Probability Theory Revision Notes

Kexing Ying

May 30, 2022

Theorem. For non-negative random variable ξ,

$$\mathbb{E}\xi^k = k \int t^{k-1}\mathbb{P}(\xi \geq t)\lambda(\mathrm{d}t).$$

Theorem. If $\mathbb{E}|\xi_n|^k$ converges to 0 for some k, then $\xi_n \to 0$ in probability.

Theorem. If $\sum \mathbb{P}(|\xi_n| \geq \epsilon) < \infty$, then $\xi_n \to 0$ almost everywhere (consider what it means for $\omega \in \{\xi_n \not\to 0\}$).

Corollary. If $\sum \mathbb{E}|\xi_n|^k < \infty$ for some k, then $\xi_n \to 0$ almost everywhere.

Theorem. $\xi_n \to \xi$ almost every where if and only if $\mathbb{P}(\sup_{k \geq n} |\xi_k - \xi| \geq \epsilon) \to 0$ for all $\epsilon > 0$ (Exercise sheet 5).

Method for showing SLLN without KSLLN (requires independence, equal mean (WLOG, mean 0), not necessary identically distributed):

- By relabelling, we have by Kolmogorov's inequality

$$\mathbb{P}\left(\max_{2^n \leq k \leq 2^{n+1}} |S_k| \geq \epsilon\right) \leq \frac{1}{\epsilon^2}\sum_{k=2^n}^{2^{n+1}} V_{\xi_k}.$$

- Choosing ϵ to be $n\epsilon$, we have

$$\mathbb{P}\left(\max_{2^n \leq k \leq 2^{n+1}} \left|\frac{1}{n}S_k\right| \geq \epsilon\right) \leq \frac{1}{n^2\epsilon^2}\sum_{k=2^n}^{2^{n+1}} V_{\xi_k}.$$

- Summing over n, we have

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\max_{2^n \leq k \leq 2^{n+1}} \left|\frac{1}{n}S_k\right| \geq \epsilon\right) \leq \sum_{n=1}^{\infty} \frac{1}{n^2\epsilon^2} V_{\xi_n} < \infty$$

  if $V_{\xi_n} \leq 1$ or some other conditions.

- Hence, by the first Borel-Cantelli lemma,

$$\mathbb{P}\left(\left|\frac{1}{n}S_k\right| \geq \epsilon \text{ i.o.}\right) = \mathbb{P}\left(\max_{2^n \leq k \leq 2^{n+1}} \left|\frac{1}{n}S_k\right| \geq \epsilon \text{ i.o.}\right) = 0.$$

- Thus, for all $\epsilon > 0$

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{m\geq n}\left|\frac{1}{n}S_k\right| < \epsilon\right) = \mathbb{P}\left\{\left|\frac{1}{n}S_k\right| \geq \epsilon \text{ i.o.}\right\}^c = 1$$

which implies convergence to 0 almost everywhere by intersecting over ϵ.

Characteristic Function

Let ξ be a random variable and let ϕ be its characteristic function

- $\phi(t) = \mathbb{E}e^{it\xi}$;
- $|\phi(t)| \leq \phi(0) = 1$;
- $\phi(-t) = \overline{\phi(t)}$;
- ϕ is uniformly continuous on $\mathbb{R}$;
- $\mathbb{E}\xi^r = i^{-r}\phi^{(r)}(0)$ if $\mathbb{E}|\xi|^n < \infty$ where $r \leq n$;
- for random variables $\xi_1, \cdots, \xi_n$, the characteristic function of $\sum \xi_i$ is $\prod \phi_i$ if and only if ξ_i are independent;
- convex linear combinations of characteristic functions is a characteristic function;
- given $\alpha \in \mathbb{R}$, $\overline{\phi}$, Re(ϕ), $|\phi|^2$ and $\phi(\alpha t)$ are all characteristic functions;
- if $|\phi(t_0)| = 1$ for some $t_0 \neq 0$, then ξ is a pure point random variable;
- if $|\phi(t)| = 1$ for all $|t| \in (-\epsilon, \epsilon)$ then ξ is degenerate;
- see also Bochner, Polya and Marcinkievicz theorems.

Theorem. If ξ is a discrete random variable with characteristic function ϕ, then

$$\mathbb{P}(\xi = k) = \frac{1}{2\pi}\int_{[-\pi,\pi]} e^{-ikt}\phi(t)\lambda(\mathrm{d}t).$$