Theory of Stochastic Processes
3. Generating functions and their applications

Tomonari Sei
sei@mist.i.u-tokyo.ac.jp

Department of Mathematical Informatics, University of Tokyo

April 20, 2017

http://www.stat.t.u-tokyo.ac.jp/~sei/lec.html

Handouts

There are 3 handouts today.

- Slides (this one)
- A copy of Sections 5.3 to 5.5 of PRP.
- A copy of end-of-chapter problems in Chapters 4 to 6. Make sure to bring it next time.

Outline today

1. Review of last week's material (slides)
2. Generating functions and their applications
   - Example: recurrence of random walk
   - Fundamental properties (slides)
   - Branching processes
3. Recommended problems

Review of last week's material

Simple random walk

A simple random walk is

$$S_n = S_0 + X_1 + \cdots + X_n,$$

where X_i are independent, $P(X_i = 1) = p$ and $P(X_i = -1) = q = 1 - p$.

A student gave a following-type question in the lecture.

Question on Markov property

Two statements are mentioned:

- S_{n+m} is independent of $S_0, \ldots, S_{n-1}$, conditional on S_n.
- The future is independent of the past, conditional on the present.

Where are $S_{n+1}, \ldots, S_{n+m-1}$?

Good question!

Conditional independence

Before giving an answer to the question, recall the notion of conditional independence.

- In the following, we only consider discrete random variables, and
- $P(Y \mid X)$ means "$P(Y = y \mid X = x)$ for any x, y".

Definition

We say that two variables X and Y are independent conditional on Z if

$$P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z) \quad \text{whenever} \quad P(Z) > 0.$$

Denote this relation by $X \perp\!\!\!\perp Y \mid Z$. (Dawid's notation)

Conditional independence

Lemma

$X \perp\!\!\!\perp Y \mid Z$ is equivalent to $P(X \mid Y, Z) = P(X \mid Z)$.

Proof.

Use the identity $P(X \mid Y, Z) = \dfrac{P(X, Y \mid Z)}{P(Y \mid Z)}$. □

Remark: One may ask what happens if $P(Z) > 0$ and $P(Y, Z) = 0$. For such cases, you have to redefine the conditional independence and study it carefully. We do not discuss this point anymore. If you get worried, refer to

- M. Studený (2005). *Probabilistic Conditional Independence Structure*, Springer.

On the Markov property

Here is an answer.

Theorem

For a process $\{S_n\}$, the following statements are equivalent to each other.

1. $S_{n+m} \perp\!\!\!\perp S_0, \ldots, S_{n-1} \mid S_n$ for any n, m. (def. of Markov property)
2. $S_{n+1}, \ldots, S_{n+m} \perp\!\!\!\perp S_0, \ldots, S_{n-1} \mid S_n$ for any n, m.
3. The joint mass function of $S_0, \ldots, S_n$ for any n is written as

$$P(S_0, \ldots, S_n) = P(S_0) \prod_{t=1}^{n} P(S_t \mid S_{t-1}).$$

Proof

You can skip.

Proof.

(2)$\to$(1) is easily proved by marginalization. Proofs of (1)$\to$(3) and (3)$\to$(2) are given below. □

Proof of (1)$\to$(3).

The statement (1) means

$$P(S_{n+1} \mid S_0, \ldots, S_n) = P(S_{n+1} \mid S_n).$$

By multiplying this equation over n's, we obtain

$$P(S_0) \prod_{i=1}^{n} P(S_i \mid S_0, \ldots, S_{i-1}) = P(S_0) \prod_{i=1}^{n} P(S_i \mid S_{i-1}).$$

The left hand side is equal to $P(S_0, S_1, \ldots, S_n)$. □

You can skip.

Proof of (3)→(2).

The statement (3) implies

$$P(S_0, \ldots, S_{n+m}) = P(S_0, \ldots, S_n) \prod_{t=n+1}^{n+m} P(S_t \mid S_{t-1}).$$

By summing up both sides with respect to $S_0, \ldots, S_{n-1}$, we have

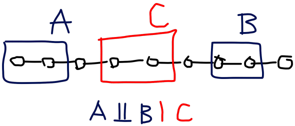
$$P(S_n, \ldots, S_{n+m}) = P(S_n) \prod_{t=n+1}^{n+m} P(S_t \mid S_{t-1}).$$

From the above two equations, we obtain the relation

$$\begin{aligned} P(S_{n+1}, \ldots, S_{n+m} \mid S_0, \ldots, S_n) &= \prod_{t=n+1}^{n+m} P(S_t \mid S_{t-1}) \\ &= P(S_{n+1}, \ldots, S_{n+m} \mid S_n). \end{aligned}$$

Remark: Graphical model

The Markov property is visualized as follows.



- But this picture is rarely used in the class since it might be confused with the transition diagram of Markov chains introduced next week.
- More generally, the following theorem is known.

Hammersley-Clifford theorem (e.g. Theorem 3.9 of Lauritzen (1996))

Let $\mathbf{X} = (X_v)_{v \in V}$ be a random vector indexed by V, and G be an undirected graph with vertices V. Suppose that the mass function $f(\mathbf{x})$ is positive everywhere. Then all the conditional independence relations implied by G hold if and only if $f(\mathbf{x}) = \prod_{C:clique} \psi_C(\mathbf{x}_C)$ for some ψ_C's.

Outline of today's lecture

1. Review of last week's material (slides)
2. Generating functions and their applications
   - Example: recurrence of random walk
   - Fundamental properties (slides)
   - Branching processes
3. Recommended problems

Recurrence of random walk

Blackboard

- Let S_n be a simple random walk with $S_0 = 0$.
- Find the probability of

$$\{\exists n \geq 1,\ S_n = 0\}.$$

Generating function

Generating functions are sometimes useful for thinking of "recurrence".

Definition

For any sequence $a = \{a_n\}_{n=0}^{\infty}$ of numbers, the (ordinal) generating function is defined by

$$G_a(s) = \sum_{n=0}^{\infty} a_n s^n.$$

Example

Consider a recurrence formula (= difference equation)

$$a_k = \frac{1}{k!} + \frac{1}{2} a_{k-1} \quad (k \geq 1), \quad a_0 = 1.$$

By multiplying s^k on both sides and summing over $k \geq 1$, we obtain

$$G_a(s) - 1 = e^s - 1 + \frac{1}{2} s G_a(s).$$

$$\begin{aligned} \Rightarrow \quad G_a(s) &= \frac{e^s}{1 - s/2} \\ &= \left(\sum_m \frac{s^m}{m!} \right) \left(\sum_n \frac{s^n}{2^n} \right). \end{aligned}$$

You may expand the right hand side to obtain each term a_k.

Properties (taken from p.150 of PRP)

Convolution

If $c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$, then $G_c(s) = G_a(s) G_b(s)$.

Convergence

There exists a radius of convergence R (≥ 0) such that the sum converges absolutely if $|s| < R$ and diverges if $|s| > R$.

Differentiation

$G_a(s)$ may be differentiated or integrated term by term any number of times at points s satisfying $|s| < R$. For example, $G_a'(s) = \sum_{n \geq 1} n a_n s^{n-1}$.

Uniqueness

If $R > 0$, the sequence $\{a_n\}$ is uniquely determined by $G_a(s)$. Explicitly,

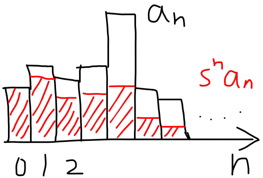
$$a_n = \frac{1}{n!} G_a^{(n)}(0) \quad \text{(note: this calculation is often unnecessary).}$$

Abel's theorem

If $a_n \geq 0$ for all n and $G_a(s) < \infty$ for $|s| < 1$, then

$$\lim_{s \uparrow 1} G_a(s) = \sum_{n=0}^{\infty} a_n,$$

where the sum is finite or $+\infty$.



- For students who know measure theory: Abel's theorem is a particular case of Lebesgue's monotone convergence theorem.

You can skip.

Proof of Abel's theorem.

Suppose first that $\sum_{n=0}^{\infty} a_n = +\infty$. Fix any large number $M > 0$. Then there is an integer N such that $\sum_{n=0}^{N} a_n > M$. Then

$$G_a(s) = \sum_{n=0}^{\infty} a_n s^n \geq \sum_{n=0}^{N} a_n s^n \to \sum_{n=0}^{N} a_n \quad \text{as } s \uparrow 1.$$

Thus $\underline{\lim}_{s\uparrow 1} G_a(s) \geq M$. Since M is arbitrary, $\lim_{s\uparrow 1} G_a(s) = \infty$.

Next suppose that $A = G_a(1) = \sum_{n=0}^{\infty} a_n$ is finite. Fix any small number $\varepsilon > 0$. Then there is an integer N such that $\sum_{n=N+1}^{\infty} a_n < \varepsilon$. Then

$$|G_a(s) - A| \leq \sum_{n=0}^{\infty} a_n |s^n - 1| \leq \sum_{n=0}^{N} a_n |s^n - 1| + \varepsilon \to \varepsilon \quad \text{as } s \uparrow 1.$$

Thus $\overline{\lim}_{s\uparrow 1} |G_a(s) - A| \leq \varepsilon$. Since ε is arbitrary, $\lim_{s\uparrow 1} |G_a(s) - A| = 0$. $\square$

Probability generating function

Definition

The (probability) generating function $G_X(s)$ of a random variable X taking values in non-negative integers is defined by

$$G_X(s) = E[s^X] = \sum_{k=0}^{\infty} s^k f(k),$$

where $f(k) = P(X = k)$ is the mass function of X.

It is obvious that $G_X(1) = 1$.

Examples

- If $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$, then $G_X(s) = (1 - p + ps)^n$.
- If $f(k) = p^k(1-p)$, then $G_X(s) = (1-p)/(1-ps)$.

Properties

We have the following properties as before.

Convolution

If X and Y are independent, then $G_{X+Y}(s) = G_X(s)G_Y(s)$.

Convergence

$G_X(s)$ absolutely converges if $|s| \leq 1$.

Differentiation

$G'_X(1) = E[X]$ and $G''_X(1) = E[X(X-1)]$.

Uniqueness

$f(n)$ is uniquely determined by $G_X(s)$. Explicitly, $f(n) = \dfrac{G^{(n)}(0)}{n!}$.

Other transform

The following is relevant. But we do not use them today.

- Moment generating function $M_X(t) = E[e^{tX}]$, $t \in \mathbb{R}$.
- Fact: If $M_X(t) < \infty$ over an open interval containing 0, then M_X is analytic over the interval and $M_X^{(n)}(0) = E[X^n]$.
- Characteristic function $\phi_X(t) = E[e^{itX}]$, $i = \sqrt{-1}$, $t \in \mathbb{R}$.
- Fact: The characteristic function is well defined for any random variable X. The distribution of X is uniquely determined by $\phi_X(t)$.

- Correspondence:
  - probabilistic generating function = Z-transform
  - moment generating function = Laplace transform
  - characteristic function = Fourier transform

Random walk again

Now let us find out the recurrence probability of the random walk using generating functions.

Blackboard

There are other approaches (exercise)

- Using absorbing probability
- Using the reflection principle

Branching processes

Blackboard

- Suppose that a population evolves in generations.
- Let Z_n be the number of members of the nth generation.
- Each member of the nth generation gives birth to a family of members of the $(n+1)$th generation.
- Assumptions:
  (a) $Z_0 = 1$.
  (b) $Z_n = X_1^{(n)} + \cdots + X_{Z_{n-1}}^{(n)}$.
  (c) $X_i^{(j)}$ are independent and have the same probability mass function f and the generating function G.
- Z_n is called a branching process (or Galton-Watson process).
- How to obtain the generating function $G_n(s)$ of Z_n using G?

Recommended problems

Recommended problems:

- §5.3, Problem 1, 3*.
- §5.4, Problem 4.
- §5.12, Problem 5, 6*, 10*, 11, 17.

The asterisk (*) shows difficulty.