Stat 310A/Math 230A Theory of Probability

Homework 5 Solutions

Andrea Montanari *Due on November 6, 2019*

Exercises on the law of large numbers and Borel-Cantelli

Exercise [2.1.5]

Let $\epsilon > 0$ and pick $K = K(\epsilon)$ finite such that if $k \geq K$ then $r(k) \leq \epsilon$. Applying the Cauchy-Schwarz inequality for $X_i - \mathbf{E}X_i$ and $X_j - \mathbf{E}X_j$ we have that

$$\mathrm{Cov}(X_i, X_j) \leq [\mathrm{Var}(X_i)\mathrm{Var}(X_j)]^{1/2} \leq r(0) < \infty$$

for all i, j. Thus, breaking the double sum in $\mathrm{Var}(S_n) = \sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j)$ into $\{(i,j) : |i-j| < K\}$ and $\{(i,j) : |i-j| \geq K\}$ gives the bound

$$\mathrm{Var}(S_n) \leq 2Knr(0) + n^2\epsilon\,.$$

Dividing by n^2 we see that $\limsup_n \mathrm{Var}(n^{-1}S_n) \leq \epsilon$. Since $\epsilon > 0$ is arbitrary and $\mathbf{E}S_n = n\overline{x}$, we have that $n^{-1}S_n \overset{L^2}{\to} \overline{x}$ (with convergence in probability as well).

Exercise [2.1.13]

We have $\mathbb{E}|X_1| = \sum_{k=2}^{\infty} 1/(ck \log k) = \infty$. On the other hand, for $n \in \mathbb{N}$

$$\begin{aligned}
n\,\mathbb{P}(|X_1| \geq n) &= \frac{n}{c}\sum_{k=n}^{\infty}\frac{1}{k^2\log k} \\
&\leq \frac{n}{c}\int_{n-1}^{\infty}\frac{1}{x^2\log x}\,\mathrm{d}x \\
&= \frac{n}{c}\int_{\log(n-1)}^{\infty}\frac{e^{-z}}{z}\,\mathrm{d}z \\
&\leq \frac{n}{c\log(n-1)}\int_{\log(n-1)}^{\infty}e^{-z}\,\mathrm{d}z = \frac{n}{c(n-1)\log(n-1)}\,.
\end{aligned}$$

In particular $n\mathbb{P}(|X_1| \geq n) \to 0$ as $n \to \infty$, which implies $\lim_{x\to\infty} x\mathbb{P}(|X_1| \geq x) = 0$. We can therefore apply Proposition 2.1.12, which yields $(S_n/n - \mu_n) \overset{p}{\to} 0$.

It is therefore sufficient to show that μ_n has a finite limit. We have, for n even

$$\begin{aligned}
\mu_n = \mathbb{E}\{X_1 I_{|X_1|\leq n} &= \frac{1}{c}\sum_{k=2}^{n}(-1)^k\frac{1}{k\log k} \\
&= \frac{1}{c}\sum_{i=1}^{n/2}\left\{\frac{1}{2i\log(2i)} - \frac{1}{(2i+1)\log(2i+1)}\right\},
\end{aligned}$$

and this series is convergent. Further, for n odd, $|\mu_n - \mu_{n-1}| = 1/(cn\log n) \to 0$. Therefore μ_n has a limit.

Exercise [2.2.9]

Fixing $1 > \lambda > 0$, define $Y_n := \sum_{k \le n} I_{A_k}$ and set $a_n = \lambda \mathbf{E} Y_n$. Since $a_n \to \infty$, we have that,

$$\mathbf{P}(A_n \text{ i.o. }) \ge \mathbf{P}(Y_n > a_n \text{ i.o. }) \ge \limsup_{n\to\infty} \mathbf{P}(Y_n > a_n)$$

where the last inequality is due to Fatou's lemma (c.f. (1.3.10), or Exercise 2.2.2). Applying Exercise 1.3.20, we have that $\mathbf{P}(Y_n > a_n) \ge (1-\lambda)^2 c_n$ for $c_n := (\mathbf{E}Y_n)^2/\mathbf{E}(Y_n^2)$. By the definition of Y_n, the assumption of the exercise is precisely that $\alpha = \limsup_n c_n$. Thus, taking first $n \to \infty$ then $\lambda \downarrow 0$ completes the proof of the Kochen-Stone lemma.

Exercise [2.2.26]

1. First note that
$$\mathrm{Var}(S_n) = \sum_{i=1}^n \mathbf{P}(A_i)(1-\mathbf{P}(A_i)) \le \sum_{i=1}^n \mathbf{P}(A_i) = \mathbf{E}S_n \,.$$
By Markov's inequality, then,
$$\mathbf{P}\Big(|\frac{S_n - \mathbf{E}S_n}{\mathbf{E}S_n}| > \epsilon\Big) \le \frac{\mathrm{Var}(S_n)}{\epsilon^2 (\mathbf{E}S_n)^2} \le \frac{1}{\epsilon^2 \mathbf{E}S_n} \,,$$
and since we assumed that $\mathbf{E}S_n = \sum_{i \le n} \mathbf{P}(A_i) \to \infty$, we are done.

2. Since $\mathbf{E}(S_{n_k}) \ge k^2$, we have from part (a) that
$$\mathbf{P}(|S_{n_k} - \mathbf{E}S_{n_k}| > \epsilon \mathbf{E}S_{n_k}) \le 1/(\epsilon^2 k^2)\,.$$
Since the series $\sum_k k^{-2}$ is finite, the first Borel-Cantelli lemma implies that $\mathbf{P}(|S_{n_k} - \mathbf{E}S_{n_k}| > \epsilon \mathbf{E}S_{n_k} \text{ i.o }) = 0$. Since $\epsilon > 0$ is arbitrary, it follows that $S_{n_k}/\mathbf{E}S_{n_k} \overset{a.s.}{\to} 1$.

3. Since $k^2 \le \mathbf{E}S_{n_k} \le k^2 + 1$ and $(k+1)^2 \le \mathbf{E}S_{n_{k+1}} \le (k+1)^2 + 1$
$$\frac{k^2}{(k+1)^2+1} \le \frac{\mathbf{E}(S_{n_k})}{\mathbf{E}(S_{n_{k+1}})} \le \frac{k^2+1}{(k+1)^2} \,,$$
so $\mathbf{E}(S_{n_k})/\mathbf{E}(S_{n_{k+1}}) \to 1$ when $k \to \infty$. Then, for $n_k \le n \le n_{k+1}$,
$$\frac{S_{n_k}}{\mathbf{E}(S_{n_k})} \frac{\mathbf{E}(S_{n_k})}{\mathbf{E}(S_{n_{k+1}})} \le \frac{S_n}{\mathbf{E}(S_n)} \le \frac{S_{n_{k+1}}}{\mathbf{E}(S_{n_{k+1}})} \frac{\mathbf{E}(S_{n_{k+1}})}{\mathbf{E}(S_{n_k})} \,.$$
Hence, by part (b) and the fact that $\mathbf{E}(S_{n_k})/\mathbf{E}(S_{n_{k+1}}) \to 1$, we conclude that $S_n/\mathbf{E}(S_n) \overset{a.s.}{\to} 1$.

Exercise [2.3.14]

1. By induction, $\log W_n = \sum_{i=1}^n X_i$ for the i.i.d. random variables $X_i = \log(qr + (1-q)V_i)$. As $\{X_i\}$ are bounded below by $\log(qr) > -\infty$, it follows that $\mathbf{E}[(X_1)_-]$ is finite, so the strong law of large numbers implies that $n^{-1} \log W_n \overset{a.s.}{\to} w(q)$, as stated.

2. Since $q \mapsto (qr + (1-q)V_1(\omega))$ is linear and $\log x$ is concave, it follows that $q \mapsto \log(qr + (1-q)V_1)$ is concave on $(0,1]$, per $\omega \in \Omega$. The expectation preserves the concavity, hence $q \mapsto w(q)$ is concave on $(0,1]$.

3. By Jensen's inequality for the concave function $g(x) = \log x$, $x > 0$, we have that

$$w(q) = \mathbf{E}\log(qr + (1-q)V_1) \le \log(qr + (1-q)\mathbf{E}V_1)\,.$$

Hence, if $\mathbf{E}V_1 \le r$ then $w(q) \le \log(qr + (1-q)r) = \log r = w(1)$.
Recall that $(\log x)_- \le 1/(ex)$ for all $x \ge 0$. Hence, if $\mathbf{E}V_1^{-1}$ is finite, then so is $\mathbf{E}[(\log V_1)_-]$. Consequently, the strong law of large numbers of part (a) also applies for $n^{-1}\log W_n$ in case $q = 0$ (i.e., for $X_i = \log V_i$). Further, when $\mathbf{E}[(\log V_1)_-]$ is finite, $w(q) = w(0) + \mathbf{E}\log(qrV_1^{-1} + 1 - q)$ and by Jensen's inequality

$$\mathbf{E}\log(qrV_1^{-1} + 1 - q) \le \log(qr\mathbf{E}V_1^{-1} + 1 - q) \le 0$$

if $\mathbf{E}V_1^{-1} \le r^{-1}$, implying that then $w(q) \le w(0)$.

4. Our assumption that $\mathbf{E}V_1^2 < \infty$ and $\mathbf{E}V_1^{-2} < \infty$ implies that $\mathbf{E}V_1 < \infty$ and $\mathbf{E}V_1^{-1} < \infty$. Further, $w(0) = \mathbf{E}\log V_1 \le \mathbf{E}V_1$ is then also finite. We have shown in part (c) that $w(q) \le w(1) = \log r$ in case $\mathbf{E}V_1 \le r$ and that $w(q) \le w(0)$ in case $\mathbf{E}V_1^{-1} \le r^{-1}$. Consequently, if suffices to show that if $\mathbf{E}V_1 > r > 1/\mathbf{E}V_1^{-1}$, then there exists $q^\star \in (0,1)$ where $w(\cdot)$ reaches its supremum (which is hence finite). The former condition is equivalent to $\mathbf{E}Y > 0$ and $\mathbf{E}Z > 0$ for $Y = rV_1^{-1} - 1 \ge -1$ and $Z = r^{-1}V_1 - 1 \ge -1$, both of which are in L^2. Further, since $q \mapsto w(q) : [0,1] \to \mathbb{R}$ is a concave function, the existence of such $q^\star \in (0,1)$ follows as soon as we check that $w(\epsilon) - w(0) = \mathbf{E}\log(1 + \epsilon Y) > 0$ and $w(1-\epsilon) - w(1) = \mathbf{E}\log(1+\epsilon Z) > 0$ when $\epsilon > 0$ is small enough. To this end, note that $\log(1+x) \ge x - x^2$ for all $x \ge -1/2$. Hence, $\mathbf{E}\log(1 + \epsilon Y) \ge \epsilon\mathbf{E}Y - \epsilon^2\mathbf{E}Y^2 > 0$ for $\epsilon \in (0, 1/2)$ small enough. As the same applies for $\mathbf{E}\log(1 + \epsilon Z)$, we are done.
We see that one should invest only in risky assets whose expected annual growth factor $\mathbf{E}V_1$ exceeds that of the risk-less asset, and that if in addition $\mathbf{E}V_1^{-1} > r^{-1}$, then a unique optimal fraction $q^\star \in (0,1)$ should be re-invested each year in the risky asset.

Exercise [2.3.9]

1. Fix $\delta > 0$ such that $p := \mathbf{P}(\tau_1 > \delta) > \delta$. Note that $\widetilde{N}_t + 1 - r$ follows the *negative Binomial distribution* of parameters p and $r = \lfloor t/\delta \rfloor + 1$. That is, for $\ell = 0, 1, 2, \ldots$,

$$\mathbf{P}(\widetilde{N}_t + 1 - r = \ell) = \mathbf{P}(\widetilde{T}_{\ell+r-1} \le t < \widetilde{T}_{\ell+r})$$

It is easy to check that $\mathbf{E}(\widetilde{N}_t) = r/p - 1$ and $\mathrm{Var}(\widetilde{N}_t) = r(1-p)/p^2$. Consequently, $\mathbf{E}[\widetilde{N}_t^2] = (r^2 + r - 3rp + p^2)/p^2$, and with $p > 0$ fixed and $r \le t/\delta + 1$ it follows that $\sup_{t\ge 1} t^{-2}\mathbf{E}\widetilde{N}_t^2 < \infty$.

2. Since $\widetilde{\tau}_i \le \tau_i$, clearly $N_t \le \widetilde{N}_t$. Hence, by part (a), $\sup_{t\ge1} t^{-2}\mathbf{E}N_t^2 < \infty$. In view of the criterion of Exercise 1.3.54 (for $f(x) = x^2$), this implies that $\{t^{-1}N_t : t \ge 1\}$ is a uniformly integrable collection of R.V. As we have seen in Exercise 2.3.7 that $t^{-1}N_t \stackrel{a.s.}{\to} 1/\mathbf{E}\tau_1$, it thus follows that also $t^{-1}N_t \stackrel{L^1}{\to} 1/\mathbf{E}\tau_1$ (c.f. Theorem 1.3.49), and in particular, $t^{-1}\mathbf{E}N_t \to 1/\mathbf{E}\tau_1$ as stated.

Exercise [2.2.24]

1. Substituting $y = x + z$ and using the bound $\exp(-z^2/2) \le 1$ yields

$$\int_x^\infty e^{-y^2/2}dy \le e^{-x^2/2}\int_0^\infty e^{-xz}dz = x^{-1}e^{-x^2/2}\,.$$

For the other direction, observe that for $x > 0$,

$$(x^{-1} - x^{-3})e^{-x^2/2} = \int_x^\infty (1 - 3y^{-4})e^{-y^2/2}dy \ge \int_x^\infty e^{-y^2/2}dy\,.$$

2. Since the probability density function for a standard normal random variable G_n is $(2\pi)^{-1/2}e^{-x^2/2}$, we get from the bounds of part (a) that

$$c_\gamma = \lim_{n\to\infty} n^\gamma \sqrt{\log n}\,\mathbf{P}\left(G_n > \sqrt{2\gamma \log n}\right),$$

exists, is finite and positive. Consequently, fixing $\epsilon > 0$ by the first Borel-Cantelli lemma we have that $\mathbf{P}(G_n/\sqrt{2\log n} > 1+\epsilon \text{ i.o. }) = 0$. Further, since G_n are mutually independent, it follows from the second Borel-Cantelli lemma that $\mathbf{P}(G_n/\sqrt{2\log n} > 1-\epsilon \text{ i.o. }) = 1$. We see that with probability one, the sequence $n \mapsto G_n(\omega)/\sqrt{2\log n}$ is infinitely often above $1-\epsilon$ but only finitely often above $1+\epsilon$, in which case $L(\omega) = \limsup_n G_n(\omega)/\sqrt{2\log n}$ must be in the interval $(1-\epsilon, 1+\epsilon]$. Considering the intersection of the relevant events for $\epsilon_k \downarrow 0$, we conclude that $\mathbf{P}(L = 1) = 1$, as stated.

3. Since $S_n/\sqrt{n}$ has the same law as G_1, the upper bound of part (a) implies that $\mathbf{P}(|S_n| \geq 2\sqrt{n\log n}) \leq Cn^{-2}$ for some $C < \infty$ and all n large enough. Since the series $\sum_n n^{-2}$ is finite, applying the first Borel-Cantelli lemma we get that $\mathbf{P}(|S_n| \geq 2\sqrt{n\log n} \text{ i.o. }) = 0$, or equivalently, that $\mathbf{P}(|S_n| < 2\sqrt{n\log n} \text{ ev. }) = 1$.

Exercise on Markov chains

Throughout this solution we let $\mathcal{X}_n \equiv \sigma(\{X_i\}_{i\leq n})$, $\mathcal{T}_n \equiv \sigma(\{X_i\}_{i\geq n})$, and $a_1^n = (a_1,\dots,a_n)$ for any sequence a. Futher we let $B(x_1^n) = \{\omega : \omega_1^n = x_1^n\}$. We will prove that $\mathcal{T}$ is independent of $\mathcal{X}^n$ for any n which implies the thesis by Lemma 1.4.9.

We start by noticing that, for any $m \leq n$, any $A \in \mathcal{T}_n$, and any $x_1^m \in \mathcal{X}^m$ we have

$$\frac{\mathbf{P}\big(B(x_1^m)\cap A\big)}{\mathbf{P}\big(B(x_1^m)\big)} = \frac{\mathbf{P}\big(\{\omega_m = x_m\}\cap A\big)}{\mathbf{P}(\omega_m = x_m)}\,. \tag{1}$$

Indeed the set functions $A \mapsto \mu_1(A)$, and $A \mapsto \mu_2(A)$ defined by the two sides of the above identity are probability measures over $\mathcal{T}_n$ with $\mu_1(\Omega) = \mu_2(\Omega) = 1$ and $\mu_1(A) = \mu_2(A)$ for any event of the form $A = \{\omega : \omega_n = x_n, \dots, \omega_{n+k} = x_{n+k}\}$ (this is an elementary calculation). Since these events form a π-system, the claim follows from the uniqueness in Carathéodory extension theorem.

Next let $m < n$, and for any $B \in \mathcal{X}_m$, we let $B_n = \{x_1^n \in \mathcal{X}^n : \omega_1^n = x_1^n \Rightarrow \omega \in B\}$. For any $A \in \mathcal{T}_n$, we have

$$\mathbf{P}(A\cap B) = \sum_{x_1^n\in B_n}\mathbf{P}\big(B(x_1^n)\cap A\big) = \sum_{x_1^n\in B_n}\mathbf{P}\big(B(x_1^n)\big)\frac{\mathbf{P}\big(\{\omega_n = x_n\}\cap A\big)}{\mathbf{P}(\omega_n = x_n)} = \sum_{x_1^n\in B_n}\mathbf{P}\big(B(x_1^n)\big)\,f_A(x_n)\,, \tag{2}$$

for some function $f_A : \mathcal{X} \to [0,1]$. Writing explicitely $\mathbf{P}\big(B(x_1^n)\big)$, using the fact that $B \in \mathcal{X}_m$, and letting $k = n - m$, we get

$$\mathbf{P}(A\cap B) = \sum_{x_m, x_n} g_B(x_m)p^k(x_m,x_n)\,f_A(x_n)\,, \tag{3}$$

$$g_B(x_m) = \mathbb{P}\big(B\cap\{\omega_m = x_m\}\big)\,. \tag{4}$$

Here p^k is the k-th power of the matrix p. By Perron-Frobenius theorem, this implies that

$$|\mathbf{P}(A\cap B) - \mathbf{P}(A)\mathbf{P}(B)| \leq C\,\lambda^k\,, \tag{5}$$

for some constant C independent of A, B, and some $\lambda \in [0,1)$. Since $A \in \mathcal{T}_n$ for any n, we can take k as large as we want, thus implying $\mathbf{P}(A\cap B) = \mathbf{P}(A)\mathbf{P}(B)$.