

$$A = \{\text{infinitely many } A_n \text{ occur}\}$$

Then

$$P(A) = 0 \quad \text{if } \textstyle\sum_n P(A_n) < \infty$$
$$P(A) = 1 \quad \text{if } \textstyle\sum_n P(A_n) = \infty \text{ and } A_n \text{ independent}$$

Asymptotic Analysis

Back to DC-in-noise example:

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x_n = \frac{1}{N}\sum_{n=0}^{N-1}(A + w_n) = A + \frac{1}{N}\sum_{n=0}^{N-1} w_n$$

If $E[w_n^2] < \infty$ and w_n independent ("weakly dependent"), then $E(\hat{A} - A)^2 \rightarrow 0$

Strong large of large numbers (later) shows $\hat{A} \rightarrow A$ w.p.1.

Central Limit Theorem (later) shows

$$\sqrt{N}(\hat{A} - A) \; \in \; \mathrm{AsN}(0, \sigma)$$

where $\sigma^2 = \sigma_w^2$ if w_n i.i.d.