Week 11: Convergence in Distribution and Central Limit Theorem

MATH697

Sahir Bhatnagar

November 14, 2017

McGill University

Convergence in Distribution

Convergence in Distribution

Definition 1 (Convergence in Distribution)

Let $X_1, X_2, \ldots$ be a sequence of random variables. Then we say that the **sequence converges in distribution** to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous and we write $X_n \xrightarrow{D} X$

Convergence in Distribution

Definition 1 (Convergence in Distribution)

Let $X_1, X_2, \ldots$ be a sequence of random variables. Then we say that the **sequence converges in distribution** to a random variable X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous and we write $X_n \xrightarrow{D} X$

- Intuitively, $X_1, X_2, \ldots$ converges in distribution to X if for large n, the distribution of X_n is close to that of X
- The importance of this, is that often the distribution of X_n is difficult to work with, while that of X is much simpler
- With X_n converging in distribution to X, however, we can approximate the distribution of X_n by that of X.

Convergence in Probability Implies in Distribution

Theorem 2 (Convergence in Probability Implies in Convergence in Distribution)

If the sequence of random variables $X_1, X_2, \ldots$ converges to a random variable X, the sequence also converges in distribution to X. That is if $X_n \xrightarrow{P} X$ then $X_n \xrightarrow{D} X$

Example 3 (Poisson Approximation to the Binomial)

Suppose that $X_n \sim Binomial(n, \lambda/n)$ and $X \sim Poisson(\lambda)$. We have previously seen that as $n \to \infty$

$$P(X_n = j) = \binom{n}{j} \left(\frac{\lambda}{n}\right)^j \left(1 - \frac{\lambda}{n}\right)^{n-j} \to e^{-\lambda} \frac{\lambda^j}{j!}$$

This implies that $F_{X_n}(x) \to F_X(x)$ and thus $X_n \xrightarrow{D} X$

Central Limit Theorem

- Let $X_1, X_2, \ldots,$ be iid with mean μ and finite variance σ^2. The law of large numbers says that as $n \to \infty$, $\bar{X}_n$ converges to the constant μ (with probability 1).
- But what is its distribution along the way to becoming a constant?
- This is addressed by the central limit theorem (CLT)

Central Limit Theorem (CLT)

- The CLT states that for large n, the distribution of $\bar{X}_n$ (after standardization) approaches a standard Normal distribution.
- By standardization, we mean that we subtract μ, the expected value of $\bar{X}_n$, and divide by $\sigma/\sqrt{n}$, the standard deviation of $\bar{X}_n$.

Theorem 4 (CLT)

As $n \to \infty$

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{D} \mathcal{N}(0,1)$$

From this we can also say that for large n

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$$

Proof:on board

Central Limit Theorem (CLT)

- The CLT states that for large n, the distribution of $\bar{X}_n$ (after standardization) approaches a standard Normal distribution.
- By standardization, we mean that we subtract μ, the expected value of $\bar{X}_n$, and divide by $\sigma/\sqrt{n}$, the standard deviation of $\bar{X}_n$.

Theorem 4 (CLT)

As $n \to \infty$

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{D} \mathcal{N}(0,1)$$

From this we can also say that for large n

$$\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$$

Proof:on board

- In words, the CDF of the left-hand side approaches Φ, the CDF of a standard Normal distribution.
- Starting from virtually no assumptions (other than independence and finite variances), **we end up with normality**

Central Limit Theorem (CLT) for the Sum

- The CLT says that the sample mean $\bar{X}_n$ is approximately Normal
- But since the sum

$$S_n = X_1 + \cdots + X_n = n\bar{X}_n$$

  is just a scaled version of $\bar{X}_n$, the CLT also implies that S_n is approximately Normal
- If X_j have mean μ and variance σ^2, S_n has mean $n\mu$ and variance $n\sigma^2$.

Theorem 5 (CLT for the Sum)

for large n

$$S_n \sim \mathcal{N}(n\mu, n\sigma^2)$$

Running Proportion of Heads Revisited

Before, we used the law of large numbers to conclude that $\bar{X}_n \to 1/2$ as $n \to \infty$. Now, using the central limit theorem, we can say more: $E(\bar{X}_n) = 1/2$ and $Var(\bar{X}_n) = 1/(4n)$, so for large n,

$$\bar{X}_n \sim \mathcal{N}\left(\frac{1}{2}, \frac{1}{4n}\right)$$

Poisson Convergence to Normal

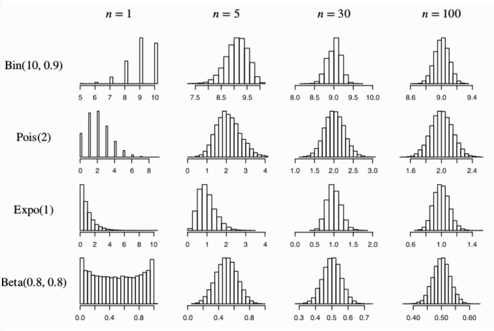
Let $Y \sim Pois(n)$. Using MGFs it can be shown that Y can be expressed as a sum of n iid Poisson(1) random variables. Therefore by CLT, for large n,

$$Y \sim \mathcal{N}(n, n)$$

Let $Y \sim \mathit{Binomial}(n, p)$. Using MGFs it can be shown that Y can be expressed as a sum of n iid Bernoulli(p) random variables. Therefore by CLT, for large n,

$$Y \sim \mathcal{N}(np, np(1-p))$$

Histograms of the distribution of $\bar{X}_n$ for different starting distributions of the X_j (indicated by the rows) and increasing values of n (indicated by the columns). Each histogram is based on 10,000 simulated values of $\bar{X}_n$. Regardless of the starting distribution of the X_j, the distribution of $\bar{X}_n$ approaches a Normal distribution as n grows.

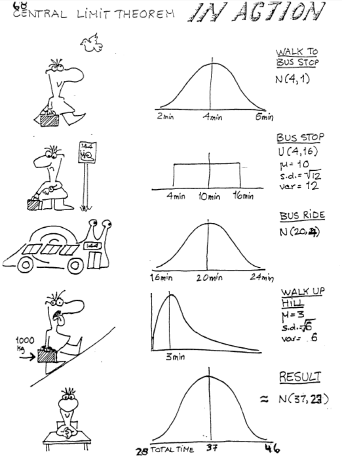


CLT Exercise

One way to visualize the CLT for a distribution of interest is to plot the distribution of $\bar{X}_n$ for various values of n. To do this, we first have to generate iid $X_1, \ldots, X_n$ a bunch of times from our distribution of interest. For example, suppose that our distribution of interest is $Unif(0, 1)$, and we are interested in the distribution of $\bar{X}_{12}$, i.e., we set $n = 12$. In the following code, we create a matrix of iid standard Uniforms. The matrix has 12 columns, corresponding to X_1 through X_{12}. Each row of the matrix is a different realization of X_1 through X_{12}.

```
nsim <- 10^4
n <- 12
x <- matrix(runif(n*nsim), nrow=nsim, ncol=n)
xbar <- rowMeans(x)
hist(xbar)
# change runif for rexp. what do you notice?
```

CLT In Action



1 http://www.medicine.mcgill.ca/epidemiology/Joseph/courses/EPIB-607/notes.pdf

CLT In Action Exercise

1. Set $n = 10$
2. Simulate data from various distributions, each representing a different part of the journey to McGill. Create a `data.frame` of the simulated times and add a column for the total transit time

```
walk <- rnorm(n, 4, 1) ; bus <- runif(n, 4, 16)
ride <- rpois(n, 8); climb <- rgamma(n, shape = 6, scale = 0.5)
fall <- rexp(n, rate = 4)
```

3. Calculate the theoretical means and variances for each of the distributions. Use the CLT to determine the mean and variance of the total transit time.

```
mean.walk=4; mean.bus=(4+16)/2; mean.ride=8; mean.climb=6*0.5; mean.fall=1/4
var.walk=1; var.bus=(16-4)^2/12; var.ride=8; var.climb=6*0.5^2; var.fall=1/4^2
```

4. Plot a histogram of the total transit times and superimpose the density of the theoretical distribution of the sum. Repeat the whole exercise for $n = 50, 100, 200, 500, 1000, 2000$.

```
hist(DT$transit_time, freq=FALSE)
curve(dnorm(x,mean=mean.transit,sd=sqrt(var.transit)),0,50,
      add=TRUE,lwd=2,col="red")
```

Session Info

```
devtools::session_info()
```

```
##  setting  value
##  version  R version 3.4.1 (2017-06-30)
##  system   x86_64, linux-gnu
##  ui       X11
##  language en_US
##  collate  en_US.UTF-8
##  tz       Canada/Eastern
##  date     2017-11-20
##
##  package     * version    date       source
##  abind         1.4-5      2016-07-21 cran (@1.4-5)
##  arm           1.9-3      2016-11-27 cran (@1.9-3)
##  assertthat    0.2.0      2017-04-11 CRAN (R 3.4.1)
##  backports     1.1.0      2017-05-22 cran (@1.1.0)
##  base        * 3.4.1      2017-07-08 local
##  bindr         0.1        2016-11-13 CRAN (R 3.4.1)
##  bindrcpp      0.2        2017-06-17 CRAN (R 3.4.1)
##  blme          1.0-4      2015-06-14 cran (@1.0-4)
##  broom         0.4.2      2017-02-13 CRAN (R 3.4.1)
```