B. Probability theory

B.1. Basic inequalities.

Proposition B.1 (Borel-Cantelli)**.** *Let $(S_n)_{n=1}^{\infty}$ be a sequence of events in a probability space $(\mathcal{X}, \mathcal{B}, \mu)$. Then,*

$$\mu\Big(\limsup_{n\to\infty} S_n\Big) \leq \lim_{N\to\infty} \sum_{n=N}^{\infty} \mu(S_n).$$

Proposition B.2 (Markov)**.** *Let X be a nonnegative random variable. Then, for every $\alpha > 0$,*

$$\mu(X \geq \alpha) \leq \alpha^{-1} \cdot \mathbb{E}[X].$$

Proposition B.3 (Hölder)**.** *Let X, Y be non-negative random variables. Then, for $p, q \in [1, \infty]$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$,*

$$\mathbb{E}(X^p)^{1/p}\mathbb{E}(Y^q)^{1/q} \geq \mathbb{E}(XY),$$

where the equality holds if and only if $X = cY$ for some $c > 0$.

Proof. Without loss of generality, assume $\mathbb{E}(X^p) = \mathbb{E}(Y^q) = 1$ to paraphrase the proposition as $1 \geq \mathbb{E}(XY)$ with equality holding if and only if $X = cY$ for some $c > 0$. Under the circumstances, it is not hard to verify the Young's inequality

$$XY \leq \frac{X^p}{p} + \frac{Y^q}{q}$$

holds if and only if $X^p = Y^q$. The paraphrased proposition then follows by integrating both sides. □

Proposition B.4 (Jensen)**.** *Let X be a real-valued random variable and $\varphi : \mathbb{R} \to \mathbb{R}$ is convex. If $\mathbb{E}(X)$ and $\mathbb{E}(\varphi(X))$ are defined, then $\mathbb{E}(\varphi(X)) \geq \varphi(\mathbb{E}(X))$ with the convention $\varphi(+\infty) = \lim_{x\to\infty} \varphi(x)$ and $\varphi(-\infty) = \lim_{x\to-\infty} \varphi(x)$.*

Proof. We make use the property of convex functions that

$$\varphi(x) = \sup\{ax + b : \varphi(x) \geq ax + b \text{ for all } x \in \mathbb{R}\}. \tag{B.1}$$

The left-hand side is by definition larger than the other, and thus it remains to show the remaining. To this end, it suffices to show that for all (x_0, y_0) satisfying $\varphi(x_0) \geq y_0$, there exists a linear function $\psi(x) = a(x - x_0) + y_0$ such that $\varphi(x) \leq \psi(x)$ for all x. Essentially, this is achieved by choosing

$$\begin{aligned} a &\in \left[\sup_{x<x_0} \frac{\varphi(x) - \varphi(x_0)}{x - x_0}, \inf_{x>x_0} \frac{\varphi(x) - \varphi(x_0)}{x - x_0}\right] \\ &= \left[\liminf_{x\to x_0} \frac{\varphi(x) - \varphi(x_0)}{x - x_0}, \limsup_{x\to x_0} \frac{\varphi(x) - \varphi(x_0)}{x - x_0}\right]. \end{aligned}$$

Now that (B.1) coincides with

$$\varphi(x) = \sup\{ax + b : \varphi(x) \geq ax + b \text{ for all } x \in \mathbb{R}, a, b \in \mathbb{Q}\}, \tag{B.2}$$

one can enumerate the linear functions on the right-hand side by $(\psi_i)_{i=1}^{\infty}$ and obtain

$$\mathbb{E}(\varphi(X)) \geq \mathbb{E}\Big(\max_{1\leq i\leq n} \psi_i(X)\Big) \geq \max_{1\leq i\leq n} \psi_i(\mathbb{E}(X)).$$

If $\mathbb{E}(X) \in \mathbb{R}$, then the proposition holds by letting $n \to \infty$. If $\mathbb{E}(X) = \infty$ (similar for $\mathbb{E}(X) = -\infty$) and $\varphi(\infty) > -\infty$, then right-hand side of the above still converges to $\varphi(\infty)$, while the proposition is trivial when $\mathbb{E}(X) = \infty$ and $\varphi(\infty) = -\infty$. □

B.2. **Radon-Nikodym theorem.**

Definition B.5 (absolute continuity)**.** Let μ and ν be defined on a common measurable space $(\mathcal{X}, \mathcal{B})$. We say ν is *absolutely continuous* with respect to μ, denoted by $\nu \ll \mu$, if $\nu(A) = 0$ for every $A \in \mathcal{B}$ satisfying $\mu(A) = 0$.

Theorem B.6 (Radon-Nikodym)**.** *Suppose ν, μ are two σ-finite measures on a common measurable space $(\mathcal{X}, \mathcal{B})$ and $\nu \ll \mu$, then there exists a $\mathcal{B}$-measurable function $f : \mathcal{X} \to [0, \infty)$ such that for every $A \in \mathcal{B}$,*

$$\nu(A) = \int_A f d\mu.$$

B.3. **Laws of large numbers.**

Theorem B.7 (Weak Law of Large Numbers)**.** *Suppose $(X_i)_{i=1}^{\infty}$ are i.i.d. and $\mathbb{E}(X_1) = 0$. Then, $n^{-1}\sum_{i=1}^n X_i \to 0$ in probability.*

Proof. Equivalently, we can assume X_i is nonnegative and $\mathbb{E}(X_1) < \infty$ and prove $n^{-1}\sum_{i=1}^n X_i \to \mathbb{E}(X_1)$. Let $M > 0$ and $X_i = Y_{i,1} + Y_{i,2}$ with $Y_{i,1} = \max\{X_i, M\}$. Immediately,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n Y_{i,1} > \varepsilon\right) \leq \frac{nM}{n^2\varepsilon^2} \to 0 \text{ as } n \to \infty.$$

On the other hand,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n Y_{i,2} > \varepsilon\right) \leq \frac{n\mathbb{E}\big(X_1 \mathbb{1}_{X_1\geq M}\big)}{n\varepsilon} \to 0 \text{ as } M \to \infty.$$

The theorem is proved by combining the above. □

Lemma B.8 (Kronecker)**.** *Suppose $a_n > 0$ and $a_n \nearrow \infty$. Then $\sum_{i=1}^{\infty} a_n^{-1} x_n < \infty$ implies $a_n^{-1}\sum_{i=1}^n x_i \to 0$.*

Proof. Writing $b_n = (a_n^{-1} - a_{n-1}^{-1})$ with $a_0^{-1} := 0$, one may use summation by part to deduce

$$a_n^{-1}\sum_{i=1}^n x_i = \sum_{i=1}^n a_i(a_i^{-1}x_i) = \sum_{i=1}^n a_i^{-1}x_i + \sum_{i=1}^{n-1} \frac{a_i - a_{i+1}}{a_n} \sum_{j=1}^i a_j^{-1}x_j.$$

The last expression converges to 0 as $n \to \infty$. □

Proposition B.9 (Kolmogorov's Criterion of SLLN)**.** *Suppose $(X_i)_{i=1}^{\infty}$ are independent such that $\mathbb{E}(X_n) = 0$ and $\sum_{i=1}^{\infty} i^{-2}\operatorname{Var}(X_i) < \infty$. Then, $n^{-1}\sum_{i=1}^n X_i \to 0$ a.s.*

Proof. By virtue of the independence, one observes that

$$\mathbb{E}\left(\left|\sum_{i=1}^{\infty}\frac{X_i}{i}\right|^2\right) \le \sum_{i=1}^{\infty}\frac{\mathrm{Var}(X_i)}{i^2} < \infty,$$

which, in particular, implies $\sum_{i=1}^{\infty}\frac{X_i}{i}$ is finite almost surely and in turn implies the conclusion by Lemma B.8. □

Proposition B.10 (Strong Law of Large Numbers)**.** *Suppose* $(X_i)_{i=1}^{\infty}$ *are i.i.d. and* $\mathbb{E}(X_1) = 0$*. Then,* $n^{-1}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_1)$ *a.s.*

Proof. It suffices to prove the case $\mathbb{E}(|X_1|) < \infty$, for if otherwise, one may still apply the result to $Y_n = \min\{M, X_n\}$ for every $M > 0$ if $\mathbb{E}(X_1^+) < \infty$ (resp., $Y_n = \max\{-M, X_n\}X_n$ if $\mathbb{E}(X_1^-) > -\infty$) so that $\mathbb{E}(|Y_n|) < \infty$ and that $Y_n \le X_n$ (resp., $Y_n \ge X_n$), leading to the conclusion by letting $M \to \infty$.

To begin, truncate X_n to define

$$Y_n = \mathbb{1}_{\{|X_n|\le n\}} X_n - \mathbb{E}\Big(X_n \mathbb{1}_{\{|X_n|\le n\}}\Big).$$

Then, verify the assumption of Proposition B.9

$$\begin{aligned}\sum_{i=1}^{\infty}\frac{\mathrm{Var}(Y_i)}{i^2} &\le \sum_{i=1}^{\infty}\frac{\mathbb{E}\Big(|X_1|^2\,\mathbb{1}_{\{|X_1|\le i\}}\Big)}{i^2} = \left(\mathbb{E}\left(|X_1|^2\sum_{i=\max\{1,|X_1|\}}^{n}\frac{1}{i^2}\right)\right)\\ &\le \mathbb{P}(|X_1|<1) + \mathbb{E}\left(|X_1|^2\sum_{i=|X_1|}^{n}\frac{2}{i(i+1)}\right)\\ &\le \mathbb{P}(|X_1|<1) + 2\mathbb{E}(|X_1|).\end{aligned}$$

to deduce

$$\frac{1}{n}\sum_{i=1}^{\infty} Y_i \to 0. \tag{B.3}$$

On the other hand, by the Borel-Cantelli lemma,

$$\mathbb{P}(|X_n| > n \text{ infinitely often}) \le \limsup_{n\to\infty}\sum_{i=n}^{\infty}\mathbb{P}(|X_i|\ge i) \le \limsup_{n\to\infty}\mathbb{E}(|X_1|). \tag{B.4}$$

Combining (B.3) and (B.4) gives that

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} X_i = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\Big(Y_i + \mathbb{E}\Big(|X_1|\,\mathbb{1}_{\{|X_1|\le i\}}\Big)\Big) = \mathbb{E}(X_1).$$

□