Limit Theorems

Ignacio Cascos

2018

Outline

Introduction

We will explore the limit behaviour of sequences of random variables. We are specifically interested in the convergece of estimators (e.g the sample mean) to a given number and their approximate distributions when built out of a large sample.

Before starting with convergence concepts, we will comment two useful probabilistic inequalities (Markov's and Chebishev's) that relate probabilities with means and variances!!

6.1 Markov and Chebishev inequalities

Markov's inequality

If X is a nonnegative random variable, for any value $a > 0$,

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \,.$$

Given $a > 0$, define r.v. $Y = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$.

Since $X \geq a$, it holds $Y \leq X/a$, and then $P(X \geq a) = \mathbb{E}[Y] \leq \mathbb{E}[X]/a$.

Example (Markov's inequality)

Example 1. Factory

The number of items produced in a factory during a week is a random variable with mean 50.
What can you say about the probability that this week's production will be at least 75?
Denote by X the production in a week,

$$P(X \geq 75) \leq \frac{\mathbb{E}[X]}{75} = \frac{50}{75} = \frac{2}{3}.$$

- What is the probability if $X \sim \mathrm{U}(50 - x, 50 + x)$? Compute in terms of x.
- What is the mean of X if $P(X = 0) = 1/3$ and $P(X = 75) = 2/3$?

Example (Markov's inequality)

Example 2. Pau Gasol

Pau Gasol averaged 10.065 points per game during the NBA regular season 2017-18. What can we say about the proportion of days at wich he scored at least 20 points?

Denote by Y the points Pau Gasol scores in a game,

$$P(Y \geq 20) \leq \frac{\mathbb{E}[Y]}{20} = \frac{10.065}{20} = 0.50325.$$

He actually scored 20 or more points in 4 games out of the 77 games he played during the regular season ($4/77 = 0.052$).

Is the bound very poor? Imagine Pau Gasol had scored 20 points in 38 games, 15 points in one game and 0 points in 38 games. He would have scored 20 or more points in 49.35% of the games he had played.

Markov and Chebishev inequalities

Chebishev's inequality

If X is a random variable with finite mean μ and variance σ^2, for any value $k > 0$,

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

The key step is to apply Markov's inequality to nonnegative random variable $(X - \mu)^2$ in order to obtain

$$P(|X - \mu| \geq k) = P((X - \mu)^2 \geq k^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}.$$

Example (Chebishev's inequality)

Example 1. Factory

The number of items produced in a factory during a week is a random variable with mean 50 and variance 25.
What can you say about the probability that this week's production will be between 40 and 60?
Denote by X the production in a week,

$$P(40 \leq X \leq 60) = P(|X-50| \leq 10) = 1 - P(|X-50| > 10)$$
$$\geq 1 - P(|X-50| \geq 10) \geq 1 - \frac{\sigma_X^2}{10^2} = 0.75.$$

- What is the probability if $X \sim \mathrm{U}(50-5\sqrt{3}, 50+5\sqrt{3})$?
- What are the mean and variance of X if $P(X=50)=3/4$ and $P(X=40)=P(X=60)=1/8$?

Example (Chebishev's inequality)

Example 2. Pau Gasol

Pau Gasol averaged 10.065 points per game during the NBA regular season 2017-18 with variance 31.6.
What can we say about the proportion of games at which he scored between 3 and 18 points?
Denote by Y the points Pau Gasol scores in a game,

$$P(3 \leq X \leq 18) = P(|Y - 10.065| \leq 8) = 1 - P(|Y - 10.065| > 8)$$
$$\geq 1 - P(|Y - 10.065| \geq 8) \geq 1 - \frac{\sigma_Y^2}{8^2} = 0.50625.$$

He actually scored between 3 and 18 points in 63 games out of the 77 games he played ($63/77 = 0.818$).

Chebishev's inequality

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

$$P(|X-\mu| < k) \geq 1-\frac{\sigma^2}{k^2}$$

k	σ^2/k^2	$1-\sigma^2/k^2$
σ	1	0
2σ	1/4	3/4
3σ	1/9	8/9
4σ	1/16	15/16
5σ	1/25	24/25

$$P(\mu - r\sigma < X < \mu + r\sigma) \geq 1-\frac{1}{r^2}$$

6.2 Weak LLN (convergence in probability)

Convergence in probability

A sequence of random variables $\{X_n\}_n$ **converges in probability** to a constant $a \in \mathbb{R}$ ($X_n \xrightarrow{Pr} a$) if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - a| \geq \varepsilon) = 0\,.$$

Weak Law of Large Numbers

If $\{X_n\}_n$ is a sequence of *independent and identically distributed* random variables with $\mathbb{E}[X_i] = \mu$, then for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \varepsilon) = 0\,,$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

6.2 Weak LLN (convergence in probability)

WLLN for a r.v. with finite second moment

If $\{X_n\}_n$ is a sequence of *independent and identically distributed* random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2$, then

$$\mathbb{E}[\overline{X}_n] = \mu;$$
$$\mathrm{Var}[\overline{X}_n] = \sigma^2/n.$$

Apply now Chebishev's inequality to $\overline{X}_n$ in order to obtain:

$$P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n\to\infty} 0.$$

6.2 Weak LLN and convergence in probability

Continuous mapping Theorem

If $X_n \xrightarrow{Pr} a$, $Y_n \xrightarrow{Pr} b$ and $g : \mathbb{R}^2 \mapsto \mathbb{R}$ is a *continuous* function, then $g(X_n, Y_n) \xrightarrow{Pr} g(a, b)$.
Consider now $X_n \xrightarrow{Pr} a$ and $Y_n \xrightarrow{Pr} b$ - $X_n + Y_n \xrightarrow{Pr} a + b$; - $X_n Y_n \xrightarrow{Pr} ab$; - $X_n / Y_n \xrightarrow{Pr} a/b$ if $b \neq 0$.

Convergence in probability to a random variable

$\{X_n\}_n$ **converges in probability** to r.v. X ($X_n \xrightarrow{Pr} X$) if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0\,.$$

6.2 Weak LLN and consistency of estimators

Given a random sample $X_1, \ldots, X_n$ drawn from some population $X \sim F_\theta$ whose distribution depends on a parameter θ. A statistic $\hat{\theta}$ that is used to *estimate* θ (approximate it) is called **estimator** of θ.

An estimator is **(weakly) consistent** if $\hat{\theta} \xrightarrow{Pr} \theta$.

- The *sample mean* is a consistent estimator of the *population mean*, $\overline{X}_n \xrightarrow{Pr} \mu$.
- The *sample variance* is a consistent estimator of the *population variance*, $S_n^2 \xrightarrow{Pr} \sigma^2$.
- The *sample proportion* is a consistent estimator of the *population proportion*, $\hat{p} \xrightarrow{Pr} p$.

6.3 Central Limit Theorem (convergence in distribution)

Convergence in distribution

A sequence of random variables $\{X_n\}_n$ with cdfs F_n **converges in distribution (or law)** to r.v. X with cdf F ($X_n \xrightarrow{d} X$) if for every continuity point x of F, $\lim_{n\to\infty} F_n(x) = F(x)$.

Central Limit Theorem (Lyapunov)

If $\{X_n\}_n$ is a sequence of *iid* r.v.s with mean μ and variance $\sigma^2 < \infty$, then

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z\,,$$

where $Z \sim \mathrm{N}(0,1)$. Under some extra condition, the identical distribution assumption of the X_is can be dropped, and if $\mathbb{E}[X_i] = \mu_i$ and $\mathrm{Var}[X_i] = \sigma_i^2$,

$$\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \xrightarrow{d} Z\,.$$

6.3 Central Limit Theorem (convergence in distribution)

If $\mu = 0$, $\sigma = 1$, and M is the MGF of X_i, then $M_{\sum_{i=1}^{n} X/\sqrt{n}}(t) = M(t/\sqrt{n})^n$.

Denote $L(t) = \log M(t)$ and observe $L(0) = 0$, $L'(0) = 0$, and $L''(0) = 1$.

$$\begin{aligned}
\lim_{n\to\infty} \frac{L(t/\sqrt{n})}{n^{-1}} &= \lim_{n\to\infty} \frac{-L'(t/\sqrt{n})n^{-3/2}t}{-2n^{-2}} \\
&= \lim_{n\to\infty} \frac{L'(t/\sqrt{n})t}{2n^{-1/2}} \\
&= \lim_{n\to\infty} \frac{-L''(t/\sqrt{n})n^{-3/2}t^2}{-2n^{-3/2}} \\
&= \lim_{n\to\infty} \frac{L''(t/\sqrt{n})t^2}{2} = \frac{t^2}{2}.
\end{aligned}$$

Applying L'Hospital's rule twice. We conclude $\lim_{n\to\infty} M(t/\sqrt{n})^n = e^{t^2/2}$.

Example: Factory (1/2)

The number of items produced in a factory during a week is a random variable with mean 50 and variance 25.

The factory is open 49 weeks every year. What can you say about the probability that this year's production will be between 2380 and 2520?

Denote by X_i the production in the i-th week and by $X = \sum_{i=1}^{49} X_i$ the total production.

Chebishev's inequality

$$\mathbb{E}[X] = 49 \times 50 = 2450\,, \quad \mathrm{Var}[X] = 49 \times 25 = 1225 = 35^2$$

$$\begin{aligned} P(2380 \le X \le 2520) &= P(|X - 2450| \le 70) = 1 - P(|X - 2450| > 70) \\ &\ge 1 - P(|X - 2450| \ge 70) \ge 1 - \frac{\sigma_X^2}{70^2} = 1 - \frac{49 \times 25}{4900} = 0.75. \end{aligned}$$

Example: Factory (2/2)

The number of items produced in a factory during a week is a random variable with mean 50 and variance 25.

The factory is open 49 weeks every year. What can you say about the probability that this year's production will be between 2380 and 2520?

Denote by X_i the production in the i-th week and by $X = \sum_{i=1}^{49} X_i$ the total production.

Central Limit Theorem

$$X \approx \mathrm{N}(2450, 35)$$

$$P(2380 \leq X \leq 2520) = P\left(\frac{2380 - 2450}{35} \leq Z \leq \frac{2520 - 2450}{35}\right)$$
$$= P(-2 \leq Z \leq 2) = 0.9545.$$

6.3 Central Limit Theorem (convergence in distribution)

Normal approximation to the Binomial and Poisson distributions

- If $X \sim \mathrm{B}(n, p)$, then $X \approx \mathrm{N}(np, \sqrt{np(1-p)})$ (good approximation if $n \geq 50$ and $0.4 < p < 0.6$ or $np > 5$ and $n(1-p) > 5$).
- If $X \sim \mathcal{P}(\lambda)$, then $X \approx \mathrm{N}(\lambda, \sqrt{\lambda})$ (good approximation if $\lambda \geq 10$).
- **Continuity corrections** for the previous *discrete* distribution models. If k is an integer
    - $P(X = k) = P(k - 0.5 < X < k + 0.5)$
    - $P(X \leq k) = P(X < k + 0.5)$
    - $P(X < k) = P(X < k - 0.5)$
    - $P(X \geq k) = P(X > k - 0.5)$
    - $P(X > k) = P(X > k + 0.5)$

Example: Potholes

Past experience suggests that there are, on average, 2 potholes per mile of highway after a certain amount of usage, and that the random variable 'number of potholes' can be modeled by means of a Poisson distribution.

A group of workers is hired to repair 100 potholes. How many miles must be inspected so that with probability 0.95 at least 100 potholes are found?

Denote $X \equiv$'number of potholes in k miles', $X \sim \mathcal{P}(\lambda = 2k)$, so $X \approx \mathrm{N}(\mu = 2k, \sigma = \sqrt{2k})$. We have the equation $P(X \geq 100) = 0.95$.

$$P(X \geq 100) = P(X > 99.5) = P(Z > (99.5 - 2k)/\sqrt{2k}) = 0.95$$

In conclusion $(99.5 - 2k)/\sqrt{2k} = -1.645$, so $k = 58.65875$ miles must be inspected.

```
1-ppois(99,lambda=2*58.65875)
```

```
## [1] 0.9529045
```

Asymptotic distribution of several estimators (1/2)

Consider $X_1, X_2, \ldots$ iid as r.v. X with mean μ and variance $\sigma^2 < \infty$.

- The *sample mean* is asymptotically normal (1/2)

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} Z.$$

Consider a population with individuals that have some given characteristic with **(population) proportion** p and a random sample for which $\hat{p}$ stands for the **sample proportion** of individuals with the characteristic.

- The *sample proportion* is asymptotically normal,

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \xrightarrow{d} Z.$$

For $Z \sim \mathrm{N}(0, 1)$.

6.3 Central Limit Theorem (convergence in distribution)

Slutsky's Theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{Pr} a$, then - $X_n + Y_n \xrightarrow{d} X + a$; - $X_n Y_n \xrightarrow{d} aX$; - $X_n / Y_n \xrightarrow{d} X/a$ if $a \neq 0$.

Asymptotic distribution of several estimators (2/2)

- The *sample mean* is asymptotically normal (2/2)

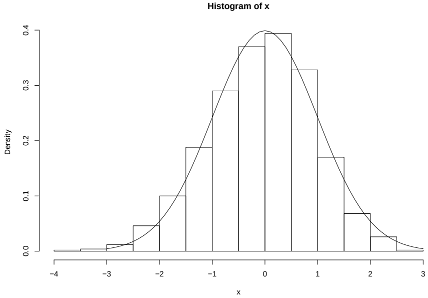
$$\sqrt{n}(\overline{X}_n - \mu)/S_n \xrightarrow{d} Z.$$

- The *sample variance* is asymptotically normal

$$\sqrt{n}(S_n^2 - \sigma^2)/\sqrt{m_4 - \sigma^4} \xrightarrow{d} Z.$$

Sample mean exponential population unknown variance

```
set.seed(1)
n=100;lambda=2;x=vector(length=1000)
for(i in 1:1000){simul=rexp(n,rate=lambda)
  x[i]=(mean(simul)-1/lambda)/sqrt(var(simul)/n)}
hist(x,probability=T)
t=seq(-3,3,by=.1)
lines(t,dnorm(t))
```



6.4 Strong LNN (almost sure convergence)

Almost sure convergence

A sequence of random variables $\{X_n\}_n$ **converges almost surely** (or with probability 1) to a constant $a \in \mathbb{R}$ if,

$$P\left(\lim_{n\to\infty} X_n = a\right) = 1\,.$$

Strong Law of Large Numbers

If $\{X_n\}_n$ is a sequence of *independent and identically distributed* random variables with $\mathbb{E}[X_i] = \mu$, then

$$P\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1\,,$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i,$.

SLNN for a r.v. with finite fourth moment

If $\{X_n\}_n$ is a sequence of *iid* r.v.s with $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^4] = \mu_4 < \infty$,

$$\begin{aligned}
\mathbb{E}\left[\frac{\left(\sum_{i=1}^n X_i\right)^4}{n^4}\right] &= \left(n\mathbb{E}[X_i^4] + 6\binom{n}{2}\mathbb{E}[X_i^2]\mathbb{E}[X_j^2]\right)/n^4 \\
&= \left(n\mu_4 + 3n(n-1)\mu_2^2\right)/n^4 \\
&\leq \frac{3n-2}{n^3}\mu_4.
\end{aligned}$$

Now

$$\mathbb{E}\sum_{n=1}^{\infty}\left[\frac{\left(\sum_{i=1}^n X_i\right)^4}{n^4}\right] = \sum_{n=1}^{\infty}\mathbb{E}\left[\frac{\left(\sum_{i=1}^n X_i\right)^4}{n^4}\right] < \infty.$$

Then $\sum_{n=1}^{\infty}\left(\sum_{i=1}^{n} X_i\right)^4/n^4 < \infty$ a.s. and $\lim_{n\to\infty}\left(\sum_{i=1}^{n} X_i\right)^4/n^4 = 0$ a.s. We conclude $\lim_n \overline{X}_n = \lim_n \sum_{i=1}^{n} X_i/n = 0$ a.s.

6.4 Strong LNN (almost sure convergence)

Convergence in probabiliy vs almost sure convergence

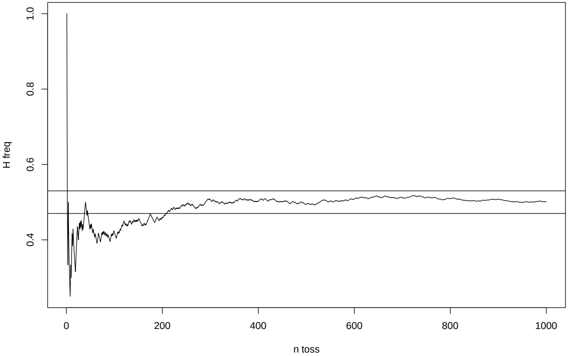
The **convergence in probabilty is implied by the almost sure convergence** while the reverse does not hold, and thus the names of **Weak** LLN and **Strong** LLN.
The almost sure convergence of an estimator to the value of the parameter is referred to as **strong consistency**.

Example of sequence of random variables converging in probability, but not almost surely.

The sequence of independent r.v.s $\{X_n\}_n$ with distributions $P(X_n = 1) = 1/n$ and $P(X_n = 0) = 1 - 1/n$ converges in probability to 0, but it does not converge to 0 almost surely.

Almost sure convergence

```
set.seed(10)
plot(cumsum(rbinom(rep(1,1000),size=1,prob=0.5))/(1:1000),
     type="l",ylab="H freq",xlab="n toss")
abline(h=c(0.53,0.47))
```



Convergence in probabiliy

```
for(i in 1:30){set.seed(i)
 points(cumsum(rbinom(rep(1,1000),size=1,prob=0.5))/(1:1000),
        type="l",col=i)}
```