1 Convergence and inequalities

Theorem 1 (Fubini's theorem). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the n-fold product of $\Omega_1, ..., \Omega_n$. If* *either* *$f \geq 0$ or $\int |f| d\mu < \infty$, then*

$$\int f d\mathbb{P} = \int_{\Omega_n} f ... (\int_{\Omega_1} f d\mathbb{P}_1) ... f d\mathbb{P}_n \tag{1}$$

Theorem 2 (Bounded convergence theorem, Durrett, p26). *Let E be a set with $\mu(E) < \infty$. Suppose f_n vanishes on $E^c, |f_n(x)| \leq M$, and $f_n \to f$ in measure. Then*

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu \tag{2}$$

Theorem 3 (Fatou's lemma). *If $f_n \geq 0$ then*

$$\int \left(\liminf_{n\to\infty} f_n \right) d\mu \leq \liminf_{n\to\infty} \int f_n d\mu \tag{3}$$

Theorem 4 (Monotone convergence theorem). *If $f_n \geq 0$ and $f_n \uparrow f$, then*

$$\int f_n d\mu \uparrow \int f d\mu \tag{4}$$

Theorem 5 (Dominated convergence theorem). *If $f_n \to f$ a.e., $|f_n| \leq g$ for all n, and g is integrable, then*

$$\int f_n d\mu \to \int f d\mu \tag{5}$$

Theorem 6 (Thm 1.6.8, Durrett). *Suppose $X_n \to X$ a.s. Let g, h be continuous function with*

(i) $g \geq 0$ *and* $g(x) \to \infty$ *as* $|x| \to \infty$

(ii) $\frac{|h(x)|}{|g(x)|} \to 0$ *as* $|x| \to \infty$

(iii) $\mathbb{E}g(X_n) \leq K < \infty$ *for all* n

Then

$$\mathbb{E}h(X_n) \to \mathbb{E}h(X) \tag{6}$$

Theorem 7 (Markov's inequality). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let X be a random variable on this space, and let $A \subseteq \mathbb{R}$ be any Borel-measurable set. Then for any non-negative real function ϕ, we have a bound:*

$$\mathbb{P}(X \in A) \leq \frac{\mathbb{E}\phi(X)}{\inf_{x\in A} \phi(x)} \tag{7}$$

and for nonnegative X

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a} \tag{8}$$

Theorem 8 (Chebyshev's inequality).

$$\mathbb{P}(|X - b| \geq a) \leq \frac{\mathbb{E}(X - b)^2}{a^2} \tag{9}$$

Lemma 1 (Kolmogorov's maximal inequality)**.** *Suppose $\{X_n\}$ are independent with mean zero and finite variance. Then*

$$\mathbb{P}\left(\max_{1\leq k\leq n}|S_k| \geq x\right) \leq x^{-2}\, Var(S_n) \tag{10}$$

Theorem 9 (Kolmogorov's Three Series Theorem)**.** *$\{X_n\}$ independent. $A > 0$ and define truncation $Y_n = X_n \mathbf{1}_{|X_n|\leq A}$. For $\sum_{n=1}^{\infty} X_n$ to converge a.e, it needs to satisfy:*

(i) $\sum_{n=1}^{\infty} \mathbb{P}(|X_n| > A)$

(ii) $\sum_{n=1}^{\infty} \mathbb{E}Y_n$

(iii) $\sum_{n=1}^{\infty} Var Y_n$

Theorem 10 (Jensen's inequality)**.** *For convex ϕ,*

$$\mathbb{E}[\phi(X)] \leq \phi(\mathbb{E}X) \tag{11}$$

as long as both expectations exist.

Theorem 11 (Hölder's inequality)**.** *For $1/p + 1/q = 1$,*

$$\mathbb{E}[XY] = ||XY||_1 \leq ||X||_p ||Y||_q \tag{12}$$

Corollary 1 (Cor 3.5 notes)**.** *X is random variable, $f(X,t)$ is differentiable in t, and $\mathbb{E}[f(X,t)]$ and $\mathbb{E}\left|\frac{\partial f(X,t)}{\partial t}\right|$ are bounded and continuous for t in an interval containing t_0, then*

$$\frac{d}{dt}\mathbb{E}f(X,t) = \mathbb{E}\frac{\partial}{\partial t}f(X,t) \tag{13}$$

Theorem 12 (Change of density, notes p.29)**.** *$f : \mathbb{R}^n \to \mathbb{R}^n$ smooth and invertible. Suppose $(X_1, \cdots, X_n)$ has law μ with density g, what is density of $f(X_1, \cdots, X_n)$ with law ν? What is density of push forward ν of μ?*

For each A, we have $\mu(A) = \int_A g\,dm$, by changing coordinates to $Y = f(X)$, we have

$$\nu(B) = \mu\left(f^{-1}B\right) = \int_{f^{-1}B} g\,dm = \int_B g \circ f^{-1}|J|^{-1}dm \tag{14}$$

where $|J|^{-1}$ is the inverse of the determinant of df at $f^{-1}(Y)$, so that

$$|J|^{-1} g \circ f^{-1} \tag{15}$$

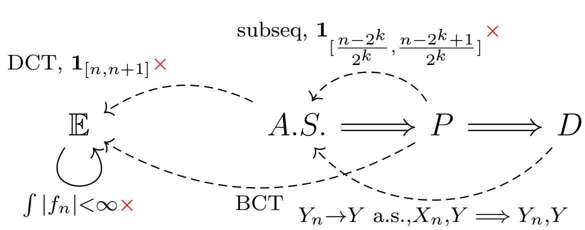
is the density of the push-forward.

2 Modes of Convergence

Definition 1 (a.s)**.** $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$

Definition 2 (p)**.** $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$

Definition 3 (d)**.** $\lim_{n\to\infty} F_n(x) = F(x)$



2.1 Counter Examples

We have gathered the following counterexamples:

(i) Convergence in probability $\not\Rightarrow$ almost surely: Typewriter sequence, which is

$$f_n(x) = \mathbf{1}_{\left[\frac{n-2^k}{2^k}, \frac{n-2^k+1}{2^k}\right]} \tag{16}$$

with $2^k \leq n < 2^{k+1}$. f_n tends to zero in probability, but not almost everywhere.

(ii) Convergence a.s. $\not\Rightarrow$ $\mathbb{E}$ converge:

(1) $\mathbf{1}_{n,n+1}$. Observe that

$$\lim_{n\to\infty} \mathbf{1}_{n,n+1} = 0 \text{ a.e. and } \mathbf{1}_{n,n+1} \leq 1 \tag{17}$$

and

$$\lim_{n\to\infty} \mathbb{E}\mathbf{1}_{n,n+1} = 1 \not\to 0 \tag{18}$$

(2) $n\mathbf{1}_{[0,\frac{1}{n}]}$. Observe that

$$\lim_{n\to\infty} f_n = 0 \text{ a.e.} \tag{19}$$

but

$$\lim_{n\to\infty} \mathbb{E}n\mathbf{1}_{[0,\frac{1}{n}]} = 1 \neq \int f dx = 0 \tag{20}$$

(iii) Three series, violated only (3): $\sum \pm n^{-\alpha}$, independent sum of mean zero powers. Note that since variance of n-th term is $n^{-2\alpha}$, then summable iff $\alpha > 1/2$.

(iv) LDP does not work: S_n sum of Cauchy IID, does $\{S_n/n\}$ satisfy LDP? No, interval J has empty interior. S_n/n has the same law as a single Cauchy variable, then there is a trivial LPD with rate function identically zero.

(v) Convergence in distribution only at points CDF is continuous: X random varaible and $X_n = X+1/n$. We must have $X_n \to X$. However, $F_n \not\to F$, where $F_n(x) = \mathbb{P}(X_n \leq x) = F(x-1/n)$, so $F_n(x) \to F(x-)$.

(vi) Convergence in distribution does not imply pair converge in distribution: $X_n = X, Y_n = Y$ and X, Y IID, then $X_n \Rightarrow X, Y_n \Rightarrow Y$, but $(X_n, Y_n) \Rightarrow (X, Y) = (X, X)$, contradiction.

3 Law of large number (weak + strong)

Theorem 13 (Best WLLN)**.** *$\{X_n\}$ IID with $t\mathbb{P}(X_1 > t) \to 0$ as $t \to \infty$. $S_n = \Sigma X_i$ *but* $\mu_n = \mathbb{E}X_1\mathbf{1}_{X_1<n}$. Then, $S_n/n - \mu_n \to 0$ in probability.*

Theorem 14 (W/SLLN)**.** *$\{X_n\}$ IID with $\mathbb{E}|X_1| < \infty$. Denote $S_n = \sum_{k=1}^n X_k$ and $\mu = \mathbb{E}X_1$, then*

$$\frac{S_n}{n} \to \mu \tag{21}$$

in probability/almost surely. (WLLN requires $VarX_n < \infty$)

Proof. **SLLN**: 1. Instead of triangular array, truncate X_n at different value; $|X_n| = n$, 2.pass subsequence in order for upper bounds on $\mathbb{P}(|S/n - \mu| > \epsilon)$ to be summable in n.,3. Can't do this with $n_j = j^\alpha$, but doable with $n_j = (1+\delta)^j$ for δ small. 4. Similar to proof of quantitative Borel-Cantelli, apply sandwich trick as long as S_n increases.,5. get a sandwiched SLLN between $1-\delta, 1+\delta$ with $\delta > 0$ small.

WLLN/L2: $\mathbb{E}(S_n/n) = \mu$, then $\mathbb{E}\left(\frac{S_n}{n} - \mu\right)^2 = \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{j=1}^n \mathrm{Var}X_j = \frac{C}{n} \to 0 \implies$ converge in probability via $\mathbb{E}|Z_n|^p \geq \epsilon^p \mathbb{P}(|Z_n| \geq \epsilon)$. □

Theorem 15 (Borel-Cantelli Lemma I). *If* $\sum_n \mathbb{P}(A_n) < \infty$, *then* $\mathbb{P}(A_n i.o.) = 0$.

Theorem 16 (Borel-Cantelli Lemma II). $\{A_n\}$ *independent. If* $\sum_n \mathbb{P}(A_n) = \infty$, *then* $\mathbb{P}(A_n i.o.) = 1$.

Theorem 17 (Borel-Cantelli Lemma II-quantitative). $\{A_n\}$ *pairwise independent. If* $\sum_n \mathbb{P}(A_n) = \infty$, *then*

$$\frac{\Sigma_1^n \mathbf{1}_{A_k}}{\Sigma_1^n \mathbb{P}(A_k)} \to \mathbf{1} \tag{22}$$

almost surely as $n \to \infty$.

Theorem 18 (HW Borel-Cantelli). *If* $\mathbb{P}(A_n) < 1$ *for all* n *and* $\mathbb{P}(\cup_n A_n) = 1$, *then* $\mathbb{P}(A_n \ i.o.) = 1$.

4 Large Deviation

The range $S_n > an$ with $a > \mu$ fixed and $n \to \infty$ is called a **large deviation**. If $\mathbb{E}\exp\{\lambda X_1\}$ exists for some $\lambda > 0$, then:

1. Compute an upper bound, depending on λ, using Markov's inequality
2. Optimize in λ, which for some positive function h, it yields

$$\mathbb{P}(S_n > an) \leq \exp(-h(a)n) \tag{23}$$

   Sharp in the sense that

$$n^{-1}\log \mathbb{P}(S_n > an) \to h(a) \tag{24}$$

3. Find an event. whose probability we can compute, contained in the event $\{S_n > an\}$ as the lower bound for $\mathbb{P}(S_n > an)$

Formally, we have

$$\mathbb{P}(S_n > an) \leq e^{-\lambda an}\mathbb{E}e^{\lambda S_n} \tag{25}$$

$$\frac{1}{n}\log \mathbb{P}(S_n > an) \leq -\lambda a + \psi(\lambda) \tag{26}$$

with $\psi(\lambda) = \log\phi(\lambda) = \log \mathbb{E}e^{\lambda X_1}$. Optimize over λ to get $\lambda_0(a)$, define the rate function $I \geq 0$ by

$$I(a) = a\lambda_0(a) - \psi(\lambda_0(a)) = \sup_\lambda a\lambda - \psi(\lambda) \tag{27}$$

It leads to the final result

$$\frac{1}{n}\log \mathbb{P}(S_n > an) \leq -I(a) \tag{28}$$

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n > an) = -I(a) \tag{29}$$

5 Central Limit Theorem

Theorem 19 (CLT for IID)**.** *$\{X_n\}$ IID with $\mathbb{E}X_1 = \mu$, $Var(X_1) = \sigma^2 \in (0, \infty)$. Then*

$$\frac{S_n - n\mu}{\sqrt{\sigma^2 n}} \implies \chi \tag{30}$$

Theorem 20 (Lindeberg-Feller CLT)**.** *$\{X_{n,k} : 1 \leq k \leq n < \infty\}$ triangular array, with indenepdence between row. Assume mean zero and*

(i) $\sum_{k=1}^{n} \mathbb{E}X_{n,k}^2 \to \sigma^2 > 0$

(ii) $\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}X_{n,k}^2 \mathbf{1}_{|X_{n,k}|>\epsilon} = 0$

Then $S_n = \sum_{k=1}^{n} X_{n,k} \to \sigma_\chi$ in distribution.

6 Total Variation distance

Definition 4 (TVD)**.** *Let μ, ν be measure on $(\Omega, \mathcal{F})$, then*

$$\|\mu - \nu\|_{TV} = \sup_{A\in\mathcal{F}} \mu(A) - \nu(A) \tag{31}$$

Remark 1. *If Ω is countable, say $\mathbb{Z}^+ \cup \{0\}$, then*

$$\|\mu - \nu\|_{TV} = \sum_{x:\mu(x)>\nu(v)} \mu(x) - \nu(x) \tag{32}$$

$$= \frac{1}{2}\sum_{x} |\mu(x) - \nu(x)| \tag{33}$$

*and note that if p is the **mean (not probability)** for each variable, then*

$$\|Ber(p) - Pois(p)\|_{TV} = p(1 - e^{-p}) \leq p^2 \tag{34}$$

Lemma 2. *μ, ν are measures, then push forward measures $\mu_f = \mu \circ f^{-1}$ and ν_f satisfy $\|\mu_f - \nu_f\|_{TV} \leq \|\mu - \nu\|_{TV}$. If μ_i, ν_i are measures on $(\Omega_i, \mathcal{F}_i)$, then*

$$\|\mu_1 * \mu_2 - \nu_1 * \nu_2\|_{TV} \leq \|\mu_1 - \nu_1\| + \|\mu_2 - \nu_2\| \tag{35}$$

7 Characteristic functions

Definition 5. *We say that a family $\{\mu_\alpha : \alpha \in A\}$ of probability measures on a space Ω is tight if for every $\epsilon > 0$ there is a compact set K such that $\mu_\alpha(K^c) < \epsilon$ simultaneously for every $\alpha \in A$.*

Theorem 21 (Equicontinuity iff tightness)**.** *$\{\mu_\alpha\}$ with corresponding $\{\phi_\alpha\}$. Then $\{\mu_\alpha\}$ is tight iff $\{\phi_\alpha\}$ is equicontinuous at zero, i.e. for all $\epsilon > 0, \exists\delta > 0$ s.t. simultaneously for all α, we have $|\phi_\alpha(t) - 1| < \epsilon$ if $|t - 0| < \delta$.*

Theorem 22. *A family of measures is tight if and only if every sequence of measures has a sub-sequential limit in distribution.*

Definition 6. *The characteristic function ϕ of a random variable X whose law μ has cdf F is the function $t \mapsto \mathbb{E}e^{itX}$. It has the following properties:*

(i) $\phi(t) = 0$, *but* $t \neq 0$ *implies random variable is discrete*

(ii) $\phi(0) = 1, |\phi(t)| \leq 1$.

(iii) $\phi_{F*G} = \phi_F \cdot \phi_G$

Theorem 23 (Inversion formula)**.** μ, ϕ_μ, *then*

$$\mu(a,b) + \frac{1}{2}\mu\{a,b\} = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\phi_\mu(t)dt \tag{36}$$

Remark 2. *In discrete case, we have*

$$\mathbb{P}(X=n) = \langle \phi, \psi_n = e^{inx}\rangle = \frac{1}{2\pi}\int_0^{2\pi}\phi(x)\overline{e^{inx}}dx \tag{37}$$

i.e., suppose we want to find $\mathbb{P}(X=n)$ *for some discrete random variable, then we would compute* $\langle \phi, e^{inx}\rangle$.

If ϕ *is integrable, i.e.,* $\int |\phi(t)|dt < \infty$, *then* μ *has continuous density*

$$f(y) = \frac{1}{2\pi}\int_{\mathbb{R}}\phi(t)e^{-ity}dt \tag{38}$$

Theorem 24 (Continuity theorem)**.** $\{\mu_n\}$ *with c.f* ϕ_n, *then*

(i) $\mu_n \to \mu$ *in distribution for some* μ, *then* $\phi_n(t) \to \phi_\infty(t)$ *pointwise, where* ϕ_∞ *is the characteristic function of* μ.

(ii) If $\phi_n \to \phi$ *pointwise for some* ϕ *that is continuous at zero, then* $\mu_n \to \mu$ *in distribution where* $\mu \sim \phi$.

8 Poisson process

Theorem 25 (Law of rare events)**.** *For each* n, *let* $X_{n,m}, 1 \leq m \leq n$ *be independent random variables with*

$$\mathbb{P}(X_{n,m}=1) = p_{n,m}, p(X_{n,m}=0) = 1 - p_{n,m} \tag{39}$$

Suppose

(i) $\sum_{m=1}^n p_{n,m} \to \lambda \in (0,\infty)$

(ii) $\max_{1\leq m\leq n} p_{n,m} \to 0$

then $S_n \Rightarrow Z$, *where* $Z \sim$ *Poisson*(λ).

Definition 7 (Poisson on $\mathbb{R}^+$)**.** $N(s,t) = N(t) - N(s)$ *is a Poisson rv with mean* $(t-s)\lambda$. *For disjoint intervals, say* $\{I_n\}$, $N(I_j), N(I_k)$ *are independent for all* j, k.

9 Simple Random Walk

Definition 8 (Stopping time)**.** τ *taking values in* $\mathbb{Z}^+ \cup \{+\infty\}$ *such that for all* $n, \{\tau \leq n\} \in \mathcal{F}_n$.

Proposition 1. τ *is a stopping time iff for all* $n, \{\tau = n\} \in \mathcal{F}_n$

Theorem 26 (Wald's first equation)**.** $\{X_n\}$ *IID with* $\mathbb{E}|X_1| < \infty$. τ *a stopping time with* $\mathbb{E}\tau < \infty$, *then*

$$\mathbb{E}S_\tau = \mathbb{E}\tau\mathbb{E}X_1 \tag{40}$$

Theorem 27 (Wald's second equation). *$\{X_n\}$ IID with $\mathbb{E}X_n = 0$ and $Var(X_1) < \infty$. If $\mathbb{E}\tau < \infty$, then*

$$\mathbb{E}\left(S_\tau\right)^2 = Var(X_1)\mathbb{E}\tau \tag{41}$$

Theorem 28 (Wald's third equation). *$\{X_n\}$ IID with $\mathbb{E}e^{\theta X_1} = \phi(\theta) < \infty$. If τ is a.s. bounded by L, that is*

$$\phi(\theta)^{-n} e^{\theta S_n} \mathbf{1}_{\tau \geq n} \leq L \tag{42}$$

then

$$\mathbb{E}\left[\phi(\theta)^{-\tau} e^{\theta S_\tau}\right] = 1 \tag{43}$$

10 Common Distribution

(i) Bernoulli: $\mathbb{P}(X=1) = p, \mathbb{E}X = \mathbb{E}X^k = p, \mathrm{Var}X = p(1-p), \phi(t) = 1 - p + pe^{it}$

(ii) Binomial: $(n,p), \mathbb{P}(X=i) = \binom{n}{i} p^i (1-p)^{n-i}, \mathbb{E}X = np, \mathrm{Var}X = np(1-p), \phi(t) = \left(1 - p + pe^{it}\right)^n$

(iii) Geometric: $(p), \mathbb{P}(X=i) = p(1-p)^{i-1}, \mathbb{E}X = \frac{1}{p}, \mathrm{Var}X = \frac{1-p}{p^2}, \phi(t) = \frac{pe^{it}}{1-(1-p)e^{it}}$

(iv) Poisson: $(\lambda), \mathbb{P}(X=i) = e^{-\lambda}\lambda^i/i!, \phi(t) = \exp\left[\lambda(e^{it}-1)\right]$

(v) Uniform: $x \in (a,b), f(x) = \frac{1}{b-a}, \mathbb{E}X = \frac{a+b}{2}, \mathrm{Var}X = \frac{(b-a)^2}{12}, \phi(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$

(vi) Normal: $(\mu, \sigma^2), x \in \mathbb{R}, f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \mathbb{E}X = \mu, \mathrm{Var}(X) = \sigma^2, \phi(t) = \exp\left(i\mu t - \sigma^2 t^2/2\right)$

(vii) Exponential: $\lambda, x > 0, f(x) = \lambda e^{-\lambda x}, \mathbb{E}X = \frac{1}{\lambda}, \mathrm{Var}X = \frac{1}{\lambda^2}, P(X > x) = e^{-\lambda x}, \phi(t) = \frac{1}{1 - it\lambda^{-1}}$

(viii) Cauchy: $x \in \mathbb{R}, f(x) = \frac{1}{\pi(1+x^2)}$, moment DNE, $\phi(t) = e^{-|t|}$

(ix) Compound Poisson c.f.: $S = \sum_{i=1}^{N} X_i, N \sim \mathrm{Poisson}(\lambda)$ and $X_1 \sim \phi(t)$, then $S \sim \exp\left(\lambda(\phi(t) - 1)\right)$

11 Trickery

(i) X is a continuous random variable with density f, then $\mathbb{E}\left(X - \mu\right)^k = \int \left(x - \mu\right)^k f(x)dx$

(ii) $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

(iii) If X, Y are independent, X, Y has pdf f, g respectively, then $X + Y$ has pdf h with $h(z) = \int f(x)g(z - x)dx$

(iv) Suppose X has pdf f or PMF P_n, then $\phi(t) = \mathbb{E}e^{itX} = \int e^{itx} f dx$ or $\sum e^{itn} P_n$