Exercise 10.11

a) For \( y \in [-c, c] \) we have that \( \frac{1}{2} \left( \frac{y}{c} + 1 \right) \in [0, 1] \), so from the convexity of \( f \) follows

\[
\begin{align*}
  f(y) &= f \left( \frac{y}{2} + c + \left( -c + \frac{y}{2} + c \right) \right) \\
  &= f \left( \frac{1}{2} \left( \frac{y}{c} + 1 \right) c + (1 - \frac{1}{2} \left( \frac{y}{c} + 1 \right)) (-c) \right) \\
  &\leq \frac{1}{2} \left( \frac{y}{c} + 1 \right) f(c) + (1 - \frac{1}{2} \left( \frac{y}{c} + 1 \right)) f(-c) \\
  &= \frac{f(c) + f(-c)}{2} + \frac{f(c) - f(-c) y}{2 c} \\
  &= \cosh(\theta c) + \sinh(\theta c) \frac{y}{c}
\end{align*}
\]

Combining this with \( P(|Y| \leq c) = 1 \) and \( EY = 0 \) gives

\[
E f(Y) \leq \cosh(\theta c) + \sinh(\theta c) \frac{EY}{c} = \cosh(\theta c).
\]

Moreover, using \( 2^n n! \leq (2n)! \) for each \( n \in \mathbb{N}_0 \) gives

\[
\cosh(\theta c) = \sum_{n=0}^{\infty} \frac{(\theta c)^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty} \frac{(\theta c)^{2n}}{2^n n!} = \sum_{n=0}^{\infty} \frac{\theta^2 c^2}{n!} = \exp \left( \frac{1}{2} \theta^2 c^2 \right).
\]

We conclude that \( E f(Y) \leq \cosh(\theta c) \leq \exp \left( \frac{1}{2} \theta^2 c^2 \right) \).

b) We write

\[
Z_n = \exp(\theta M_n) = \exp(\theta \sum_{k=0}^{n} \Delta M_k) = \prod_{k=1}^{n} \exp(\theta \Delta M_k) = \prod_{k=1}^{n} f(\Delta M_k),
\]

from which we see that \( Z_n = Z_{n-1} f(\Delta M_n) \) for each \( n \in \mathbb{N} \). Since each \( Z_n \) is \( \mathcal{F}_n \)-measurable, we obtain from this (using Theorem 8.10(iv) and Theorem 8.12(ii)) that \( Z_n = E[Z_n | \mathcal{F}_n] = Z_{n-1} E[f(\Delta M_n) | \mathcal{F}_n] \). For each \( n \in \mathbb{N} \) we know by assumption that \( P(|\Delta M_n| \leq c_n) = 1 \) and (since \( M \) is a martingale) \( E \Delta M_n = 0 \), so (a) gives (using \( E[\Delta M_n | \mathcal{F}_n] = 0 \), because \( M \) is a martingale)

\[
E[f(\Delta M_n) | \mathcal{F}_n] \leq E[\cosh(\theta c_n) + \sinh(\theta c_n) \frac{\Delta M_n}{c_n} | \mathcal{F}_n] = \cosh(\theta c_n) \leq \exp \left( \frac{1}{2} \theta^2 c_n^2 \right).
\]

Combining all this with \( Z_0 = \exp(\theta M_0) = 1 \) gives (recursively) for each \( n \in \mathbb{N} \) that

\[
Z_n \leq \prod_{k=1}^{n} \exp \left( \frac{1}{2} \theta^2 c_k^2 \right) = \exp \left( \frac{1}{2} \theta^2 \sum_{k=1}^{n} c_k^2 \right).
\]

We conclude for each \( n \in \mathbb{N} \) that

\[
EZ_n \leq \exp \left( \frac{1}{2} \theta^2 \sum_{k=1}^{n} c_k^2 \right).
\]
c) We obtain for each \( \theta > 0 \) that

\[
P(\sup_{k \leq n} M_k \geq x) = P(\sup_{k \leq n} Z_k \geq e^{\theta x})
\]

(exp(⋅) is increasing)

\[
\leq \exp(-\theta x) \cdot E Z_n
\]

((Proposition 10.12)

\[
\leq \exp(-\theta x) \cdot \exp\left(\frac{1}{2} \theta^2 \sum_{k=1}^{n} c_k^2\right)
\]

(see (b))

\[
\exp\left(-\theta x + \frac{1}{2} \theta^2 \sum_{k=1}^{n} c_k^2\right)
\]

(1)

Setting \( f(\theta) = -\theta x + \frac{1}{2} \theta^2 \sum_{k=1}^{n} c_k^2 \), we see that \( f \) is a parabola that opens up, so solving \( f'(\theta) = 0 \) (i.e. \( -x + \theta \sum_{k=1}^{n} c_k^2 = 0 \), we see that \( \theta = \frac{x}{\sum_{k=1}^{n} c_k^2} \) is a minimum of \( f \). Substituting \( \theta = \frac{x}{\sum_{k=1}^{n} c_k^2} \) into (1) we obtain the Azuma-Hoeffding inequality:

\[
P(\sup_{k \leq n} M_k \geq x) \leq \exp\left(-\frac{x^2}{2 \sum_{k=1}^{n} c_k^2}\right).
\]

Exercise 12.1

Suppose \( X_n \xrightarrow{p} X \). In order to show \( X_n \xrightarrow{w} X \), it follows from Definition 12.1 and Proposition 4.27 that we must show \( E(f(X_n)) \to E(f(X)) \) for all \( f \in C_b(\mathbb{R}) \).

So let \( f \in C_b(\mathbb{R}) \). For each \( \delta > 0 \), \( K > 0 \) and \( n \in \mathbb{N} \) we have

\[
|E(f(X_n)) - E(f(X))| \leq E|f(X_n) - f(X)|
\]

\[
= E\left[|f(X_n) - f(X)|P(|X_n - X| < \delta, |X| < K) + E\left[|f(X_n) - f(X)|P(|X_n - X| < \delta, |X| \geq K)\right]
\]

\[
+ E\left[|f(X_n) - f(X)|P(|X_n - X| \geq \delta, |X| < K)\right]
\]

\[
+ 2\|f\|_{\infty}E(|X| \geq K) + 2\|f\|_{\infty}E(|X_n - X| \geq \delta)
\]

Now let \( \varepsilon > 0 \). Take \( K(\varepsilon) > 0 \) such that \( P(|X| \geq K(\varepsilon)) \leq \frac{\varepsilon}{6\|f\|_{\infty}} \) and \( \delta(\varepsilon) \in (0, 1) \) such that if \( x, y \in \mathbb{R} \) are such that \( |x|, |y| \leq K(\varepsilon) + 1 \) and \( |x - y| < \delta(\varepsilon) \), then \( |f(x) - f(y)| < \frac{\varepsilon}{3} \) (we can take such a \( \delta(\varepsilon) \) because \( f \) is uniformly continuous on the compact set \( [-K(\varepsilon) - 1, K(\varepsilon) + 1] \)). Then from the previous follows \( \forall n \in \mathbb{N} \)

\[
|E(f(X_n)) - E(f(X))| \leq \frac{\varepsilon}{3} P(|X_n - X| < \delta(\varepsilon), |X| < K(\varepsilon))
\]

\[
+ 2\|f\|_{\infty} \cdot \frac{\varepsilon}{6\|f\|_{\infty}} + 2\|f\|_{\infty}P(|X_n - X| \geq \delta(\varepsilon))
\]

\[
\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + 2\|f\|_{\infty}P(|X_n - X| \geq \delta(\varepsilon))
\]

Since \( X_n \xrightarrow{p} X \), we can take \( N(\varepsilon) \in \mathbb{N} \) such that \( P(|X_n - X| \geq \delta(\varepsilon)) \leq \frac{\varepsilon}{6\|f\|_{\infty}} \) for all \( n \in \mathbb{N} \) with \( n \geq N(\varepsilon) \). We take such a \( N(\varepsilon) \in \mathbb{N} \). Then from the previous follows for all \( n \in \mathbb{N} \) with \( n \geq N(\varepsilon) \) that

\[
|E(f(X_n)) - E(f(X))| \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.
\]

Since \( \varepsilon > 0 \) was arbitrary, we obtain \( E(f(X_n)) \to E(f(X)) \). We conclude \( X_n \xrightarrow{w} X \).
Now suppose $X_n \xrightarrow{w} X$, and let $g : \mathbb{R} \to \mathbb{R}$ be continuous. Then, by Definition 12.1 (and Proposition 4.27), we have $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded continuous functions $f$ on $\mathbb{R}$. In particular, for each $f \in C_b(\mathbb{R})$ we have that $f \circ g \in C_b(\mathbb{R})$ and thus $\mathbb{E}f(g(X_n)) \to \mathbb{E}f(g(X))$. We conclude from Definition 12.1 (using Proposition 4.27) that $g(X_n) \xrightarrow{w} g(X)$.

Finally, suppose $X_n \xrightarrow{w} X$, where $X$ is a constant random variable taking the value $x \in \mathbb{R}$. Combining our assumption with Theorem 12.12(i)⇒(ii) gives for each $\varepsilon > 0$ that (with $B_{\varepsilon/2}(x)$ the open ball around $x$ with radius $\varepsilon/2$)

$$\limsup_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) \leq \limsup_{n \to \infty} \mathbb{P}(|X_n - X| \geq \varepsilon/2)$$

$$= \limsup_{n \to \infty} \mathbb{P}(X_n \in (B_{\varepsilon/2}(x))^\complement)$$

$$\leq \mathbb{P}(X \in (B_{\varepsilon/2}(x))^\complement)$$

$$= \mathbb{P}(x \in (B_{\varepsilon/2}(x))^\complement)$$

$$= 0$$

We conclude that $\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$ for each $\varepsilon > 0$, so $X_n \xrightarrow{p} X$.

**Exercise 12.4**

Suppose $X_n \xrightarrow{w} X$. Then, by Definition 12.1 (and Proposition 4.27), we have $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded continuous functions $f$ on $\mathbb{R}$. In particular, we have $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded uniformly continuous functions $f$ on $\mathbb{R}$ (since uniform continuity implies continuity).

Conversely, suppose that $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded uniformly continuous functions $f$ on $\mathbb{R}$. Denoting the distributions of $X, X_1, X_2, \ldots$ by $\mu, \mu_1, \mu_2, \ldots$, respectively, and their distributions with $F,F_1,F_2,\ldots$, respectively, then in order to show $X_n \xrightarrow{w} X$ it remains according to Corollary 12.5 to show that $F_n(x) \to F(x)$ for all $x \in C_F$.

Let $x \in \mathbb{R}, \varepsilon > 0$ and define $g$ by $g(y) = 1$ for $y \leq x$, $g(y) = 0$ for $y \geq x + \varepsilon$ and by linear interpolation on $(x, x + \varepsilon)$. Then $g$ is in fact bounded and uniformly continuous on $\mathbb{R}$ (even more, $g$ is Lipschitz continuous) and we can follow the steps of the proof of Prop 12.3. Since $1_{(-\infty, x]} \leq g \leq 1_{(-\infty, x+\varepsilon]}$, we have

$$F_n(x) \leq \mu_n(g) \leq F_n(x + \varepsilon),$$

$$F(x) \leq \mu(g) \leq F(x + \varepsilon).$$

By the assumption, this leads to

$$\limsup_{n} F_n(x) \leq \limsup_{n} \mu_n(g) = \mu(g) \leq F(x + \varepsilon).$$

Letting $\varepsilon \to 0$ we get by the right-continuity of $F$ that $\limsup_{n} F_n(x) \leq F(x)$. Similarly,

$$\liminf_{n} F_n(x + \varepsilon) \geq \liminf_{n} \mu_n(g) = \mu(g) \geq F(x).$$

Since this holds for all $x$, we rename $x + \varepsilon$ as $x$ to obtain $\liminf_{n} F_n(x) \geq F(x - \varepsilon)$. The final statement follows by letting $\varepsilon \to 0$ and observing that $x \in C_F$. 
