CLASSICAL PROBABILITY 2008
2. MODES OF CONVERGENCE AND INEQUALITIES

JOHN MORIARTY

In many interesting and important situations, the object of interest is influenced by many random factors. If we can construct a probability model for the individual factors, then the limit theorems of classical probability may give useful statistical information about their cumulative effect. Examples are:

- stock prices are affected by many individual trades and pieces of information
- the flow of traffic on a motorway or of a crowd though a stadium is the result of many individual decisions.

We will see that in such situations, recognisable patterns can occur. In other words, we see convergence to some limiting object. The limiting object may be, among other things:

- a number (the law of large numbers)
- a random variable (the central limit theorem).

The simulations given in the lecture illustrate these two possiblities. Other limiting objects are possible, but we do not study them in this course.

Depending on the purpose of our probability model, we may have different types of convergence in mind. For example, we may wish to know:

- Does this convergence always happen?
- If not, what is the probability that it does not?
- What is the distribution of the limiting object?
- How big is the average error between the actual and limiting objects?

—and even these questions are still imprecise. It should be clear that we need a range of definitions of convergence to random objects.

We have seen above that, in the statements of the law of large numbers and the central limit theorem, the limiting objects are of different types. Appropriately, they use different notions or 'modes' of convergence of a sequence of random variables. Below four different modes of convergence are defined, and certain relationships between them are proven.

First, however, we state and prove some useful inequalities.

1. Inequalities

Question. Suppose you make a sequence of 10 investments, and I offer to either:

(1) take your average gain, then square it; or
(2) square each of your gains, then average it,

and pay you the result. Which should you choose?

Theorem 1.1. (Jensen's inequality) *Let X be a random variable with $E(X) < \infty$, and let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Then*

$$f(E(X)) \le E(f(X)).$$

Remark Recall that $f : \mathbb{R} \to \mathbb{R}$ is convex if for any $x_0 \in \mathbb{R}$ there exists a $\lambda \in \mathbb{R}$ such that

$$f(x) \ge \lambda(x - x_0) + f(x_0)$$

for all $x \in \mathbb{R}$. If f is twice differentiable, then f is convex if and only if $f'' \ge 0$. Examples of convex functions: $y \mapsto x, x^2, e^x, |x|$.

Proof. (Jensen's inequality) Let f be convex, and let $\lambda \in \mathbb{R}$ be such that

$$f(x) \ge \lambda(x - E(X)) + f(E(X))$$

for all x. Then

$$\begin{aligned} E(f(X)) &\ge E\left(\lambda(x - E(X)) + f(E(X))\right) \\ &= f(E(X)). \end{aligned}$$

□

Question. Suppose we are designing a flood defence. Let X be the (random) level of high tide. We have historical data giving an estimate for $E(X)$, but no information on the distribution of X. How high should the flood defence be to ensure just 5 percent chance of flooding?

Theorem 1.2. *Let X be a random variable, and let $f : \mathbb{R} \to [0, \infty)$. Then*

$$P(f(X) \ge a) \le \frac{E(f(X))}{a}$$

for all $a > 0$.

Proof. Let $A = \{f(X) \ge a\}$. Then

$$f(X) \ge a 1_A,$$

where

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

is the indicator function of A. Taking expectations gives

$$E(f(X)) \ge E(a1_A) = aP(A) = aP(f(X) \ge a),$$

which finishes the proof. □

Choosing $f(x) = |x|$ and $f(x) = x^2$ gives **Markov's inequality**

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

and **Chebychev's inequality**

$$P(|X| \geq a) \leq \frac{E(X^2)}{a^2},$$

respectively.

Question. Suppose that X is a random variable with finite mean and variance. Using all the results proved so far, which inequality gives more information about $P(|X| \geq a)$? Is it

(1) Markov's,
(2) Chebyshev's, or
(3) neither?

Question. Suppose that we have two random variables X and Y, and want some information on the average of the size of their product, $|XY|$. What information about X and Y might we need?

Recall that the <u>nth moment</u> of a random variable X is defined to be $E(X^n)$.

Theorem 1.3. (Hölder's inequality) *Assume $p > 1$ and $q > 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$. Let X and Y be two random variables. Then*

$$E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}.$$

Proof. If $E|X|^p = 0$, then $P(X = 0) = 1$ (you can use Theorem 1.2), so the inequality clearly holds (and also if $E|Y|^q = 0$). Thus we may assume that $E|X|^p > 0$ and $E|Y|^q > 0$.

Note that the function g defined by

$$g(t) = \frac{t^p}{p} + \frac{t^{-q}}{q}, \qquad t > 0$$

satisfies $g(t) \geq 1$ (you can examine the derivative for t around the point 1). Inserting

$$t = \left(\frac{|X|}{(E|X|^p)^{1/p}}\right)^{1/q} \left(\frac{|Y|}{(E|Y|^q)^{1/q}}\right)^{-1/p}$$

gives

$$1 \leq g(t) = \frac{1}{p}\left(\frac{|X|}{(E|X|^p)^{1/p}}\right)^{p/q} \frac{E(|Y|^q)^{1/q}}{|Y|} + \frac{1}{q}\frac{(E|X|^p)^{1/p}}{|X|}\left(\frac{|Y|}{(E|Y|^q)^{1/q}}\right)^{q/p}$$

for $\omega \in \Omega$ such that $X(\omega)Y(\omega) \neq 0$. Consequently,

$$\begin{aligned} \frac{|XY|}{(E|X|^p)^{1/p}(E|Y|^q)^{1/q}} &\leq \frac{1}{p}\frac{|X|^{1+p/q}}{(E|X|^p)^{1/q+1/p}} + \frac{1}{q}\frac{|Y|^{1+q/p}}{(E|Y|^q)^{1/p+1/q}} \\ &= \frac{1}{p}\frac{|X|^p}{(E|X|^p)} + \frac{1}{q}\frac{|Y|^q}{(E|Y|^q)} \end{aligned}$$

(this inequality also holds if $X(\omega)Y(\omega) = 0$). Taking expectations of both sides gives

$$E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}.$$

□

Using $p = q = 2$ in Hölder's inequality gives the Cauchy-Schwartz inequality.

Consequence 1.4. (Cauchy-Schwartz inequality)

$$(E|XY|)^2 \leq (E|X|^2)(E|Y|^2).$$

Question. Let X, Y be random variables with means μ_1, μ_2 and variances σ_1, σ_2 respectively, all finite. What can you now say about the average size of their product?

Our final inequality gives a similar estimate, in terms of the moments of X and Y, of the average of the size of their sum, $|X + Y|$.

Theorem 1.5. (Minkowski's inequality) *Let $p \geq 1$. Then*

$$(E|X + Y|^p)^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}.$$

Proof. If $p = 1$, the inequality follows directly from the triangle inequality. Thus we assume that $p > 1$, and we let $q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $Z = |X + Y|$. Then

$$\begin{aligned} EZ^p &= E(ZZ^{p-1}) \\ &\leq E(|X|Z^{p-1}) + E(|Y|Z^{p-1}) \\ &\leq (E|X|^p)^{1/p}(EZ^{q(p-1)})^{1/q} + (E|Y|^p)^{1/p}(EZ^{q(p-1)})^{1/q} \\ &= \left((E|X|^p)^{1/p} + (E|Y|^p)^{1/p}\right)(EZ^{q(p-1)})^{1/q}, \end{aligned}$$

where we in the first inequality used $|X + Y| \leq |X| + |Y|$, and in the second inequality we used Hölder's inequality. Since $q(p - 1) = p$ we have

$$(EZ^p)^{1-1/q} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p},$$

which finishes the proof since $1 - 1/q = 1/p$. □

2. Different modes of convergence

In simulations we saw convergence of random objects to limiting objects. Those limiting objects were either random or deterministic (which is a special case of random!). We also discussed different possible criteria for convergence to a random object. In order to do calculations or prove theorems, we of course need precise technical definitions of these different modes of convergence.

Let $X_1, X_2, X_3, ...$ be a sequence of random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$, and let X be a random variable on this probability space. We want to make precise the statement

$$\text{“}X_n \to X \text{ as } n \to \infty\text{”}.$$

More precisely, we will consider four different modes of convergence.

Definition 2.1. *We say that*

- X_n *converges to* X *almost surely if*
$$\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}$$
is an event with probability 1.
- X_n *converges to* X *in r:th mean if* $E|X_n|^r < \infty$ *for all* n *and*
$$E|X_n - X|^r \to 0$$
as $n \to \infty$.
- X_n *converges to* X *in probability if for all* $\epsilon > 0$ *we have*
$$P(|X_n - X| > \epsilon) \to 0$$
as $n \to \infty$.
- X_n *converges to* X *in distribution if* $F_n(x) \to F(x)$ *as* $n \to \infty$ *for all* x *where* $F(x)$ *is continuous.*

Question. Match each of the following four statements to the corresponding mode of convergence:
As $n \to \infty$,

(1) For any $0 < a < b < 1$, the probability that X_n lies between the values a and b tends to $(b - a)$.
(2) X_n always tends to $Y/2$.
(3) The probability that X_n is more than distance $1/n^2$ from e^Z is $1/n$.
(4) X_n has mean value 1 and the variance of X_n is e^{-n}.

An obvious question is whether we really need so many modes of convergence. The answer is yes, although the situation does simplify a little: using inequalities proved in this chapter we can establish some relationships between the modes.

It turns out that these four different modes of convergence are distinct, i.e. no two modes of convergence are equivalent. We show below that the following set of implications between them holds:

Theorem 2.2. *Let X_n, $n = 1, 2, 3...$ and X be random variables on some probability space. We then have*

$$\begin{array}{ccc} X_n \to X \text{ in distribution} & & \\ \Uparrow & & \\ X_n \to X \text{ in probability} & \Leftarrow & X_n \to X \text{ almost surely} \\ \Uparrow & & \\ X_n \to X \text{ in } r\text{:th mean for some } r \geq 1. & & \end{array}$$

Also,

$$X_n \to X \text{ in } r\text{:th mean} \implies X_n \to X \text{ in } s\text{:th mean}$$

for $r > s \geq 1$.

Theorem 2.2 follows from the lemmas below.

Question. Can you convince yourself that the next lemma is true (without a formal proof)?

Lemma 2.3. *Convergence in probability implies convergence in distribution.*

Proof. Assume $X_n \to X$ in probability. If $\epsilon > 0$, then

$$\begin{aligned} F_n(x) &= P(X_n \leq x) \\ &= P(X_n \leq x \text{ and } X \leq x + \epsilon) + P(X_n \leq x \text{ and } X > x + \epsilon) \\ &\leq F(x + \epsilon) + P(|X - X_n| > \epsilon). \end{aligned}$$

Similarly,

$$\begin{aligned} F(x - \epsilon) &\leq P(X \leq x - \epsilon \text{ and } X_n \leq x) + P(X \leq x - \epsilon \text{ and } X_n > x) \\ &\leq F_n(x) + P(|X_n - X| > \epsilon). \end{aligned}$$

Thus we have

$$F(x - \epsilon) - P(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x + \epsilon) + P(|X_n - X| > \epsilon).$$

Technical point: We now want to let $n \to \infty$ so that the $P(\ )$ terms disappear; however we don't yet know whether $F_n(x)$ has a limit. So we write

$$F(x - \epsilon) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x + \epsilon).$$

If x is a point of continuity of F, then letting $\epsilon \to 0$ shows that the $\liminf$ and $\limsup$ are equal, so $F_n(x)$ does have a limit and

$$\lim_{n\to\infty} F_n(x) = F(x),$$

which finishes the proof. □

Lemma 2.4. *i. Convergence in rth mean implies convergence in sth mean for $1 \leq s < r$.*

ii. Convergence in mean ($r = 1$) implies convergence in probability.

Proof. i. Use Jensen's inequality to show that

$$(E|Z|^s)^{1/s} \leq (E|Z|^r)^{1/r},$$

see Problem 3 on Exercise Sheet 2. Using this inequality with $Z = X_n - X$ gives

$$(E|X_n - X|^s)^{1/s} \leq (E|X_n - X|^r)^{1/r}.$$

It follows (by continuity) that convergence in rth mean implies convergence in sth mean.

ii. Markov's inequality gives

$$P(|X_n - X| \geq \epsilon) \leq \frac{E|X_n - X|}{\epsilon},$$

which shows that convergence in mean implies convergence in probability. □

Question. Could you have also seen that part ii above was true before seeing the proof? If so, is your argument different to the formal proof?

Lemma 2.5. *$X_n \to X$ almost surely if and only if*

$$P\left(\sup_{k\geq n} |X_k - X| \geq \epsilon\right) \to 0$$

as $n \to \infty$ for all $\epsilon > 0$.

Proof. Let $A_k(\epsilon) = \{|X_k - X| \geq \epsilon\}$, and let

$$A(\epsilon) = \{A_k(\epsilon), \text{ i.o.}\}.$$

We claim that

(1) $$P(A(\epsilon)) = 0 \text{ for all } \epsilon > 0 \iff P(X_n \to X) = 1.$$

To see this, first note that

$$X_k(\omega) \to X(\omega)$$

implies $\omega \notin A(\epsilon)$ for all $\epsilon > 0$. Thus

$$P(X_n \to X) = 1 \implies P(A(\epsilon)^C) = 1 \implies P(A(\epsilon)) = 0 \text{ for all } \epsilon > 0.$$

Moreover,

$$\begin{aligned} P\left(\{X_n \to X\}^C\right) &= P\left(\bigcup_{\epsilon>0} A(\epsilon)\right) = P\left(\bigcup_{m=1}^{\infty} A(1/m)\right) \\ &\leq \sum_{m=1}^{\infty} P(A(1/m)), \end{aligned}$$

which proves the other implication, and thus finishes the proof of (1).

Now, let

$$B_n(\epsilon) = \bigcup_{k=n}^{\infty} A_k(\epsilon) = \{\sup_{k\geq n} |X_k - X| \geq \epsilon\}.$$

Then

$$B_1(\epsilon) \supseteq B_2(\epsilon) \supseteq B_3(\epsilon) \supseteq ... \supseteq A(\epsilon)$$

with

$$\lim_{n\to\infty} B_n(\epsilon) = A(\epsilon).$$

Thus

$$P(A(\epsilon)) = 0 \iff P(\lim_{n\to\infty} B_n(\epsilon)) = \lim_{n\to\infty} P(B_n(\epsilon)) = 0$$

(the second last equality is justified by Problem 2 on Exercise Sheet 1). From (1) we conclude that

$$\begin{aligned} P(X_n \to X) = 1 \quad &\iff \quad \lim_{n\to\infty} P(B_n(\epsilon)) = 0 \text{ for all } \epsilon > 0 \\ &\iff \quad \lim_{n\to\infty} P(\sup_{k\geq n} |X_k - X| \geq \epsilon) = 0 \text{ for all } \epsilon > 0, \end{aligned}$$

which finishes the proof. □

Lemma 2.5 has two important consequences:

Consequence 2.6. *Convergence almost surely implies convergence in probability.*

Proof. Let $\epsilon > 0$. Then

$$P(|X_n - X| > \epsilon) \leq P(\sup_{k\geq n} |X_k - X| \geq \epsilon) \to 0$$

as $n \to \infty$ if $X_n \to X$ almost surely. □

Consequence 2.7. *If*

$$\sum_{k=1}^{\infty} P(|X_k - X| > \epsilon) < \infty$$

for all $\epsilon > 0$*, then* $X_k \to X$ *almost surely.*

Proof. If

$$\sum_{k=1}^{\infty} P(|X_k - X| > \epsilon) < \infty,$$

then

$$P(\sup_{k\geq n} |X_k - X| \geq \epsilon) \leq \sum_{k=n}^{\infty} P(|X_k - X| \geq \epsilon) \to 0$$

as $n \to \infty$. Thus $X_k \to X$ almost surely follows from Lemma 2.5. □

All implications in Theorem 2.2 are strict—that is, their reverse implications do not hold. For example, there exists a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ such that $X_n \to X$ in probability but not almost surely (the strictness of the implications in Theorem 2.2 follows from the problems in Example Sheet 2 and 3). However, under some extra assumptions we have the following converses to the implications in Theorem 2.2.

Theorem 2.8. *Let X_n, $n = 1, 2, ...$ and X be random variables.*

(i) *$X_n \to C$ in distribution, where C is a constant, implies $X_n \to C$ in probability.*

(ii) *$X_n \to X$ in probability and $P(|X_n| \leq M) = 1$ for some constant M imply $X_n \to X$ in rth mean for $r \geq 1$.*

(iii)

$$\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty \text{ for all } \epsilon > 0 \implies X_n \to X \text{ a.s.}$$

(iv) *$X_n \to X$ in probability implies the existence of a subsequence $\{n_k\}_{k=1}^{\infty}$ with $\lim n_k \to \infty$ such that $X_{n_k} \to X$ as $k \to \infty$ almost surely.*

Proof. (i) is Problem 1 on Exercise Sheet 4, and (iii) is Consequence 2.7 above.

To prove (ii), we first claim that

$$P(|X| \leq M) = 1. \tag{2}$$

To see this, let $\epsilon > 0$. Then

$$\begin{aligned} P(|X| \geq M + \epsilon) &= P(|X - X_n + X_n| \geq M + \epsilon) \\ &\leq P(|X - X_n| + |X_n| \geq M + \epsilon) \\ &= P(|X - X_n| \geq M - |X_n| + \epsilon) \\ &\leq P(|X - X_n| \geq \epsilon) \to 0 \end{aligned}$$

since $X_n \to X$ in probability. Letting $\epsilon \to 0$ and using the continuity property (for increasing families of events) proves the claim (2). Now, for $\epsilon > 0$ let

$$A_n = \{|X_n - X| > \epsilon\}.$$

Then

$$|X_n - X|^r \leq \epsilon^r 1_{A_n^C} + (2M)^r 1_{A_n},$$

so

$$\begin{aligned} E|X_n - X|^r &\leq E(\epsilon^r 1_{A_n^C} + (2M)^r 1_{A_n}) \\ &\leq \epsilon^r + (2M)^r P(A_n) \\ &= \epsilon^r + (2M)^r P(|X_n - X| > \epsilon) \to \epsilon^r \end{aligned}$$

as $n \to \infty$ if $X_n \to X$ in probability. Since $\epsilon > 0$ is arbitrary, it follows that $X_n \to X$ in rth mean.

To prove (iv), pick an increasing sequence n_k such that

$$P(|X_{n_k} - X| > 1/k) \leq 1/k^2$$

(this can be done since $X_n \to X$ in probability). Then, if $\epsilon > 0$, we have

$$\begin{aligned}\sum_{k=1}^{\infty} P(|X_{n_k} - X| > \epsilon) &\leq \sum_{k=1}^{1/\epsilon} P(|X_{n_k} - X| > \epsilon) \\ &\quad + \sum_{k \geq 1/\epsilon} P(|X_{n_k} - X| > 1/k) \\ &\leq \sum_{k=1}^{1/\epsilon} P(|X_{n_k} - X| > \epsilon) + \sum_{k=1}^{\infty} 1/k^2 < \infty.\end{aligned}$$

Consequently, $X_{n_k} \to X$ almost surely as $k \to \infty$ according to (iii) above. □