Measure Theoretic Probability Exam

Shakeel Gavioli-Akilagun

April 2019

Question 1

To show that with pairwise independence varainces add over independent summands, let $\mathbf{x} = (x_1, ..., x_n)$ be a collection of random variables with $\mathbb{E}(x_i) = \mu_i < \infty$ and $\mathbb{V}(x_i) = \sigma_i^2 < \infty$ for each x_i, and let each x_i have marginal distribution function $F_i(x_i)$. Pairwise independence means that the joint distribution of any (x_j, x_k) with $j \neq k$ satisfies $F_{j,k}(x_j, x_k) = F_j(x_j) \cdot F_k(x_k)$. I first show that this implies zero covariance for all pairs (x_j, x_k) with $j \neq k$:

$$
\begin{aligned}
\text{cov}(x_j, x_k) &= \mathbb{E}\left((x_j - \mu_j)(x_k - \mu_k)\right) \\
&= \int \int (x_j - \mu_j)(x_k - \mu_k) dF_{j,k}(x_j, x_k) \\
&= \int (x_j - \mu_j) dF_j(x_j) \cdot \int (x_k - \mu_k) dF_k(x_k) \\
&= 0
\end{aligned}
$$

Letting $S_n = \sum_{i=1}^n x_i$ be the sum of all xs, using the above I now go on to show the main result:

1

$$\mathbb{V}(S_n) = \mathbb{V}(\sum_{i=1}^{n} x_i)$$

$$= \sum_{i=1}^{n} \mathbb{V}(x_i) + \sum_{j=1}^{n}\sum_{k=1}^{n} \text{cov}(x_j, x_k) \cdot \mathcal{I}(j \neq k)$$

$$= \sum_{i=1}^{n} \sigma_i^2$$

Question 2

The direct half of the SLLN states that for a collection of i.i.d. random variables $(x_1, ..., x_n)$ with $\mathbb{E}|x_i| < \infty$, $\mathbb{E}(x_i) = \mu < \infty$, and $\mathbb{V}(x_i) = \sigma^2 < \infty$ for each i, defining S_n as above it holds that $\frac{1}{n}S_n \xrightarrow{a.s.} \mu$. I first define two types of convergence:

Convergence completely: $X_n \xrightarrow{c} X \Leftrightarrow \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \delta) < \infty$ for some $\delta > 0$.

Convergence almost surely: $X_n \xrightarrow{a.s.} X \Leftrightarrow \lim_{n\to\infty} \mathbb{P}(|X_m - X| > \delta) = 0$ for some $\delta > 0$ and some $m \geq n$.

To show that the SLLN holds under pairwise independence only, I present an adapted proof of Theorem 1 in Etemadi (1981). The approach will be to show that $\frac{1}{n}S_n \xrightarrow{c} \mu$, and then to show that convergence completely implies convergence almost surely. Letting (x_i, x_j) be pairwise independent as defined in question 1, since $x_i = x_i^+ - x_i^-$ where both parts satisfy the assumptions above without loss of generality let $x_i \geq 0$ for each i. Notice that we can split:

$$S_n - \mathbb{E}(S_n) = \left(\sum_{i=1}^{n} x_i \cdot \mathcal{I}(x_i \leq i) - \mathbb{E}(\sum_{i=1}^{n} x_i \cdot \mathcal{I}(x_i \leq i)) \right) + \left(\sum_{i=1}^{n} x_i \cdot \mathcal{I}(x_i > i) - \mathbb{E}(\sum_{i=1}^{n} x_i \cdot \mathcal{I}(x_i > i)) \right)$$

Let A_i be the event $\{x_i > i\}$. By the first Borel-Cantelli lemma if we can show that A_i occurs only finitely many times in an infinite sequence of $\{x_i\}$s then we can conclude that

$\mathbb{P}\left(\lim_{i\to\infty}\sup A_i\right) = 0$. The proof is as follows:

$$\sum_{n=1}^{\infty}\mathbb{P}(A_n) = \sum_{n=1}^{\infty}\int_{n}^{\infty} dF(x)$$

$$= \sum_{n=1}^{\infty}\sum_{i=1}^{\infty}\int_{i}^{i+1} dF(x) = \sum_{i=1}^{\infty} i \int_{i}^{i+1} dF(x)$$

$$\leq \sum_{i=1}^{\infty}\int_{i}^{i+1} x dF(x)$$

$$\leq \mathbb{E}(x_1) = \mu < \infty$$

The event $A := \lim_{i\to\infty}\sup A_i$ therefore has measure zero, and it is sufficient to consider the behaviour of $\sum_{i=1}^{n} x_i \cdot \mathcal{I}(x_i \leq i) - \mathbb{E}(\sum_{i=1}^{n} x_i \cdot \mathcal{I}(x_i \leq i))$. Let $y_i = x_i \cdot \mathcal{I}(x_i \leq i)$ and define $S_n^* = \sum_{i=1}^{n} y_i$, the proof that $\frac{1}{n}S_n^* \xrightarrow{c} \mathbb{E}(S_n^*)$ is as follows:

$$\sum_{n=1}^{\infty}\mathbb{P}\left\{\left|\frac{S_n^* - \mathbb{E}S_n^*}{n}\right| > \varepsilon\right\} \leq c\sum_{n=1}^{\infty}\frac{\mathbb{V}S_n^*}{n^2} = c\sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(y_i) \tag{1}$$

$$\leq c\sum_{i=1}^{\infty}\frac{\mathbb{E}(y_i^2)}{i^2} = c\sum_{i=1}^{\infty}\frac{1}{i^2}\int_{0}^{i} x^2 dF(x) \tag{2}$$

$$= c\sum_{i=1}^{\infty}\frac{1}{i^2}\left(\sum_{k=0}^{i-1}\int_{k}^{k+1} x^2 dF(x)\right) \tag{3}$$

$$\leq c\sum_{k=0}^{\infty}\frac{1}{k+1}\int_{k}^{k+1} x^2 dF(x) \tag{4}$$

$$\leq c\sum_{k=0}^{\infty}\int_{k}^{k+1} x dF(x) \tag{5}$$

$$= c \cdot \mathbb{E}(x_1) < \infty \tag{6}$$

Note that the constant c is allowed to change from line to line. In line (1) the inequality holds because of Chebyshev's inequality and the equality holds because under pairwise

independence variances add over indecent summands. In line (2) the fact that $\left\{\frac{1}{n^2}\right\}_{n=1}^{\infty}$ is a convergent geometric series, and so is at most a multiple of its first term, is used. Line (3) splits each integral into i parts, and line (4) uses the fact that since the integral is finite the total value of the integral is at most a multiple of one of the parts defined in (3).

It remains to show that convergence completely implies convergence almost surely. To show this note that if $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \delta) < \infty$ then it is always possible to find some $\varepsilon > 0$ such that $\sum_{m=n+1}^{\infty} \mathbb{P}(|X_m - X| > \delta) < \varepsilon$. We then have that:

$$
\begin{aligned}
\mathbb{P}\left(|X_m - X| > \delta \text{ for some m} \geq \text{n}\right) = \mathbb{P}\left(\bigcup_{m=n+1}^{\infty} |X_m - X| > \delta\right) \\
\leq \sum_{m=n+1}^{\infty} \mathbb{P}\left(|X_m - X| > \delta\right) \\
< \epsilon
\end{aligned}
$$

Since the condition for convergence almost surely is bounded from above by the condition for convergence completely this completes the proof.

Question 3

The converse half of the SLLN states if $(x_1, ..., x_n)$ is a collection of i.i.d. random variables and it is known that $n^{-1} \sum_{i=1}^{n} x_i \xrightarrow{a.s.} \mu$ then we can claim that $n^{-1} \sum_{i=1}^{n-1} x_i \xrightarrow{a.s} \mu$. To show why this may not hold under pairwise independence only note that:

$$\frac{1}{n}\sum_{i=1}^{n-1} x_i = \frac{x_1}{n} + \frac{x_2}{n} + ... + \frac{x_{n-1}}{n} + \frac{x_n}{n} - \frac{x_n}{n}$$

$$= \left(\frac{x_1}{n} + \frac{x_2}{n} + ... + \frac{x_{n-1}}{n} + \frac{x_n}{n}\right) - \frac{x_n}{n}$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{x_n}{n}$$

For the converse half to hold we require $\frac{x_n}{n} \xrightarrow{a.s.} 0$. With i.i.d. random variables we can show this to be true with the second Borel-Cantelli lemma, however with pairwise independent variables this may not hold. The crucial difference is that the first Borel-Cantelli lemma used to prove the direct part of the SLLN does not require independence, while the second does.

Question 4

In practice, stock prices tend to behave like random walks. We could model this as $P_t = +P_{t-1} + u_t$ with a initial price P_0 being fixed. We can let $u_t \sim$ i.i.d.$(0, \sigma^2)$, however in a sense the errors are not independent since each price depends on the previous day's price, which in turn depends on the previous day's error. By back substitution we have that $P_t = \alpha t + P_0 + \sum_{i=0}^{t-1} u_{t-i}$ so $\mathbb{E}(P_t) = P_0 + \alpha t$. To show the effect of strongly dependent errors consider $Q_t = P_t - \mathbb{E}(P_t)$, then:

$$\overline{Q} = \frac{1}{T}\sum_{t=1}^{T} Q_t = \frac{T}{T}u_1 + \frac{T-1}{T}u_2 + ... + \frac{1}{T}u_T$$

$$\mathbb{V}(\overline{Q}) = \left(\frac{T}{T}\right)^2 \sigma^2 + \left(\frac{T-1}{T}\right)^2 \sigma^2 + ... + \left(\frac{1}{T}\right)^2 \sigma^2$$

$$= \left(\frac{T(T+1)(1+2T)}{6T^2}\right) \sigma^2$$

$$> \left(\frac{T}{3}\right) \sigma^2$$

Since the variance diverges, $\overline{Q}$ does not even converge in probability to its mean. This shows that with strongly dependent errors averaging does not reveal any useful information.