Solutions to Selected Exercises in Rick Durrett's
Probability: Theory and Examples, 4th edition

Emad Zadegan
emadzadegan.github.io
Last updated 10-04-2019

1.1.4

A σ-field $\mathcal{F}$ is said to be countably generated if there is a countable collection $\mathcal{C} \subset \mathcal{F}$ so that $\sigma(\mathcal{C}) = \mathcal{F}$. Show that $\mathcal{R}^d$ is countably generated.

Since the rationals, $\mathbf{Q}$ are dense in the real line, $\mathbf{R}$, it follows that any $x \in \mathbf{R}$ can be represented as the limit of a sequence $\{q_n\}_{n \in \mathbf{N}}$ where $q_n \in \mathbf{Q}\ \forall n$.

Let $\mathbf{C} = \{(a, b) : -\infty \leq a < b \leq \infty; a, b \in \mathbf{Q}\}$. Any open set in $\mathbf{R}$ may be constructed as a countable union of the elements in $\mathbf{C}$

Let $\mathcal{R}$ be the Borel sets on $\mathbf{R}$. By definition, $\mathcal{R}$ is the smallest sigma-algebra containing the open sets. Therefore, $\mathbf{C} \subset \mathcal{R}$ since every element of $\mathbf{C}$ is an open set. Then,

$\mathbf{C} \subset \mathcal{R} \implies \sigma(\mathbf{C}) \subset \sigma(\mathcal{R}) = \mathcal{R}$

By the definiton of sigma-algebras and $\mathbf{C}$, we have that every open set is contained in $\sigma(\mathbf{C})$. By the definition of Borel sets, $\mathcal{R}$ is the smallest sigma-algebra containing the open sets. Therefore $\mathcal{R} \subset \sigma(\mathbf{C})$.

Thus, $\mathcal{R} = \sigma(\mathbf{C})$.

To extend the proof to $\mathcal{R}^d$, we use $\mathbf{C}^d = \{(a_1, b_2) \times ... \times (a_d, b_d) : -\infty \leq a_i < b_i \leq \infty; a_i, b_i \in \mathbf{Q} \forall i\}$. Note that $\mathbf{C}^d$ is a countable set.

Lemma 2: For any open set $G \subset \mathcal{R}^d$, there exists a countable collection $\{G_i\}$ of open sets such that $G = \bigcup_{i=1}^{\infty} G_i$

Proof of Lemma 2: Since G is open, for every $x \in G$ there exists an open ball centred around x contained in G. Within this open ball, there exists an open rectangle with rational endpoints containing x. We will denote this box by $\mathbf{C}_x^d$ and note that $\mathbf{C}_x^d \in \mathbf{C}^d$. Furthermore, $G = \bigcup\limits_{x \in G} \mathbf{C}_x^d \subset \mathbf{C}^d$.

$\bigcup\limits_{x \in G} \mathbf{C}_x^d$ is an uncountable union of the elements of a countable set $\mathbf{C}^d$. Therefore, $G = \bigcup\limits_{x \in G} \mathbf{C}_x^d$ may be re-indexed as a countable union, for example $G = \bigcup_{i=1} G_i$ where $G_i \in \mathbf{C}^d$. $\square$

The result for $\mathcal{R}^d$ follows from the $\mathcal{R}$ case and *Lemma 2*.

$\square$

1.2.3

Show that a distribution function has at most countably many discontinuities.

This a consequence of the monotonicity of the distribution function. Let F be a distribution function. And let D be the set of points at which F is discontinuous. For each $d \in D$, $F(d-) = \underset{x \uparrow d}{lim} F(x) < \underset{x \downarrow d}{lim} F(x) = F(d+)$. Therefore for all $d \in D$ there exists a unique rational number q_d such that $F(d-) < q_d < F(d+)$. The collection $\{q_d\}_{d \in D}$ is consists of but D is uncountable. This is a contradiction and therefore a distribution function cannot have uncountably many discontinuities.

$\square$

1.2.5

Suppose X has continuous density $f, P(\alpha \leq X \leq \beta) = 1$ and g is a function that is strictly increasing and differentiable on (α, β). Then $g(X)$ has density $f(g^{-1}(y))/g'(g^{-1}(y))$ for $y \in (g(\alpha), g(\beta))$ and 0 otherwise.

This follows from the chain rule, the fact that

$$\frac{d}{dx}[g^{-1}](x) = \frac{1}{g'(g^{-1}(x))} \tag{1}$$

and the fact that g is invertible since it is strictly increasing and differentiable.

Let F_g and f_g be the distribution and densitiy functions for $g(X)$ respectively; and let F_X and f_X be the distribution and density functions for X.

$$\begin{align} F_g(y) &= P(g(X) \leq y) \tag{2} \\ &= P(g^{-1}(g(X)) \leq g^{-1}(y)) \tag{3} \\ &= F_X(g^{-1}(y)) \tag{4} \end{align}$$

Then by the definition of density,

$$\begin{align} f_g(y) &= \frac{d}{dy} F_g(y) \tag{5} \\ &= \frac{d}{dy} F_X(g^{-1}(y)) \tag{6} \\ &= f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) \tag{7} \\ &= f_X(g^{-1}(y)) \frac{1}{g'(g^{-1}(y))} \tag{8} \end{align}$$

Going from (7) to (8) uses (1). $\square$

1.3.5

Show that f is lower semicontinuous if and only if $x : f(x) \leq a$ is closed for each $a \in \mathbb{R}$ and conclude that semicontinuous functions are measurable.

f is l.s.c. $\Rightarrow \{x : f(x) \leq a\}$ is closed for all $a \in \mathbf{R}$. *Proof:*

Suppose f is l.s.c., then by definition: $\liminf_{y \to x} f(y) \geq f(x)$. For an arbitrary $a \in \mathbf{R}$ let $X_a = \{x : f(x) \leq a\}$. Let $\{x_n\}$ be a sequence such that it converges to some limit η and $x_n \in X_a \forall n$. If we show that $\eta \in X_a$ then we are done.

Suppose for a contradiction that $\eta \notin X_a$. Then $f(\eta) > a$.

$\liminf_{n \to \infty} f(x_n) \leq a$ since $x_n \in X_a \forall n$

$\liminf_{x \to \eta} f(x) \geq f(\eta)$ since by assumption f is l.s.c.

Then putting it all together, we have:

$$\liminf_{n \to \infty} f(x_n) \leq a < f(\eta) \leq \liminf_{x \to \eta} f(x)$$

This is a contradiction because $\liminf_{n \to \infty} f(x_n) \geq \liminf_{x \to \eta} f(x)$

f is l.s.c. $\Leftarrow \{x : f(x) \leq a\}$ is closed for all $a \in \mathbf{R}$. *Proof:*

Fix x and suppose $A_{x,\epsilon} = \{y : f(y) \leq f(x) - \epsilon\}$ is a closed set. Then $A^C_{x,\epsilon} = \{y : f(y) > f(x) - \epsilon\}$ is open.

By definition, $\liminf_{y \to x} f(y) = \lim_{\delta \to 0} [\inf\{f(y) : y \in B(x, \delta) \setminus x\}]$ where $B(x, \delta)$ is an open ball of radius δ centered at x.

By definition $A^C_{x,\epsilon}$ is an open set containing of x and so $\forall \epsilon > 0 \; \exists \delta > 0$ such that $B(x, \delta) \subset A^C_{x,\epsilon}$. Thus we can say

$$\liminf_{y \to x} f(y) = \lim_{\delta \to 0} [\inf\{f(y) : y \in B(x, \delta) \setminus x\}] \tag{9}$$
$$= \lim_{\epsilon \to 0} [\inf\{f(y) : y \in A^C_{x,\epsilon} \setminus x\}] \tag{10}$$

By definition, $f(y) > f(x) - \epsilon \; \forall \epsilon > 0 \; \forall y \in A^C_{x,\epsilon}$; therefore $\inf\{f(y) : y \in A^C_{x,\epsilon} \setminus x\} > f(x) - \epsilon$

It follows that

$$\liminf_{y \to x} f(y) = \lim_{\epsilon \to 0} [\inf\{f(y) : y \in A^C_{x,\epsilon} \setminus x\}] \tag{11}$$
$$\geq f(x) \tag{12}$$

To show that f is measurable, we will apply *Theorem 1.3.1.*

Suppose f is l.s.c. and let $A = (-\infty, a]$, $a \in \mathbf{R}$, where $\mathbf{R}$ is the real line.

Then $f^{-1}(A) = \{x : f(x) \in A\}$ is a closed a set and $A \in \mathcal{R}$ where $\mathcal{R}$ is the Borel sets.

$f^{-1}(A) \subset \mathcal{R}$ because $f^{-1}(A)$ is closed, $\mathcal{R}$ is the smallest sigma-algebra generated from the open sets, and $\mathcal{R}$ is closed under complements.

Next, to prove that $\sigma(\{(-\infty, a] : a \in \mathbf{R}\}) = \mathcal{R}$, it suffices to show that any open interval (p, q) may be expressed as the countable unions and intersections of the sets $\{(-\infty, a] : a \in \mathbf{R}\}$ and its complements.

$$(p, q) = (-\infty, p] \bigcap \Big(\bigcup_{i \geq 1} (-\infty, q - \frac{1}{n}] \Big)$$

Then the result follows from exercise 1.1.4. □

1.6.3

Chebyshev's inequality is and is not sharp.

(i) **Show that for fixed** $0 < b \leq a$ **there exists an** X **with** $\mathbb{E}X^2 = b^2$ **for which** $\mathbb{P}\big(|X| \geq a\big) = b^2/a^2$

$$X = \begin{cases} a, & \text{with probability } b^2/a^2 \\ 0, & \text{with probability } 1 - b^2/a^2 \end{cases}$$

Then $\mathbb{P}\big(|X| \geq a\big) = \mathbb{P}\big(|X| = a\big) = b^2/a^2$ and $\mathbb{E}X^2 = 0^2\big(1 - b^2/a^2\big) + a^2\big(b^2/a^2\big) = b^2$

□

(ii) **Show that if** $0 < \mathbb{E}X^2 < \infty$ **then** $\lim_{a \to \infty} a^2 \mathbb{P}\big(|X| \geq a\big)/\mathbb{E}X^2 = 0$

$$a^2 1_{\{|X| \geq a\}} \leq X^2 \text{ and } X^2 \text{ is integrable, and}$$

$$a^2 1_{\{|X| \geq a\}} \overset{a\to\infty}{\to} 0$$

$$\text{Therefore, } a^2 \mathbb{P}\big(|X| \geq a\big) = \int a^2 1_{\{|X| \geq a\}} dP \underset{DCT}{\overset{a\to\infty}{\to}} 0$$

$$\text{Hence, } a^2 \mathbb{P}\big(|X| \geq a\big)/\mathbb{E}X^2 \overset{a\to\infty}{\to} 0 \text{ as desired}$$

□

1.6.5

Two nonexistent lower bounds.

(i) **Show that if** $\epsilon > 0$**, then** $\inf\{\mathbb{P}(|X| > \epsilon) : \mathbb{E}X = 0, \mathbf{var}(X) = 1\} = 0$

Let

$$X_n = \begin{cases} -n, & \text{with probability } 1/(2n^2) \\ n, & \text{with probability } 1/(2n^2) \\ 0, & \text{with probability } 1 - 1/(n^2) \end{cases}$$

Then $\mathbb{E}X_n = (-n)\frac{1}{2n^2} + (n)\frac{1}{2n^2} + (0)\frac{1}{n^2} = 0$ and

$\text{var}(X_n) = \mathbb{E}[X_n - \mathbb{E}X_n]^2 = \mathbb{E}(X_n)^2 = (-n)^2\frac{1}{2n^2} + (n)^2\frac{1}{2n^2} + (0)^2(1 - \frac{1}{n^2}) = 1$

Note that $\mathbb{P}(|X_n| > \epsilon) = 1/n^2 \ \forall n > 1/\epsilon$

Since $1/n^2$ may be arbitrarily small, the result follows.

(ii) **Show that if** $y \geq 1$**,** $\sigma^2 \in (0,1)$**, then** $\inf\{\mathbb{P}(|X| > y) : \mathbb{E}X = 1, \mathbf{var}(X) = \sigma^2\} = 0$

Let $Y_n = \sigma X_n + 1$

Then $\mathbb{E}Y_n = \mathbb{E}(\sigma X_n + 1) = \sigma\mathbb{E}X_n + 1 = 1$ and

$\text{var}(Y_n) = \mathbb{E}[Y_n - \mathbb{E}Y_n]^2 = \mathbb{E}Y_n^2 - (\mathbb{E}Y_n)^2 = \big((-\sigma n + 1)^2\frac{1}{2n^2} + (\sigma n + 1)^2\frac{1}{2n^2} + (0+1)(1 - \frac{1}{n^2})\big) - 1^2 = \frac{\sigma^2 n^2 - 2\sigma n + 1}{2n^2} + \frac{\sigma^2 n^2 + 2\sigma n + 1}{2n^2} + (1 - \frac{1}{n^2}) - 1 = \sigma^2$

Similar to (i), note that $\mathbb{P}(|Y_n| > \epsilon + 1) = 1/n^2 \ \forall n > 1/\epsilon$

Since $1/n^2$ may be arbitrarily small, the result follows. $\square$

1.6.8

Suppose the probability measure μ **has** $\mu(A) = \int_A f(x)dx$ **for all** $A \in \mathcal{R}$**. Use the proof technique in Theorem 1.6.9 to show that for any** g **with** $g \geq 0$ **or** $\int |g(x)|\mu(dx) < \infty$ **we have**

$$\int g(x)\mu(dx) = \int g(x)f(x)dx$$

(i) *Indicator Functions*: if $B \in \mathcal{S}$ and $g = 1_B$ then

$$\int g(x)\mu(dx) = \int_B \mu(dx) = \mu(B) = \int_B f(x)dx = \int g(x)f(x)dx$$

(ii) *Simple Functions*: let $g = \sum_{i=1}^{n} c_i 1_{B_i}$, where $c_i \in \mathbb{R}$ and $B_i \in \mathcal{R}$. Using (i) and the linearity of integration, we have:

$$\int g(x)\mu(dx) = \int \sum_{i=1}^{n} c_i 1_{B_i} \mu(dx) \tag{13}$$
$$= \sum_{i=1}^{n} c_i \int 1_{B_i} \mu(dx) \tag{14}$$
$$= \sum_{i=1}^{n} c_i \int_{B_i} \mu(dx) \tag{15}$$
$$= \sum_{i=1}^{n} c_i \mu(B_i) \tag{16}$$
$$= \sum_{i=1}^{n} c_i \int_{B_i} f(x)dx \tag{17}$$
$$= \int_{B_i} \sum_{i=1}^{n} c_i f(x)dx \tag{18}$$
$$= \int \sum_{i=1}^{n} c_i 1_{B_i} f(x)dx \tag{19}$$
$$= \int g(x)f(x)dx \tag{20}$$

(iii) *Non-negative Functions* let $g \geq 0$, and $g_n = \min\Big(\big([2^n f(x)]/2^n\big), n\Big)$, where $[x] = \text{floor}(x)$. Then g_n is simple for all n, and $g_n \uparrow g$. By applying MCT, we have

$$\int g(x)\mu(dx) = \int \lim_{n\to\infty} g_n(x)\mu(dx) \tag{21}$$
$$= \lim_{n\to\infty} \int g_n(x)\mu(dx) \tag{22}$$
$$= \lim_{n\to\infty} \int g_n(x)f(x)dx \tag{23}$$
$$= \int g(x)f(x)dx \tag{24}$$

(iv) *Integrable Functions* for integrable g, let $g^+ = max(g, 0)$ and $g^- = max(-g, 0)$; then $g = g^+ - g^-$, and

$$\int g(x)\mu(dx) = \int g^+(x) - g^-(x)\mu(dx) = \int \big(g^+(x) - g^-(x)\big)f(x)dx = \int g(x)f(x)dx$$

$\square$

1.7.3

Let F, G be Stieltjes measure functions and let μ, ν be the corresponding measures on $(\mathbb{R}, \mathcal{R})$. Show that:

(i) $\int_{(a,b]}\{F(y)-F(a)\}dG(y) = (\mu\times\nu)\big(\{(x,y): a<x\le y\le b\}\big)$

Note that, by the definition of Stieltjes measure functions,

$$F(y)-F(a) = \mu((a,y]) = \int_{(a,y]} d\mu = \int 1_{(a,y]}(x)d\mu(x)$$

Then, using indicator functions and applying Fubini's theorem we have:

$$\begin{aligned}
\int_{(a,b]}\{F(y)-F(a)\}dG(y) &= \int_{(a,b]}\int 1_{(a,y]}d\mu(x)dG(y) && (25)\\
&= \int_{(a,b]}(y)\int 1_{(a,y]}(x)d\mu(x)d\nu(y) && (26)\\
&= \int 1_{(a,b]}(y)\int 1_{(a,y]}(x)d\mu(x)d\nu(y) && (27)\\
&= \int 1_{(a,b]}(y)\int 1_{(a,y]}(x)d\mu(x)d\nu(y) && (28)\\
&= \int\int 1_{(a,b]}(y)1_{(a,y]}(x)d\mu(x)d\nu(y) && (29)\\
&= \int\int 1_{(a<x\le y\le b]}(x,y)d\mu(x)d\nu(y) && (30)\\
&= (\mu\times\nu)\big(\{(x,y): a<x\le y\le b\}\big) && (31)\\
& && (32)
\end{aligned}$$

$\square$

(ii) $\int_{(a,b]} F(y)dG(y) + \int_{(a,b]} G(y)dF(y) = F(b)G(b) - F(a)G(a) + \sum_{x \in (a,b]} \mu(\{x\})\nu(\{x\})$

That by adding and subtracting $F(a)$, using part (i), and simplifying we get

$$
\begin{aligned}
\int_{(a,b]} F(y)dG(y) &= \int_{(a,b]} F(y) - F(a) + F(a)dG(y) && (33)\\
&= \int_{(a,b]} F(y) - F(a)dG(y) + \int_{(a,b]} F(a)dG(y) && (34)\\
&= (\mu \times \nu)\big(\{(x,y) : a < x \leq y \leq b\}\big) + \int_{(a,b]} F(a)dG(y) && (35)\\
&= (\mu \times \nu)\big(\{(x,y) : a < x \leq y \leq b\}\big) + F(a)\int_{(a,b]} dG(y) && (36)\\
&= (\mu \times \nu)\big(\{(x,y) : a < x \leq y \leq b\}\big) + F(a)\int_{(a,b]} d\nu && (37)\\
&= (\mu \times \nu)\big(\{(x,y) : a < x \leq y \leq b\}\big) + F(a)[G(b) - G(a)] && (38)\\
& && (39)
\end{aligned}
$$

And similarly, for the second term, we have

$$
\begin{aligned}
\int_{(a,b]} G(y)dF(y) &= (\mu \times \nu)\big(\{(x,y) : a < y \leq x \leq b\}\big) + G(a)[F(b) - F(a)] && (40)\\
& && (41)
\end{aligned}
$$

Note that

$$
\begin{aligned}
&(\mu \times \nu)\big(\{(x,y) : a < x \leq y \leq b\}\big) + (\mu \times \nu)\big(\{(x,y) : a < y \leq x \leq b\}\big) && (42)\\
&\quad = \int\int 1_{(a,b]}(x)1_{(a,b]}(y)dF(x)dG(y) + \int\int 1_{(a<x=y\leq b]}(x,y)dF(x)dG(y) && (43)
\end{aligned}
$$

And that

$$
\begin{aligned}
\int\int 1_{(a,b]}(x)1_{(a,b]}(y)dF(x)dG(y) &= \big(F(b) - F(a)\big)\big(G(b) - G(a)\big) && (44)\\
&= F(b)G(b) - F(b)G(a) - F(a)G(b) + F(a)G(a) && (45)
\end{aligned}
$$

Also, since F, and G are Stieltjes measure functions (i.e. non-decreasing and right-continuous), they may have at most countably many discontinuities, as proven in exercise 1.2.3 in Homework 1. Hence $\mu(x)$ and $\nu(x)$ are non-zero on at most countably many $x \in (a,b]$. It follows that

$$\begin{align}
\int\int 1_{(a<x=y\le b]}(x,y)dF(x)dG(y) &= (\mu\times\nu)\big(\{(x,y) : a < x = y \le b\}\big) \tag{46}\\
&= (\mu\times\nu)\big(\{(x,x) : a < x \le b\}\big) \tag{47}\\
&= \sum_{x\in(a,b]} \mu\{x\}\times\nu\{x\} \tag{48}
\end{align}$$

Putting it all together, we have

$$\begin{align}
\int_{(a,b]} F(y)dG(y) + \int_{(a,b]} G(y)dF(y) &= \int\int 1_{(a,b]}(x)1_{(a,b]}(y)dF(x)dG(y) \tag{49}\\
&\quad + \int\int 1_{(a<x=y\le b]}(x,y)dF(x)dG(y) \tag{50}\\
&\quad + F(a)[G(b)-G(a)] + G(a)[F(b)-F(a)] \tag{51}\\
&= F(b)G(b) - F(b)G(a) - F(a)G(b) + F(a)G(a) \tag{52}\\
&\quad + F(a)G(b) - F(a)G(a) + F(b)G(a) - F(a)G(a) \tag{53}\\
&\quad + \sum_{x\in(a,b]} \mu\{x\}\times\nu\{x\} \tag{54}\\
&= F(b)G(b) - F(a)G(a) + \sum_{x\in(a,b]} \mu\{x\}\times\nu\{x\} \tag{55}
\end{align}$$

□

(iii) if $F = G$ is continuous then $\int_{(a,b]} 2F(y)dF(y) = F^2(b) - F^2(a)$

$F = G \implies \mu = \nu$. Furthermore, continuity of F implies that $\mu(\{x\}) = 0$ for all x. Substituting these into part (ii), we get

$$\begin{align}
\int_{(a,b]} F(y)dG(y) + \int_{(a,b]} G(y)dF(y) &= F(b)G(b) - F(a)G(a) + \sum_{x\in(a,b]} \mu\{x\}\times\nu\{x\} \tag{56}\\
&= F(b)G(b) - F(a)G(a) + \sum_{x\in(a,b]} \mu\{x\}\times\mu\{x\} \tag{57}\\
&= F(b)F(b) - F(a)F(a) + 0 \tag{58}\\
&= F^2(b) - F^2(a) \tag{59}
\end{align}$$

□

2.1.12

Let $\Omega = \{1, 2, 3, 4\}$ and $\mathcal{F} = 2^{\Omega}$, and $\mathbb{P}(\{i\}) = 1/4$. Give an example of two collections of sets $\mathcal{A}_1$ and $\mathcal{A}_2$ that are independent but such that $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are not independent.

Let $\mathcal{A}_1 = \{\{1, 2\}, \{2, 3\}\}$ and $\mathcal{A}_2 = \{\{1, 3\}\}$. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent since

$\mathbb{P}(\{1, 2\} \cap \{1, 3\}) = \mathbb{P}(\{1\}) = 1/4 = 1/2 \times 1/2 = \mathbb{P}(\{1, 2\})\mathbb{P}(\{1, 3\})$, and

$\mathbb{P}(\{2, 3\} \cap \{1, 3\}) = \mathbb{P}(\{3\}) = 1/4 = 1/2 \times 1/2 = \mathbb{P}(\{2, 3\})\mathbb{P}(\{1, 3\})$

But $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are not indepndent since

$\{1, 2\} \cup \{2, 3\} = \{1, 2, 3\} \in \sigma(\mathcal{A}_1)$, and $\{1, 3\} \in \sigma(\mathcal{A}_2)$ and

$\mathbb{P}(\{1, 2, 3\} \cap \{1, 3\}) = \mathbb{P}(\{1, 3\}) = 1/2 \neq 3/4 \times 1/2 = \mathbb{P}(\{1, 2, 3\})\mathbb{P}(\{1, 3\})$ □

Let F_n, F be distribution functions of random variables X_n, X. Show that if $X_n \Rightarrow X$ and $P(X_n = x) \to P(X = x)$ for all x, then $F_n(x) \to F(x)$ for all x.

Proof

By **theorem 3.2.5** in Durrett, the following statements are equivalent:

(i) $X_n \Rightarrow X$

(ii) For all open sets G, $\liminf_{n\to\infty} P(X_n \in G) \geq P(X \in G)$

(iii) For all closed sets K, $\limsup_{n\to\infty} P(X_n \in K) \leq P(X \in K)$

Therefore, using that (i) $\Rightarrow$ (iii), we have

$$\limsup_{n\to\infty} P(X_n \leq x) \leq P(X \leq x) \text{ for all } x \tag{60}$$

Then, using that (i) $\Rightarrow$ (ii), we have

$$\liminf_{n\to\infty} P(X_n < x) \geq P(X < x) \text{ for all } x \tag{61}$$

Furthermore, we are given $P(X_n = x) \to P(X = x)$ which implies that $\liminf_{n\to\infty} P(X_n = x) = P(X = x)$, which implies that

$$\liminf_{n\to\infty} P(X_n = x) \geq P(X = x) \tag{62}$$

Then, adding (2) and (3) we get that

$$\liminf_{n\to\infty} P(X_n \leq x) \geq P(X \leq x) \text{ for all } x \tag{63}$$

Then, combining (1) and (4), we have

$$\limsup_{n\to\infty} P(X_n \leq x) \leq P(X \leq x) \leq \liminf_{n\to\infty} P(X_n \leq x) \text{ for all } x \tag{64}$$

which implies that $P(X_n = x) \to P(X = x)$ as desired. □

3.2.1

Give an example of randm variables X_n with densities f_n so that $X_n \Rightarrow$ a uniform distribution on $(0,1)$ but $f_n(x)$ does not converge to 1 for any $x \in [0,1]$

Let X_n be the random variable with density $f_n(x) = 1 + 2\pi\cos(2\pi nx)$.

Then $\int_0^1 f_n(x)dx = 1$ and $F_n = \int f_n(x)dx = x + \frac{sin(2\pi x)}{n} \to x$ for all $x \in [0,1]$ while $f_n(x) \nrightarrow x$ for all $x \in [0,1]$ □

3.2.9

If $F_n \Rightarrow F$ and F is continuous then $\sup_x |F_n(x) - F(x)| \to 0$

Fix $\epsilon > 0$ and choose k such that $\frac{1}{k} < \frac{\epsilon}{2}$, and for $i = 0, 1, ..., k-1, k$, choose $x_i \in \mathbb{R} \cup \{-\infty, \infty\}$ such that $F(x_i) = \frac{i}{k}$. Such x_i exist by continuity of F.

Since $F_n \Rightarrow F$, $\exists N_i \in \mathbb{N}$ such that $|F_n(x_i) - F(x_i)| \leq \frac{\epsilon}{2}$ for $i = 0, 1, ..., k-1, k$, for all $n > N_i$.

Fix i and note that for all $x \in (x_{i-1}, x_i]$, we have $F_n(x) \leq F_n(x_i)$, and $F_n(x_{i-1}) \leq F(x)$ by monotonicity of F and F_n. Hence,

$$F_n(x) - F(x) \leq F_n(x_i) - F(x_{i-1}) = F_n(x_i) - (F(x_i) - \frac{1}{k}) < \epsilon \tag{65}$$

$$F_n(x) - F(x) \geq F_n(x_{i-1}) - F(x_i) = F_n(x_{i-1}) - (F(x_{i-1}) + \frac{1}{k}) > -\epsilon \tag{66}$$

so $|F_n(x) - F(x)| < \epsilon$ for all $n > N_i$, $i = 0, 1, ..., k-1, k$. Now choose $N^* = \max\{N_i : i = 0, 1, ..., k-1, k\}$ to get $|F_n(x) - F(x)| < \epsilon$ for all x, for all $n > N_*$, and we have the desired result. □

3.2.11

Let $X_n, 1 \leq n \leq \infty$ be integer valued. Show that $X_n \Rightarrow X$ if and only $P(X_n = m) \to P(X_\infty = m)$ for all m.

Proof

$(\Rightarrow)$

Suppose $X_n \Rightarrow X$. Then $F_n(x) \to F(x)$ for all continuity points x of F. Fix m and let $a \in [m, m+1)$ and $b \in [m-1, m)$ be continuity points. Note that F may have at most countably many discontinuity points so a and b must always exist.

$$P(X_n = m) = F_n(a) - F_n(b) \to F(a) - F(b) = P(X_\infty = m)$$

$(\Leftarrow)$

Suppose $P(X_n = m) \to P(X_\infty = m)$. Then $F_n(x) = \sum_{m \leq x} P(X_n = m) \to \sum_{m \leq x} P(X_\infty = m) = F(x)$

□

3.3.2

(i) Show that $\mu(\{a\}) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-ita}\phi(t)dt$

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-ita}\phi(t)dt &= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-ita}\int_{\mathbb{R}} e^{itx}d\mu(x)dt \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T}\int_{\mathbb{R}} e^{-ita}e^{itx}d\mu(x)dt \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{\mathbb{R}}\int_{-T}^{T} e^{-ita}e^{itx}dtd\mu(x) \qquad \text{by Fubini} \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{\mathbb{R}}\int_{-T}^{T} e^{it(x-a)}dtd\mu(x) \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{x=a}\int_{-T}^{T} e^{it(x-a)}dtd\mu(x) + \lim_{T\to\infty} \frac{1}{2T}\int_{x\neq a}\int_{-T}^{T} e^{it(x-a)}dtd\mu(x) \\
&= \mu(\{a\}) + 0
\end{aligned} \tag{67}$$

since

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{2T}\int_{x=a}\int_{-T}^{T} e^{it(x-a)}dtd\mu(x) &= \lim_{T\to\infty} \frac{1}{2T}\int_{\mathbb{R}}\int_{-T}^{T} e^{0}\mathbf{1}_{\{a\}}(x)dtd\mu(x) \\
&= \lim_{T\to\infty} \frac{1}{2T}\int_{\mathbb{R}} 2T\mathbf{1}_{\{a\}}(x)d\mu(x) \\
&= \int_{\mathbb{R}} \mathbf{1}_{\{a\}}(x)d\mu(x) = \mu(\{a\})
\end{aligned} \tag{68}$$

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{2T}\int_{x\neq a}\int_{-T}^{T} e^{it(x-a)}dtd\mu(x) &= \lim_{T\to\infty} \frac{1}{2T}\int_{\mathbb{R}}\int_{-T}^{T} \cos(t(x-a)) + i\sin(t(x-a))dtd\mu(x) \\
&= \lim_{T\to\infty} \frac{1}{2T}\frac{1}{x-a}\int_{x\neq a} \sin(t(x-a))|_{-T}^{T} - i\cos(t(x-a))|_{-T}^{T}d\mu(x) \\
&= \lim_{T\to\infty} \frac{1}{2T}\frac{1}{x-a}\int_{x\neq a} \sin(T(x-a)) - \sin(-T(x-a)) \\
&\qquad - i\big(\cos(T(x-a)) - \cos(-T(x-a))\big)d\mu(x) \\
&= \lim_{T\to\infty} \frac{1}{2T}\frac{1}{x-a}\int_{x\neq a} \sin(T(x-a)) + \sin(T(x-a)) \\
&\qquad - i\big(\cos(T(x-a)) - \cos(T(x-a))\big)d\mu(x) \\
&= \lim_{T\to\infty} \frac{1}{2T}\frac{1}{x-a}\int_{x\neq a} 2\sin(T(x-a)))d\mu(x) \\
&= \lim_{T\to\infty}\int_{x\neq a} \frac{\sin(T(x-a)))}{T(x-a)}d\mu(x) \\
&= \int_{x\neq a}\lim_{T\to\infty} \frac{\sin(T(x-a)))}{T(x-a)}d\mu(x) \qquad \text{by DCT} \\
&= 0
\end{aligned} \tag{69}$$

(ii) If $P(X \in h\mathbb{Z}) = 1$ where $h > 0$, then its ch.f. has $\phi(2\pi/h + t) = \phi(t)$, so $P(X = x) = \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-itx}\phi(t)dt$

Note that I'm using $y \in h\mathbb{Z}$ in place of the $x \in h\mathbb{Z}$ that Durett uses for this question.

$$\begin{aligned}
\frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-ity}\phi(t)dt &= \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h}\int_{\mathbb{R}} e^{-ity}e^{itx}d\mu(x)dt \\
&= \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h}\int_{\mathbb{R}} e^{it(x-y)}d\mu(x)dt \\
&= \frac{h}{2\pi}\int_{\mathbb{R}}\int_{-\pi/h}^{\pi/h} e^{it(x-y)}dtd\mu(x) \qquad \text{by Fubini} \\
&= \frac{h}{2\pi}\int_{x=y}\int_{-\pi/h}^{\pi/h} e^{it(x-y)}dtd\mu(x) + \frac{h}{2\pi}\int_{x\neq y}\int_{-\pi/h}^{\pi/h} e^{it(x-y)}dtd\mu(x)
\end{aligned} \tag{70}$$

$$\begin{aligned}
\frac{h}{2\pi}\int_{x=y}\int_{-\pi/h}^{\pi/h} e^{it(x-y)}dtd\mu(x) &= \frac{h}{2\pi}\int_{\mathbb{R}}\int_{-\pi/h}^{\pi/h} e^{it(y-y)}\mathbf{1}_{\{y\}}(x)dtd\mu(x) \\
&= \frac{h}{2\pi}\int_{\mathbb{R}}\int_{-\pi/h}^{\pi/h} \mathbf{1}_{\{y\}}(x)dtd\mu(x) \\
&= \frac{h}{2\pi}\frac{2\pi}{h}\int_{\mathbb{R}} \mathbf{1}_{\{y\}}(x)d\mu(x) \\
&= \mu(\{y\}) = \mathbb{P}(X = y)
\end{aligned} \tag{71}$$

$$\begin{aligned}
\frac{h}{2\pi}\int_{x\neq y}\int_{-\pi/h}^{\pi/h} e^{it(x-y)}dtd\mu(x) &= \frac{h}{2\pi}\int_{x\neq y}\int_{-\pi/h}^{\pi/h} \cos(t(x-y)) + i\sin(t(x-y))dtd\mu(x) \\
&= \frac{h}{2\pi}\frac{1}{x-y}\int_{x\neq y} \sin(t(x-y))|_{-\pi/h}^{\pi/h} - i\cos(t(x-y))|_{-\pi/h}^{\pi/h}d\mu(x) \\
&= \int_{x\neq y}\frac{h}{2\pi}\frac{2\sin(\frac{\pi}{h}(x-y))}{x-y}d\mu(x) \\
&= \int_{x\neq y}\frac{h}{\pi}\frac{\sin(\frac{\pi}{h}(x-y))}{x-y}d\mu(x) = (*)
\end{aligned} \tag{72}$$

If show $(*) = 0$ then the proof is complete. To see why it is zero, first note that $\sin(n\pi) = 0$ for all $n \in \mathbb{Z}$; second, note that $y \in h\mathbb{Z}$ by assumption − this implies that $y/h \in \mathbb{Z}$; and third, note that $\mathbb{P}(X \in h\mathbb{Z}) = 1$ by assumption. Therefore, $\frac{x-y}{h} \in \mathbb{Z}$ and thus $\sin(\pi\frac{(x-y)}{h}) = 0$.

(iii) If $X = Y + b$ then $Eexp(itX) = eitbEexp(itY)$. So if $P(X \in b + h\mathbb{Z}) = 1$, the inversion formula in (ii) is valid for $x \in b + h\mathbb{Z}$.

$\phi_X(t) = e^{itb}\phi_Y(t)$ and so

$$\begin{aligned}
P(X = x) &= \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-itx}\phi_X(t)dt \\
&= \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-itx}e^{itb}\phi_Y(t)dt \\
&= \frac{h}{2\pi}\int_{-\pi/h}^{\pi/h} e^{-it(x-b)}\phi_Y(t)dt \\
&= P(Y = x - b)
\end{aligned} \tag{73}$$

□

3.3.8

Use the last result to conclude that if $X_1, X_2, ...$ are independent and have the Cauchy distribution, then $(X_1 + ... + X_n)/n$ has the same distribution as X_1.

$\phi_{X_i}(t) = \exp(-|t|) \Rightarrow \phi_{\frac{X_i}{n}}(t) = \phi_{X_i}(\frac{t}{n}) = \exp(-|\frac{t}{n}|) \Rightarrow \phi_{(\frac{X_1}{n}+...+\frac{X_n}{n})}(t) = \exp(-|\frac{t}{n}|)^n = \exp(-|t|)$ for iid X_i, by theorem 3.3.2 □

3.3.9

Suppose $X_n \Rightarrow X$ and X_n has a normal distribution with mean 0 and variance σ_n^2. Prove that $\sigma_n^2 \to \sigma \in [0, \infty)$

$X_n \xrightarrow{d} X \Rightarrow \phi_{X_n}(t) \to \phi_X(t)$ for all $t \in \mathbb{R}$.

Since $X_n \sim \mathcal{N}(0, \sigma_n^2)$, we have $\phi_{X_n}(t) = \exp(-\sigma_n^2 t^2/2)$ for all n.

$\phi_{X_n}(0) = 1$ for all n, which implies that $\phi_X(0) = 1$. And also, $\phi_{X_n}(t) \leq 1$ for all t, n, which implies that $\phi_X(t) \leq 1$ for all t.

By theorem 3.3.1, $\phi_X(t)$ is uniformly continuous. Therefore for all $\epsilon > 0$ there exists $\delta > 0$ such that $|\phi_X(r) - \phi_X(r)| < \epsilon$ whenever $|r - s| < \delta$.

Set $\epsilon = 1/2$ and $r = 0$, then by uniform continuity and the fact that $\phi_X(0) = 1$, we have $|1 - \phi_X(s)| < 1/2$. A little manipulation gives $1 - \phi_X(s) \leq |1 - \phi_X(s)| < 1/2 \Rightarrow 1/2 < \phi_X(s)$ for $s \in (-\delta, \delta)$.

Note that $\sigma_n^2 = -\frac{2}{t^2}\log(\phi_{X_n}(t))$ for all $t \neq 0 \in \mathbb{R}$. And by continuity of logarithm and the fact that $\phi_{X_n}(t) \to \phi_X(t)$, we have $-\frac{2}{t^2}\log(\phi_{X_n}(t)) \to -\frac{2}{t^2}\log(\phi_X(t))$

Fix $t = s^* \in (0, s)$, then $1/2 < \phi_X(s^*) \leq 1$.

then $\sigma_n^2 = -\frac{2}{s^{*2}}\log(\phi_{X_n}(s^*)) \to -\frac{2}{s^{*2}}\log(\phi_X(s^*)) = \sigma^2$

$\sigma^2 \in [0, \infty)$ because $1/2 < \phi_X(s^*) \leq 1$. □

3.3.10

Show that if X_n and Y_n are independent for $1 \leq n \leq \infty$, $X_n \Rightarrow X_\infty$, and $Y_n \Rightarrow Y_\infty$, then $X_n + Y_n \Rightarrow X_\infty + Y_\infty$

Since X_n and Y_n are independent, we have

$\phi_{X_n+Y_n}(t) = \phi_{X_n}(t)\phi_{Y_n}(t) \to \phi_{X_\infty}(t)\phi_{Y_\infty}(t) = \phi_{X_\infty+Y_\infty}(t)$ □

3.3.13

Let $X_1, X_2, ...$ be independent taking values 0 and 1 with probability 1/2 each. $X = \sum_{j\geq 1} X_j/3^j$ **has the Cantor distribution. Compute the ch.f. ϕ of X and notice that ϕ has the same value at $t = 3^k\pi$ for $k = 1, 2, 3, ...$**

$\phi_{X_i}(t) = \frac{1}{2} + \frac{1}{2}e^{it}$

$\phi_{\frac{2X_j}{3^j}}(t) = \frac{1}{2} + \frac{1}{2}e^{i2t/3^j}$

$\phi_X(t) = \prod_{j=1}^{\infty} \phi_{\frac{2X_j}{3^j}}(t) = \prod_{j=1}^{\infty} \left[\frac{1}{2} + \frac{1}{2}e^{i2t/3^j}\right]$

$$\begin{aligned}
\phi_X(3^k\pi) &= \prod_{j=1}^{\infty} \phi_{\frac{2X_j}{3^j}}(3^k\pi) \\
&= \prod_{j=1}^{\infty} \left[\frac{1}{2} + \frac{1}{2}e^{i2(3^k\pi)/3^j}\right] \\
&= \prod_{j=1}^{\infty} \left[\frac{1}{2} + \frac{1}{2}e^{i2\pi 3^{k-j}}\right] \\
&= \prod_{j=1}^{\infty} \left[\frac{1}{2} + \frac{1}{2}\cos(2\pi 3^{k-j}) + i\frac{1}{2}\sin(2\pi 3^{k-j})\right] \\
&= \prod_{j\leq k}^{\infty} \left[\frac{1}{2} + \frac{1}{2}\cos(2\pi 3^{k-j}) + i\frac{1}{2}\sin(2\pi 3^{k-j})\right] \prod_{j>k}^{\infty} \left[\frac{1}{2} + \frac{1}{2}\cos(2\pi 3^{k-j}) + i\frac{1}{2}\sin(2\pi 3^{k-j})\right] \\
&= \prod_{j\leq k}^{\infty} \left[1 + 0\right] \prod_{j>k}^{\infty} \left[\frac{1}{2} + \frac{1}{2}\cos(2\pi 3^{k-j}) + i\frac{1}{2}\sin(2\pi 3^{k-j})\right] \qquad (*) \\
&= \prod_{j>0}^{\infty} \left[\frac{1}{2} + \frac{1}{2}\cos(2\pi 3^{-j}) + i\frac{1}{2}\sin(2\pi 3^{-j})\right] \\
&= \prod_{j=1}^{\infty} \left[\frac{1}{2} + \frac{1}{2}e^{it3^{-j}}\right]
\end{aligned} \tag{74}$$

which does not depend on k.

Note that (*) holds because for all $j \leq k$, $\cos(2\pi 3^{k-j}) = 1$ and $\sin(2\pi 3^{k-j}) = 0$. □

3.3.23

If $X_1, X_2, ...$ are independent and have characteristic function $exp(-|t|^\alpha)$ then $(X_1 + ... + X_n)/n^{1/\alpha}$ has the same distribution as X_1.

$\phi_{X_i}(t) = \exp(-|t|^\alpha) \Rightarrow \phi_{\frac{X_i}{n^{1/\alpha}}}(t) = \exp(-|\frac{t}{n^{1/\alpha}}|^\alpha) = \exp(-\frac{|t|^\alpha}{n}) \Rightarrow \phi_{\frac{X_1}{n^{1/\alpha}}+...+\frac{X_n}{n^{1/\alpha}}}(t) = \left[\exp(-\frac{|t|^\alpha}{n})\right]^n = \exp(-|t|^\alpha) = \phi_{X_1}(t)$ □

3.3.26

Show that if X and Y are independent and $X+Y$ and X have the same distribution then $Y=0$ a.s.

$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \phi_X(t)$

$\Rightarrow \phi_X(t)(1-\phi_Y(t)) = 0$ for all t. So either $\phi_X(t) = 0$ or $(1-\phi_Y(t)) = 0$ for all t.

Since, by definition, $\phi_X(0) = 1$, it follows that $\phi_X(t) \neq 0$ for all t, and therefore, $\phi_Y(t) = 1$ for all t.

$\phi_Y(t) = 1 \Rightarrow \lim_{h\downarrow 0} \frac{\phi_Y(h)-2\phi_Y(0)+\phi_Y(-h)}{h^2} = \lim_{h\downarrow 0} \frac{1-2+1}{h^2} = 0 \Rightarrow E|Y|^2 < \infty$ by theorem 3.3.9.

Then we apply theorem 3.3.8 (since $E|Y|^2 < \infty$) to get $\phi_Y(t) = 1 + itEY - t^2E(Y^2)/2 + o(t^2)$. But since we know that $\phi_X(0) = 1$, it follows that $EY = 0$ and $E(Y^2)$. Therefore $Y = 0$ a.s. □

5.1.3

Prove Chebyshev's inquality. If a>0 then $P(|X| \geq |\mathcal{F}) \leq a^{-2}E(X^2|\mathcal{F})$

Note that $\mathbf{1}_{|X|\geq a} = \mathbf{1}_{|X|/a\geq 1} = \mathbf{1}_{X^2/a^2\geq 1} \leq X^2/a^2$. Thus,

$$\begin{aligned} P(X \geq a|\mathcal{F}) &= E[\mathbf{1}_{|X|\geq a}|\mathcal{F}] \\ &= E[\mathbf{1}_{X^2/a^2\geq 1}|\mathcal{F}] \\ &\leq E[X^2/a^2|\mathcal{F}] \\ &= a^{-2}E[X^2|\mathcal{F}] \end{aligned} \tag{75}$$

□

5.1.4

Suppose $X \geq 0$ and $EX = \infty$. Show that there exists a unique $\mathcal{F}$-measurable Y with $0 \leq Y \leq \infty$ such that for all $A \in \mathcal{F}$ we have

$$\int_A X dP = \int_A Y dP$$

Hint: Let $X_M = X \wedge M, Y_M = E(X_M|\mathcal{F})$, and let $M \to \infty$

We will use the hint and let $X_M = X \wedge M, Y_M = E(X_M|\mathcal{F})$. Note that $X_M \uparrow X$ as $M \to \infty$ and therefore $Y_M \uparrow$ to some limit. Let $Y = \lim_{M\to\infty} Y_M$.

By the definition of conditional expectation, we have

$$\int_A X \wedge M dP = \int_A Y_M dP \quad \text{for all } A \in \mathcal{F}$$

Taking the limit as $M \to \infty$ and applying MCT to both sides of the equation (since $X_M \uparrow X$ and $Y_M \uparrow Y$) we get

$$\int_A X dP = \int_A \lim_{M\to\infty} X \wedge M dP = \lim_{M\to\infty} \int_A X \wedge M dP = \lim_{M\to\infty} \int_A Y_M dP = \int_A \lim_{M\to\infty} Y_M dP = \int_A Y dP$$

□

5.1.7

Show that when $E|X|$, $E|Y|$, and $E|XY|$ are finite, each statement implies the next one and give examples with $X, Y \in \{-1, 0, 1\}$ a.s. that show the reverse implications are false: (i) X and Y are independent, (ii) $E(Y|X) = EY$, (iii) $E(XY) = EXEY$.

First note that $E(Y|X) = E(Y|\sigma(X))$

(i) $\Rightarrow$ (ii)

Let $A \in \sigma(X)$. Then by the definition of conditional expectation we have $\int_A Y dP = \int_A E(Y|\sigma(X))dP$

Then $\int_A Y dP = E[Y\mathbf{1}_A] = EYE\mathbf{1}_A = \int_A EY dP$, where the second equality follows from independence.

Example to demonstrate the converse is false:

Let $X, Y \in \{-1, 0, 1\}$ with the following joint distribution:

X/Y	-1	0	1
-1	0	0.25	0
0	0.25	0	0.25
1	0	0.25	0

Then, $P(X = 0, Y = 0) = 0 \neq 1/4 = 1/2 \times 1/2 = P(X = 0)P(Y = 0)$. Therefore X and Y are not independent.

(ii) $\Rightarrow$ (iii)

CLAIM: $E[E(Z|\mathcal{G})] = EZ$
PROOF: By the second property of conditional expectation, for all $A \in \mathcal{G}$, $\int_A Z dP = \int_A E(Z|\mathcal{G})dP$.
Let $A = \Omega$, and we get $EZ = \int_\Omega Z dP = \int_\Omega E(Z|\mathcal{G})dP = E[E(Z|\mathcal{G})]$

By theorem 5.1.7, if $X \in \mathcal{F}$ and $E|Y|$, and $E|XY|$ are finite, then $E(XY|\mathcal{F}) = XE(Y|\mathcal{F})$

$X \in \sigma(X)$ and the expectations are finite, so $E(XY|\sigma(X)) = XE(Y|\sigma(X)) = XEY$, where the second equality follows from (ii). Then taking expectation of left and right sides of the equality we get: LHS: $E(XY) = E[E(XY|\sigma(X))]$ and RHS: $E[XEY] = EXEY$

Example to demonstrate the converse is false:

Let $X \in \{-1, 0, 1\}$ and $X \in \{-1, 1\}$ with the following joint distribution:

X/Y	-1	1
-1	0	0.25
0	0.5	0
1	0	0.25

Then $EXY = 0 = EXEY$ but $E(Y|X = -1) = 1 \neq 0 = EY$

□

5.1.9

Let $var(X|\mathcal{F}) = E(X^2|\mathcal{F}) - E(X|\mathcal{F})^2$**. Show that** $var(X) = E(var(X|\mathcal{F})) + var(E(X|\mathcal{F}))$

$$\begin{aligned} E(var(X|\mathcal{F})) &= E[E(X^2|\mathcal{F}) - E(X|\mathcal{F})^2] \quad = E[E(X^2|\mathcal{F})] - E[E(X|\mathcal{F})^2] \\ &= E[X^2] - E[E(X|\mathcal{F})^2] \end{aligned} \tag{76}$$

$$\begin{aligned} var(E(X|\mathcal{F})) &= E[E(X|\mathcal{F})^2] - E[E(X|\mathcal{F})]^2 \\ &= E[E(X|\mathcal{F})^2] - E[X]^2 \end{aligned} \tag{77}$$

Then,

$$\begin{aligned} E(var(X|\mathcal{F})) + var(E(X|\mathcal{F})) &= E[X^2] - E[E(X|\mathcal{F})^2] + E[E(X|\mathcal{F})^2] - E[X]^2 \\ &= E[X^2] - E[X]^2 \\ &= var(X) \end{aligned} \tag{78}$$

□

5.1.11

Show that if X **and** Y **are random variables with** $E(Y|\mathcal{G}) = X$ **and** $EY^2 = EX^2 < \infty$**, then** $X = Y$ **a.s.**

By theorem 5.1.7, if $X \in \mathcal{G}$ and $E|Y|, E|XY| < \infty$ then $E(XY|\mathcal{G}) = XE(Y|\mathcal{G})$

Note that $X = E(Y|\mathcal{G}) \in \mathcal{G}$ by definition of conditional expectation. And $(EXY)^2 \leq EX^2EY^2 < \infty$ by the Cauchy-Schwarz inequality and the assumption of the question. Therefore, we can apply the theorem to get $E(XY|\mathcal{G}) = XE(Y|\mathcal{G})$. Taking expectation on both sides gives:

$$\begin{aligned} E[E(XY|\mathcal{G})] &= E[XE(Y|\mathcal{G})] \\ &= E[X^2] \end{aligned} \tag{79}$$

Then

$$\begin{aligned} E(X-Y)^2 &= E(X^2) - 2E(XY) + E(Y^2) \\ &= E(X^2) - 2E(X^2) + E(Y^2) \\ &= E(X^2) - 2E(X^2) + E(X^2) \\ &= 0 \end{aligned} \tag{80}$$

Then by Markov's inequality we have for all $\epsilon > 0$ $P(|X-Y| \geq \epsilon) \leq \frac{E(X-Y)^2}{\epsilon^2} = 0$. Therefore $X = Y$ a.s.

5.1.12

The result in the last exercise implies that if $EY^2 < \infty$ and $E(Y|\mathcal{G})$ has the same distribution as Y, then $E(Y|\mathcal{G}) = Y$ a.s. Prove this under the assumption $E|Y| < \infty$. Hint: The trick is to prove that sgn$(X) =$ sgn$(E(X|\mathcal{G}))$ a.s. and then take $X = Y - c$ to get the desired result.

Suppose X is a random variable with $E|X| < \infty$ such that $X \overset{d}{=} E(X|\mathcal{G})$. Let $A = \{\omega : X(\omega) > 0\}$, $B = \{\omega : E(X|\mathcal{G})(\omega) > 0\}$.

By the definition of conditional expectation we have $\int_B XdP = \int_B E(X|\mathcal{G})dP$.

And since $X \overset{d}{=} E(X|\mathcal{G})$, we have $\int_B E(X|\mathcal{G})dP = \int_A XdP$. Therefore $\int_B XdP = \int_A XdP$.

Furthermore,

$$\begin{aligned}\int_{A\cap B} XdP + \int_{A\cap B^C} XdP &= \int_A XdP \\ &= \int_B XdP \\ &= \int_{A\cap B} XdP + \int_{A^C\cap B} XdP\end{aligned} \tag{81}$$

which implies that $\int_{A\cap B^C} XdP = \int_{A^C\cap B} XdP$

Note that $X(A \cap B^C) > 0$ and $X(A^C \cap B) \leq 0$. Therefore, both integrals above are =0, and hence $\{X > 0, E(X|\mathcal{G}) \leq 0\} = A \cap B^C = \phi$. Since $X \overset{d}{=} E(X|\mathcal{G})$ it follows that $\{E(X|\mathcal{G}) > 0, X \leq 0\} = \phi$ also. Thus sgn$(X) =$ sgn$(E(X|G))$ a.s.

Now suppose that Y is a random variable such that $E|Y| < \infty$ such that $X \overset{d}{=} E(Y|\mathcal{G})$. Let $X = Y - c$. By the above arguments, we have that $\{Y - c > 0, E(Y - c|\mathcal{G}) \leq 0\} = \{E(Y - c|\mathcal{G}) > 0, Y - c \leq 0\} = \phi$ or equivalently $\{Y > c, E(Y|\mathcal{G}) \leq c\} = \{E(Y|\mathcal{G}) > c, Y \leq c\} = \phi$.

Since $\mathbb{Q}$ is dense in $\mathbb{R}$ we have that

$$\{Y \neq E(Y|G)\} = \bigcup_{c\in\mathbb{Q}} \Big(\{Y > c, E(Y|G) \leq c\} \cup \{E(Y|\mathcal{G}) > c, Y \leq c\}\Big) = \phi$$

Therefore $E(Y|\mathcal{G}) = Y$ a.s. as desired □

5.3.2.

Give an example of a martingale X_n with $\sup_n |X_n| < 1$ and $P(X_n = a\text{i.o.}) = 1$ for $a = -1, 0, 1$. This example shows that it is not enough to have $\sup |X_{n+1} - X_n| < \infty$ in Theorem 5.3.1.

Let $U_i \overset{iid}{\sim} Uniform(0, 1)$, $i \in \mathbb{N}$.

If $X_n = 0$ then let $X_{n+1} = 1$ if $U_{n+1} \geq 1/2$, and $X_{n+1} = -1$ if $U_{n+1} < 1/2$.

If $X_n \neq 0$ then let $X_{n+1} = 0$ if $U_{n+1} > 1/n^2$, and $X_{n+1} = n^2 X_n$ if $U_{n+1} < 1/n^2$.

$\sum 1/n^2 < \infty$ so by the Borel Cantelli lemma we evenentually just go from 0 to ± 1 and then back to 0 again, so $\sup |X_n| < \infty$.

5.4.7

Let X_n and Y_n be martingales with $EX_n^2 < \infty$ and $EY_n^2 < \infty$.

$$EX_nY_n - EX_0Y_0 = \sum_{m=1}^{n} E(X_m - X_{m-1})(Y_m - Y_{m-1})$$

Note that

$$\begin{aligned} E(X_m - X_{m-1})(Y_m - Y_{m-1}) &= E[(X_m - X_{m-1})Y_m] - E[(X_m - X_{m-1})Y_{m-1})] \\ &= EX_mY_m - EX_{m-1}Y_m - EX_mY_{m-1} + EX_{m-1}Y_{m-1} \end{aligned} \tag{82}$$

By theorem 5.4.6 (orthogonality of martingale increments) we have that $E[(X_m - X_{m-1})Y_{m-1}] = 0$ Then we have

$$E[(X_m - X_{m-1})Y_{m-1}] = 0 \Rightarrow E[X_mY_m - X_{m-1}Y_{m-1}] = 0 \Rightarrow EX_mY_{m-1} = EX_{m-1}Y_{m-1}$$

and similarly

$$EY_mX_{m-1} = EX_{m-1}Y_{m-1}$$

Substituting these into (1) and summing from $m = 1$ to n we have

$$\begin{aligned} \sum_{m=1}^{n} E(X_m - X_{m-1})(Y_m - Y_{m-1}) &= \sum_{m=1}^{n} E[(X_m - X_{m-1})Y_m] - E[(X_m - X_{m-1})Y_{m-1})] \\ &= \sum_{m=1}^{n} EX_mY_m - EX_{m-1}Y_m - EX_mY_{m-1} + EX_{m-1}Y_{m-1} \\ &= \sum_{m=1}^{n} EX_mY_m - EX_{m-1}Y_{m-1} - EX_{m-1}Y_{m-1} + EX_{m-1}Y_{m-1} \\ &= \sum_{m=1}^{n} EX_mY_m - EX_{m-1}Y_{m-1} \\ &= EX_nY_n - EX_0Y_0 \end{aligned} \tag{83}$$

(The last equality follows by telescoping the sum) □

5.4.8

Let $X_n, n \geq 0$ be a martingale and let $\xi_n = X_n - X_{n-1}$ for $n \geq 0$. Show that if $EX_0^2, \sum_{m=1}^{\infty} \xi_m^2 < \infty$ for $n \geq 1$ then $X_n \to X_\infty$ a.s. and in L^2

In the previous exercise, let $X = Y$ and note that

$$\begin{aligned}\sum_{m=1}^{n} \xi_m^2 &= \sum_{m=1}^{n} E(X_m - X_{m-1})^2 \\ &= \sum_{m=1}^{n} E(X_m - X_{m-1})(X_m - X_{m-1}) \\ &= EX_nX_n - EX_0X_0 \\ &= EX_n^2 - EX_0^2\end{aligned} \tag{84}$$

Rearranging (3) and using that $EX_0^2, \sum_{m=1}^{n} \xi_m^2 < \infty$ we get $EX_n^2 = EX_0^2 + \sum_{m=1}^{n} \xi_m^2 < \infty$.

Thus $\sup EX_n^2 < \infty$, and so $X_n \longrightarrow X$ a.s. and in L^p by theorem 5.4.5 (L^p convergence theorem). □

5.4.10

Let $\xi_1, \xi_2, ...$ be i.i.d. with $E\xi_i = 0$ and $E\xi_i^2 < \infty$. Let $S_n = \xi_1 + ... + \xi_n$. Theorem 5.4.1 implies that for any stopping time $N, $ES_{N^n} = 0$. Use Theorem 5.4.12 to conclude that if $EN^{1/2} < \infty$ then $ES_N = 0$.

S_n is a martingale.

By theorem 5.2.6, for any stopping time N, and martingale S_n, $S_{N \wedge n}$ is a martingale.

Note that $EN^{1/2} < \infty \Rightarrow N^{1/2} < \infty$ a.s. $\Rightarrow N < \infty$ a.s.

Furthermore, $ES_{N \wedge n}$=0, because $\{N \wedge n = n\} \in \mathcal{F}_n$ for all n so $N \wedge n$ is a stopping time, and, $P(N \wedge n \leq n) = 1$ so by theorem 5.4.1, we have $0 = ES_0 = ES_{N \wedge n} = ES_n = 0$.

Also, note that $S_{N \wedge m} - S_{N \wedge m-1} = \begin{cases} S_m - S_{m-1}, & \text{if } N \geq m \\ 0, & \text{otherwise} \end{cases}$

and $A_n = \sum_{m=1}^{n} E((S_{N \wedge m} - S_{N \wedge m-1})^2 | \mathcal{F}_{m-1}) = \sum_{m=1}^{N \wedge n} E(\xi_m^2 | \mathcal{F}_{m-1}) = \sum_{m=1}^{N \wedge n} E\xi_m^2 < \infty$

by the assumptions of independence and finiteness of second moments of ξ_i's.

Letting $n \to \infty$, we have that $A_\infty < \infty$ since $N < \infty$ a.s..

Now by theorem 5.4.12, we have $E(\sup |S_{N \wedge n}|) \leq 3EA_\infty^{1/2} < \infty$.

Finally, since i. $|S_{N \wedge n}| \leq \sup |S_{N \wedge n}|$ for all n, and ii. $E(\sup |S_{N \wedge n}|) < \infty$, and iii. $ES_{N \wedge n} = 0$, we can use the dominated convergence theorem to get

$ES_N = E(\lim_{n \to \infty} S_{N \wedge n}) = \lim_{n \to \infty} E(S_{N \wedge n}) = \lim_{n \to \infty} 0 = 0$ □

5.5.2

Let $Z_1, Z_2, ...$ be i.i.d. with $E|Z_i| < \infty$, let θ be an independent r.v. with finite mean, and let Yi $= Z_i + \theta$. If Z_i is normal(0,1) then in statistical terms we have a sample from a normal population with variance 1 and unknown mean. The distribution of θ is called the prior distribution, and $P(\theta \in ů|Y_1, ..., Y_n)$ is called the posterior distribution after n observations. Show that $E(\theta|Y_1, ..., Y_n) \to \theta$ a.s.

Let $\mathcal{F}_n = \sigma(Y_1, ..., Y_n)$ and $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$.Then

$$E(\theta|\mathcal{F}_n) \xrightarrow{a.s.} E(\theta|\mathcal{F}_\infty) = \theta$$

where the a.s. convergence on the left is due to theorem 5.5.7, and the equality on the right is due to the fact that $\theta \in \mathcal{F}_\infty$ which is a consequence of the strong law of large numbers as follows:

$$\frac{Y_1 + ... + Y_n}{n} = \theta + \frac{Z_1 + ... + Z_n}{n} \xrightarrow{a.s.} \theta + EZ_i = \theta + 0$$

□

5.5.5

Let X_n be a r.v.'s taking values in $[0, \infty)$. Let $D = \{X_n = 0$for some n $\geq 1\}$ and assume that $P(D|X_1, ..., X_n) \geq \delta(x) > 0$ on $\{X_n \leq x\}$. Use theorem 5.5.8 to conclude that

$$P(D \cup \{\lim_{n\to\infty} X_n = \infty\}) = 1$$

By theorem 5.5.8, if $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and $A \in \mathcal{F}_\infty$, then $E(\mathbf{1}_A|\mathcal{F}_n) \to \mathbf{1}_A$

Let $\mathcal{F}_n = \sigma(X_1, ..., X_n)$ and $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$. Then $D \in \mathcal{F}_\infty$.

Let $B = \{\omega : \lim_{n\to\infty} X_n(\omega) \neq \infty\}$. Note that $P(B \cup B^c) = 1$.

Let $B_M = \{\omega : \liminf_{n\to\infty} X_n(\omega) \leq M\}$ for $M \in \mathbb{N}$. Note that $B = \bigcup_M B_M$,

Fix some $M \in \mathbb{N}$, and note that for all $\omega \in B_M$, we have $X_n(\omega) \leq M + 1$ infinitely often.

In other words, $B_M \subset \{\omega : X_n(\omega) \leq M + 1 \text{ i.o.}\}$.

Then by theorem 5.5.8, we have

$$0 < \delta(M + 1) \leq P(D|X_1, ..., X_n) = P(D|\mathcal{F}_n) = E(\mathbf{1}_D|\mathcal{F}_n) \longrightarrow \mathbf{1}_D \tag{85}$$

which implies that $B_M \subset \{\omega : X_n(\omega) \leq M + 1 \text{ i.o.}\} \subset D$ (because $0 < E(\mathbf{1}_D|\mathcal{F}_\infty) \in \{0, 1\}$)

Since $B_M \subset D$, and $B = \bigcup_M B_M$, we have $B \subset D$, and it follows that

$$1 \geq P(D \cup \{\lim_{n\to\infty} X_n = \infty\}) = P(D \cup B^c) \geq P(B \cup B^c) = 1$$

which gives the desired result. □

5.7.2

Let S_n be an asymmetric simple random walk with $1/2 < p < 1$, and let $\sigma^2 = pq$. Use the fact that $X_n = (S_n - (p-q)n)^2 - \sigma^2 n$ is a martingale to show $var(T_b) = b\sigma^2/(p-q)^3$.

Since $X_n = (S_n - (p-q)n)^2 - \sigma^2 n$ is a martingale, and $T_b = \inf\{n : S_n = b\}$ is a stopping time, it follows by theorem 5.2.6 that $X_{T_b \wedge n} = (S_{T_b \wedge n} - (p-q)(T_b \wedge n))^2 - \sigma^2 (T_b \wedge n)$ is also a martingale.

Since $X_{T_b \wedge 0} = X_0 = (S_0 - (p-q)0)^2 - \sigma^2 0 = 0$, and $(T_b \wedge n)$ is a bounded stopping time, it follows by theorem 5.4.1 that

$$EX_{T_b \wedge 0} = EX_{T_b \wedge n} = E[(S_{T_b \wedge n} - (p-q)(T_b \wedge n))^2 - \sigma^2 (T_b \wedge n)] = 0$$

Then we have

$$E[(S_{T_b \wedge n} - (p-q)(T_b \wedge n))^2] = E[\sigma^2 (T_b \wedge n)]$$

Using Fatou's lemma (since $(S_{T_b \wedge n} - (p-q)(T_b \wedge n))^2 \geq 0$), and bounded convergence theorem since $(T_b \wedge n$ is bounded) we have

$$\begin{aligned} E[(b - (p-q)T_b)^2] &= E[\lim_{n\to\infty} (S_{T_b \wedge n} - (p-q)(T_b \wedge n))^2] \\ &= E[\liminf_{n\to\infty} (S_{T_b \wedge n} - (p-q)(T_b \wedge n))^2] \\ &\leq \liminf_{n\to\infty} E[(S_{T_b \wedge n} - (p-q)(T_b \wedge n))^2] \\ &= \liminf_{n\to\infty} E[\sigma^2 (T_b \wedge n)] \\ &= \lim_{n\to\infty} E[\sigma^2 (T_b \wedge n)] \\ &= E[\lim_{n\to\infty} \sigma^2 (T_b \wedge n)] = \sigma^2 ET_b = b/(2p-1) < \infty \end{aligned} \tag{86}$$

Thus $ET_b^2 \leq b/(2p-1) < \infty$. Then,

$$\begin{aligned} 0 = EX_{T_b \wedge n} &= E[(S_{T_b \wedge n} - (p-q)(T_b \wedge n))^2 - \sigma^2 (T_b \wedge n)] \\ &= E[S_{T_b \wedge n}^2 - 2S_{T_b \wedge n}(p-q)(T_b \wedge n) + (p-q)^2 (T_b \wedge n)^2 - \sigma^2 (T_b \wedge n)] \end{aligned} \tag{87}$$

Note that ${S_{T_b \wedge n}}^2 \leq b^2 < \infty$. Moreover, $E[T_b \wedge n] \leq E[T_b] = b/(2p-1) < \infty$ by theorem 5.7.7 (d). And $ES_{T_b \wedge n}(T_b \wedge n) \leq \left[E[S_{T_b \wedge n}^2] E[(T_b \wedge n)^2]\right]^{1/2} < \infty$ by the Cauchy-Schwarz inequality and (5).

Thus we can apply dominated convergence theorem to (6) to get

$$\begin{aligned} 0 &= \lim_{n\to\infty} E[S_{T_b \wedge n} - (p-q)(T_b \wedge n))^2 - \sigma^2 (T_b \wedge n)] \\ &= \lim_{n\to\infty} E[S_{T_b \wedge n}^2 - 2S_{T_b \wedge n}(p-q)(T_b \wedge n) + (p-q)^2 (T_b \wedge n)^2 - \sigma^2 (T_b \wedge n)] \\ &= E \lim_{n\to\infty} [S_{T_b \wedge n}^2 - 2S_{T_b \wedge n}(p-q)(T_b \wedge n) + (p-q)^2 (T_b \wedge n)^2 - \sigma^2 (T_b \wedge n)] \\ &= E[b^2 - 2b(p-q)T_b + (p-q)^2 T_b^2 - \sigma^2 T_b] \\ &= b^2 - 2b(p-q)b/(2p-1) + (p-q)^2 ET_b^2 - \sigma^2 b/(2p-1) \\ &= b^2 - 2b(p-q)b/(p-q) + (p-q)^2 ET_b^2 - \sigma^2 b/(p-q) \\ &= -b^2 + (p-q)^2 ET_b^2 - \sigma^2 b/(p-q) \end{aligned} \tag{88}$$

Rearranging (7) gives $ET_b^2 = \frac{\sigma^2 b/(p-q) + b^2}{(p-q)^2} = \sigma^2 b/(p-q)^3 + b^2/(p-q)^2$

Finally, using the fact that $(p-q) = 2p-1$ and $E[T_b] = b/(2p-1)$ (theorem 5.7.7 d), we get the desired result:

$Var(T_b) = E[T_b^2] - E[T_b]^2 = \sigma^2 b/(p-q)^3 + b^2/(p-q)^2 - b^2/(2p-1)^2 = \sigma^2 b/(p-q)^3$ $\square$