Probability and Statistics

Radjesvarane ALEXANDRE

November 6, 2016

Chapter 1

Probability Spaces

Chapter 2

Random Variables

Chapter 3

2d Random Vectors

Chapter 4

Limit Theorems

Theorem 4.0.1

Weak law of large numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d r.v. such that $E(X_1) = \mu \in \mathbb{R}$. Let

$$S_n = \sum_{k=1}^{n} X_k \text{ and } M_n \equiv \frac{S_n}{n}$$

Then for any constant $\varepsilon > 0$, we have

$$\lim_{n\to+\infty} P(|M_n - \mu| < \varepsilon) = 1$$

Statistics Vocabulary

We say that μ is the mean of the population and that S_n/n is the mean of a random sample of size n of the population. The above theorem says that the mean of the sample converges towards the mean of the population.

In practise, if the mean μ is unknown, we can estimate it by using the mean of a sample of the population. Bigger is the size of the sample, better would be the approximation of μ by the numerical value taken by S_n/n.

Theorem 4.0.2

Strong law of large numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d r.v. such that $E(X_1) = \mu \in \mathbb{R}$ and $Var(X_1) < +\infty$. Then

$$P[\lim_{n\to+\infty} \frac{S_n}{n} = \mu] = 1$$

Remarks 4.0.1

i) In the case of the weak law: we say that S_n/n converges in probability. In the case of the

strong law: we say that S_n/n converges almost surely (towards μ in both cases).
ii) If we apply the strong law of large numbers to the relative frequencies, we find that the relative frequency of an event A converges towards the probability of A with probability 1.

Exemple: Let X_1, X_2, ... be a sequence of i.i.d. r.v. following an exponential law with parameter λ. Let us introduce the characteristic function

$$I_k = 1 \text{ if } X_k > 1 \text{ and } 0 \text{ otherwise}$$

for all k. As I_k follows a Bernoulli law with parameters $p = P(X_k > 1) = e^{-1}$, the strong law of large numbers says that

$$\lim_{n \to +\infty} \sum_{k=1}^{n} \frac{I_k}{n} = E(I_k) = p = e^{-1}$$

with probability 1.

⊙

Theorem 4.0.3

Central Limit Theorem

Let X_1, ... X_n, ... be i.i.d. r.v. with finite mean μ and variance σ^2. Set

$$S_n = X_1 + ... + X_n$$

and

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{S_n/n - \mu}{\sigma/\sqrt{n}}$$

Then the distribution function of Z_n converges towards the distribution function of a law $N(0,1)$.

This may be written also as:

$$S_n \sim N(n\mu, n\sigma^2) \text{ for } n \text{ large}$$

or

$$\frac{S_n}{n} \sim N(\mu, \sigma^2/n) \text{ for } n \text{ large}$$

In general, if $n \geq 30$, we may use the gaussian distribution to approximate the exact distribution of Z_n.

One may generalize CLT to the following case:

if X_1, ..., X_n, ... are independent r.v., then S_n/n follows approximatively, for n suffiently large, a gaussian law with parameters

$$\mu = \frac{1}{n} \sum_{k=1}^{n} E(X_k) \text{ and } \sigma^2 = \frac{1}{n^2} \sum_{k=1}^{n} Var(X_k)$$

Thus, we do not need that all r.v. X_k are i.d.

Exemple: Let $X_1, ..., X_n, ...$ be i.i.d. r.v. with same law as a discrete rv X, whose mass function is given by

x	-1	0	2	Σ
$p_X(x)$	$1/2$	$1/8$	$3/8$	1

Here is the table for S_2:

x	-2	-1	0	1	2	4	Σ
$p_{S_2}(x)$	$1/4$	$1/8$	$1/64$	$3/8$	$3/32$	$9/64$	1

Try to compute other cases.

⊙

Exemple: A computer, when adding numbers, rounds up each number to the nearest integer. Assume that round up errors are independent and follow an uniform law over $(-1/2, 1/2)$. If 1500 numbers are added, what is the probability that the total error, in absolute value, will be bigger that 15?

We introduce E for the total error obtained by rounding the 1500 numbers. Then we may write $E = E_1 + ... + E_{1500}$ where E_k is the round up error for the k-th number. As $E_k \sim U(-1/2, 1/2)$, and the E_k are independent, we have

$$E \sim N(1500(0), 1500(1/12)) \text{ approximatively}$$

by CLT: $E(E_k) = 0$ and $var(E_k) = [1/2 - (-1/2)]^2/12$.

We look for

$$P(|E| > 15) \simeq 2[1 - \Phi(1,34)] \simeq_{table} 2(1 - 0,91) \simeq 0,18$$

Careful: it is not true that $E \sim U(-750, 750)$. If it was the case, then we would have obtained the wrong result that

$$P(|E| > 15) = ... = 0,98$$

In fact, the sum of two independent uniform r.v. is no more uniform (otherwise, there would be a contradiction with CLT if we were to add up a large number of uniform i.i.d. r.v.).

Theorem 4.0.4

Another form of CLT *Let X_i be independent r.v., and set*

$$X = X_1 + ... + X_n$$

with mean $\mu = \mu_1 + ... + \mu_n$ and variance $\sigma^2 = \sigma_1^2 + ...\sigma_n^2$. Then, under some general assumptions, the distribution $F_X(x)$ of X converges towards the normal distribution, with the same mean and same variance:

$$F_X(x) \simeq \Phi(\frac{x - \mu}{\sigma})$$

when n increases. That is, if $Z = \frac{X-\mu}{\sigma}$ then

$$F_Z(z) \to \Phi(z) \equiv \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Sufficient conditions for applying this result are:
a) $\sigma_1^2 + ... + \sigma_n^2 \to +\infty$;
b) there exists a number $\alpha > 2$ and a constant K such that

$$\int x^{\alpha} f_i(x)dx < K, \forall i.$$

4.1 Approximation of a binomial law by a gaussian law

Let $X \sim B(n,p)$.

As we may represent X by the sum of n i.i.d. Bernoulli r.v., we may use CLT to approximate the distribution of X.

Indeed we may write $X = \sum_{k=1}^{n} X_k$, with X_k being 1 if the k-th trial is a success. That is the binomial law counts for the number of 1 obtained after n Bernoulli trials.

de Moivre-Laplace approximation

If n is large enough and p close enough to $1/2$, we may write

$$p_X(k) \simeq f_Z(k)$$

where $Z \sim N(np, npq)$, as $E(X) = np$ and $Var(X) = npq$).

This approximation is good if $min\{np, nq\} \geq 5$.

Thus, if $p = 1/2$, then $n \geq 10$ is enough to get a good approximation.

If $p = 1/100$, then $n \geq 100$! In fact, if p is too small, or too close to 1, Poisson approximation (see below) is used instead.

Exemple : If 20% of diodes manufactured by a specific machine are defective, what is the probability that in a lot of 100 diodes taken at random (and without reset) produced by this machine, we have exactly 15 defective?

If X denotes the number of defective diodes, among the 100 examined, then it follows a binomial law with parameters $n = 100$ and $p = 0,20$. We look for

$$P(X = 15) = P(14,5 \leq X \leq 15,5) \simeq P(14,5 \leq Z \leq 15,5)$$

with $Z \simeq N(20, 16)$ and we find

$$= \Phi(1,375) - \Phi(1,125) \simeq 0,9155 - 0,8697 \simeq 0,0458$$

If we had used the formula for the density, we would have obtained

$$f_Z(15) = \frac{1}{\sqrt{2\pi}.4} exp\{-\frac{1}{2}\frac{(15-20)^2}{16}\} \simeq 0,0457$$

In fact, one can show that with the exact binomial law, we get nearly $0,0481$. If we use Poisson instead we find $0,0516$. That is, the success probability $p = 0,20$ is too large in order to use Poisson.

We may be a little more precice mathematically, without using CLT (but in fact one of its various proofs), and based on Stirling formula

$$n! \simeq n^n e^{-n}\sqrt{2\pi n} \text{ when } n \to +\infty$$

Proposition 4.1.1

De Moivre-Laplace If npq is large w.r.t. 1, then

$$C_k^n p^k q^{n-k} \text{ is equivalent to } \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq}$$

for k in an interval of size $\sqrt{npq}$ centered at np, and for $n \to +\infty$.

Exemple : We throw a coin 1000 times. We look for the probability p_a to obtain 500 times "head" and the probability p_b to obtain 510 times "head".

Here $p = q = 0,5$, $n = 1000$ and $npq = 5\sqrt{10}$. For p_a, we have $k = 500$, then $k - np = 0$ and then de Moivre-Laplace gives

$$p_a \simeq \frac{1}{\sqrt{2\pi npq}} = \frac{1}{10\sqrt{5\pi}} \simeq 0,0252$$

For p_b, we obtain $p_b \simeq 0,0207$.

⊙

It follows from de Moivre-Laplace (using integral calculus) that

$$\sum_{k=k_1}^{k_2} C_k^n p^k q^{n-k} \simeq \Phi(\frac{k_2 - np}{\sqrt{npq}}) - \Phi(\frac{k_1 - np}{\sqrt{npq}})$$

This approximation is good if npq is very large compared to 1 and if the differences $k_1 - np$ and $k_2 - np$ are of the same order as $\sqrt{npq}$.

In fact, one can show that

$$\sum_{k=k_1}^{k_2} C_k^n p^k q^{n-k} \simeq \Phi(\frac{k_2 + 0,5 - np}{\sqrt{npq}}) - \Phi(\frac{k_1 - 0,5 - np}{\sqrt{npq}})$$

Exemple : We throw a coin 10000 times. We look for the proability to obtain a number of "head" between 4900 and 5100.

Here $n = 10000$, $p = q = 0,5$, $k_1 = 4900$ and $k_2 = 5100$. As $\frac{k_2+0,5-np}{\sqrt{npq}} = 100/50$ and $\frac{k_1-0,5-np}{\sqrt{npq}} = -100/50$, we conclude with the previous formula that the sought probability is $\Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \simeq 0,9545$.

Remarque : One may also show that if $n >> 1$ and$np >> 1$, then we have

$$\sum_{k=0}^{k_2} C_k^n p^k q^{n-k} \simeq \Phi(\frac{k_2 - np}{\sqrt{npq}})$$

■

Approximation de type Poisson

We just have seen how to approximate $C_k^n p^k q^{n-k}$, that is the probability that an event A occurs k times among n trials.

Now we are going to find an approximation for this probability, when $p << 1$.

If n is sufficiently large so that $np \simeq npq >> 1$, we can still use de Moivre Laplace.

However, if np is of the order of 1, it does not work anymore. We have to use instead the following apprximation: for k of the order of np,

$$C_k^n p^k q^{n-k} \simeq e^{-np} \frac{(np)^k}{k!}$$

More precisely, we have

Proposition 4.1.2

Poisson If $n \to +\infty$, $p \to 0$ and $np \to a$, then

$$C_k^n p^k q^{n-k} \to e^{-a} \frac{a^k}{k!}$$

We also deduce that

$$P(k_1 \leq k \leq k_2) \simeq e^{-np} \sum_{k=k_1}^{k_2} \frac{(np)^k}{k!}$$

Exemple : A system contains 1000 components. Each component fails eventually independently from each other, and the probability of a failure per month is of 10^{-3}. We look for the probability that the system works (that is no component fails) at the end of one month.

Here, we may consider that is is a repeated trials problem, with $p = 10^{-3}$, $n = 10^3$ and $k = 0$. Thus

$$P(k = 0) = q^n = 0,999^{1000}$$

As $np = 1$, Poisson type approximation gives

$$P(k = 0) \simeq e^{-np} = e^{-1} = 0,368$$

$\odot$