Solutions to Homework 10

Exercise 1.*

a) Consider

$$\log(Y_n) = \frac{1}{n}\sum_{j=1}^{n}\log(X_j)$$

As $\log(X_j)$ are i.i.d. bounded random variables, the strong law of large numbers applies, so

$$\frac{1}{n}\sum_{j=1}^{n}\log(X_j) \underset{n\to\infty}{\to} \mathbb{E}(\log(X_1)) \quad \text{almost surely}$$

so $\mu = \exp(\mathbb{E}(\log(X_1)))$.

b) In this case $\mathbb{E}(\log(X_1)) = \frac{\log(a)+\log(b)}{2}$, so $\mu = \sqrt{ab}$.

c) Observing that $\log(X_j) \in [\log(a), \log(b)]$ and using (the generalized version of) Hoeffding's inequality, we obtain

$$\mathbb{P}(\{Y_n \geq t\}) = \mathbb{P}(\{\log(Y_n) - \log(\mu) \geq \log(t) - \log(\mu)\}) \leq \exp\left(-\frac{2n\,(\log(t)-\log(\mu))^2}{(\log(b)-\log(a))^2}\right)$$

so $\mathbb{P}(\{Y_n \geq t\}) \leq C^n$ for every $t > \mu = \sqrt{ab}$, and a possible value for C is $\exp(-\frac{2\,(\log(t)-\log(\mu))^2}{(\log(b)-\log(a))^2})$ (note that the same result may be obtained by a direct computation and the use of the inequality $\cosh(x) \leq \exp(x^2/2)$).

And a weaker result can be obtained also via the inequality

$$\mathbb{P}(\{Y_n \geq t\}) \leq \frac{\mathbb{E}(Y_n^n)}{t^n} = \left(\frac{\mathbb{E}(X_1)}{t}\right)^n = \left(\frac{a+b}{2t}\right)^n$$

which shows only that concentration holds for every $t > \frac{a+b}{2}$ (and not $t > \mu$).

(For homework grading, any non-trivial bound with valid justification is accepted.)

Exercise 2.

a)

$$\begin{aligned}
\mathbb{P}\left(\left\{\frac{T_N}{\sqrt{N}} \leq t\right\}\right) &= 1 - \mathbb{P}\left(\left\{\frac{T_N}{\sqrt{N}} > t\right\}\right) = 1 - \mathbb{P}\left(\left\{T_N > \sqrt{N}t\right\}\right) \\
&= 1 - \prod_{k=1}^{\lfloor\sqrt{N}t\rfloor-1}\left(1-\frac{k}{N}\right) \\
&\approx 1 - \prod_{k=1}^{\lfloor\sqrt{N}t\rfloor-1}\left(e^{-k/N}\right) \\
&= 1 - e^{-S_{\lfloor\sqrt{N}t\rfloor}} \text{ where } S_n = \frac{1}{N}\sum_{k=1}^{n-1}k = \frac{n(n-1)}{2N}
\end{aligned}$$

where the approximation holds for small k/N. As $N \to \infty$, $S_{\lfloor \sqrt{N} t \rfloor} \to t^2/2$. Thus,

$$\mathbb{P}\left(\left\{\frac{T_N}{\sqrt{N}} \leq t\right\}\right) \underset{N\to\infty}{\to} 1 - e^{-t^2/2}.$$

b)

$$\mathbb{P}\left(\{T_{365} \leq 22\}\right) \approx \mathbb{P}\left(\left\{T_{365} \leq 1.15\sqrt{365}\right\}\right) \approx 1 - e^{-1.15^2/2} \approx 0.484$$

and

$$\mathbb{P}\left(\{T_{365} \leq 50\}\right) \approx \mathbb{P}\left(\left\{T_{365} \leq 3.07\sqrt{365}\right\}\right) \approx 1 - e^{-3.07^2/2} \approx 0.991.$$

Exercise 3.

a) Let $(Y_n, n \geq 1)$ be random variables such that $Y_i = X_i/\sqrt{V_N}$ for some fixed N. Let $(Z_n, n \geq 1)$ be independent normal random variables such that $\mathbb{E}(Z_i) = 0, \mathbb{E}(Z_i^2) = \sigma_i^2/V_N$. We know the following:

$$\begin{aligned}
&\mathbb{E}(Y_i) = \mathbb{E}(Z_i) = 0 \\
&\mathbb{E}(Y_i^2) = \mathbb{E}(Z_i^2) = \sigma_i^2/V_N \\
&\mathbb{E}(|Y_i|^3) \leq K((\sigma_i/\sqrt{V_N})^3) \\
&\mathbb{E}(|Z_i|^3) = O((\sigma_i/\sqrt{V_N})^3).
\end{aligned}$$

From Lemma 9.12, we know

$$\begin{aligned}
|\mathbb{E}(g(Y_1 + \cdots + Y_N)) - \mathbb{E}(g(Z_1 + \cdots + Z_N))| &\leq \frac{C}{6}\sum_{i=1}^{N}(\mathbb{E}(|Y_i|^3) + \mathbb{E}(|Z_i|^3)) \\
&\leq \frac{C}{6}\sum_{i=1}^{N}(K((\sigma_i/\sqrt{V_N})^3) + O((\sigma_i/\sqrt{V_N})^3)) \\
&\leq O\left(\frac{\sum_{i=1}^{N}(\sigma_i)^3}{\sqrt{V_N}^3}\right) \\
&\leq O\left(\frac{\sum_{i=1}^{N}(\sigma_i)^3}{\left(\sum_{i=1}^{N}(\sigma_i)^2\right)^{3/2}}\right)
\end{aligned}$$

Following the ideas in the lecture notes, we know that $\frac{1}{\sqrt{V_n}}(X_1 + \ldots + X_n) \underset{n\to\infty}{\overset{d}{\to}} Z \sim \mathcal{N}(0,1)$ if $|\mathbb{E}(g(Y_1 + \cdots + Y_N)) - \mathbb{E}(g(Z_1 + \cdots + Z_N))| \underset{n\to\infty}{\to} 0$. This holds when the sequence $(\sigma_n, n \geq 1)$ satisfies the condition

$$\frac{\sum_{i=1}^{N}(\sigma_i)^3}{\left(\sum_{i=1}^{N}(\sigma_i)^2\right)^{3/2}} \underset{N\to\infty}{\to} 0.$$

b) Only the first one satisfies the condition. For $\sigma_n = n$,

$$\frac{\sum_{i=1}^{n} i^3}{(\sum_{i=1}^{n} i^2)^{3/2}} = \frac{O(n^4)}{O(n^3)^{3/2}} = O(n^{-1/2}) \underset{n\to\infty}{\to} 0.$$

For $\sigma_n = 1/n$,

$$\frac{\sum_{i=1}^{n} i^{-3}}{(\sum_{i=1}^{n} i^{-2})^{3/2}} = \frac{O(1)}{O(1)^{3/2}} = O(1) \underset{n\to\infty}{\to} C.$$

For $\sigma_n = 2^n$,

$$\frac{\sum_{i=1}^{n} 8^{n}}{(\sum_{i=1}^{n} 4^{n})^{3/2}} = \frac{O(2^{3n})}{O(2^{2n})^{3/2}} = O(1) \underset{n\to\infty}{\to} C.$$

Exercise 4.

We are given i.i.d random variables $(X_n, n \geq 1)$, with $\mathbb{E}(X_1) = 1$ and $\mathrm{Var}(X_1) = \sigma^2$. Thus,

$$\mathbb{E}(S_n) = n$$
$$\mathrm{Var}(S_n) = n\sigma^2$$

Therefore, from the central limit theorem (CLT), we have

$$\frac{S_n - n}{\sigma\sqrt{n}} \xrightarrow{d} X \tag{1}$$

where X is a standard Gaussian random variable i.e., $\mathcal{N}(0, 1)$.

Now, note that

$$\frac{2}{\sigma}(\sqrt{S_n} - \sqrt{n}) = \Big(\frac{S_n - n}{\sigma\sqrt{n}}\Big)\Big(\frac{2\sqrt{n}}{\sqrt{S_n} + \sqrt{n}}\Big) \tag{2}$$

Now, to show that the L.H.S converges to the random variable X, we need to show that the product on the R.H.S converges to X.

Let's denote $\left(\frac{S_n - n}{\sigma\sqrt{n}}\right)$ and $\left(\frac{2\sqrt{n}}{\sqrt{S_n}+\sqrt{n}}\right)$ by U_n and V_n, respectively. From the Equation 1, we have that $U_n \xrightarrow{d} X$. Independently, from the strong law of large numbers (SLLN), we know that $S_n \xrightarrow{a.s} n$ (implying also $S_n \xrightarrow{d} n$). From the continuous mapping theorem (see below), we therefore have that $V_n \xrightarrow{d} 1$.

So, we have $U_n \xrightarrow{d} X$ and $V_n \xrightarrow{d} 1$, the question we seek to answer next is whether the product sequence i.e., $U_n.V_n \xrightarrow{d} X.1$?

The answer is **YES!**, due to a result from Slutsky, which is stated below (and can be proven with minimal effort):

Slutsky's Theorem: Let $(X_n, n \geq 1)$ and $(Y_n, n \geq 1)$ be a sequence of i.i.d random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $X_n \xrightarrow{d} X$ (where X is a random variable on the same probability space) and $Y_n \xrightarrow{d} c$ (where c is a real and invertible constant). Then, the following is always true:

1. $X_n + Y_n \xrightarrow{d} X + c$
2. $X_n - Y_n \xrightarrow{d} X - c$
3. $X_n.Y_n \xrightarrow{d} X.c$

4. $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$

Hint for the proof (Only if you'd like to try!): The main idea behind proving the above result is showing that the joint vector (X_n, Y_n) converges to (X, c) in distribution. It could be done in two steps. First, try showing that the sequence $((X_n, Y_n) - (X_n, c)), n \geq 1)$ converges to 0 in probability (which implies convergence in distribution). Then, show that $((X_n, c), n \geq 1)$ converges to (X, c) in distribution.

Thus, on using the Slutsky's theorem for the Equation 2, we have the desired result i.e.,

$$\frac{2}{\sigma}(\sqrt{S_n} - \sqrt{n}) \xrightarrow{d} X$$

Note: The **continuous mapping theorem** extends the Heine's definition of continuous functions on deterministic sequences to sequence of random variables. More precisely, it states that for any sequence of random variables $(X_n, n \geq 1)$ and random variable X all defined on the same probability space $(\omega, \mathcal{F}, \mathbb{P})$ such that $X_n \xrightarrow{d} X$. Then, for any continuous function on X_n, the convergence $g(X_n) \xrightarrow{d} g(X)$ is true as well. It also holds for convergence in distribution and almost sure convergence.