1. Random Variable:

   - **Given a sample space Ω, and the corresponding set of possible outcomes, *a random variable* associate a particular number with each outcome**
   - It is a real-value function of the outcomes in sample space
   - **A function of a random variable** defines another random variable
   - **mean and variance** are "averages" of a random variable
   -

2. Discrete Random Variable: **if its range is either finite or countable finite**

   - Associated function:

     (a) **PMF: probability mass function $p_X(x)$ is the probability of set $\{X = x\}$ consisting of all outcomes that give rise to a value of $X = x$ $p_X(x) = P(\{X = x\})$** $\rightarrow$ gives the probability of each numerical value that the random variable can take

     For any set S of possible values of X

     $$P(X \in S) = \sum_{x \in S} p_X(x)$$

     (b) Expectation(expected value or mean):

     (c) Variance:

       – Definition $var(X)$ of RV X is

       $$var(X) = E[(X - E[X])^2]$$

       – Formula for calculation:

       $$var(x) = \sum_{x} (X - E[X])^2 p_X(x)$$

       – $var(X) \geq 0$

   - Different discrete RVs

     (a) **Bernoulli RV:**

     (b) **Binomial RV,** $B(n, p)$

     (c) **Geometric RV**

       – The probability distribution of the number of Bernoulli trials needed to get one success
       – The probability distribution of the number $Y = X - 1$ of failure the first success

     (d) **Poisson RV**: useful in the cases the success has very small probability p in a very large size of sample space

3. General Random Variable

   - **Probability Density Function of X: PDF**

     (a) Ideas:
     -In other words, while the absolute likelihood for a continuous random variable to take on any particular value is 0 (since there are an infinite set of possible values to begin with), the value of the PDF at two different samples can be used to infer, in any particular draw of the random variable, how much more likely it is that the random variable would equal one sample compared to the other sample.
     -the PDF is used to specify the probability of the random variable falling within a particular range of values, as opposed to taking on any one value.

4. Further Topic on Random Variable

- **Convolution** of the PMFs of X and Y
- **estimator** of X given Y
- **estimation error**
- **moment generating function:** $M_X(s)$

5. Limit Theorems

- Markov Inequality [about small mean]
  If a RV X such that $x \geq 0$, $\forall x$. Then
  $$P(X \geq a) \leq \frac{E[X]}{a}$$
  (a) It asserts that if a nonnegative random variable has a **small mean**, then the probability that it takes a large value must also be small
- Chebyshev Inequalities [about small variance]
  If X is a RV with mean $E[X] = \mu$ and variance $var(X) = \sigma^2$, then
  $$P(|X - \mu| \geq c) = P(|X - E[X]| \geq c) \leq \frac{var(X)}{c^2}$$
  (a) It asserts that if a random variable has **small variance**, then the **probability** that it takes a value far from its mean is also **small**
  (b) Alternative Form
  Letting $c = k\sigma$, where k is positive,
  $$P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$
  Means: the probability that a random variable takes a value more than k standard deviations away from its mean is at most $\frac{1}{k^2}$
  (c) Ex 1
  Let X be $Exp(1)$, so that $E[X] = var(X) = 1$. For $c > 1$, by the Chebyshev inequality
  $$P(X \geq c) = P(X - 1 \geq c - 1) \leq P(|X - 1| \geq c - 1) \leq \frac{1}{(c-1)^2}$$
- The Weak Law of Large Numbers
  Let $X_1, ...$ be independent identically distributed random variables with mean μ. $\forall \epsilon > 0$, we have
  $$P(|M_n - \mu| \geq \epsilon) = P(|\frac{X_1 + ... + X_n}{n} - \mu| \geq \epsilon) \rightarrow 0, as\ n \rightarrow \infty$$
  Corollary:
  $$P(|M_n - \mu| < \epsilon) \rightarrow 0,\ as\ n \rightarrow \infty$$
  1. It asserts that the sample mean of a large number of independent identically distributed random variables is very close to the true mean, with high probability
  2. It also states that for large n, the bulk of the distribution of M_n is concentrated near μ. That is, if we consider a interval $[\mu - \epsilon, \mu + \epsilon]$ around μ, then there is high probability that M_n will fall in that interval; as $n \rightarrow \infty$, this probability converges to 1
  3. If ϵ is very small, that the value of n shoulf be larger for M_n to highly likely to fall into that interval

- The Strong Law of Large Numbers
  Let $X_1, ...$ be a sequence of independent identically distributed random variables with mean μ. Then, the sequence of sample means $M_n = \frac{(X_1 + ... + M_n)}{n}$ converges to μ, with probability 1, in the sense that
  $$P\left(\lim_{n\to\infty} M_n = \mu\right) = 1$$
  All of the probability is concentrated on this particular subset A of the sample space Ω; the collection of outcomes that do not belong to A has probability 0*, where the* $\lim_{n\to\infty} M_n \neq \mu$