Convergence Concepts

Limit Theory

Convergence in Probability

Definition

We saw that the estimator of p, $\hat{p}$, from the Binomial example seemed to be observed closer and closer to p for larger sample sizes. Additionally, we saw a good large-sample distribution for $\hat{p}$ is

$$\hat{p} \overset{\bullet}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

Does this large-sample distribution support the 'convergence' of $\hat{p}$ to p idea?

More formally, we're going to take on the idea of **convergence in probability to a constant**. First, let's define convergence in probability generally.

Convergence in Probability A sequence of RVs $Y_1, ..., Y_n, ...$ converges in probability to a RV Y if for every $\epsilon > 0$

$$\lim_{n\to\infty} P(|Y_n - Y| \geq \epsilon) = 0 \iff \lim_{n\to\infty} P(|Y_n - Y| < \epsilon) = 1$$

This is denoted as

$$Y_n \overset{p}{\to} Y$$

We'll mostly care about convergence in probability to a constant, call it c. We can see the definition in this case can be simplied to the following:

$$\lim_{n\to\infty} P(|Y_n - c| < \epsilon) = \lim_{n\to\infty} P(-\epsilon < Y_n - c < \epsilon) = \lim_{n\to\infty} P(c - \epsilon < Y_n < c + \epsilon) = 1$$

$Y_n \overset{p}{\to} c$ if the *probability* we observe Y_n close to c goes to 1 in the limit.

Example - We can visualize this idea.

Assume that $Y_i \overset{iid}{\sim} N(0,1)$. Let's investigate the behavior of

$$X = \frac{1}{n^2}\sum_{i=1}^{n} Y_i$$

To put this in the context of the definition, let's refer to X explicitly as a function of n:

$$X_n = \frac{1}{n^2}\sum_{i=1}^{n} Y_i$$

We want to understand the behavior of X_n as n grows. We'll see that $X_n \xrightarrow{p} 0$, which implies that for any $\epsilon > 0$ we have

$$\lim_{n\to\infty} P(-\epsilon < X_n < \epsilon) = 0$$

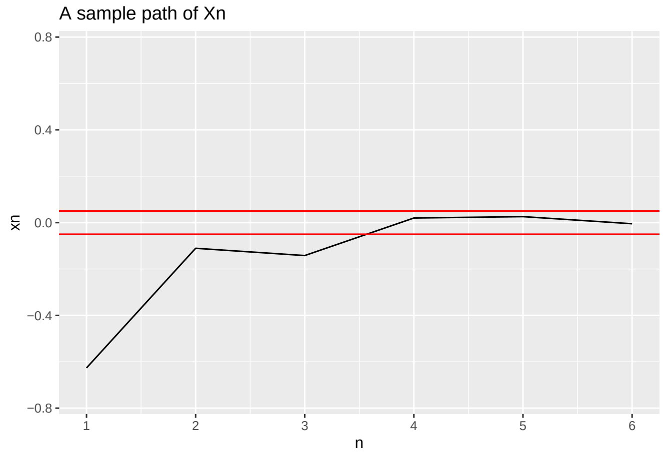
To visualize this, we can consider **sample paths** of X_n. That is, we can look at a particular sequence of y_i's that will generate a sequence of x and see how the values change.

Consider the following 6 values randomly sampled from a $N(0,1)$ and the corresponding sequence of x_n values.

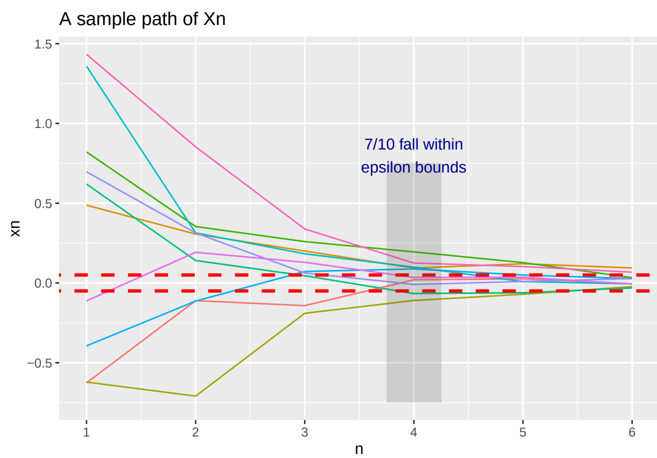
y sequence	x sequence
y_1 = -0.6264538	x_1 = -0.6264538/1^2 = -0.6264538
y_2 = 0.1836433	x_2 = (-0.6264538+0.1836433)/2^2 = -0.1107026
y_3 = -0.8356286	x_3 = (-0.6264538+0.1836433+-0.8356286)/3^2 = -0.1420488
y_4 = 1.5952808	x_4 = (-0.6264538+…+1.5952808)/4^2 = 0.0198026
y_5 = 0.3295078	x_5 = (-0.6264538+…+0.3295078)/5^2 = 0.025854
y_6 = -0.8204684	x_6 = (-0.6264538+…+-0.8204684)/6^2 = -0.0048366

If we consider multiple sample paths, then convergence in probability to 0 of this sequence implies that the proportion of sample paths outside of $\pm\epsilon$ should go to zero.

Let's plot our sample path with an $\epsilon = 0.05$:



Now let's add 9 more sample paths:



What we hope to see is that the proportion of lines falling outside of the ϵ bars goes to 0!

Example - Suppose we have a random sample from a Normal distribution with mean 10 and standard deviation 1. What do you think $W = (\bar{Y})^2$ converges to in probability? Take an educated guess and use the app below to explore!

- Select the value c that you guess W converges to in probability.
- Choose a sample size to go up to (start smaller and then get larger once you have a good idea).
- Select an ϵ range.
- Look for the proportion of lines (50 sample paths are generated) falling outside of the ϵ bars to go to 0!

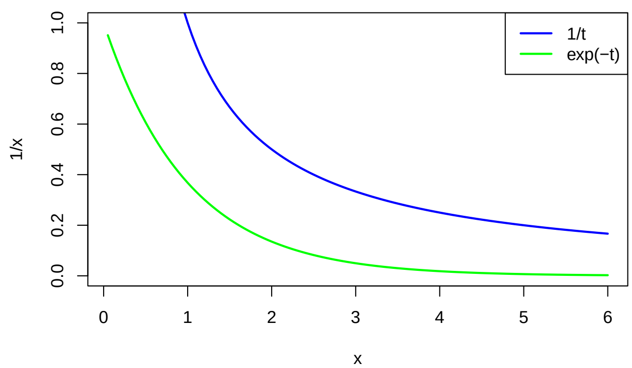
Inequalities

To prove convergence in probability, we'll sometimes rely on some very famous inequalities. These will help us to show the probability goes to 0 or 1.

Markov's Inequality If X is a nonnegative RV (support has no negative values) for which $E(X)$ exists, then for $t > 0$

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Example: If $X \sim exp(1)$ then $P(X \geq t) = e^{-t}$ and $E(X)/t = 1/t$.



Chebychev's Inequality Let X be a RV with mean $= \mu$ and variance $= \sigma^2$, then for $t > 0$

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

Example: If $t = \sigma k$ for $k > 0$, we can apply Chebychev's to get

$$P\left(|X - \mu| \geq k\sigma\right) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

For $k = 2$ we have $P\left(|X - \mu| \geq 2\sigma\right) \leq 1/4$.

\~\

Practically, what can we take home from this?

- At least 75% of a RVs distribution lies within 2 standard deviations of the mean (if these moments exist)
- Regardless of distribution! (if moments exist)
- If $X \sim N(\mu, \sigma^2)$ we know $P(|X - \mu| \geq 2\sigma) \approx 0.05$. The bound isn't always very tight!

WLLN

One of the most important results regarding convergence in probability is called the Law of Large Numbers (LLN).

(Weak) Law of Large Numbers (WLLN) Suppose $Y_i \overset{iid}{\sim} f$ where the mean and variance of Y_i exist. Then

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \overset{p}{\to} E(Y) = \mu$$

- Big picture goal is to estimate parameters such as μ
- If we get a RS we know that $\bar{Y}$ will be a 'close' to μ for 'large' samples
- Applies to the **average of any independent random variables with the same finite mean**

Note that the variance assumption is actually not needed but will help us facilitate an easy proof. Let's use our inequalities to prove this result!

Example - Let $Y_i \overset{iid}{\sim} N(\mu, \sigma^2)$. What does $\bar{Y}$ converge to? What does $\frac{1}{n}\sum_{i=1}^{n} Y_i^2$ converge to?

Continuity Theorems

The WLLN is also quite useful when combined with the continuity theorem.

Continuity Theorem If $Y_1, Y_2, Y_3, ...$ converge to Y and $g()$ is a continuous function then $g(Y_1), g(Y_2), g(Y_3)...$ converge to $g(Y)$.

Example (exploration example proved) - Suppose we have a random sample from a Normal distribution with mean 10 and standard deviation 1. Consider $W = (\bar{Y})^2$. What does this converge to in probability?

Note: The continuity theorem also works for convergence in distribution!

Example - Suppose that $Y_i \overset{iid}{\sim} gamma(\alpha, \lambda)$. We have that

$$\frac{\bar{Y} - \alpha/\lambda}{\frac{\alpha}{\lambda^2 \sqrt{n}}} \overset{d}{\to} Z$$

where $Z \sim N(0, 1)$. By the continuity theorem we have that

$$\left(\frac{\bar{Y} - \alpha/\lambda}{\frac{\alpha}{\lambda^2 \sqrt{n}}}\right)^2 \overset{d}{\to} Z^2$$

and recall that a standard Normal squared is distributed as a χ^2_1 or a $gamma(1/2, 1/2)$.

Other Standard Limit Results Work Too!\ Most of the common limit theorem ideas from calculus follow here as well (θ and λ are constants below):

$$\text{If } Y \overset{p}{\to} \theta, X \overset{p}{\to} \lambda \text{ then } Y \pm X \overset{p}{\to} \theta \pm \lambda$$

Example - Consider the 'biased' version of the sample variance, $S_n^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2$. Let's show $S_n^2 \overset{p}{\to} \sigma^2$

$\xrightarrow{d}$ & $\xrightarrow{p}$ Relationship

Convergence in probability implies convergence in distribution. However, the converse is not true generally (**convergence in distribution does not imply convergence in probability**).

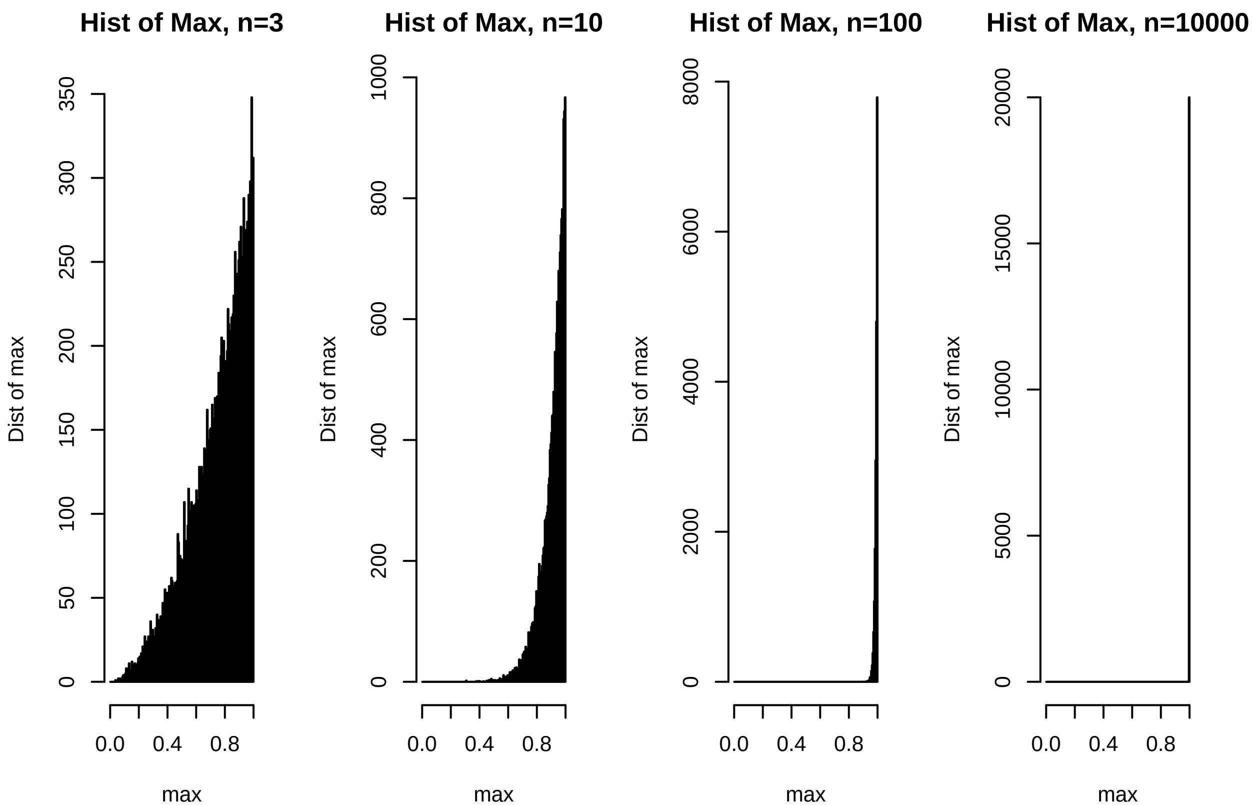
Example - Suppose $X \sim Beta(2,2)$ then $1 - X$ is also distributed as Beta(2,2) (recall the symmetry of the Beta distribution with equal α and β).

Define a sequence of RVs to be $X_n = X$ for all n. Then $X_n \xrightarrow{d} 1 - X \sim Beta(2,2)$.

Now consider convergence in probability, does $X_n \xrightarrow{p} 1 - X$?

Convergence in distribution to a constant - If $Y_n \xrightarrow{d} c$ then $Y_n \xrightarrow{p} c$.

Why does it makes sense that convergence in distribution to a constant implies convergence in probability to that constant? Consider our example where we look at the maximum from a random sample of $U(0, 1)$ RVs. Below are plots of the distribution of the sample max for varying n values.



Another really useful theorem relating convergence results is called Slutsky's Theorem.

Slutsky's Theorem If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, then

- $X_n Y_n \xrightarrow{d} aX$
- $X_n + Y_n \xrightarrow{d} X + a$

Slutsky's theorem is extremely useful for creating hypothesis tests and confidence intervals! Recall the example we talked about when discussing the importance of the CLT:

Example:

- Suppose we know σ and we want inference for μ.
- If we have a random sample $Y_1, ..., Y_n$, we know $\bar{Y} \overset{\bullet}{\sim} N(\mu, \sigma^2/n)$ (μ only unknown)
- We can make an approximate claim about μ via a confidence interval derived from an argument similar to that below:

$$
\begin{aligned}
P(-1.96 < Z < 1.96) &= 0.95 \\
\Leftrightarrow P\left(-1.96 < \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < 1.96\right) &= 0.95 \\
\Leftrightarrow P\left(\bar{Y} - 1.96\sigma/\sqrt{n} < \mu < \bar{Y} + 1.96\sigma/\sqrt{n}\right) &= 0.95
\end{aligned}
$$

- That is, there is a 95% probability the RVs $\bar{Y} - 1.96\sigma/\sqrt{n}$ and $\bar{Y} + 1.96\sigma/\sqrt{n}$ capture μ!

Of course, σ won't be known. Slutsky's theorem allows us to substitute a 'consistent' estimator of σ (i.e. an estimator of σ^2 that converges in probability to σ) and obtain a similar result!

Delta Method

A common place where we'd use the CLT, LLN, and Slutsky's theorem together is when looking at **Delta Method Normality**.

Large Sample Normality and the Delta Method Let $Y_1, Y_2, ...$ be a sequence of RVs such that

$$\sqrt{n}(Y_n - \theta_0) \xrightarrow{d} N(0, \sigma^2) \qquad or \qquad Y_n \overset{\bullet}{\sim} N(\theta_0, \sigma^2/n)$$

For a function g and value θ_0 where $g^{'}(\theta_0)$ exists and is not 0 we have

$$\sqrt{n}(g(Y_n) - g(\theta_0)) \xrightarrow{d} N(0, (g^{'}(\theta_0))^2\sigma^2) \qquad or \qquad g(Y_n) \overset{\bullet}{\sim} N(g(\theta_0), (g^{'}(\theta_0))^2\sigma^2/n)$$

Example - Suppose $Y \sim gamma(n, \lambda)$. Goal: make inference on $\frac{1}{\mu}$. Provide an approximate distribution for $1/Y$ an **estimator** of $1/\mu$.

Example - Let $Y_i \stackrel{iid}{\sim} Ber(p)$ then $\bar{Y} \stackrel{\bullet}{\sim} N(p, \frac{p(1-p)}{n})$. Goal: make inference for $\frac{p}{1-p}$ using $\frac{\bar{Y}}{1-\bar{Y}}$.

Example - Suppose $Y_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$ where $E(Y_i) = \mu \neq 0$. Goal: make inference on $\frac{1}{\mu}$. Provide an approximate distribution for $1/\bar{Y}$ an **estimator** of $1/\mu$.

Recap

We have two big ideas:

- convergence in distribution
- convergence in probability

There are two big theorems:

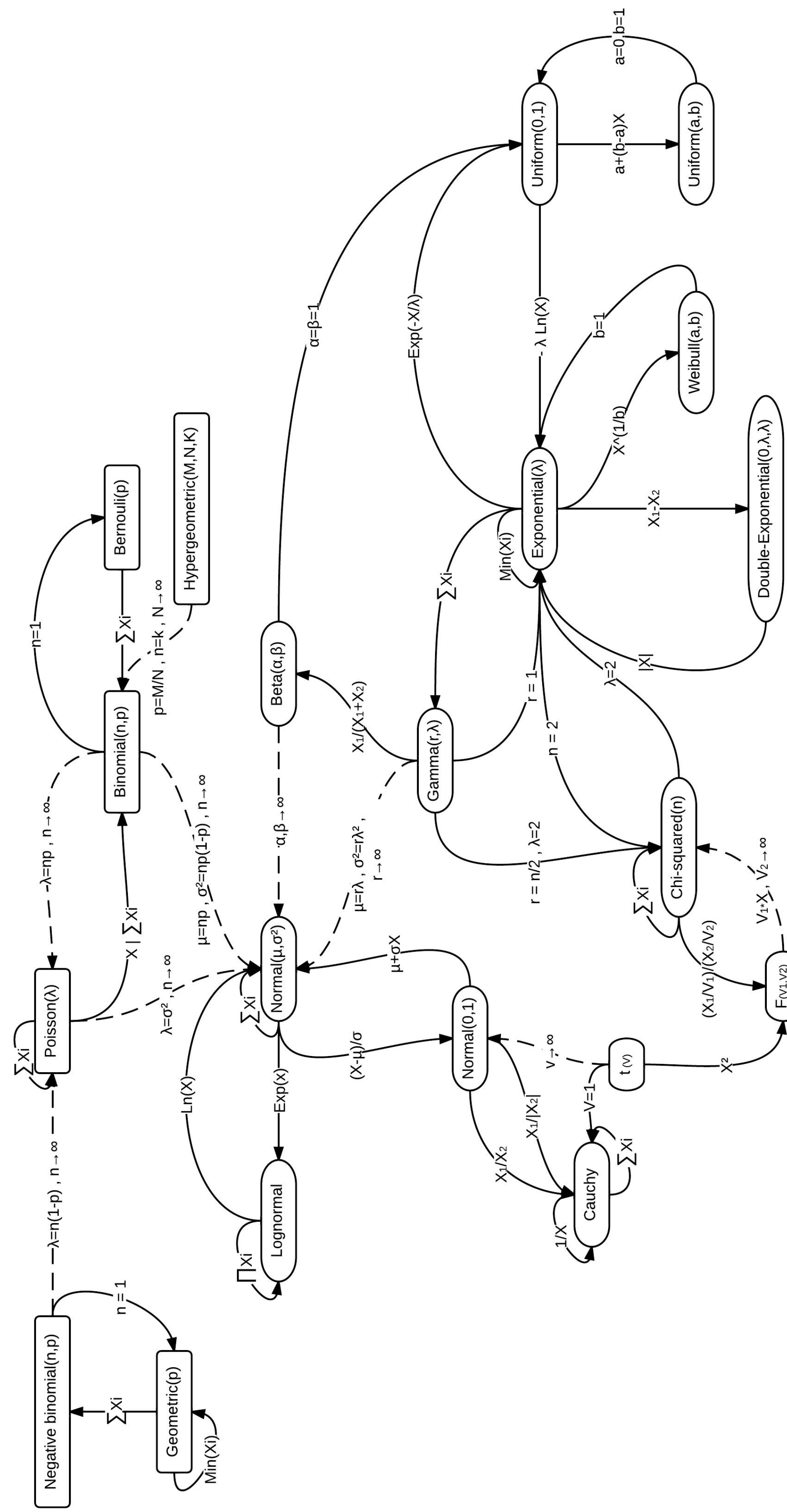
- CLT
- WLLN

Strategies for proving convergence in distribution:

- CLT
- Delta Method Normality
- CDF convergence
- MGF convergence
- Convergence in probability implies convergence in distribution
- Continuity theorem applied to some result

Strategies for proving convergence in probability:

- LLN
- Continuity theorem
- Convergence in distribution to a constant implies convergence in probability
- Resort to the definition of convergence in probability and directly find the probability or use inequalities (Markov's or Chebychev's)


Uniform(0,1)
Uniform(a,b)
a+(b-a)X
a=0,b=1
- λ Ln(X)
Exp(-X/λ)
α=β=1
Weibull(a,b)
b=1
X^(1/b)
Exponential(λ)
Double-Exponential(0,λ,λ)
X1-X2
Min(Xi)
∑Xi
|X|
λ=2
n = 2
r = 1
Gamma(r,λ)
Beta(α,β)
X1/(X1+X2)
Chi-squared(n)
r = n/2 , λ=2
μ=rλ , σ²=rλ² , r→∞
α,β→∞
(X1/V1)/(X2/V2)
F(V1,V2)
V1*X , V2→∞
X²
t(V)
V→∞
V=1
Cauchy
1/X
X1/X2
X1/|X2|
Normal(0,1)
μ+σX
(X-μ)/σ
Normal(μ,σ²)
Exp(x)
Ln(X)
Lognormal
∏Xi
λ=σ² , n→∞
μ=np , σ²=np(1-p) , n→∞
Binomial(n,p)
Bernoulli(p)
n=1
Hypergeometric(M,N,K)
p=M/N , n=k , N→∞
λ=np , n→∞
X | ∑Xi
Poisson(λ)
λ=n(1-p) , n→∞
Negative binomial(n,p)
n = 1
Geometric(p)