Probability 2

Hadrien Montanelli

1 Convergence of random variables, and limit theorems

Notation Let $X_1, X_2, \dots$ and X be random variables with cdfs $F_1, F_2, \dots$ and F. Note that

$$\{X_n \to X \text{ as } n \to \infty\}$$

is an event.

Convergence almost surely $X_n \xrightarrow{as} X$ *almost surely* if

$$\mathrm{P}(X_n \to X \text{ as } n \to \infty) = 1.$$

Convergence in probability $X_n \xrightarrow{p} X$ *in probability* if, for any $\epsilon > 0$,

$$\mathrm{P}(|X_n - X| < \epsilon) \to 1 \text{ as } n \to \infty.$$

Convergence in distribution $X_n \xrightarrow{d} X$ *in distribution* if, for any x such that F is continuous at x,

$$F_n(x) \to F(x) \text{ as } n \to \infty.$$

(Note that $X_n \xrightarrow{d} c$ means that X_n converges to a random variable with cdf $F(x) = 0$ for $x < c$ and $F(x) = 1$ for $x \geq c$.)

Comparison of modes of convergence The following implications hold:

$$\text{almost surely} \;\Rightarrow\; \text{in probability} \;\Rightarrow\; \text{in distribution.}$$

The reverse implications do not hold in general.

Week law of large numbers (WLLN) Suppose that $X_1, \dots, X_n$ are iid random variables with finite mean μ. Let $S_n = X_1 + \dots + X_n$. Then,

$$\frac{S_n}{n} \xrightarrow{p} \mu \text{ as } n \to \infty.$$

Strong law of large numbers (SLLN) Suppose that $X_1, \dots, X_n$ are iid random variables with mean μ. Let $S_n = X_1 + \dots + X_n$. Then,

$$\frac{S_n}{n} \xrightarrow{as} \mu \text{ as } n \to \infty.$$

Central limit theorem (CLT) Let $X_1, \dots, X_n$ be a random sample of size n of any distribution with mean μ and finite variance σ^2. Let $S_n = X_1 + \dots + X_n$. Then,

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathrm{N}(0,1) \text{ as } n \to \infty.$$

2 Generating functions

Probability generating function (pgf) Let X be a non-negative integer-valued random variable. The *probability generating function* of X is

$$G_X(s) = \mathrm{E}(s^X) = \sum_{k=0}^{\infty} s^k p_X(k).$$

If X and Y are non-negative integer-valued random variables with the same pgf, then they have the same pmf.

Properties of pgf

$$\begin{aligned}
&G_{aX+b}(s) = s^b G_X(s^a),\\
&G_{X+Y}(s) = G_X(s)G_Y(s) \quad \text{(independent)},\\
&G_X^{(k)}(0) = k!\, p_X(k), \quad k \geq 0,\\
&G_X^{(k)}(1) = \mathrm{E}(X[X-1]\dots[X-k+1]), \quad k \geq 1.
\end{aligned}$$

Sum of iid random variables Let $X_1, X_2, \dots$ be iid non-negative integer-valued random variables with pgf G_X. Let N be another non-negative integer-valued random variable, independent of $X_1, X_2, \dots$, and with pgf G_N. Then, the pgf of $\sum_{i=1}^N X_i$ is $G_N \circ G_X$.

Moment generating function (mgf) Let X be a random variable. The *moment generating function* of X is

$$M_X(t) = \mathrm{E}(e^{tX}).$$

If X and Y are random variables with the same mgf, which is finite on $[-t_0, t_0]$ for some $t_0 > 0$, then they have the same pmf/pdf.

Properties of mgf

$$\begin{aligned}
&M_{aX+b}(t) = e^{bt} M_X(at),\\
&M_{X+Y}(t) = M_X(t)M_Y(t) \quad \text{(independent)},\\
&M_X(t) = \sum_{k=0}^{\infty} \frac{t^k \mathrm{E}(X^k)}{k!}, \quad M_X^{(k)}(0) = \mathrm{E}(X^k).
\end{aligned}$$

Sum of iid random variables Let $X_1, X_2, \dots$ be iid random variables with mgf M_X. Let N be a non-negative integer-valued random variable, independent of $X_1, X_2, \dots$, and with pgf G_N. Then, the mgf of $\sum_{i=1}^N X_i$ is $G_N \circ M_X$.

Continuity result for mgf Suppose Y and $X_1, X_2, \dots$ are random variables whose moment generating functions M_Y and $M_{X_1}, M_{X_2}, \dots$ are all finite on $[-t_0, t_0]$ for some $t_0 > 0$. If

$$M_{X_n}(t) \to M_Y(t) \text{ as } n \to \infty, \text{ for all } t \in [-t_0, t_0],$$

then $X_n \xrightarrow{d} Y$ as $n \to \infty$.

Characteristic function Let X be a random variable. The *characteristic function* of X is defined by

$$\phi_X(t) = \mathrm{E}(e^{itX}) = \mathrm{E}(\cos(tX)) + i\mathrm{E}(\sin(tX)).$$

3 Joint distribution of continuous random variables

Jacobian Suppose $T : (x, y) \mapsto (u, v)$ is a one-to-one mapping. The *Jacobian* is defined as

$$J(u, v) = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}.$$

Change of variables If X, Y have joint pdf $f_{X,Y}$, then the random variables U, V defined by $(U, V) = T(X, Y)$ are jointly continuous with joint pdf $f_{U,V}$ given by

$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v))|J(u, v)|.$$

Covariance matrix Let $Z_1 \ldots, Z_n$ be idd standard normal random variables. Define $X_1, \ldots, X_n$ by

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = A \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix},$$

for some $n \times n$ invertible matrix A. The matrix $\Sigma = AA^T$ with entries $\Sigma_{i,j} = \mathrm{cov}(X_i, X_j)$ is the *covariance matrix*.

Multivariate normal distribution $(n \geq 2)$ The *multivariate normal distribution* has pdf

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right).$$

Standard normal distribution $(n = 2)$ In this case,

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

with $\rho = \mathrm{cov}(X, Y)/(\sigma_X\sigma_Y) = \mathrm{cov}(X, Y)$, which yields

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1 - \rho^2)}\right).$$

Conditioning by an event Suppose that B is an event such that $\mathrm{P}(B) > 0$. Then, the *conditional distribution function* (or *conditional cdf*) of X given B is

$$F_{X|B}(x) = \frac{\mathrm{P}(\{X \leq x\} \cap B)}{\mathrm{P}(B)},$$

the *conditional density function* of X given B (or *conditional pdf*) is $f_{X|B}(x) = F'_{X|B}(x)$, and the *conditional expectation* of X given B is

$$\mathrm{E}(X|B) = \int_{-\infty}^{\infty} x f_{X|B}(x)dx.$$

Conditioning by a continuous variable The *conditional distribution function* of X given $Y = y$ is

$$F_{X|Y=y}(x) = \int_{-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(y)}du,$$

the *conditional density function* of X given $Y = y$ is

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

and the *conditional expectation* of X given $Y = y$ is

$$\mathrm{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y=y}(x)dx.$$

Law of total expectation Suppose Y has pdf f_Y. Then,

$$E(X) = \int_{-\infty}^{\infty} E(X|Y = y)f_Y(y)dy.$$

4 Markov chains

Set-up Let I be finite or countably infinite. A probability distribution on I is a (possibly infinite) row vector $\lambda = (\lambda_i, i \in I)$ with $\lambda_i \geq 0$ for all i, and $\sum_{i \in I} \lambda_i = 1$. We will say that a random variable Y taking values in I has distribution λ if $\mathrm{P}(Y = i) = \lambda_i$ for all i.

Markov chain Let $X = (X_n, n \geq 0)$ be a sequence of random variables taking values in I. The process X is called a *Markov chain* if, for any $n \geq 0$ and $i_0, \ldots, i_{n+1}$,

$$\begin{aligned} &\mathrm{P}(X_{n+1} = i_{n+1} \,|\, X_n = i_n, \ldots, X_0 = i_0), \\ &= \mathrm{P}(X_{n+1} = i_{n+1} \,|\, X_n = i_n). \end{aligned}$$

Homogeneous The Markov chain is called *homogeneous* if the *transition probability* $p_{ij} = \mathrm{P}(X_{n+1} = j \,|\, X_n = i)$ depends only on i and j, not on n.

Transition matrix The *transition matrix* P is the (possible infinite) matrix $P = (p_{ij})_{i,j \in I}$. The ith row of P is the probability distribution of X_{n+1} given $X_n = i$.

Initial distribution Let λ be the distribution of X_0, called the *initial distribution*. Then,

$$\mathrm{P}(X_0 = i_0, \ldots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n}.$$

Chapman–Kolmogorov equations Consider the *n-step transition probability* $p_{ij}^{(n)} = \mathrm{P}(X_{k+n}| \; X_k = i)$. Then,

$$p_{ij}^{(n+m)} = \sum_{k \in I} p_{ik}^{(n)} p_{kj}^{(m)}, \quad p_{ij}^{(n)} = (P^n)_{ij}.$$

As a consequence, the distribution of X_n is λP^n.

Function of a Markov chain Suppose that the random process $Y = (Y_n, n \geq 0)$ is such that $Y_{n+1} = f(Y_n, X_{n+1})$ for each n, where X_{n+1} is independent of $Y_0, \ldots, Y_n$. Then, Y is a Markov chain.

Closed class A class C is called *closed* if $p_{ij} = 0$, whenever $i \in C$ and $j \notin C$. A closed class cannot be escaped.

Absorbing class If $\{i\}$ is a closed class, then $p_{ii} = 1$, and i is called an *absorbing* state.

Irreducible chain A chain for which I consists of a single communicating class is called *irreducible*.

Period The greatest common divisor of $\{n : p_{ii}^{(n)} > 0\}$.

Hitting probability The *hitting probability* of A starting from i is $h_i^A = \mathrm{P}(X_n \in A \text{ for some } n \geq 0 \,|\, X_0 = i)$, and is the minimal non-negative solution to

$$h_i^A = \begin{cases} 1, & \text{if } i \in A, \\ \sum_{j \in I} p_{ij} h_j^A, & \text{if } i \notin A. \end{cases}$$

Transience and recurrence The state i is called *transient* if

$$\begin{aligned} &\mathrm{P}(X_n = i \text{ for some } n \geq 1 \,|\, X_0 = i) = p < 1, \\ \Rightarrow\ &\mathrm{P}(\text{hit } i \text{ infinitely often} \,|\, X_0 = i) = 0. \end{aligned}$$

The state i is called *recurrent* if

$$\begin{aligned} &\mathrm{P}(X_n = i \text{ for some } n \geq 1 \,|\, X_0 = i) = 1, \\ \Rightarrow\ &\mathrm{P}(\text{hit } i \text{ infinitely often} \,|\, X_0 = i) = 1. \end{aligned}$$

Note that in a communicating class either all states are recurrent or all transient, that every recurrent class is closed, and that every finite closed class is recurrent.

Mean hitting time The *mean hitting time* of A starting from i is $k_i^A = \mathrm{E}(\inf\{n : X_n \in A\} \,|\, X_0 = i)$, and is the minimal non-negative solution to

$$k_i^A = \begin{cases} 0, & \text{if } i \in A, \\ 1 + \sum_{j \in I} p_{ij} k_j^A, & \text{if } i \notin A. \end{cases}$$

Mean return time The *mean return time* to a state i is

$$\begin{aligned} m_i &= \mathrm{E}(\inf\{n \geq 1 : X_n = i\} \,|\, X_0 = i), \\ &= 1 + \sum_{j \in I} p_{ij} k_j^{\{i\}}. \end{aligned}$$

Stationary distribution We say that $\pi = (\pi_i, i \in I)$ is a *stationary distribution* if $\pi P = \pi$.

Existence and uniqueness Let P be an irreducible transition matrix. Then, P has a stationary distribution if and only if P is positive recurrent, and the stationary distribution π is unique and given by $\pi_i = 1/m_i$ for all i.

Convergence to equilibrium Suppose P is irreducible and aperiodic, with stationary distribution π. If X_n is a Markov chain with transition matrix P and any initial distribution, then for all $j \in I$,

$$\mathrm{P}(X_n = j) \to \pi_j \text{ as } n \to \infty.$$

In particular, $p_{ij}^{(n)} \to \pi_j$ as $n \to \infty$, for all i and j.

Ergodic theorem Let P be irreducible. Let $V_i(n)$ be the number of visits to state i before time n, that is

$$V_i(n) = \sum_{\ell=0}^{n-1} \mathbf{1}(X_\ell = i).$$

Then, for any initial distribution, and for all $i \in I$,

$$\frac{V_i(n)}{n} \xrightarrow{as} \frac{1}{m_i} \text{ as } n \to \infty.$$

5 Poisson process

Set-up Consider a random process $(N_t, t \geq 0)$ with a continuous time variable t. Such a process is called a counting process if N_t takes non-negative integer values, and $N_s \leq N_t$ whenever $s \leq t$.

Interarrival time If N_t describes an arrival process, then $N_t = k$ means that there have been k arrivals in the time interval $[0, t]$. Let $T_k = \inf\{t \geq 0 : N_t \geq k\}$ for $k \geq 0$. Then, $T_0 = 0$ and T_k is the k-th arrival time, for $k \geq 1$. We also define $Y_k = T_k - T_{k-1}$, for $k \geq 1$, the *interarrival time* between arrivals $k-1$ and k.

Increment For $s < t$, we write $N(s, t]$ for $N_t - N_s$, which we can think of as the number of points of the process which occur in the time-interval $(s, t]$. This is also called the *increment* of the process N on the interval $(s, t]$.

Poisson process $(N_t, t \geq 0)$ is a *Poisson process* of rate λ if its interarrival times $Y_1, Y_2, \ldots \overset{\text{iid}}{\sim} \mathrm{Exp}(\lambda)$.

Poisson process (increment) $(N_t, t \geq 0)$ is a *Poisson process* of rate λ if

(i) $N_0 = 0$;

(ii) if $(s_1, t_1), \ldots, (s_k, t_k)$ are disjoint intervals in $\mathbb{R}_+$, then $N(s_1, t_1], \ldots, N(s_k, t_k]$ are independent;

(iii) for any $s < t$, the increment $N(s, t]$ has Poisson distribution with mean $\lambda(t - s)$.

Superposition Let L_t and M_t be independent Poisson processes of rate λ and μ respectively. Then, $L_t + M_t$ is a Poisson process of rate $\lambda + \mu$.

Thinning Let N_t be a Poisson process of rate λ. Mark independently each point of the process with probability p. Let M_t be the counting process of the marked points. Then, M_t is a Poisson process of rate $p\lambda$.