

Asymptotic Statistics

Daniel Frederico Lins Leite

May 1, 2019

Contents

1	Introduction	5
2	Asymptotic Behaviour of Estimators	7
2.1	Introduction	7
2.2	Markov's Inequality Intuition	9
2.3	Markov's Inequality	11
2.4	Chebyshev's Inequality	13
2.5	Markov's and Chebyshev's Inequalities	14
2.6	68-95-99.7 rule	14
2.7	Summary	14
2.7.1	Markov's Inequality	14
2.7.2	Chebyshev's Inequality	15
3	Convergence	17
3.1	Convergence in probability of a random variable	17
3.2	Convergence in probability to a random variable	17
3.3	Convergence in probability of a random variable to a constant . .	17
3.4	Convergence in distribution of a random variable	17
3.5	Slutsky's Theorem	17
4	Characteristic Functions	19
4.1	Introduction	19
5	Weak Law of Large Numbers	21
5.1	Introduction	21
5.2	Proof of the weak law of large numbers	21
5.3	Proof using characteristic functions	21
6	Central Limit Theorems	23
6.1	Introduction	23
6.2	Proof	23
6.3	The Characteristic Function of a Normal Random Variable . . .	23
7	Applications	25
7.1	Two-sided, Two-sample Tests	25

Chapter 1

Introduction

The idea of this crash course on asymptotic Statistics is to have all proofs and ideas related with the one of the three most important aspects of statistics, being estimation. The other two being confidence interval and hypothesis testing, off course.

The idea is to present as gentle as possible the needed concepts to follow a more thorough course. Here one will find the concepts of "asymptotic", "estimators", Markov and Chebyshevs inequalities, convergence in the sense for Statistics, characteristic functions, the weak law of large numbers and the Central Limit Theorem.

Chapter 2

Asymptotic Behaviour of Estimators

2.1 Introduction

As "Asymptotic Statistics" we understand the study of Probabilities and Statistics as one of its parameters go to infinity. We will study for example:

$$\lim_{n \rightarrow \infty} P(X_n \in A) \quad (2.1)$$

or

$$\lim_{n \rightarrow \infty} P((X_n - X) \in A) \quad (2.2)$$

Here $P(x \in A) = \int_A f(x)dx$. The lower-case f has a especial meaning because that is how we label the PDF of a distribution. We use upper-case F to mean de CDF. We also know that to be a valid "random variable":

$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad f(x) \geq 0 \quad (2.3)$$

Where X_n means off course a "random variable", pure in the sense that is obeys a known distribution or a "transformed random variable". We will be interested to know if a "random variable" converge to a constant, to another "random variable" or if it diverge. Throughout this book we will see that we can conclude different things if any of these cases are proved as true.

One of the most interesting, or one the most useful, "random variables" is the "estimator". We call "estimator" any "statistic" that does not depend on any of the "population parameters" an possible estimator. If on top of that we define that this "statistic" will be used as an "estimation" of the "population parameter", we have a "estimator".

A famous statistic is the "sample mean" that we will define as:

$$\begin{aligned}
X_i &\sim P_\theta \\
S_n &= \sum_{\forall i} X_i \\
\bar{X}_n &= \frac{S_n}{n}
\end{aligned}$$

We understand the expressions above as:

- X_i is a "random variable" that follows a distribution described with a parameter θ , that can be a number or various numbers;
- The set of all X_i are i.i.d.
- S_n is another "random variable". Being the sum of n "random variables" allows us to know the distribution of S_n in some cases
- \bar{X}_n is another "random variable" that applies the "transformation" $1/x$ to the "random variable" S_n

Being random variables, X_i , S_n , \bar{X}_n , allow them to have "expected values", "variations" and any other statistics themselves. This is a powerfull source of confusion. So we can have:

$$\begin{aligned}
E[X_i] &= a \text{ or } f_a(i) \\
E[S_n] &= b \text{ or } f_b(n) \\
E[\bar{X}_n] &= c \text{ or } f_c(n)
\end{aligned}$$

In these cases, a , b , c are not "random variables". They are constants, or functions that depend on the "random variables" "parameter". The same can be said about their "variance" and other statistics.

Remember that "Expected Value" is:

$$E[X] = \int_{-\infty}^{\infty} x * f(x) dx \quad (2.4)$$

Actually this is just the "First Moment" of X . We can the n -th "moment" as:

$$\int_{-\infty}^{\infty} x^n * f(x) dx \quad (2.5)$$

or

$$\begin{aligned}
E[X^1] &= \int_{-\infty}^{\infty} x^1 * f(x) dx \\
E[X^2] &= \int_{-\infty}^{\infty} x^2 * f(x) dx \\
&\dots \\
E[X^n] &= \int_{-\infty}^{\infty} x^n * f(x) dx
\end{aligned}$$

And this allows us to calculate the variance as, the "second moment" minus the square of the "first moment":

$$\text{VAR}[X] = E[X^2] - E[X]^2 \quad (2.6)$$

$$(2.7)$$

or

$$\text{VAR}[X] = E[(X - E[X])^2] \quad (2.8)$$

And we also have the "standard deviation".

$$\text{SD}[X] = \sqrt{\text{VAR}[X]} \quad (2.9)$$

2.2 Markov's Inequality Intuition

What can you say about a "Probability distribution" if you only have its mean? Probably not much. But not much is more than nothing. If we imagine a standard dice with six faces, but not necessarily fair, each face having the same probability. The question is, knowing just the mean, can we affirm with a 100% certainty that the dice is not fair?

For example, if someone tell us that the $E[X_i]$, X_i being a roll of the dice is 6. We know that the only way of this happening is when $P(X = 6) = 1$ and $P(X < 6) = 0$ because of the "Expected Value" definition.

$$\begin{aligned}
E[X] &= \sum_{i=1}^6 i * P(X = i) = 6 \\
&= 6 * P(X = 6) = 6 \\
P(X = 6) &= \frac{6}{6} \\
P(X = 6) &= 1
\end{aligned}$$

So it seems that knowing only the $E[X]$ allows us to take some conclusions of the distribution. But this is a extreme case. Can we take conclusions if for example the $E[X]$ is the same as the "expected value" when the dice is fair? Let us see.

We know that the $E[X]$ is:

$$E[X] = \sum_{i=1}^6 i * P(X = i)$$

And we know that $i \in 1, 2, 3, 4, 5, 6$ or $i > 0$. We know that i is positive. We also know that $P(X = i) \in [0, 1]$. It is also positive. The multiplication of two positive values is also positive. And the sum of two positive values is also positive.

$$\begin{aligned} i &> 0 \\ P(X = i) &> 0 \\ a * b &> 0, \text{ if } a > 0 \text{ and } b > 0 \\ a + b &> 0, \text{ if } a > 0 \text{ and } b > 0 \end{aligned}$$

Which give us that

$$\begin{aligned} E[X] &= \sum_{i=1}^6 i * P(X = i) \\ &= \sum_{i=1}^5 i * P(X = i) + 6 * P(X = 6) \end{aligned}$$

We know that this summation is going to be greater than zero. Which make the $E[X]$ greater than or equal $6 * P(X = 6)$. In the extreme case above, it was equal because the summation was zero. But the summation can be greater than zero (never lower).

$$\begin{aligned} E[X] &= \sum_{i=1}^6 i * P(X = i) \\ E[X] &= \sum_{i=1}^5 i * P(X = i) + 6 * P(X = 6) \\ E[X] &\geq 6 * P(X = 6) \\ \frac{E[X]}{6} &\geq P(X = 6) \\ P(X = 6) &\leq \frac{E[X]}{6} \end{aligned}$$

Which is quite surprising if you think. But looking from another perspective, it is quite obvious that there is a relationship with the $P(X = 6)$ and the "expected value".

2.3 Markov's Inequality

This relationship was discovered by two mathematicians:
Andrey Markov



and Pafnuty Chebyshev



And it is generalized as:

- X is a nonnegative random variable;
- $a > 0$

We now that:

$$\begin{aligned}
 E[X] &= \int_{-\infty}^{\infty} x * f(x) dx \\
 &= \int_0^{\infty} x * f(x) dx && \text{because of } X > 0 \\
 &= \int_0^a x * f(x) dx + \int_a^{\infty} x * f(x) dx && \text{because of } a > 0
 \end{aligned}$$

We know that $\int_0^a x * f(x) dx > 0$, so

$$\begin{aligned}
 E[X] &= \int_0^a x * f(x) dx + \int_a^{\infty} x * f(x) dx \\
 E[X] &= b + \int_a^{\infty} x * f(x) dx && b > 0 \\
 E[X] - b &= \int_a^{\infty} x * f(x) dx \\
 E[X] &\geq \int_a^{\infty} x * f(x) dx && b > 0
 \end{aligned}$$

Before the next step, we must realize that:

$$\begin{aligned}
 \int_a^\infty x * f(x) dx &= \int_a^\infty (a + (x - a)) * f(x) dx \\
 &= \int_a^\infty a * f(x) dx + \int_a^\infty (x - a) * f(x) dx
 \end{aligned}$$

We must realize the in the second integration, $x \geq a$ and $f(x) > 0$ which makes the second integration greater than or equal to zero. This allow us to:

$$\begin{aligned}
 \int_a^\infty x * f(x) dx &= \int_a^\infty a * f(x) dx + \int_a^\infty (x - a) * f(x) dx \\
 \int_a^\infty x * f(x) dx &= \int_a^\infty a * f(x) dx + c \\
 \int_a^\infty x * f(x) dx - c &= \int_a^\infty a * f(x) dx \\
 \int_a^\infty x * f(x) dx &\geq \int_a^\infty a * f(x) dx
 \end{aligned}$$

We can use this fact with the fact that if $a > b$ and $b > c$ then $a > c$, and finish our proof

$$\begin{aligned}
 E[X] &\geq \int_a^\infty x * f(x) dx && b > 0 \\
 E[X] &\geq \int_a^\infty a * f(x) dx && \text{see above} \\
 E[X] &\geq a * \int_a^\infty f(x) dx \\
 E[X] &\geq a * P(X \geq a) \\
 \frac{E[X]}{a} &\geq P(X \geq a)
 \end{aligned}$$

$$P(X \geq a) \leq \frac{E[X]}{a} \tag{2.10}$$

Which is the obvious generalization we did in the intuition section. But remember that we are studying "Asymptotic Statistics", this is useful for us, because in some cases, the right side will depend on n , and in the limit the right side will converge or diverge. This will give even more information about the "random variable". All of this with just its "expected value".

2.4 Chebyshev's Inequality

One of the interesting aspects of the "Markov Inequality" is that it only contains two assumptions and both are very generic. It only depends on X being non-negative and a being nonnegative. We know that a nonnegative function, when sifted on the X -Axis is still nonnegative. So $X - a$ is still non-negative and the "Markov Inequality" still applies.

$$P((X - b) \geq a) \leq \frac{E[X - b]}{a}$$

Another "transformation" that keeps the non-negativeness of the function is squaring its value.

$$P((X - b)^2 \geq a^2) \leq \frac{E[(X - b)^2]}{a^2}$$

But we do not have any restriction on b here, and we can easily choose the "expected value" of X .

$$P((X - E[X])^2 \geq a^2) \leq \frac{E[(X - E[X])^2]}{a^2}$$

Which gives us the $\text{VAR}[X]$. Although this inequality is often written in a different form.

$$P((X - E[X])^2 \geq a^2) \leq \frac{\text{VAR}[X]}{a^2}$$

But this is still not the most "famous" way this formula is written, and the reason is going to appear in future chapters. To derive the most famous form we are going to apply some transformations only to the left side, because they will not change the boundness of the inequality.

$$P(\sqrt{(X - E[X])^2} \geq \sqrt{a}) \leq \frac{\text{VAR}[X]}{a^2}$$

$$P(|X - E[X]| \geq a) \leq \frac{\text{VAR}[X]}{a^2} \quad \text{no modulus on } a \text{ because } a > 0$$

$$P(|X - E[X]| \geq a) \leq \frac{\text{VAR}[X]}{a^2} \tag{2.11}$$

Which is quite fascinating, because it gives you a known bound when you know the distribution "expected value" and "variance". This formula will demonstrate itself very powerful and useful in the study of sampling and convergence.

We can also interpret that it is "harder", less probable, of being n "standard deviations" from the "expected value". What is very intuitive, if you stop to

think about it. What you can not do, is forget the "standard deviation" and say that it is "harder" to be far from the "expected value". This is not true. Without the "standard deviation", or "variance" this phrase, does not make any sense. Stop to think of a distribution with "infinite" "variance". It is very "easy" to be "infinitely" far from the "expected value".

2.5 Markov's and Chebyshev's Inequalities

If you search you will probably find another formula for both inequalities. They are all equivalent, off course.

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

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

$$P(X \geq a * E[X]) \leq \frac{1}{a} \quad (2.12)$$

$$P((X - E[X])^2 \geq a^2) \leq \frac{E[(X - E[X])^2]}{a^2}$$

$$P((X - E[X])^2 \geq a^2 * \text{VAR}[X]) \leq \frac{E[(X - E[X])^2]}{a^2 * \text{VAR}[X]}$$

$$P((X - E[X])^2 \geq a^2 * \text{VAR}[X]) \leq \frac{\text{VAR}[X]}{a^2 * \text{VAR}[X]}$$

$$P((X - E[X])^2 \geq a^2 * \text{VAR}[X]) \leq \frac{1}{a^2}$$

$$P(\sqrt{(X - E[X])^2} \geq \sqrt{a^2 * \text{VAR}[X]}) \leq \frac{1}{a^2}$$

$$P(|(X - E[X])| \geq a * \text{SD}[X]) \leq \frac{1}{a^2} \quad (2.13)$$

2.6 68-95-99.7 rule

2.7 Summary

2.7.1 Markov's Inequality

$$P(X \geq a) \leq \frac{E[X]}{a} \quad (2.14)$$

or

$$P(X \geq aE[X]) \leq \frac{1}{a} \quad (2.15)$$

2.7.2 Chebyshev's Inequality

$$P(|X - E[X]| \geq a) \leq \frac{\text{VAR}[X]}{a^2} \quad (2.16)$$

or

$$P(|X - E[X]| \geq a * \text{SD}[X]) \leq \frac{1}{a^2} \quad (2.17)$$

Chapter 3

Convergence

3.1 Convergence in probability of a random variable

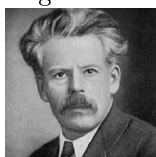
3.2 Convergence in probability to a random variable

3.3 Convergence in probability of a random variable to a constant

3.4 Convergence in distribution of a random variable

3.5 Slutsky's Theorem

Eugen Slutsky



and Harald Cramér



If:

$$X \sim P \tag{3.1}$$

$$c \in \mathbb{R} \tag{3.2}$$

$$X_n \xrightarrow{d} X \tag{3.3}$$

$$Y_n \xrightarrow{d} c \tag{3.4}$$

$$\tag{3.5}$$

then:

$$X_n + Y_n \xrightarrow{d} X + c \tag{3.6}$$

$$X_n * Y_n \xrightarrow{d} X * c \tag{3.7}$$

$$\tag{3.8}$$

Chapter 4

Characteristic Functions

4.1 Introduction

Chapter 5

Weak Law of Large Numbers

5.1 Introduction

5.2 Proof of the weak law of large numbers

5.3 Proof using characteristic functions

Chapter 6

Central Limit Theorems

6.1 Introduction

6.2 Proof

6.3 The Characteristic Function of a Normal Random Variable

Chapter 7

Applications

7.1 Two-sided, Two-sample Tests

In the scenario:

$$X_1, \dots, X_n \sim P_1 \quad (7.1)$$

$$E[X_i] = \mu_1 \quad (7.2)$$

$$VAR[X_i] = \sigma_1^2 \quad (7.3)$$

$$(7.4)$$

$$Y_1, \dots, Y_n \sim P_2 \quad (7.5)$$

$$E[Y_i] = \mu_2 \quad (7.6)$$

$$VAR[Y_i] = \sigma_2^2 \quad (7.7)$$

$$(7.8)$$

the test statistic T_n .

$$\sqrt{n} \frac{(\bar{X}_n - \bar{Y}_n)}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sim N_{\mu, \sigma^2}(0, 1) \quad (7.9)$$

If both variances σ_1^2 and σ_2^2 are unknown we can use the "Plug-In" method using "Slutsky Theorem here".

Given that T_n converges to $N(0, 1)$. We can use the fact that:

$$\frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \xrightarrow{d} 1 \quad (7.10)$$

and

$$\sqrt{n} \frac{(\bar{X}_n - \bar{Y}_n)}{\sqrt{\sigma_1^2 + \sigma_2^2}} * \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \sim N_{\mu, \sigma^2}(0, 1) * 1 \quad (7.11)$$

$$\sqrt{n} \frac{(\bar{X}_n - \bar{Y}_n)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} * \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sim N_{\mu, \sigma^2}(0, 1) * 1 \quad (7.12)$$

$$\sqrt{n} \frac{(\bar{X}_n - \bar{Y}_n)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} * 1 \sim N_{\mu, \sigma^2}(0, 1) * 1 \quad (7.13)$$

$$\sqrt{n} \frac{(\bar{X}_n - \bar{Y}_n)}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \sim N_{\mu, \sigma^2}(0, 1) \quad (7.14)$$