Lecture #6

Review: Probability Theory

**Random Variable:** \( x : \Omega \rightarrow \mathbb{R} \)

- A real valued function on
- A sample space, \( \Omega \)
- A set of possible outcomes.

**Random Outcome**

\( \Rightarrow \) Property or measurements on
- the random outcomes.

**Probability Space** \((\Omega, \mathcal{F}, \mathbb{P})\)

- \( \Omega \): Set of possible outcomes [No additional Structure.
- \( \mathcal{F} \): \( \sigma \)-field of subsets of \( \Omega \) [Nonempty, closed under complements and countable union.
- \( \mathbb{P} \): Measure on
- the measurable space \((\Omega, \mathcal{F})\).

\[ \mathbb{P}(\Omega) = 1 \Rightarrow \mathbb{P}(\emptyset) = 0 \]

\[ A \in \mathcal{F} \Rightarrow \mathbb{P}(A) \in [0, 1] \]
\[ \{ \omega \in \Omega : X(\omega) < a \} \in \mathcal{F} \quad \text{Event.} \]
B = Borel subset of the real line.
\[ P_x(B) = P(X^{-1}(B)) \]

**HISTORY.**

a) **Hilbert's Sixth Problem (1900)**
b) **Borel's Paradox**
   (Related to Baye's Theorem).
c) **Banach-Tarski Paradox**
   (Also Axiom of Choice).
d) **Savage's Utility Problem**

**Solution → Kolmogorov**

1) **Kolmogorov's 0-1 Law**
   
   "Events in the asymptotic σ-field has probability 0 or 1."

2) **Law of Large Numbers**
   Strong Law of Large Numbers.

\[ \{ x_n, n \geq 1 \} = \text{Sequence of centered independent r.v.s.} \]
\[ S_n = x_1 + x_2 + \ldots + x_n \]
\[ \sum_{n \geq 1} \frac{E(x_n^2)}{n} < \infty \Rightarrow \frac{S_n}{n} \to 0 \quad \text{a.s.} \]
\[ \{ E(1x,1) < \infty \Rightarrow \frac{S_n}{n} \to E(x) \quad \text{a.s.} \}
\[ \{ E(1x,1) = \infty \Rightarrow \lim \sup_{n \to \infty} \frac{1}{n} = +\infty \quad \text{a.s.} \]
Probability Distribution Function (PDF)

\[ P_x(B) : P(x'(B)) \leftarrow \text{TDF.} \]

\[ P[a < x < b] = P_x(a, b) = P\{\omega \in \Omega \mid a < x(\omega) < b\} \]

Open Interval = Borel.

Sample Space, \( \Omega \) = Set of outcomes of a Probability Experiment

Example

\( n \): coin tosses \( \rightarrow \) \( (H+T)^n \)

= \( 2^n \) \( n \)-long strings over the alphabet \( \Sigma = \{H,T\} \)

random genomes of length \( n \) \( \rightarrow \) \( (a+t+c+g)^n \)

= \( 4^n \) \( n \)-long strings over the alphabet \( \Sigma = \{a,t,c,g\} \)

Real Random Variable \( X : \Omega \rightarrow \mathbb{R} \)

Real-valued functions on \( \Omega \)

Example: Number of heads among the \( n \)-coin tosses

\( X = k \) \( \binom{n}{k} \) possibilities

\[
\frac{n!}{k! \cdot (n-k)!} \left\{ \begin{array}{c}
\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \\
k = \text{Head} \\
k' = \text{Tail}
\end{array} \right.
\]

\( X = k-1 \) \( X' = k \)
Event: Subset of \( \Omega \). In the \( \sigma \)-field, if \( \Omega \) is not finite.

Example: A specific subset of 3 tosses with even number of heads.
\( \{ TTT, HHT, HTH, THH \} \)

\[ \# \binom{3}{0} + \binom{3}{2} = 1 + 3 = 4. \]

Conditional Probability (of events \( F \) conditional on event \( G \)).
\[ \Pr(F|G) \Pr(G) = \Pr(F \& G) \]

Baye's Rule:
\[ \Pr(F|G) = \frac{\Pr(F \& G)}{\Pr(G)} = \frac{\Pr(F \& G)}{\Pr(G \& F) + \Pr(G \& \neg F) \Pr(\neg F)} \]

\[ = \frac{\Pr(G|F) \Pr(F)}{\Pr(G|F) \Pr(F) + \Pr(G|\neg F) \Pr(\neg F)} \]

Independence (of events \( F \) and \( G \)).
\[ \Pr(F|G) = \Pr(F) \]
\[ \Pr(F \& G) = \Pr(F|G) \Pr(G) = \Pr(F) \Pr(G) \]
Mutual Independence (of events $F_1, F_2, \ldots, F_n$)

$$\Pr(F_1 \land F_2 \land \ldots \land F_n)$$

$$= \Pr(F_1 \mid F_2 \land \ldots \land F_n) \Pr(F_2 \land \ldots \land F_n) \cdots \Pr(F_n)$$

$$= \Pr(F_1) \Pr(F_2) \cdots \Pr(F_n).$$

Indicator Variable

$$1_{\text{event}} = \begin{cases} 1 & \text{if event happens} \\ 0 & \text{otherwise} \end{cases}$$

Independence of Random Variables $X$ & $Y$

$\forall A, B \subseteq \mathbb{R}$

$$\Pr(x \in A \land y \in B) = \Pr(x \in A) \Pr(y \in B)$$

$\forall a, b \in \mathbb{R}$

$$\Pr(x = a \land y = b) = \Pr(x = a) \Pr(y = b)$$

Mutual Independence of r.v.s. $X_1, \ldots, X_n$

(1) $\forall A_1, A_2, \ldots, A_n$

$$\Pr(x_1 \in A_1 \land x_2 \in A_2 \land \ldots \land x_n \in A_n)$$

$$= \Pr(x_1 \in A_1) \Pr(x_2 \in A_2) \cdots \Pr(x_n \in A_n)$$

(2) $\forall a_1, a_2, \ldots, a_n$

$$\Pr(x_1 = a_1 \land x_2 = a_2 \land \ldots \land x_n = a_n)$$

$$= \Pr(x_1 = a_1) \Pr(x_2 = a_2) \cdots \Pr(x_n = a_n).$$
General Random Variable

Function from $\Omega$ to any arbitrary set $X: \Omega \rightarrow S$

Example: Random Graph

$G(n, p): (\binom{n}{2})$ coin tosses $\rightarrow$

$G = (V, E)$; $|V| = n$

$A_g = \begin{cases} 1, & i < j \text{ (i,j}^\text{th} \text{ coin toss = 1)} \\
0, & \text{otherwise.} 
\end{cases}$

Probability Density Function pdf $p(x)$ such that

$P_r (a < x < b) = \int_a^b p(x) \, dx.$

Cumulative Density Function cdf $F(x)$

$f(x) = \int_{-\infty}^a p(x) \, dx = P_r (x \leq a).$
Mean or Expectation. $E(X)$ of a r.v. $X$.

$$E(X) = \int_{-\infty}^{\infty} x \ p(x) \ dx = \mu(x)$$

or $$\sum_{x} x \ p(x) \ 	ext{if} \ -\infty < \text{finite}$$

$$E(1_{a < x < b}) = \int_{a}^{b} 1_{a < x < b} \ p(x) \ dx$$

$$= \int_{a}^{b} p(x) \ dx = Pr[a < x < b]$$

Let $x$ be a non-negative r.v. Then

$$t < x \iff 1 < \frac{x}{t}$$

$$1 \leq \frac{x}{t}$$

$$E(1_{x \geq t}) \leq E(\frac{x}{t}) = \frac{E(x)}{t}$$

$$Pr[x \geq t] \leq \frac{E(x)}{t} \quad \text{Markov Inequality}$$
Variance: \[ \text{Var}(x) = \mathbb{E}(x - \mathbb{E}(x))^2 = \sigma^2(x) \]

\[ k^{th} \text{ Moment} \quad \mu_k(x) = \mathbb{E}(x^k) \]

\[ \text{Var}(x) = \mu_2(x) - \mu_1(x)^2 = \mathbb{E}(x^2) - \mathbb{E}(x)^2 = \sigma^2. \]

\[ \Pr\left[ (x - \mu)^2 > a^2\sigma^2 \right] \leq \frac{\mathbb{E}\left[ (x - \mu)^2 \right]}{a^2\sigma^2} = \frac{1}{a^2} \]

(by Markov Ineq.)

\[ \Pr\left[ |x - \mu| > a\sigma \right] \leq \frac{1}{a^2} \]

(chebyshev Inequality)

**Linearity of Expectations**

(No independence assumed.)

\[ \mathbb{E}(\Sigma x_i) = \Sigma \mathbb{E}(x_i) \]

Expectation of sum of r.v.'s is sum of expectations.

But,

\[ \text{Var}(\Sigma x_i) = \Sigma \text{Var} (x_i) \quad \text{if } x_i \text{'s are pairwise independent r.v.s.} \]
Probability Distributions

- **Bernoulli**
  \( x \sim \text{Bernoulli}(p) \)
  \( x \in \{0, 1\} \)
  \( \Pr[x = 1] = p \)
  \( \Pr[x = 0] = q = 1 - p \)

- **Binomial**
  \( x \sim \text{Binomial}(n, p) \)
  \( x \in \{0, 1, 2, ..., n\} \)
  \( \Pr[x = k] = \binom{n}{k} p^k (1-p)^{n-k} \)

In \( n \) independent trials, exactly \( k \) successes are observed:

\[ \mu = np \quad \sigma^2 = npq \]

- **Poisson**
  \( x \sim \text{Poisson}(\lambda) \)
  \( \Pr(x = k) = e^{-\lambda} \frac{\lambda^k}{k!} \)

Average rate of an event per unit time is \( \lambda \)

Exactly \( k \) events are observed in a unit time:

\( x \sim \text{B}(n, \frac{\lambda^k}{k!}) \)

\[ \Pr[x = k] = \binom{n}{k} \left( \frac{\lambda^k}{k!} \right) \left( 1 - \frac{\lambda}{n} \right)^{n-k} \]

\[ = \frac{n^k}{k!} \frac{\lambda^k}{n^k} \left( 1 - \frac{\lambda}{n} \right)^n \left( 1 - \frac{\lambda}{n} \right)^{-k} \]

\[ = e^{-\lambda} \frac{\lambda^k}{k!} \frac{n^k (\frac{\lambda}{n})^k}{n^k (1 - \frac{\lambda}{n})^k} \]

\[ \lim_{n \to \infty} \left( \frac{n^k (\frac{\lambda}{n})^k}{n^k (1 - \frac{\lambda}{n})^k} \right) = e^{-\lambda} \frac{\lambda^k}{k!} \]

\[ \mu = \lambda \quad \sigma^2 = \lambda \]
Let \( x_1, x_2, \ldots, x_n \) be independent Bernoulli random variables
\( x_i \sim \text{Bernoulli}(p) \)

**Random Walk.**
\[
S_n = \sum_{i=1}^{n} x_i \quad m_n = \mu(S_n) = np
\]

**Chernoff Bound:**
\[
\forall \delta > 0 \quad \Pr \left[ S_n > (1+\delta) m_n \right] \leq \left[ \frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \right]^{m_n}
\]
\[
\forall \delta > 2e-1 \quad \Pr \left[ S_n > (1+\delta) m_n \right] \leq \left( \frac{e^{\delta}}{1+\delta} \right)^{(1+\delta)m_n}
\]
\[
\forall 0 < \nu < 1 \quad \Pr \left[ S_n < (1-\nu) m_n \right] < \left[ \frac{e^{-\nu}}{(1+\nu)^{(1+\nu)}} \right]^{m_n} < e^{-nm^2/2}
\]

**Hoeffding Bounds**
\[
\Pr \left[ S_n - m_n \geq \epsilon \right] < e^{-2\epsilon^2/m}
\]
Gaussian

\[ x \sim N(\mu, \sigma^2) \]

\[ p(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

\[ \mu: \mu, \quad \sigma^2 = \sigma^2 \]

CLT (Central Limit Theorem)

\[ x_1, x_2, \ldots, x_n \text{ iid with } E(x_i) = \mu \quad \text{Var}(x_i) = \sigma^2 \]

Then

\[ \frac{\sum_{i=1}^{n} x_i - n\mu}{\sqrt{n}\sigma} \sim N(0, 1) \]

Random Graph \ G(n,p) Model \ \{ \text{Erdős-Rényi} \}

A graph \ G = (V,E) \ is constructed by connecting every pair of nodes uniformly randomly.

\[ \forall u, v \in V \quad [ (u, v) \in E ] \sim \text{Bernoulli}(p) \text{ iid.} \]

For every pair of nodes \( u, v \in V \), an edge \( (u, v) \in E \) is included in the graph with probability \( p \) independent from every other edge.
\[ |V| = n \quad \langle 1E1 \rangle = \binom{n}{2} p \]

Expected number of edges.

Expected degree of a graph \( G \in G(n, p) \)

\[
\langle d \rangle = \frac{2 \langle 1E1 \rangle}{\langle 1V1 \rangle} = \frac{2 \binom{n}{2} p}{n} = (n-1)p
\]

\( d(v) \sim \text{Binomial} \left( n-1, p \right) \)

\[
\Pr \left[ d(v) = k \right] = \binom{n-1}{k} p^k (1-p)^{n-k-1} \]

\[
x \frac{(np)^k}{k!} e^{-np}
\]

If expected degree \( d \) is held constant (independent of \( n \)): \( n \) large, \( np = \text{const.} = \lambda \)

\( d(v) \sim \text{Poisson} (\lambda) \quad \lambda = d = np \)

\textit{Poisson Approximation.}

\textit{Phase Transition: 0.1 Laws.}

Small \( p \) \quad \( p \leq \frac{(1-e) \ln n}{n} \) \quad \( G(n,p) = \text{Disconnected a.s.} \)

Large \( p \) \quad \( p \geq \frac{(1+e) \ln n}{n} \) \quad \( G(n,p) = \text{Connected a.s.} \)