Advanced Probability Theory
Supplementary Materials

Xinyi Li

BICMR, Peking University

November 10, 2020

These slides contain materials covered in-class but not included in Durrett's textbook.

Percolation

Example

Fix $p \in [0,1]$ and consider the d-dimensional lattice $\mathbb{Z}^d$. Assign to each edge $e \in \mathbb{E}$ a independent Bernoulli r.v. $I(e)$ with parameter p. If $I(e) = 1$, we say that this edge is **open**, *otherwise* **closed**. *Consider the connected components of open edges, then for any $p \in [0,1]$,*

$$P_p(A) = 0 \text{ or } 1, \text{ where } A = \{\exists \text{ infinite open clusters}\}.$$

Hint: Order the edges in a certain manner and fit A into a tail σ-field.

Remark: Actually we can go further and show that (this requires the knowledge of ergodicity) for any $N = 0, 1, \ldots, \infty$

$$P_p[A(N)] = 0 \text{ or } 1, \text{ where } A(N) = \{\exists N \text{ infinite open clusters}\}.$$

Or even further: for $N = 2, 3, \ldots$ (relatively easy) and $N = \infty$ (requires a clever argument), $P_p[A(N)] = 0$.

Percolation

Theorem

Let $p_c = p_c(d) = \sup\{p : P_p(A) = 0\}$. *Then one can show that*

$$1/3 \le p_c(2) \le 2/3.$$

More generally, $p_c(1) = 1$ *and for* $d \ge 2$*,*

$$1/(2d-1) \le p_c(d) \le p_c(2)(= 1/2).$$

Remark: By knowledge of Galton-Watson tree and the analogy between $\mathbb{Z}^d$ and $2d$-regular tree in high dimensions, we can take an educated guess that

$$p_c(d) \sim \frac{1}{2d} \quad \text{as } d \to \infty.$$

The very beautiful proof of $p_c = 1/2$ is much longer so we do not work it out here.

Percolation

Proof: The one-dimensional case $p_c(1) = 1$ is obvious.
Let $A_0 = \{0 \leftrightarrow \infty\}$. Then,

$$p_c = \sup\{p : P_p(A_0) = 0\}.$$

For the lower bounds, think that for any $n \geq 0$

$$P_p(A_0) \leq P_p(\exists \text{ self-avoiding paths from 0 of length } n).$$

Observe that RHS $\leq p^n(2d-1)^n \to 0$ as $n \to \infty$ if $p < \frac{1}{2d-1}$.
For the upper bounds, first observe that $p_c(d)$ decreases in d.

Percolation

For $d = 2$, write $B(n)$ for the box of side length $2n$. Note that

$$P_p(A_0) \geq P_p(D_n \cap E_n) = P_p(D_n)P_p(E_n)$$

where $D_n = \cap_{e \in B(n)}\{I(e) = 1\}$ and $E_n = \{\partial B(n) \leftrightarrow \infty\}$. Note that D_n and E_n are independent because one can write E_n as $\{\partial B(n) \overset{\text{in } B(n)^c}{\longleftrightarrow} \infty\}$ and hence they depend on sets of edges that are non-intersecting. It is easy to see that

$$P_p(D_n) \geq p^{c(n)}$$

for some $c(n)$. Considering the dual graph of $\mathbb{Z}^2$ and the self-avoiding loops in the dual graph that encircles $B(n)$, we can prove that

$$P_p(E_n^c) \leq \sum_{k=8n+4} Ck^2(1-p)^k(2d-1)^k \text{ for some fixed } C.$$

It is not hard to see that if $p \geq 2/3$, RHS $\to 0$ as $n \to \infty$. Now pick some n that is sufficiently large such that $P_p(E_N) \geq 1/2$. Finally, $P_p(A_0) \geq p^{c(n)}/2 > 0$ and hence $p_c(2) \leq 2/3$. $\square$

Example (Coupon collector's problem)

Let $X_1, X_2, \cdots$ *be i.i.d. uniform on* $\{1, 2, \cdots, n\}$ *and* $\tau_k^n = \inf\{m : |\{X_1, \cdots, X_n\}| = k\}$*. As* $n \to \infty$*, we want to consider the asymptotic of* $T_n = \tau_n^n$*.*

Remark: This is (almost) same as the **cover time** problem for a simple random walk on complete graph.

Solution: Let $X_{n,k} = \tau_k^n - \tau_{k-1}^n$. It is easy to see $X_{n,k}$ are independent, and that $\sum_{i=1}^n X_{n,k} = T_n$. Moreover $X_{n,k} \sim \text{Geom}(1 - (k-1)/n)$. Thus,

$$E[T_n] = n \sum_{m=1}^{n} \frac{1}{m} \sim n \log(n);$$

$$\text{var}(T_n) \le n^2 \sum_{m=1}^{n} \frac{1}{m^2} = O(n^2) = o((n \log n)^2).$$

Let $b_n = n \log n$, one can see that $T_n / b_n \to 1$ in probability.

Weak Laws of Large Numbers

- **Question:** How close is T_n to b_n?
- As $n \to \infty$,
$$\frac{T_n}{n} - \log(n) \Rightarrow \eta$$
where η has density and CDF as follows:
$$f_\eta(x) = \exp(-x - e^{-x}) \quad \text{and} \quad F(x) = \exp(-\exp(-x)),$$
which is usually referred to as the **Gumbel distribution**.
- **Remark:** cf. Exercise 3.2.2 (iii): Let $X_1, X_2, \dots$ be i.i.d. with CDF F and let $M_n = \max_{m \le n} X_m$. Then
$P(M_n \le x) = F(x)^n$.
If $X_i \sim \text{Exp}(1)$, then for any $y \in (-\infty, \infty)$,
$$P(M_n - \log n \le y) \to \exp(-\exp(-y)).$$
- **Question:** For what other Markov chains we have the same limiting distribution?

A counterpart of Borel-Cantelli lemma

Lemma

For a sequence of **increasing** *events* A_n,

$$P(A_n \ i.o.) = 1 \quad \text{iff} \quad \sum_n P(A_n | A_{n-1}^c) = \infty.$$

Remark: The event $\{A_n \text{ i.o.}\}$ is equivalent to "at least one of A_n's happens".

Proof.

By telescoping we have

$$P(A_n^c) = (A_1^c) \prod_{i=1}^{n-1} P(A_{i+1}^c | A_i^c) = P(A_1^c) \prod_{i=1}^{n-1} \left[1 - P(A_{i+1} | A_i^c)\right].$$

Again note that $\prod_{n=1}^{\infty}(1 - a_n) = 0 \Leftrightarrow \sum_{n=1}^{\infty} a_n = \infty$ and we are done. □

Uniform law of large numbers

We now state without proof another SLLN result with the same flavor.

Theorem

Suppose $f(x, \theta)$ is continuous in $\theta \in \Theta$ for some compact Θ. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. If

- *f is continuous at θ for a.s. all $x \in \mathbb{R}$ and measurable of x at each θ;*
- *There exists some function $d(x)$ such that $E[d(X_i)] < \infty$ and for all $\theta \in \Theta$,*

$$|f(x, \theta)| \leq d(x).$$

Then

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i, \theta)/n - E[f(X_1, \theta)] \right| \xrightarrow{\text{a.s.}} 0.$$

Shannon's theorem and lossless compression

Let $X_1, X_2, \ldots \in \{1, \ldots, r\}$ be i.i.d. with $P(X_i = k) = p(k) > 0$ for $1 \le k \le r$. Here we are thinking of $\{1, \ldots, r\}$ as the letters of an alphabet, and $X_1, X_2, \ldots$ are the successive letters produced by an information source. Let

$$\pi_n(\omega) = p(X_1(\omega)) \cdots p(X_n(\omega))$$

be the probability of the realization we observed in the first n trials. It follows from SLLN that

$$-n^{-1} \log \pi_n(\omega) \to H \equiv -\sum_{i=1}^{r} p(k) \log p(k).$$

The constant H is called the entropy of the source and is a measure of how random it is.

Theorem (asymptotic equipartition property)

If $\epsilon > 0$ then as $n \to \infty$,

$$P\left[2^{-n(H+\epsilon)} \le \pi_n(\omega) \le 2^{-n(H-\epsilon)}\right] \to 1.$$

Shannon's theorem and lossless compression

Theorem (asymptotic equipartition property)

If $\epsilon > 0$ then as $n \to \infty$,

$$P\left[2^{-n(H+\epsilon)} \leq \pi_n(\omega) \leq 2^{-n(H-\epsilon)}\right] \to 1.$$

We call a sequence of letters $X_1 X_2 \cdots X_n$ (ϵ-)typical if

$$2^{-n(H+\epsilon)} \leq p(X_1(\omega)) \cdots p(X_n(\omega)) \leq 2^{-n(H-\epsilon)}.$$

Theorem

There is an algorithm such that for every $\epsilon > 0$, there exists some big N such that for all $n \geq N$, the expected total bits needed to encode a message of length n is less than $n(H + \epsilon)$.

Proof: If the message is typical we can encode it with $n(H + \epsilon + \epsilon')$ bits, otherwise we keep the original sequence of $2^r n$ bits, but this happens with very small probability (say $< \epsilon H / 2^r$) if n is sufficiently large. □

Kronecker's lemma

Theorem (Kronecker's lemma)

If $a_n \uparrow \infty$ *and* $\sum_{n=1}^{\infty} x_n/a_n$ *converges, then*

$$a_n^{-1} \sum_{m=1}^{n} x_m \to 0.$$

Remark: The converse is not true! E.g., one can take $x_n = (\log n)^{-1}$ and $a_n = n$. In fact any $(x_n)_n$ with $x_n \to 0$ and $\sum x_n/n$ diverges would suffice.

Cauchy Distribution

Example

For an example where the weak law does not hold, suppose $X_1, X_2, \cdots$ *are independent and have a* **Cauchy distribution***:*

$$P(X_i \leq x) = \int_{-\infty}^{x} \frac{dt}{\pi(1+t^2)}.$$

- One can see that, as $x \to \infty$,

$$P(|X_1| > x) = 2\int_{x}^{\infty} \frac{dt}{\pi(1+t^2)} \sim \frac{2}{\pi} x^{-1}.$$

- In fact, we can show later[1] that $S_n/n =^d X_1$.

---

[1] These are the contents of Chap. 3.8 and 3.9, which we might not have time to cover in this course.

Infinitely divisible distributions

- Recall that $S_n/n =^d X_1$.
- In other words, Cauchy distribution is **infinitely divisible**:

Definition

The distribution of X is inf. div. iff for any $n \in \mathbb{N}$, there exists i.i.d. Y_i's such that $X = \sum_{i=1}^n Y_i$.

- Examples of infinitely divisible distributions include: Poisson, geometric, gamma, normal, etc.
- In fact, all **(discrete) stable distributions** are infinitely divisible.

Definition

The distribution of X is stable if for all $a, b > 0$, and X_1, X_2 i.i.d. copies of X, $aX_1 + bX_2 =^d cX + d$ for some $c > 0$ and d.

St. Petersburg Paradox

Example

A game goes as follows: you keep on tossing a coin until you get a heads. If it takes k tosses you will win 2^k dollars. Your expected revenue is the expectation of the random variable X with $P(X = 2^j) = 2^{-j}$ for $j > 0$, which is **infinite**. *How much would you pay to play this game?*

Nicolas Bernoulli wrote in 1713, "there ought not to exist any even halfway sensible person who would not sell the right of playing the game for 40 ducats (per play)."

Recall: In weak law for triangular arrays, in order to make $(S_n - a_n)/b_n \to 0$ in probability, we need

1. $\sum_{k=1}^n P(|X_{n,k}| > b_n) \to 0$.
2. $b_n^{-2} \sum_{k=1}^n \mathrm{var}(\bar{X}_{n,k}) \to 0$.

where $\bar{X}_{n,k} = X_{n,k} 1_{|X_{n,k}| \le b_n}$, $S_n = \sum_{k=1}^n X_{n,k}$ and $a_n = \sum_{k=1}^n E[\bar{X}_{n,k}]$.

Infinite mean

Recall St. Petersburg paradox:

Example

Let X_i be i.i.d. with distribution $P(X_1 = 2^j) = 2^{-j}$ for $j > 0$, which has **infinite** *mean. We have proved that*

$$S_n/(n \log_2 n) \to 1 \text{ in probability.}$$

but we can prove via Borel-Cantelli lemma (see Exercise 2.3.20)

$$\limsup S_n/(n \log_2 n) = \infty \text{ a.s.}$$

Theorem

Let $X_1, X_2, \ldots$ be i.i.d. with $E[|X_1|] = \infty$ and let $S_n = X_1 + \cdots + X_n$. Let a_n be a sequence of positive numbers with a_n/n increasing. Then $\limsup_{n\to\infty} |S_n|/a_n = 0$ or ∞ according as $\sum_n P(|X_1| \geq a_n) < \infty$ or $= \infty$.

A list of sources

- A short presentation of what I covered in-class on percolation is Chap. 2.4 of “Probability Theory” by Achim Klenke.
- For infinitely divisible distributions, see Chap. 3.9 of Durrett.
- The converse of Kronecker’s lemma is an easy exercise.
- Uniform law of large numbers: just search with the keyword.
- Shannon’s Theorem: any textbook on information theory.