Probability and Random Processes

Lecture 5

- Probability and random variables
- The law of large numbers

Why Measure Theoretic Probability?

- Stronger limit theorems
- Conditional probability/expectation
- Proper theory for continuous and mixed random variables

Probability Space

- A probability space is a measure space $(\Omega, \mathcal{A}, P)$
  - the sample space Ω is the 'universe,' i.e. the set of all possible outcomes
  - the event class $\mathcal{A}$ is a σ-algebra of measurable sets called events
  - the probability measure is a measure on events in $\mathcal{A}$ with the property $P(\Omega) = 1$

Interpretation

- A random experiment generates an outcome $\omega \in \Omega$
- For each $A \in \mathcal{A}$ either $\omega \in A$ or $\omega \notin A$
- An event A in $\mathcal{A}$ occurs if $\omega \in A$ with probability $P(A)$
  - since $\mathcal{A}$ is the σ-algebra of measurable sets, we are ensured that all 'reasonable' combinations of events and sequences of events are measurable, i.e., have probabilities

With Probability One

- An event $E \in \mathcal{A}$ occurs with probability one if $P(E) = 1$
  - almost everywhere, almost certainly, almost surely,...

Independence

- E and F in $\mathcal{A}$ are independent if $P(E \cap F) = P(E)P(F)$
- The events in a collection $A_1, \ldots, A_n$ are
  - pairwise independent if A_i and A_j are independent for $i \neq j$
  - mutually independent if for any $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

- An infinite collection is mutually independent if any finite subset of events is mutually independent
- 'mutually' $\Rightarrow$ 'pairwise' but not vice versa

Eventually and Infinitely Often

- A probability space $(\Omega, \mathcal{A}, P)$ and an infinite sequence of events $\{A_n\}$, define

$$\liminf A_n = \bigcup_{n=1}^{\infty} \left(\bigcap_{k=n}^{\infty} A_k \right), \quad \limsup A_n = \bigcap_{n=1}^{\infty} \left(\bigcup_{k=n}^{\infty} A_k \right)$$

- $\omega \in \liminf A_n$ iff there is an N such that $\omega \in A_n$ for all $n > N$, that is, the event $\liminf A_n$ occurs eventually,

$$\{A_n \text{ eventually}\}$$

- $\omega \in \limsup A_n$ iff for any N there is an $n > N$ such that $\omega \in A_n$, that is, the event $\limsup A_n$ occurs infinitely often

$$\{A_n \text{ i.o.}\}$$

Borel–Cantelli

- The Borel–Cantelli lemma: A probability space $(\Omega, \mathcal{A}, P)$ and an infinite sequence of events $\{A_n\}$
  1. if $\sum_n P(A_n) < \infty$, then

$$P(\{A_n \text{ i.o}\}) = 0$$

  2. if the events $\{A_n\}$ are mutually independent and $\sum_n P(A_n) = \infty$, then

$$P(\{A_n \text{ i.o}\}) = 1$$

Random Variables

- A probability space $(\Omega, \mathcal{A}, P)$. A real-valued function $X(\omega)$ on Ω is called a random variable if it's measurable w.r.t. $(\Omega, \mathcal{A})$
  - Recall: *measurable* $\Rightarrow X^{-1}(O) \in \mathcal{A}$ for any *open* $O \subset \mathbb{R}$

$\iff$ $X^{-1}(A) \in \mathcal{A}$ for any $A \in \mathcal{B}$ (the Borel sets)

- Notation:
  - the event $\{\omega : X(\omega) \in B\} \to$ '$X \in B$'
  - $P(\{X \in A\} \cap \{X \in B\}) \to$ '$P(X \in A, X \in B)$', etc.

Distributions

- X is measurable $\Rightarrow P(X \in B)$ is well-defined for any $B \in \mathcal{B}$
- The distribution of X is the function $\mu_X(B) = P(X \in B)$, for $B \in \mathcal{B}$
  - μ_X is a probability measure on $(\mathbb{R}, \mathcal{B})$
- The probability distribution function of X is the real-valued function

$$F_X(x) = P(\{\omega : X(\omega) \leq x\}) = \text{(notation)} = P(X \leq x)$$

- F_X is (obviously) the distribution function of the finite measure μ_X on $(\mathbb{R}, \mathcal{B})$, i.e.

$$F_X(x) = \mu_X((-\infty, x])$$

Independence

- Two random variables X and Y are pairwise independent if the events $\{X \in A\}$ and $\{Y \in B\}$ are independent for any A and B in $\mathcal{B}$
- A collection of random variables $X_1, \ldots, X_n$ is mutually independent if the events $\{X_i \in B_i\}$ are mutually independent for all $B_i \in \mathcal{B}$

Expectation

- For a random variable on $(\Omega, \mathcal{A}, P)$, the expectation of X is defined as
$$E[X] = \int_{\Omega} X(\omega) dP(\omega)$$
- For any Borel-measurable real-valued function g
$$E[g(X)] = \int g(x) dF_X(x) = \int g(x) d\mu_X(x)$$
in particular
$$E[X] = \int x d\mu_X(x)$$

Variance

- The variance of X,

$$\mathrm{Var}(X) = E[(X - E[X])^2]$$

- Chebyshev's inequality: For any $\varepsilon > 0$,

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

- Kolmogorov's inequality: For mutually independent random variables $\{X_k\}_{k=1}^n$ with $\mathrm{Var}(X_k) < \infty$, set $S_j = \sum_{k=1}^j X_k$, $1 \leq j \leq n$, then for any $\varepsilon > 0$

$$P\left(\max_j |S_j - E[S_j]| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(S_n)}{\varepsilon^2}$$

($n = 1 \Rightarrow$ Chebyshev)

The Law of Large Numbers

- A sequence $\{X_n\}$ is iid if the random variables X_n all have the same distribution and are mutually independent
- For any iid sequence $\{X_n\}$ with $\mu = E[X_n] < \infty$, the event

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \mu$$

occurs with probability one

- Toward the end of the course, we will generalize this result to stationary and ergodic random processes. . .

- $S_n = n^{-1} \sum_n X_n \to \mu$ with probability one $\Rightarrow S_n \to \mu$ in probability, i.e.,

$$\lim_{n\to\infty} P(\{|S_n - \mu| \geq \varepsilon\}) = 0$$

for each $\varepsilon > 0$

- in general 'in probability' does not imply 'with probability one' (convergence in measure does not imply convergence a.e.)

The Law of Large Numbers: Proof

- Lemma 1: For a nonnegative random variable X

$$\sum_{n=1}^{\infty} P(X \geq n) \leq E[X] \leq \sum_{n=0}^{\infty} P(X \geq n)$$

- Lemma 2: For mutually independent random variables $\{X_n\}$ with $\sum_n \mathrm{Var}(X_n) < \infty$ it holds that $\sum_n (X_n - E[X_n])$ converges with probability one
- Lemma 3 (Kronecker's Lemma): Given a sequence $\{a_n\}$ with $0 \leq a_1 \leq a_2 \leq \cdots$ and $\lim a_n = \infty$, and another sequence $\{x_k\}$ such that $\lim \sum_k x_k$ exists, then

$$\lim_{n\to\infty} \frac{1}{a_n} \sum_{k=1}^{n} a_k x_k = 0$$

- Assume without loss of generality (why?) that $\mu = 0$
- Lemma 1 $\Rightarrow$

$$\sum_{n=1} P(|X_n| \geq n) = \sum_{n=1} P(|X_1| \geq n) < \infty$$

- Let $E = \{|X_k| \geq k \text{ i.o.}\}$, Borel–Cantelli $\Rightarrow P(E) = 0 \Rightarrow$ we can concentrate on $\omega \in E^c$
- Let $Y_n = X_n \chi_{\{|X_n|<n\}}$; if $\omega \in E^c$ then there is an N such that $Y_n(\omega) = X_n(\omega)$ for $n \geq N$, thus for $\omega \in E^c$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = 0 \iff \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = 0$$

- Note that $E[Y_n] \to \mu = 0$ as $n \to \infty$

- Letting $Z_n = n^{-1} Y_n$, it can be shown that $\sum_{n=1}^{\infty} \mathrm{Var}(Z_n) < \infty$ (requires some work). Hence, according to Lemma 2 the limit

$$Z = \lim_{n\to\infty} \sum_{k=1}^{n} (Z_k - E[Z_k])$$

exists with probability one.
- Furthermore, by Lemma 3

$$\frac{1}{n} \sum_{k=1}^{n} (Y_k - E[Y_k]) = \frac{1}{n} \sum_{k=1}^{n} k(Z_k - E[Z_k]) \to 0$$

where also

$$\frac{1}{n} \sum_{k=1}^{n} E[Y_k] \to 0$$

since $E[Y_k] \to E[X_k] = E[X_1] = 0$

Proof of Lemma 2

- Assume w.o. loss of generality that $E[X_n] = 0$, set $S_n = \sum_{k=1}^{n} X_k$
- For $E_n \in \mathcal{A}$ with $E_1 \subset E_2 \subset \cdots$ it holds that

$$P\left(\bigcup_n E_n\right) = \lim_{n\to\infty} P(E_n)$$

Therefore, for any $m \geq 0$

$$P\left(\bigcup_{k=1}^{\infty} \{|S_{m+k} - S_m| \geq \varepsilon\}\right) = \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} \{|S_{m+k} - S_m| \geq \varepsilon\}\right)$$
$$= \lim_{n\to\infty} P\left(\max_{1\leq k\leq n} |S_{m+k} - S_m| \geq \varepsilon\right)$$

- Let $Y_k = X_{m+k}$ and

$$T_k = \sum_{j=1}^{k} Y_j = S_{m+k} - S_m,$$

then Kolmogorov's inequality implies

$$P\left(\max_{1\leq k\leq n} |T_k - E[T_k]| \geq \varepsilon\right) =$$
$$P\left(\max_{1\leq k\leq n} |S_{m+k} - S_m| \geq \varepsilon\right) \leq \frac{\mathrm{Var}(S_{m+n} - S_m)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \sum_{k=m+1}^{m+n} \mathrm{Var}(X_k)$$

- Hence

$$P\left(\bigcup_{k=1}^{\infty} \{|S_{m+k} - S_m| \geq \varepsilon\}\right) \leq \frac{1}{\varepsilon^2} \sum_{k=m+1}^{\infty} \mathrm{Var}(X_k)$$

- Since $\sum_n \text{Var}(X_n) < \infty$, we get for any $\varepsilon > 0$

$$\lim_{m\to\infty} P\left(\bigcup_{k=1}^{\infty} \{|S_{m+k} - S_m| \geq \varepsilon\}\right) = 0$$

- Now, let $E = \{\omega : \{S_n(\omega)\} \text{ does not converge}\}$. Then $\omega \in E$ iff $\{S_n(\omega)\}$ is not a Cauchy sequence $\Rightarrow$ for any n there is a k and an r such that $|S_{n+k} - S_n| \geq r^{-1}$. Hence, equivalently,

$$E = \bigcup_{r=1}^{\infty}\left(\bigcap_n \left(\bigcup_k \left\{|S_{n+k} - S_n| \geq \frac{1}{r}\right\}\right)\right)$$

- For $F_1 \supset F_2 \supset F_3 \cdots$, $P(\cap_k F_k) = \lim P(F_k)$, hence for any $r > 0$

$$P\left(\bigcap_{n=1}^{\infty}\left(\bigcup_k \left\{|S_{n+k} - S_n| \geq \frac{1}{r}\right\}\right)\right) = P\left(\bigcap_{n=1}^{\infty}\left(\bigcap_{\ell=1}^{n}\left(\bigcup_k \left\{|S_{\ell+k} - S_\ell| \geq \frac{1}{r}\right\}\right)\right)\right)$$

$$= \lim_{n\to\infty} P\left(\bigcap_{\ell=1}^{n}\left(\bigcup_k \left\{|S_{\ell+k} - S_\ell| \geq \frac{1}{r}\right\}\right)\right) \leq \lim_{n\to\infty} P\left(\bigcup_k \left\{|S_{n+k} - S_n| \geq \frac{1}{r}\right\}\right)$$