Theorems

Sum of Geometric Series and Sequences

$$S_n = \sum_{i=1}^{n} a_i r^{i-1} = a_1 \left(\frac{1-r^n}{1-r} \right)$$

$$S = \sum_{i=0}^{\infty} a_i r^i = \frac{a_1}{1-r}$$

Newton Binomial Theorem

$$(x+y)^n = \sum_{k=0}^{n} C(n,k) x^k y^{n-k}$$

Maclaurin Series of e^x

$$e^x = \sum_{n \geq 0} \frac{x^n}{n!} = \sum_{n \geq 1} \frac{x^{n-1}}{(n-1)!}$$

Arrangment, Permutation, Combination, Conditional Probability

Permutation

If Ω has n elements, then there are $n!$ number of permutations of Ω.

Arrangement

$$A_n^k = n(n-1)(n-2)...(n-k+1) = \frac{n!}{(n-k)!}$$

Combination

$$C(n,k) = \frac{n!}{k!(n-k)!}$$

Conditional Probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Law of Total Probability

Let $\{B_1, \ldots, B_n\}$ be a complete system of events, and A be any event. Then:

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A|B_i)$$

Bayes' Formula

Let A and B be two events of non-zero probability. Then we have the identity:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

If the family $\{A_1, \ldots, A_n\}$ is a complete system of events, then for $1 \leq i \leq n$ we have:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

Independence

$$\forall A \in \mathcal{F}_1, \forall B \in \mathcal{F}_2 : P(A \cap B) = P(A)P(B)$$

Random Variables

Atom

$$P(X = x) > 0$$

CDF

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt$$

If $F_X = F_Y$, then $E[X] = E[Y]$

PDF

$$f_X = (F_X)' = \frac{d}{dx} F_X$$

$$f_X(x)\varepsilon \approx P(X \in [x, x+\varepsilon])$$

$$P(a \leq X \leq b) = \int_{a}^{b} f_X(t)\, dt, \quad a < b$$

Expectation, Variance, Median

Expectation Discrete

$$\sum_{n=0}^{\infty} |x_n| P(X = x_n) < +\infty$$

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

Expectation Density

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < +\infty$$

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

Transfer Theorem

Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$, and g be a function from $\mathbb{R}$ to $\mathbb{R}$. Then, the quantity $E[g(X)]$ depends only on the function g and the distribution of X.

$$E[g(X)] = \sum_{n=0}^{\infty} g(x_n)P(X = x_n)$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

Markov's Inequality

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha}, \quad \alpha > 0$$

$$P(X \geq \alpha) \leq \frac{E[g(X)]}{g(\alpha)}, \quad g \text{ is positive and strictly increasing}$$

Median

$$P(X \leq m) \geq \frac{1}{2} \text{ and } P(X \geq m) \geq \frac{1}{2}$$

Variance

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

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

$$\sigma(X) = \sqrt{\mathrm{Var}(X)} - \text{ standard deviation}$$

Let X be a random variable with finite variance, and λ a real number. We have:

- $\mathrm{Var}(\lambda X) = \lambda^2 \mathrm{Var}(X)$
- $\mathrm{Var}(X + \lambda) = \mathrm{Var}(X)$
- X is almost surely (a.s.) constant if and only if $\mathrm{Var}(X) = 0$
- $\mathrm{Var}(X) = E[X^2] - E[X]^2$

Chebyshev's Inequality

$$P(|X - E[X]| \geq \beta) \leq \frac{\mathrm{Var}(X)}{\beta^2}, \quad \beta \in R \backslash 0$$

Probability Distributions

Discrete

Bermoulli $\mathcal{B}(\theta)$

$$P(X = 0) = 1 - \theta \text{ and } P(X = 1) = \theta$$

$$E[X] = \theta$$

$$\mathrm{Var}(X) = \theta(1 - \theta)$$

Poisson $\mathcal{P}(\theta)$

$$P(X = k) = \frac{e^{-\theta}\theta^k}{k!}$$

$$E[X] = \theta$$

$$\mathrm{Var}(X) = \theta$$

Binomial $\mathbb{B}(n, \theta)$

$$P(X = k) = \binom{n}{k}\theta^k(1 - \theta)^{n-k}$$

$$\text{if } S_n = \sum_{i=1}^{n} X_i, X_i \sim \mathcal{B}(\theta) => S_n \sim \mathbb{B}(n, \theta)$$

$$E[X] = n\theta$$

$$\mathrm{Var}(X) = n\theta(1 - \theta)$$

Uniform $\mathbb{U}(a, b)$

$$n = b - a + 1$$

$$k \in \{1, ..., n\}$$

$$P(X = k) = \frac{1}{n}$$

$$E[X] = \frac{n + 1}{2}$$

$$\mathrm{Var}(X) = \frac{n^2 - 1}{12}$$

Geometric $\mathcal{G}(\theta)$

$$k \in N$$

$$P(X = k) = \theta(1 - \theta)^{k-1}$$

$$P(X \leq k) = 1 - (1 - \theta)^k$$

$$P(X > k) = (1 - \theta)^k$$

$$E[X] = \frac{1}{\theta}$$

$$\mathrm{Var}(X) = \frac{1 - \theta}{\theta^2}$$

Density

Normal $\mathcal{N}(\theta, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(x - \theta)^2)$$

$$\theta = E[X]$$

$$\sigma^2 = \mathrm{Var}(X)$$

Exponential (Memoryless) $\varepsilon(\lambda)$

$$P(X \leq x) = 1 - e^{-\lambda x} \cdot 1_{[0,+\infty]}$$
$$P(X \geq x) = e^{-\lambda x} \cdot 1_{[0,+\infty]}$$
$$P(X \geq x + s | X \geq x) = P(X \geq s)$$
$$f_X(t) = \lambda e^{-\lambda t} 1_{[0,+\infty]}(t)$$
$$E[X] = \frac{1}{\lambda}$$
$$\mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Uniform $\mathcal{U}([a,b])$

$$a \leq c \leq d \leq b$$
$$P(c < X < d) = \frac{d-c}{b-a}$$
$$f_X(x) = \frac{1}{b-a} 1_{[a,b]}$$
$$E[X] = \frac{b+a}{2}$$
$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Pair of Random Variables

Discrete

Joint Law

$$P(X = x_k, Y = y_j)$$

If you know the distribution of the pair (X, Y), then you can deduce the Marginal Distributions of X and Y.

Marginal Law

$$P(X = x_k) = \sum_j P(X = x_k, Y = y_j)$$
$$P(Y = y_j) = \sum_k P(X = x_k, Y = y_j)$$

Knowing marginals does not allow us to compute the joint, unless we add some hypothesis, like independence.

Continious Case

Joint Law

$$P(X \leq x, Y \leq y)$$

Joint Density

$$P(X \in I, Y \in J) = \iint_{I \times J} f(X,Y)(x,y)\, dx\, dy$$

If you have joint density, you can compute marginal densities.

Marginal Density

$$f_X(x) \to \int_R f_{(X,Y)}(x,y)\, dy$$
$$f_Y(y) \to \int_R f_{(X,Y)}(x,y)\, dx$$

Independent Random Variables

$\forall A \in G, \forall B \in H, P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$

Discrete Case

$$P(X = x_k, Y = y_j) = P(X = x_k)P(Y = y_j)$$

Continious Case

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

Joint Density

$$f_{(X,Y)}(x,y) \to f_X(x) f_Y(y)$$

Expected Values

$$E[XY] = E[X]E[Y]$$
$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$
$$E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n]$$

Variances

$$Var(X+Y) = Var(X) + Var(Y)$$
$$Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n)$$

Covariance

$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$

PMF ($Z = X + Y, X \perp Y$)

Discrete

$$P(Z = n) = \sum_{k=0}^{n} P(X = k)P(Y = n-k)$$

Continious

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(u) f_Y(z-u)\, du$$

Limit Theorems

Modes of Convergence of Random Variables

Almost Sure Convergence

We say that $(X_n)_{n\geq 0}$ converges to X **almost surely** if, for almost every $\omega \in \Omega$

$$\lim_{n\to+\infty} X_n(\omega) = X(\omega)$$

$$X_n \to^{a.s.}_{n\to+\infty} X$$

A simple case of almost sure convergence occurs when dealing with an **increasing** and **bounded sequence** of random variables. In that case, it converges almost surely to a limiting random variable.

Convergence in Law L^p

We say that $(X_n)_{n\geq 0}$ converges to X **in Law** L^p if

$$\lim_{n\to+\infty} E[|X_n - X|^p] = 0$$

$$X_n \to^{L^p}_{n\to+\infty} X$$

When $p = 2$ it is called quadratic convergence.

In other words, as we compute distances for each ω for $(X_n)_{n\geq 0}$ and X, each ω should be close for $E[|X_n - X|^p]$ to go to 0.

Convergence in Probability

We say that $(X_n)_{n\geq 0}$ converges to X **in probability** if, for every $\varepsilon > 0$

$$\lim_{n\to+\infty} P[|X_n - X| > \varepsilon] = 0$$

$$X_n \to^{P}_{n\to+\infty} X$$

In other words, a sequence of random variables converges in probability to its limit when the probability that the sequence takes values far from its limit tends to 0.

Convergence in Distribution

We say $(X_n)_{n\geq 0}$ (with distribtuion functions $(F_n)_{n\geq 0}$) **converges in distribution** to X (with distribution function F) if, at every continuity point x of function F we have

$$\lim_{n\to+\infty} F_n(x) = F(x)$$

$$X_n \to^{L}_{n\to+\infty} X$$

In other words, if $(X_n)_{n\geq 0}$ is Gaussian, or Bernoulli, or etc., then X is distributed the same way.

Instead of X we can put a distribution, for example

$$X_n \to^{L}_{n\to+\infty} \mathcal{N}(0,1)$$

Example $(X_n)_{n\geq 0}$ can converge in **distribution** to X, but not converge in other modes of convergence. Take sequence $(X, -X, X, -X, ...)$. We have Gaussian in every X_i, so sequence converges to Gaussian. Convergence in distribution generally does not imply other modes of convergence. Convergence in distribution is **weak mode of convergence**.

Convergence to Constant

$$\text{if } (X_n)_{n\geq 0} \to^{L}_{n\to+\infty} C, C \in R$$

$$=> (X_n)_{n\geq 0} \to^{P}_{n\to+\infty} C$$

Convergence Implications

- Almost Sure Convergence => Convergence in Probability => Convergence in Distribution
- Convergence in L^p => Convergence in Probability => Convergence in Distribution

Independent and Identically Distributed Sequence of Random Variables (i.i.d.)

If $(X_n)_{n\geq 1}$ are i.i.d., then they all have the same distribution.

LLN | Monte Carlo | CLT

Empirical Mean

We define the sequence of empirical means $\overline{(X_n)_{n\geq 1}}$ of i.i.d. random variables as

$$\forall n \geq 0 \quad \overline{X_n} = \frac{X_1 + ... + X_n}{n}$$

$$\sqrt{Var(\overline{X_n})} = \frac{\sqrt{Var(X)}}{\sqrt{n}}$$

$$Var(\overline{X_n}) = \frac{Var(X)}{n}$$

Law of Large Numbers

If $(X_n)_{n \geq 1}$ are i.i.d. and expectation of X is finite, then

$$\overline{X_n} \to^{\text{a.s.}}_{n \to +\infty} E[X]$$

For example, the law of large numbers asserts that, on average, after a large number of dice rolls, I would have rolled a 5 approximately one in six times.

Monte Carlo

An application of the law of large numbers is the Monte Carlo method for estimating the value of an integral. This method is particularly useful in dimensions $d \geq 1$.

Let $f : [0,1]^d \to R$ be an integrable function, and $(U_n)_{n \geq 1}$ be an i.i.d. sequence of uniform random variables on $[0,1]^d$. Then

$$\frac{f(U_1) + ... + f(U_n)}{n} \to_{n \to +\infty} \int_{[0,1]^d} f(x)dx$$

Example

Define $X_n = f(U_n)$. The sequence $(X_n)_{n \geq 1}$ is a sequence of real r.v. i.i.d. with

$$\overline{f(U_n)} \to^{\text{a.s.}}_{n \to \infty} E[f(U)] = \int_{[0,1]^d} f(x)dx$$

Central Limit Theorem (CLT)

Assuming that the variance of X is finite, then

$$\frac{\overline{X_n} - E[\overline{X_n}]}{\sqrt{Var(\overline{X_n})}} \to^{L}_{n \to +\infty} \mathcal{N}(0,1)$$

$$Var(X) = \sigma^2$$

$$E[\overline{X_n}] = E[X]$$

$$Var(\overline{X_n}) = \frac{\sigma^2}{n}$$

$$\sqrt{n}(\overline{X_n} - E[X]) \to^{L}_{n \to +\infty} \mathcal{N}(0, \sigma^2)$$

$$\frac{(X_1 + ... + X_n) - nE[X]}{\sqrt{n}\sigma} \to^{L}_{n \to +\infty} \mathcal{N}(0,1)$$

$$\overline{X_n} \approxeq \mathcal{N}(E[X], \frac{\sigma^2}{n})$$

The CLT asserts that the error between the theoretical and empirical means is of the order of $\sqrt{Var(\overline{X_n})} = \frac{\sigma}{\sqrt{n}}$. This allows the calculation of asymptotic confidence interval.

Asymptotic Confidence Interval at level α

Let Z be a standard normal random variable, and $0 < \alpha < 1$. We define the number $q_{\alpha/2}$ such that

$$P(-q_{\alpha/2} \leq Z \leq q_{\alpha/2}) = 1 - \alpha$$

$$\lim_{n \to +\infty} P\left(\overline{X_n} - \frac{\sigma}{\sqrt{n}} q_{\alpha/2} \leq E[X] \leq \overline{X_n} + \frac{\sigma}{\sqrt{n}} q_{\alpha/2}\right) = 1 - \alpha$$

$$\lim_{n \to +\infty} P\left(\left|\frac{\overline{X_n} - E[\overline{X_n}]}{\sqrt{Var(\overline{X_n})}}\right| \leq q_{\alpha/2}\right) \approxeq 1 - \alpha$$

$$\lim_{n \to +\infty} P\left(\left|\frac{\overline{X_n} - E[X]}{\frac{1}{\sqrt{n}}\sqrt{Var(X)}}\right| \leq q_{\alpha/2}\right) \approxeq 1 - \alpha$$