Basic Concepts

Modes of Convergence

Definitions

1. The strongest mode of convergence is actually weaker than pointwise. We say a sequence of random variables $X_1, X_2, \ldots$ converges *almost surely* to L and write $X_n \overset{(a.s.)}{\to} L$ if
$$\mathbb{P}(X_n \to L) = 1,$$
i.e., the set of points for which the sequence does *not* converge is negligible.
2. Another concept of convergence is *convergence in quatratic mean*, which is simply L^2 convergence in the probability space.
3. We say a sequence of random variables $X_1, X_2, \ldots$ converges *in probability* to L and write $X_n \overset{(p)}{\to} L$ if for all $\epsilon > 0$,
$$\mathbb{P}(|X_n - L| > \epsilon) \to 0,$$
where here convergence is that of real numbers. Both almost sure convergence and L^p convergence imply convergence in probability.
4. We say $X_n \to L$ *in distribution* if the sequence of CDF functions F_{X_n} converges pointwise to F_L. Convergence in probability implies convergence in distribution. The converse is true when L is a point mass.
5. We briefly talk about *uniform* convergence later in the class, which is simply convergence in the L^∞ space.

Counterexamples

TODO... Some day_

Limits of Sequences of Random Variables

1. **Law of Large Numbers**: Let $X_1, X_2, \ldots$ be an *iid* sequence of random variables with finite mean μ. Then $X_n \to \mu$ in probability. Or maybe only in distribution. I think in most well behaved examples, the convergence is in probability. There is also a strong law of large numbers, which asserts almost sure convergence, but I have not used it in the class and don't remember the conditions.
2. **Central Limit Theorem**: Let $X_1, X_2, \ldots$ be an *iid* sequence of random variables with finite mean μ and finite variance σ^2. Then
$$\sqrt{n}(\bar{X}_n - \mu) \to N(0, \sigma^2),$$

where convergence is in distribution. It is important to note that CLT holds for multivariate distributions as well, with the obvious generalization to a multivariate Gaussian limit.

The Delta Method

1. **The Delta Method**: Let $X_1, X_2, \ldots$ be a sequence of d-dimensional random vectors and suppose

$$\sqrt{n}(\bar{X}_n - \mu) \stackrel{(d)}{\to} N(0, \Sigma).$$

   Then for any continuously differentiable function $g : \mathbb{R}^d \to \mathbb{R}$, we have

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \stackrel{(d)}{\to} N\big(0, (\nabla g(\mu))^T \cdot \Sigma \cdot \nabla g(\mu)\big).$$

2. The most obvious application of the Delta method is in parameter estimation for distributions whose parameter is something other than a simple mean (e.g., exponential distribution). We will see the Delta method again when we derive the maximum likelihood estimator.