Lecture 3

Last updated by Serik Sagitov: December 2, 2012

Abstract

A course based on the book Probability and Random Processes by Geoffrey Grimmett and David Stirzaker. Chapter 5. Generating functions and their applications (5.6-5.10). Chapter 6. Markov chains (6.1-6.3).

1 Characteristic functions

If X takes values $k = 0, 1, 2, \ldots$ with probabilities p_k and $\sum_{k=0}^{\infty} p_k = 1$, then the distribution of X is fully described by its probability generating function

$$G(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} p_k s^k.$$

Key property: if X and Y are independent, then $G_{X+Y}(s) = G_X(s)G_Y(s)$. Examples

- Bernoulli distribution $G(s) = q + ps$,
- binomial distribution $G(s) = (q + ps)^n$,
- Poisson distribution $G(s) = e^{\lambda(s-1)}$.

Moment generating function of X is $M(t) = \mathbb{E}(e^{tX})$. In the continuous case $M(t) = \int e^{tx} f(x)dx$. Computing moments

$$\mathbb{E}(X) = G'(1), \quad \mathbb{E}(X(X-1)) = G''(1),$$
$$\mathbb{E}(X) = M'(0), \quad \mathbb{E}(X^k) = M^{(k)}(0).$$

Examples of moment generating functions

- Normal distribution $M(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$,
- exponential distribution $M(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$,
- Gamma(α, λ) distribution $M(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}$ for $t < \lambda$,
- Cauchy distribution $M(0) = 1$, $M(t) = \infty$ for $t \neq 0$.

The characteristic function of X is complex valued $\phi(t) = \mathbb{E}(e^{itX})$. The joint characteristic function for $\mathbf{X} = (X_1, \ldots, X_n)$ is $\phi(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{t}\mathbf{X}^{\mathrm{t}}})$. Examples of characteristic functions

- normal distribution $\phi(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}$,
- gamma distribution $\phi(t) = \left(\frac{\lambda}{\lambda - it}\right)^{\alpha}$,
- Cauchy distribution $\phi(t) = e^{-|t|}$,
- multivariate normal distribution $\phi(\mathbf{t}) = e^{i\mathbf{t}\boldsymbol{\mu}^{\mathrm{t}} - \frac{1}{2}\mathbf{t}\mathbf{V}\mathbf{t}^{\mathrm{t}}}$.

From the last example it follows that given a vector $\mathbf{X} = (X_1, \ldots, X_n)$ with a multivariate normal distribution any linear combination $\mathbf{a}\mathbf{X}^{\mathrm{t}} = a_1X_1 + \ldots + a_nX_n$ is normally distributed since

$$\mathbb{E}(e^{t\mathbf{a}\mathbf{X}^{\mathrm{t}}}) = \phi(t\mathbf{a}) = e^{it\mu - \frac{1}{2}t^2\sigma^2}, \quad \mu = \mathbf{a}\boldsymbol{\mu}^{\mathrm{t}}, \quad \sigma^2 = \mathbf{a}\mathbf{V}\mathbf{a}^{\mathrm{t}}.$$

2 Weak law of large numbers

Definition 1 *Convergence in distribution* $X_n \stackrel{d}{\to} X$ *means*

$$\mathbb{P}(X_n \le x) \to \mathbb{P}(X \le x) \text{ for all } x \text{ such that } \mathbb{P}(X = x) = 0.$$

This is equvalent to the weak convergence $F_n \stackrel{d}{\to} F$ *of distribution functions when* $F_n(x) \to F(x)$ *at each point* x *where* F *is continuous.*

Properties of characteristic functions

- two r.v. have the same characteristic function iff they have the same distribution function,
- if $X_n \stackrel{d}{\to} X$, then $\phi_n(t) \to \phi(t)$ for all t,
- conversely, if $\phi(t) = \lim_{n\to\infty} \phi_n(t)$ exists and continuous at $t = 0$, then ϕ is cf of some F, and $F_n \stackrel{d}{\to} F$.

Theorem 2 *If* $X_1, X_2, \ldots$ *are iid with finite mean* μ *and* $S_n = X_1 + \ldots + X_n$*, then*

$$S_n/n \stackrel{d}{\to} \mu, \quad n \to \infty.$$

Proof. Let F_n and ϕ_n be the df and cf of $n^{-1}S_n$. To prove $F_n(x) \stackrel{d}{\to} 1_{\{x\ge\mu\}}$ we have to see that $\phi_n(t) \to e^{it\mu}$ which is obtained using a Taylor expansion

$$\phi_n(t) = \Big(\phi_1(tn^{-1})\Big)^n = \Big(1 + i\mu tn^{-1} + o(n^{-1})\Big)^n \to e^{it\mu}.$$

Remarks. Statistical application: the sample mean is a consistent estimate of the population mean. Counterexample: if X_i has the Cauchy distribution, then $S_n/n \stackrel{d}{=} X_1$ since $\phi_n(t) = \phi_1(t)$.

3 Central limit theorem

According the LLN $|S_n - n\mu|$ is much smaller than n. This difference is of order $\sqrt{n}$.

Theorem 3 *If* $X_1, X_2, \ldots$ *are iid with finite mean* μ *and positive finite variance* σ^2*, then for any* x

$$\mathbb{P}\Big(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\Big) \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy, \quad n \to \infty.$$

Proof. Let ψ_n be the cf of $\frac{S_n - n\mu}{\sigma\sqrt{n}}$. Using a Taylor expansion we obtain

$$\psi_n(t) = \Big(1 - \frac{t^2}{2n} + o(n^{-1})\Big)^n \to e^{-t^2/2}.$$

Remarks. Important example: simple random walks. 280 years ago de Moivre (1733) obtained the first CLT in the symmetric case with $p = 1/2$.

Statistical application: the standardized sample mean has the sampling distribution which is approximately $\mathrm{N}(0, 1)$. Approximate 95% confidence interval formula for the mean $\bar{X} \pm 1.96\frac{s}{\sqrt{n}}$.

4 Markov chains

Conditional on the present value, the future of the system is independent of the past. A Markov chain $\{X_n\}_{n=0}^{\infty}$ with countably many states and transition matrix $\mathbf{P}$ with elements p_{ij}

$$\mathbb{P}(X_n = j|X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) = p_{ij}.$$

The n-step transition matrix with elements $p_{ij}^{(n)} = \mathbb{P}(X_{n+m} = j|X_m = i)$ equals $\mathbf{P}^n$. Given the initial distribution $\mathbf{a}$ as the vector with components $a_i = \mathbb{P}(X_0 = i)$, the distribution of X_n is given by the vector $\mathbf{aP}^n$ since

$$\mathbb{P}(X_n = j) = \sum_{i=-\infty}^{\infty} \mathbb{P}(X_n = j|X_0 = i)\mathbb{P}(X_0 = i) = \sum_{i=-\infty}^{\infty} a_i p_{ij}^{(n)}.$$

Examples:

- IID chain has transition probabilities $p_{ij} = p_i$,
- simple random walk has transition probabilities $p_{ij} = p1_{\{j=i+1\}} + q1_{\{j=i-1\}}$,
- Bernoulli process has transition probabilities $p_{ij} = p1_{\{j=i+1\}} + q1_{\{j=i\}}$ and state space $S = \{0, 1, 2, \ldots\}$.

Let $T_i = \min\{n \geq 1 : X_n = i\}$ and put $f_{ij}^{(n)} = \mathbb{P}(T_j = n|X_0 = i)$. Define the generating functions

$$P_{ij}(s) = \sum_{n=0}^{\infty} s^n p_{ij}^{(n)}, \qquad F_{ij}(s) = \sum_{n=1}^{\infty} s^n f_{ij}^{(n)}.$$

It is not difficult to see that

$$\begin{aligned} P_{ij}(s) &= 1_{\{j=i\}} + F_{ij}(s)P_{jj}(s) \\ P_{ii}(s) &= \frac{1}{1 - F_{ii}(s)}. \end{aligned}$$

Classification of states

- state i is called recurrent (persistent), if $\mathbb{P}(T_i < \infty|X_0 = i) = 1$, otherwise i is called a transient state,
- a recurrent state i is called null-recurrent, if $\mathbb{E}(T_i|X_0 = i) = \infty$,
- state i is called positive-recurrent, if $\mathbb{E}(T_i|X_0 = i) < \infty$.

Since $F_{ii}(1) = \mathbb{P}(T_i < \infty|X_0 = i)$, we conclude that state i is recurrent iff $P_{ii}(1) = \infty$.

Theorem 4 *State i is recurrent iff*

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty. \tag{1}$$

A recurrent state i is null-recurrent iff $p_{ii}^{(n)} \to 0$.

Example. For a simple random walk

$$\mathbb{P}(S_{2n} = i|S_0 = i) = \binom{2n}{n}(pq)^n.$$

Using the Stirling formula $n! \sim n^n e^{-n}\sqrt{2\pi n}$ we get

$$p_{ii}^{(2n)} \sim \frac{(4pq)^n}{\sqrt{\pi n}}, \quad n \to \infty.$$

Criterium of recurrence (1) holds only if $p = 0.5$ when $p_{ii}^{(2n)} \sim \frac{1}{\sqrt{\pi n}}$. The one and two-dimensional symmetric simple random walks are null-recurrent but the three-dimensional walk is transient!

Definition 5 *The period $d(i)$ of state i is the greatest common divisor of n such that $p_{ii}^{(n)} > 0$. We call i periodic if $d(i) \geq 2$ and aperiodic if $d(i) = 1$.*

If two states i and j communicate with each other, then

- i and j have the same period,
- i is transient iff j is transient,
- i is null-recurrent iff j is null-recurrent.

Definition 6 *A chain is called irreducible if all states communicate with each other.*

All states in an irreducible chain have the same period d. It is called the period of the chain. Example: a simple random walk is periodic with period 2. Irreducible chains are classified as transient, recurrent, positively recurrent, or null-recurrent.

Definition 7 *State i is absorbing if $p_{ii} = 1$. More generally, C is called a closed set of states, if $p_{ij} = 0$ for all $i \in C$ and $j \notin C$.*

The state space S can be partitioned uniquely as

$$S = T \cup C_1 \cup C_2 \cup \ldots,$$

where T is the set of transient states, and the C_i are irreducible closed sets of recurrent states. If S is finite, then at least one state is recurrent and all recurrent states are positively recurrent.