Chapter 8

Laws of large numbers

We first discuss the notion of "tail equivalence" of a sequence of random variables. Here, the Borel-Cantelli Lemma is very useful. Recall that it says that if $\{E_n\}_{n\in\mathbb{N}}$ is a sequence of events with $\sum_{n=1}^{\infty} P(E_n) < \infty$, then $P\left(\limsup_{n\to\infty} E_n\right) = 0$.

Definition 8.1. *Two sequences of random variables $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are tail equivalent if*

$$\sum_{n=1}^{\infty} P\left(\{\omega : X_n(\omega) \neq Y_n(\omega)\}\right) = \sum_{n=1}^{\infty} P(\{\omega : X_n(\omega) - Y_n(\omega) \neq 0\}) = \sum_{n=1}^{\infty} P(A_n) < \infty,$$

where $A_n = \{\omega : X_n(\omega) - Y_n(\omega) \neq 0\}$.

Theorem 8.1. *Suppose $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are tail equivalent. Then,*

1. *$\sum_{n=1}^{\infty}(X_n - Y_n)$ converges almost surely,*

2. *$\sum_{n=1}^{\infty} X_n$ converges as $\iff$ $\sum_{n=1}^{\infty} Y_n$ converges as,*

3. *If there exists $a_n \to \infty$ and a random variable X such that $\frac{1}{a_n}\sum_{j=1}^{n} X_j \xrightarrow{as} X$, then $\frac{1}{a_n}\sum_{j=1}^{n} Y_j \xrightarrow{as} X$.*

Proof. 1. By tail equivalence and the Borel-Cantelli Lemma $P\left(\limsup_{n\to\infty} A_n\right) = 0$. Now, recall that $\limsup_{n\to\infty} A_n = \cap_{n=1}^{\infty}\cup_{m=n}^{\infty} A_m = \cap_{n=1}^{\infty} C_n$, where $C_n := \cup_{m=n}^{\infty} A_m$. Consequently,

$$\left(\limsup_{n\to\infty} A_n\right)^c = (\cap_{n=1}^{\infty} C_n)^c = \cup_{n=1}^{\infty} C_n^c = \cup_{n=1}^{\infty}(\cup_{m=n}^{\infty} A_m)^c = \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m^c = \liminf_{n\to\infty} A_n^c.$$

Thus,

$$P\left(\liminf_{n\to\infty}\{\omega: X_n(\omega)=Y_n(\omega)\}\right) = P\left(\liminf_{n\to\infty} A_n^c\right) = 1 - P\left(\limsup_{n\to\infty} A_n\right) = 1.$$

Since $\liminf_{n\to\infty} A_n^c = \{\omega: \sum_{n=1}^{\infty} I_{A_n}(\omega) < \infty\}$, $P\left(\{\omega: \sum_{n=1}^{\infty} I_{A_n}(\omega) < \infty\}\right) = 1$. Hence, there exists a set of ω's which occurs with probability 1, and in this set $X_n(\omega) = Y_n(\omega)$ for all but finitely many n. That is, for $\omega \in \{\omega: \sum_{n=1}^{\infty} I_{A_n}(\omega) < \infty\}$ there are only finitely many n for which $I_{\{X_n(\omega) \neq Y_n(\omega)\}}(\omega) = 1$. That is, there exists $N(\omega)$ such that for all $n > N(\omega), I_{\{X_n(\omega) \neq Y_n(\omega)\}}(\omega) = 0$. Hence, in this same set,

$$\sum_{n=1}^{\infty} X_n(\omega) - \sum_{n=1}^{\infty} Y_n(\omega) = \sum_{n=1}^{N(\omega)} (X_n(\omega) - Y_n(\omega)) < \infty \text{ almost surely .}$$

2. Note that

$$\sum_{n=1}^{\infty} Y_n(\omega) = \sum_{n=1}^{\infty} X_n(\omega) + \sum_{n=1}^{\infty} Y_n(\omega) - \sum_{n=1}^{\infty} X_n(\omega)$$

$$= \sum_{n=1}^{\infty} X_n(\omega) - \sum_{n=1}^{\infty} (X_n(\omega) - Y_n(\omega)).$$

If $\sum_{n=1}^{\infty} X_n(\omega)$ converges as and X_n and Y_n are tail equivalent, then both terms on the right side of the equality converge as, hence $\sum_{n=1}^{\infty} Y_n(\omega) < \infty$ as. Similarly, writing

$$\sum_{n=1}^{\infty} X_n(\omega) = \sum_{n=1}^{\infty} X_n(\omega) + \sum_{n=1}^{\infty} Y_n(\omega) - \sum_{n=1}^{\infty} Y_n(\omega)$$

$$= \sum_{n=1}^{\infty} Y_n(\omega) - \sum_{n=1}^{\infty} (Y_n(\omega) - X_n(\omega)).$$

we conclude $\sum_{n=1}^{\infty} X_n(\omega) < \infty$ as.

3. Write

$$\frac{1}{a_n} \sum_{j=1}^{n} Y_j(\omega) = \frac{1}{a_n} \sum_{j=1}^{n} (Y_j(\omega) - X_j(\omega) + X_j(\omega))$$

$$= \frac{1}{a_n} \sum_{j=1}^{n} (Y_j(\omega) - X_j(\omega)) + \frac{1}{a_n} \sum_{j=1}^{n} X_j(\omega)$$

$$= \frac{1}{a_n} \sum_{j=1}^{N-1} (Y_j(\omega) - X_j(\omega)) + \frac{1}{a_n} \sum_{j=N}^{n} (Y_j(\omega) - X_j(\omega)) + \frac{1}{a_n} \sum_{j=1}^{n} X_j(\omega).$$

As $n \to \infty$ the last term converges as to $X(\omega)$ by assumption. The second term converges to zero since $Y_j(\omega)$ and $X_j(\omega)$ are tail equivalent (and by 1), and the first term goes to 0 as $a_n \to \infty$. Hence, $\frac{1}{a_n} \sum_{j=1}^{n} Y_j(\omega) \overset{as}{\to} X(\omega)$.

∎

The following definition and associated notation will be useful.

Definition 8.2. *Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables defined on $(\Omega, \mathcal{F}, P)$ and $\{s_n\}_{n \in \mathbb{N}}$ be a sequence in $(0, \infty)$. We write,*

1. $X_n = O_p(s_n)$ if for all $\epsilon > 0$ and $n \in \mathbb{N}$, there exists $B_\epsilon > 0$ such that

$$P\left(\left\{\omega : \frac{|X_n(\omega)|}{s_n} > B_\epsilon\right\}\right) < \epsilon$$

2. $X_n = o_p(s_n)$ if $\frac{X_n}{s_n} \overset{P}{\to} 0$.

Theorem 8.2. *(General Law of Large Numbers) Suppose $\{X_n\}_{n \in \mathbb{N}}$ is a sequence of independent random variables defined on $(\Omega, \mathcal{F}, P)$ and $S_n = \sum_{j=1}^{n} X_j$. If*

1. $\sum_{j=1}^{n} P(\{\omega : |X_j(\omega)| > n\}) \to 0$ as $n \to \infty$ and

2. $\frac{1}{n^2} \sum_{j=1}^{n} E(X_j^2 I_{\{\omega : |X_j| \leq n\}}) \to 0$ as $n \to \infty$,

then $\frac{S_n}{n} - \frac{1}{n} \sum_{j=1}^{n} E(X_j I_{\{\omega : |X_j| \leq n\}}) \overset{P}{\to} 0$.

Proof. Let $T_{n,j}(\omega) = X_j(\omega) I_{\{\omega : |X_j| \leq n\}}$ and $S'_n(\omega) = \sum_{j=1}^{n} T_{n,j}(\omega)$. Note that $\{\omega : X_j(\omega) \neq T_{n,j}(\omega)\} = \{\omega : |X_j(\omega)| > n\}$ and by assumption $\sum_{j=1}^{n} P(\{\omega : T_{n,j}(\omega) \neq X_j(\omega)\}) \to 0$ as $n \to \infty$. Note also that

$$|S_n(\omega) - S'_n(\omega)| = \left|\sum_{j=1}^{n} X_j(\omega) - \sum_{j=1}^{n} T_{n,j}(\omega)\right| \leq \sum_{j=1}^{n} |X_j(\omega) - T_{n,j}(\omega)|.$$

Thus, for all $\epsilon > 0$,

$$\{\omega : |S_n(\omega) - S'_n(\omega)| > \epsilon\} \subset \left\{\omega : \sum_{j=1}^{n} |X_j(\omega) - T_{n,j}(\omega)| > \epsilon\right\}$$

$$\subset \bigcup_{j=1}^{n} \{\omega : |X_j(\omega) - T_{n,j}(\omega)| > \epsilon/n\}.$$

Consequently,

$$P(\{\omega : |S_n(\omega) - S'_n(\omega)| > \epsilon\}) \leq \sum_{j=1}^{n} P\left(\{\omega : |X_j(\omega) - T_{n,j}(\omega)| > \epsilon/n\}\right)$$

$$\leq \sum_{j=1}^{n} P\left(\{\omega : |X_j| > n\}\right).$$

Taking limits on both sides as $n \to \infty$, we have that $S_n - S'_n \xrightarrow{p} 0$ since by assumption 1 $\sum_{j=1}^{n} P\left(\{\omega : |X_j| > n\}\right) \to 0$.

Now, since $\{X_n\}_{n \in \mathbb{N}}$ is an independent sequence $E\left((T_{n,k} - E(T_{n,k}))(T_{n,l} - E(T_{n,l}))\right) = 0$ and consequently $V(S'_n) = \sum_{j=1}^{n} V(T_{n,j}) \leq \sum_{j=1}^{n} E(T_{n,j}^2)$. Note also that for given n

$$E(T_{n,j}^2) = \int_{\Omega} X_j^2 I_{\{\omega : |X_j| \leq n\}} dP \leq n^2 \int_{\Omega} dP = n^2.$$

Consequently, since $V(S'_n)$ exists for every n, by Chebyshev's Inequality (Remark 5.1),

$$P\left(\left\{\omega : \left|\frac{S'_n - E(S'_n)}{n}\right| > \epsilon\right\}\right) \leq \frac{V(S'_n)}{n^2 \epsilon^2} \leq \frac{1}{n^2 \epsilon^2} \sum_{j=1}^{n} E\left(X_j^2 I_{\{\omega : |X_j| \leq n\}}\right).$$

Taking limits on both sides as $n \to \infty$ and by the assumption that $\frac{1}{n^2} \sum_{j=1}^{n} E\left(X_j^2 I_{\{\omega : |X_j| \leq n\}}\right) \to 0$, we have $\frac{S'_n}{n} - \frac{E(S'_n)}{n} \xrightarrow{p} 0$. Now, since

$$\frac{S_n}{n} - E\left(\frac{S'_n}{n}\right) = \frac{S_n}{n} - \frac{S'_n}{n} + \frac{S'_n}{n} - E\left(\frac{S'_n}{n}\right)$$

we can immediately conclude that $\frac{S_n}{n} - E\left(\frac{S'_n}{n}\right) = o_p(1)$. Finally, from the definition of S'_n we have that $\frac{S_n}{n} - \frac{1}{n} \sum_{j=1}^{n} E(X_j I_{\{\omega : |X_j| \leq n\}}) = o_p(1)$. ∎

We note $E(X_j) < \infty$ or $E(X_j^2) < \infty$ are not required for Theorem 8.2. The following are examples of how Theorem 8.2 can be used.

140

Example 8.1. *Let $\{X_n\}_{n\in\mathbb{N}}$ be an independent and identically distributed sequence of random variables with $E(X_n) = \mu$, $E(X_n^2) \leq C < \infty$. Then, we verify condition 1 by noting that the identical distribution assumption and Markov's Inequality give*

$$\sum_{j=1}^{n} P(|X_j| > n) = nP(|X_1| > n) \leq n\frac{E(X_1^2)}{n^2} = \frac{1}{n}E(X_1^2) \leq \frac{C}{n}.$$

Taking limits on both sides as $n \to \infty$ gives $\lim_{n\to\infty} \sum_{j=1}^{n} P(|X_j| > n) = 0$. For condition 2, note that by the identical distribution assumption

$$\frac{1}{n^2}\sum_{j=1}^{n} E(X_j^2 I_{\{|X_j|\leq n\}}) = \frac{1}{n}E(X_1^2 I_{\{|X_1|\leq n\}}) \leq \frac{1}{n}E(X_1^2) \leq \frac{C}{n}.$$

Again, taking limits on both sides as $n \to \infty$ gives $\lim_{n\to\infty} \frac{1}{n^2} \sum_{j=1}^{n} E(X_j^2 I_{\{|X_j|\leq n\}}) = 0$. Finally, observe that

$$\frac{\sum_{j=1}^{n} E(X_j I_{\{|X_j|\leq n\}})}{n} = E(X_1 I_{\{|X_1|\leq n\}}) \to E(X_1) = \mu$$

as $n \to \infty$ by Lebesgue's dominated convergence theorem. Thus, $\frac{1}{n}S_n \xrightarrow{p} \mu$.

Example 8.2. *Let $\{X_n\}_{n\in\mathbb{N}}$ be an independent and identically distributed sequence with $E(|X_1|) \leq C < \infty$ and let $E(X_1) = \mu$. For condition 1, note that*

$$\sum_{j=1}^{n} P(|X_j| > n) = nP(|X_1| > n) = E(nI_{\{\omega:|X_1|>n\}}).$$

But since $nI_{\{\omega:|X_1|>n\}} \leq |X_1|I_{\{\omega:|X_1|>n\}}$, we have that

$$\sum_{j=1}^{n} P(|X_j| > n) \leq E(|X_1|I_{\{\omega:|X_1|>n\}})$$

Consequently, $\lim_{n\to\infty} \sum_{j=1}^{n} P(|X_j| > n) \leq \lim_{n\to\infty} E(|X_1|I_{\{\omega:|X_1|>n\}})$. And since $E(|X_1|) < C$, $\lim_{n\to\infty} E(|X_1|I_{\{\omega:|X_1|>n\}}) = 0$.

For condition 2, note that by the identical distribution assumption

$$\frac{1}{n^2}\sum_{j=1}^{n} E\left(X_j^2 I_{\{\omega:|X_j|\leq n\}}\right) = \frac{1}{n}E\left(X_1^2 I_{\{\omega:|X_1|\leq n\}}\right)$$

$$= \frac{1}{n}\left(E\left(X_1^2 I_{\{\omega:|X_1|\leq \epsilon\sqrt{n}\}}\right) + E\left(X_1^2 I_{\{\omega:\epsilon\sqrt{n}\leq|X_1|\leq n\}}\right)\right) \text{ for any } \epsilon \in (0,1)$$

141

Since $E(X_1^2 I_{\{\omega:|X_1|\leq\epsilon\sqrt{n}\}}) = \int_\Omega X_1^2 I_{\{\omega:|X_1|\leq\epsilon\sqrt{n}\}} dP \leq n\epsilon^2 \int_\Omega dP = n\epsilon^2$, we have

$$\frac{1}{n^2}\sum_{j=1}^{n} E\left(X_j^2 I_{\{\omega:|X_j|\leq n\}}\right) \leq \epsilon^2 + \frac{1}{n}E\left(|X_1||X_1|I_{\{\omega:\epsilon\sqrt{n}\leq|X_1|\leq n\}}\right) \leq \epsilon^2 + \frac{1}{n}E(n|X_1|I_{\{\omega:\epsilon\sqrt{n}\leq|X_1|\leq n\}}))$$

$$\leq \epsilon^2 + E\left(|X_1|I_{\{\omega:\epsilon\sqrt{n}\leq|X_1|\}}\right)$$

Taking limits on both sides as $n\to\infty$, and noting that $E(|X_1|) < C$, we have that

$$\lim_{n\to\infty} E(|X_j|I_{\{\omega:\epsilon\sqrt{n}\leq|X_j|\}}) = 0.$$

And, since ϵ can be made arbitrarily small, $\lim_{n\to\infty}\frac{1}{n^2}\sum_{j=1}^{n} E(X_j^2 I_{\{\omega:|X_j|\leq n\}}) = 0$. Conse-quently, $\frac{S_n}{n} - E(X_1 I_{\{\omega:|X_1|\leq n\}}) \xrightarrow{p} 0$. Lastly, note that

$$\lim_{n\to\infty}\left(\int_\Omega X_1 dP - \int_\Omega X_1 I_{\{|X_1|\leq n\}} dP\right) = \int_\Omega X_1 dP - \lim_{n\to\infty}\int_\Omega X_1 I_{\{|X_1|\leq n\}} dP = E(X_1) - E(X_1) = 0$$

by the previous example. Hence,

$$\frac{S_n}{n} - E(X_1) = \frac{S_n}{n} + E(X_1 I_{\{|X_1|\leq n\}}) - E(X_1 I_{\{|X_1|\leq n\}}) - E(X_1) = o_p(1) + o(1) = o_p(1).$$

Example 8.3. *Suppose $\{X_n\}_{n\in\mathbb{N}}$ is an independent and identically distributed sequence with $\lim_{x\to\infty} xP(|X_1| > x) = 0$. For condition 1, given the identically distributed assumption, we have*

$\sum_{j=1}^{n} P(|X_j| > n) = nP(|X_j| > n) \to 0$ *by assumption. For condition 2, note that*

$$\frac{1}{n^2} \sum_{j=1}^{n} E(X_j^2 I_{\{\omega : |X_j| \leq n\}}) = \frac{1}{n} E(X_1^2 I_{\{\omega : |X_j| \leq n\}}) = \frac{1}{n} \int_{|x| \leq n} x^2 dF_{X_1}(x)$$

$$= \frac{2}{n} \int_{|x| \leq n} \left(\int_0^{|x|} s ds \right) dF_{X_1}(x) = \frac{2}{n} \int_0^n s \left(\int_{s < |x| \leq n} dF_{X_1}(x) \right) ds$$

$$= \frac{2}{n} \int_0^n s(P(|X_1| \leq n) - P(|X_1| < s)) ds$$

$$= \frac{2}{n} \int_0^n s(1 - P(|X_1| > n) - 1 + P(|X_1| \geq s)) ds$$

$$= \frac{2}{n} \int_0^n s(P(|X_1| \geq s) - P(|X_1| > n)) ds$$

$$= \frac{2}{n} \int_0^n \tau(s) ds - 2P(|X_1| > n) \frac{1}{n} \int_0^n s ds, \ \ where \ \tau(s) = sP(|X_1| > s)$$

$$= \frac{2}{n} \int_0^n \tau(s) ds - 2P(|X_1| > n) \frac{1}{n} \frac{n^2}{2}$$

$$= \frac{2}{n} \int_0^n \tau(s) ds - nP(|X_1| > n) = \frac{2}{n} \int_0^n \tau(s) ds - \tau(n).$$

Since, $\tau(n) \to 0$ as $n \to \infty$, we have that for all $\epsilon > 0$ there exists N_ϵ such that if $n > N_\epsilon$, $\tau(n) \leq \epsilon$. Consequently,

$$\frac{1}{n} \int_0^n \tau(s) ds = \frac{1}{n} \int_0^{N_\epsilon} \tau(s) ds + \frac{1}{n} \int_{N_\epsilon}^n \tau(s) ds \leq \frac{1}{n} \int_0^{N_\epsilon} \tau(s) ds + \epsilon.$$

Taking limits on both sides as $n \to \infty$ gives $\frac{1}{n} \int_0^n \tau(s) ds \to 0$. Then, $\frac{S_n}{n} - E(X_1 I_{|X_1| \leq n}) \xrightarrow{p} 0$.

If $\{X_j\}_{j \in \mathbb{N}}$ with $E(X_j) < \infty$, $E(X_j^2) < M < \infty$ for all j, we have that

$$\frac{1}{n} S_n - \frac{1}{n} E(S_n) = \frac{1}{n} \sum_{j=1}^{n} (X_j - E(X_j)) := \frac{1}{n} \sum_{j=1}^{n} Z_j$$

where $E(Z_j) = 0$. If $E(Z_i Z_j) = 0$ for all $i \neq j$, then

$$E\left(\left(\frac{1}{n} \sum_{j=1}^{n} Z_j \right)^2 \right) = \frac{1}{n^2} \sum_{j=1}^{n} E(Z_j^2) < \frac{M}{n} \to 0 \text{ as } n \to \infty.$$

Hence, $\frac{1}{n} \sum_{j=1}^{n} Z_j \xrightarrow{\mathcal{L}^2} 0$, and by Theorem 7.8 $\frac{1}{n} \sum_{j=1}^{n} Z_j \xrightarrow{p} 0$. In fact, $\frac{1}{n} \sum_{j=1}^{n} Z_j \xrightarrow{as} 0$ as shown in the next theorem.

Theorem 8.3. *Let* $\{X_j\}_{j\in\mathbb{N}}$ *with* $E(X_j) < \infty$, $E(X_j^2) < M < \infty$ *for all* j, *and assume that* $E((X_j - E(X_j))(X_i - E(X_i))) = 0$ *for all* $i \neq j$. *Then, letting* $Z_j = X_j - E(X_j)$ *and* $S_n = \sum_{j=1}^n Z_j$, *we have*

$$\frac{1}{n} S_n \xrightarrow{as} 0.$$

Proof. For all $\epsilon > 0$ and by Chebyshev's Inequality

$$P(|S_n| > n\epsilon) \leq \frac{M}{n\epsilon^2}.$$

Since $\sum_{n=1}^\infty \frac{1}{n}$ diverges we can't use the Borel-Cantelli Lemma directly. However, if we consider the subsequence S_{n^2}, we have

$$\sum_{n=1}^\infty P(\{\omega : |S_{n^2}| > n^2\epsilon\}) \leq \sum_{n=1}^\infty \frac{M}{n^2\epsilon^2} < \infty,$$

since $\sum_{n=1}^\infty \frac{1}{n^2} = \frac{\pi^2}{6}$. Hence, $P\left(\limsup_{n\to\infty}\{\omega : |S_{n^2}(\omega)| > n^2\epsilon\}\right) = 0$ and we have $\frac{S_{n^2}}{n^2} \xrightarrow{as} 0$. Now, let

$$D_n := \max_{n^2 \leq k < (n+1)^2} |S_k - S_{n^2}|$$

and note that

$$\frac{|S_k|}{k} \leq \frac{|S_k|}{n^2} = \frac{|S_k - S_{n^2} + S_{n^2}|}{n^2} \leq \frac{|S_k - S_{n^2}|}{n^2} + \frac{|S_{n^2}|}{n^2} \leq \frac{D_n}{n^2} + \frac{|S_{n^2}|}{n^2}.$$

Now, since $P(\max_{1\leq k\leq m} |W_k| \geq \epsilon) \leq \sum_{k=1}^m P(|W_k| \geq \epsilon)$ and using Markov's Inequality

$$P(D_n \geq n^2\epsilon) \leq \sum_{l=1}^{2n} P\left(\left|\sum_{j=1}^l Z_{n^2+j}\right| \geq n^2\epsilon\right) \leq \sum_{l=1}^{2n} \frac{1}{n^4\epsilon^2} E\left(\left(\sum_{j=1}^l Z_{n^2+j}\right)^2\right)$$

$$= \sum_{l=1}^{2n} \frac{1}{n^4\epsilon^2} \sum_{j=1}^l E(Z_{n^2+j}^2) \leq \frac{4n^2 M}{n^4\epsilon^2} = \frac{4M}{n^2\epsilon^2}.$$

Then, we have $\sum_{n=1}^\infty P(D_n \geq n^2\epsilon) \leq \frac{4M}{\epsilon^2} \sum_{n=1}^\infty \frac{1}{n^2} < \infty$, and by the Borel-Cantelli Lemma $\frac{D_n}{n^2} \xrightarrow{as} 0$. Since, as $n \to \infty$ we have that $k \to \infty$, $\frac{|S_k|}{k} \xrightarrow{as} 0$. ∎

We now state Markov's Law of Large Numbers.

Theorem 8.4. *(Markov's LLN) Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variables with $E(X_n) = \mu_n$. If for some $\delta > 0$ we have $\sum_{n=1}^{\infty} \frac{E|X_n - \mu_n|^{1+\delta}}{n^{1+\delta}} < \infty$, then*

$$\frac{1}{n}S_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i \overset{as}{\to} 0.$$

Proof. Chung (1974, A Course in Probability Theory, pp. 125-126). ∎

8.1 Exercises

1. Let U and V be two points in an n-dimensional unit cube, i.e., $[0,1]^n$ and X_n be the Euclidean distance between these two points which are chosen independently and uniformly. Show that $\frac{X_n}{\sqrt{n}} \overset{p}{\to} \frac{1}{\sqrt{6}}$.

2. Show that if $\{X_j\}_{j\in\mathbb{N}}$ be a sequence of random variables with $E(X_j) = 0$ and $\sum_{j=1}^{\infty} \frac{1}{a_j^p}E(|X_j|^p) < \infty$ for some $p \geq 1$ and a sequence of positive constants $\{a_j\}_{j\in\mathbb{N}}$. Then,

$$\sum_{j=1}^{\infty} P(|X_j| > a_j) < \infty \text{ and } \sum_{j=1}^{\infty}\frac{1}{a_j}|E(X_j I_{\{\omega:|X_j|\leq a_j\}})| < \infty.$$

Furthermore, for any $r \geq p$,

$$\sum_{j=1}^{\infty}\frac{1}{a_j^r}E(|X_j|^r I_{\{\omega:|X_j|\leq a_j\}}) < \infty.$$

Use this result to prove Theorem 8.4 in your class notes with convergence in probability.

3. Let $\{X_i\}_{i=2,3,\ldots}$ be a sequence of independent random variables such that

$$P(X_i = i) = P(X_i = -i) = \frac{1}{2i\log i}, \quad P(X_i = 0) = 1 - \frac{1}{i\log i}$$

Show that $\frac{1}{n}\sum_{i=2}^{n} X_i \overset{p}{\to} 0$.

Chapter 9

Conditional expectation

9.1 Inner product spaces

There are several ways to introduce the notion of conditional expectation. We begin by introducing inner-product spaces and motivate a definition of conditional expectation by using the Projection Theorem.

Definition 9.1. *A real vector space $\mathbb{X}$ is called an inner-product space if for all $x, y \in \mathbb{X}$, there exists a function $\langle x, y \rangle$, called an inner-product, such that for all $x, y, z \in \mathbb{X}$ and $a \in \mathbb{R}$[1]*

1. *$\langle x, y \rangle = \langle y, x \rangle$*

2. *$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$*

3. *$\langle ax, y \rangle = a \langle x, y \rangle$, $a \in \mathbb{R}$*

4. *$\langle x, x \rangle \geq 0$, for all x*

5. *$\langle x, x \rangle = 0 \iff x = \theta$, where θ is the null vector in $\mathbb{X}$.*

The following theorem shows that a general version of the Cauchy-Schwarz Inequality holds for inner-product spaces.

---

[1]If the vector space $\mathbb{X}$ is associated with a complex field, property 1 becomes $\langle x, y \rangle = \overline{\langle y, x \rangle}$, where for $x \in \mathbb{C}$, $\bar{x}$ is the complex conjugate of x, and in property 3 $a \in \mathbb{C}$.

Theorem 9.1. *Let* $\mathbb{X}$ *be an inner-product space and* $x, y \in \mathbb{X}$. *Then,*

$$|\langle x, y \rangle| \leq \langle x, x \rangle^{1/2} \langle y, y \rangle^{1/2}.$$

Proof. Let $y \neq \theta$ and note that for all $a \in \mathbb{R}$,

$$
\begin{aligned}
0 &\leq & \langle x - ay, x - ay \rangle = \langle x, x \rangle - 2a\langle x, y \rangle + a^2\langle y, y \rangle \\
&\leq & \langle x, x \rangle - \frac{\langle x, y \rangle^2}{\langle y, y \rangle} \text{ by letting } a = \langle x, y \rangle / \langle y, y \rangle.
\end{aligned}
$$

The last inequality is equivalent to $\langle x, y \rangle^2 \leq \langle x, x \rangle \langle y, y \rangle$ or $|\langle x, y \rangle| = \langle x, x \rangle^{1/2} \langle y, y \rangle^{1/2}$. Lastly, if $y = \theta$ then the inequality holds with equality and $\langle x, \theta \rangle = 0$. ∎

It can be easily shown that the function $\| \cdot \| : \mathbb{X} \to [0, \infty)$ defined as $\|x\| = \langle x, x \rangle^{1/2}$ is a norm on $\mathbb{X}$. Thus, every inner-product space can be taken to be a normed space with this induced norm. Another important property in inner-product spaces is the Parallelogram Law, which is given in the next theorem.

Theorem 9.2. *In an inner-product space* $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$.

Proof. $\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \langle y, y \rangle + 2\langle x, y \rangle$ and $\|x - y\|^2 = \langle x - y, x - y \rangle = \langle x, x \rangle + \langle y, y \rangle - 2\langle x, y \rangle$. Hence, we obtain

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

∎

Example 9.1. *Let* $x, y \in \mathbb{R}^n$ *and define* $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$. *It can be easily shown that* $\langle x, y \rangle$ *is an inner-product for* $\mathbb{R}^n$ *and* $\langle x, x \rangle^{1/2} = \|x\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$ *is a norm.*

Example 9.2. *Consider the space* $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ *of random variables* $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ *such that* $\int_{\Omega} X^2 dP < \infty$. *By Theorem* 5.10.1 $XY \in \mathcal{L}(\Omega, \mathcal{F}, P)$ *and by Theorem* 5.10.3

$\mathcal{L}^2(\Omega, \mathcal{F}, P)$ is a vector space. Now, define $\langle X, Y \rangle = E(XY) = \int_{\Omega} XY \, dP$. Using the properties of integrals, conditions 1-4 in Definition 9.1 are easily verified. However, condition 5 does not hold. Whereas it is true that $X(\omega) = 0$ for all ω, the null vector in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, gives $\langle X, X \rangle = \int_{\Omega} X^2(\omega) dP = 0$, $\int_{\Omega} X^2(\omega) dP = 0$ does not imply $X(\omega) = 0$ for all ω. This is true since a random variable Z that takes non-zero values in sets of measure zero and is equal to 0 elsewhere will be such that $\int_{\Omega} Z^2(\omega) dP = 0$. If we treat any two variables X and Z in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ as being identical if they differ only in a set of measure zero, that is if $P(\{\omega : X(\omega) \neq Z(\omega)\}) = 0$, then condition 5 is met and $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ is an inner product space with $\|X\|_2 = \left(\int_{\Omega} X^2 dP \right)^{1/2}$. We know from the Riez-Fisher Theorem that $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ is a Banach space, viz., a complete vector space. Hence, $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ is a Hilbert space.

Theorem 9.3. *Let $\{X_n\}_{n=1,2,\dots}$ and $\{Y_n\}_{n=1,2,\dots}$ be sequences in a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$. Let $X_n \to X$ in that $\|X_n - X\| \to 0$ as $n \to \infty$ and $Y_n \to Y$. Then, $\langle X_n, Y_n \rangle \to \langle X, Y \rangle$.*

Proof. By the Cauchy-Schwarz inequality (Theorem 9.1), $|\langle X, Y \rangle| \leq \|X\| \|Y\|$. Therefore,

$$|\langle X, Y \rangle - \langle X_n, Y_n \rangle| = |\langle X, Y_n \rangle - \langle X_n, Y_n \rangle + \langle X, Y \rangle - \langle X, Y_n \rangle - \langle X_n, Y \rangle + \langle X_n, Y_n \rangle$$

$$+ \langle X_n, Y \rangle - \langle X_n, Y_n \rangle|$$

$$= |\langle X - X_n, Y_n \rangle + \langle X - X_n, Y - Y_n \rangle + \langle X_n, Y - Y_n \rangle|$$

$$\leq |\langle X - X_n, Y_n \rangle| + |\langle X - X_n, Y - Y_n \rangle| + |\langle X_n, Y - Y_n \rangle|$$

$$\leq \|X - X_n\| \|Y_n\| + \|X - X_n\| \|Y - Y_n\| + \|X_n\| \|Y - Y_n\|.$$

By convergence, $\|X - X_n\|, \|Y - Y_n\| \to 0$ and since $\|X_n\|, \|Y_n\| < \infty$ for all n, $|\langle X, Y \rangle - \langle X_n, Y_n \rangle| \to 0$, as $n \to \infty$. $\blacksquare$

Definition 9.2. *Let S be a closed subset of a Hilbert space $\mathcal{H}$. The distance from $Y \in \mathcal{H}$ to S is denoted by*

$$d(Y, S) = \inf\{\|Y - X\| : X \in S\}.$$

If $Y \in S$, $d(Y, S) = 0$.

Theorem 9.4. *(Projection Theorem): Let S be a closed subspace of a Hilbert space $\mathcal{H}$ and $Y \in \mathcal{H}$. There exists a unique $X \in S$ such that $\|Y - X\| := \inf\{\|Y - X'\| : X' \in S\}$. Furthermore, $\langle Y - X, s \rangle = 0$, for all $s \in S$.*

Proof. First, consider existence of X. If $Y \in S$, put $X = Y$. If $Y \notin S$, we would like to obtain $X \in S$ such that $\|Y - X\| = \inf_{X' \in S}\{\|Y - X'\|\} = \delta > 0$.

Let $\{X_i\}_{i \in \mathbb{N}} \in S$ such that $\|X_i - Y\| \to \delta$. Now, if X_i and Y are in a Hilbert space, we have by the Parallelogram Law

$$\|(X_j - Y) + (Y - X_i)\|^2 + \|(X_j - Y) - (Y - X_i)\|^2 = 2\|X_j - Y\|^2 + 2\|Y - X_i\|^2$$

and

$$\|X_j - X_i\|^2 = 2\|X_j - Y\|^2 + 2\|Y - X_i\|^2 - 4\|Y - \frac{X_i + X_j}{2}\|^2.$$

For all i, j the vector $\frac{X_i + X_j}{2} \in S$ (since S is a subspace). Therefore, by definition of δ, $\|Y - \frac{X_i + X_j}{2}\| \geq \delta$ and we obtain $\|X_j - X_i\|^2 \leq 2\|X_j - Y\|^2 + 2\|Y - X_i\|^2 - 4\delta^2$. Since $\|X_i - Y\|^2 \to \delta^2$ by continuity of inner product (Theorem 9.3), $\|X_j - X_i\|^2 \to 0$ as $i, j \to \infty$. Hence, $\{X_i\}$ is a Cauchy sequence. Since S is closed, $\{X_i\}$ converges to $\tilde{X} \in S$. Furthermore, $\delta \leq \|Y - \tilde{X}\| \leq \|Y - X_i\| + \|X_i - \tilde{X}\| \leq \delta$. Hence, $\tilde{X} = X$ which we wanted to show existed.

Now, consider the proof of $\langle Y - X, s \rangle = 0$ for all $s \in S$. Suppose there exists $s \in S$ such that $\langle Y - X, s \rangle \neq 0$. Without loss of generality assume that $\|s\| = 1$ and that $\langle Y - X, s \rangle = \delta \neq 0$ and define $s_1 \in S$ such that $s_1 = X + \delta s$. Then,

$$\|Y - s_1\|^2 = \|Y - X - \delta s\|^2 \text{by definition of } s_1$$
$$= \|Y - X\|^2 - \langle Y - X, \delta s \rangle - \langle \delta s, Y - X \rangle + \delta^2 \|s\|^2$$
$$= \|Y - X\|^2 - \delta^2 - \delta^2 + \delta^2$$
$$= \|Y - X\|^2 - \delta^2 < \|Y - X\|^2$$

150

Hence, if $\langle Y - X, s\rangle \neq 0$, then X is not the minimizing element of S and it must be that for all $s \in S$, $\langle Y - X, s\rangle = 0$.

Lastly, let's prove uniqueness. For all $s \in S$, the theorem of Pythagoras says that $\|Y - s\|^2 = \|Y - X + X - s\|^2 = \|Y - X\|^2 + \|X - s\|^2$. (Note that $\langle Y - X, X - s\rangle = 0$ due to the fact that $\langle Y - X, s\rangle = 0$, $\forall s \in S$). Hence, $\|Y - s\| > \|Y - X\|$ for $s \neq X$. $\blacksquare$

As a matter of terminology, we call any two elements X and Y of a Hilbert space orthogonal if $\langle X, Y\rangle = 0$.

9.2 Conditional expectation for random variables in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$

Now consider the Hilbert space $\mathcal{L}^2$ composed of all random variables defined on $(\Omega, \mathcal{F}, P)$ and for precision denote this space by $\mathcal{L}^2(\Omega, \mathcal{F}, P)$. Let X be a random vector taking values in $\mathbb{R}^n$ defined in the same probability space with $\sigma(X) \subset \mathcal{F}$. Then, $\mathcal{L}^2(\Omega, \sigma(X), P) \subset \mathcal{L}^2(\Omega, \mathcal{F}, P)$ is a Hilbert space with the same inner product. Furthermore, $\mathcal{L}^2(\Omega, \sigma(X), P)$ is a closed subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, P)$. We now define conditional expectation.

Definition 9.3. *Let $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. The conditional expectation of Y given X is the unique element $\hat{Y} \in \mathcal{L}^2(\Omega, \sigma(X), P)$ such that*

$$E((Y - \hat{Y})s) = 0, \ \text{for all } s \in \mathcal{L}^2(\Omega, \sigma(X), P).$$

We write $\hat{Y} = E(Y|X)$ or $\hat{Y} = E(Y|\sigma(X))$.

Recall that if $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^n, \mathcal{B}^n)$ is a random vector, then $X^{-1}(\mathcal{B}^n) \subset \mathcal{F}$ is a σ-algebra and we wrote $X^{-1}(\mathcal{B}^n) = \sigma(X)$, the σ-algebra generated by X. Consider a random variable $Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$. It is legitimate to ask when Y is measurable (a random variable) with respect to $\sigma(X)$.[2] The following theorem provides a useful characterization.

---

[2] More generally, for $\mathcal{G} \subset \mathcal{F}$ a σ-algebra, we say that X is $\mathcal{G}$-measurable if for all $B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{G}$. There may be many of these $\mathcal{G}$'s. The intersection of all of them, i.e. $\sigma(X) := \cap_{i \in I} \mathcal{G}_i$ is called the σ-algebra generated by X.

Theorem 9.5. *Let $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^n, \mathcal{B}^n)$ be a random vector and $Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ be a random variable. Y is $\sigma(X)$-measurable if, and only if, there exists $f : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}, \mathcal{B})$ such that $Y = f(X)$ and f is $\mathcal{B}^n$-measurable.*

Proof. ($\Longleftarrow$) We want to show that for every $B \in \mathcal{B}$ we have $Y^{-1}(B) \in \sigma(X)$. But $Y^{-1}(B) = X^{-1}(f^{-1}(B))$ and by measurability of f, $f^{-1}(B) \in \mathcal{B}^n$ and since X is a random vector $X^{-1}(f^{-1}(B)) \in \sigma(X)$. Thus, Y is $\sigma(X)$-measurable.

($\Longrightarrow$) Suppose $Y^{-1}(B) \in \sigma(X)$ for all $B \in \mathcal{B}$. First, assume that Y is simple. Then, for $k \in \mathbb{N}$ we have $Y = \sum_{i=1}^{k} a_i I_{A_i}$ for a_i distinct and A_i pairwise-disjoint. In this case, $Y^{-1}(\{a_i\}) = A_i$ and by assumption $A_i \in \sigma(X)$. Hence there exists $B_i \in \mathcal{B}^n$ such that $X^{-1}(B_i) = A_i$ (definition of $\sigma(X)$). Let $f(x) = \sum_{i=1}^{k} a_i I_{B_i}(x)$, then $Y = f(X)$, f $\mathcal{B}^n$-measurable. Thus, the implication is proved for every Y simple that is $\sigma(X)$-measurable.

If $Y : (\Omega, \mathcal{F}, P) \to [0, \infty)$ then, by Theorem 4.4, there exist $Y_n(\omega)$ simple such that

$$Y(\omega) = \lim_{n \to \infty} Y_n(\omega), \ 0 \le Y_n(\omega) \le Y_{n+1}(\omega).$$

Each Y_n is $\sigma(X)$-measurable and $Y_n = f_n(X)$ from the first part of the proof. Now, set $f(x) = \limsup_{n \to \infty} f_n(x)$ and note $Y = \lim_{n \to \infty} Y_n = \lim_{n \to \infty} f_n(X)$.

Given that $(\limsup_{n \to \infty} f_n)(X) = \limsup_{n \to \infty} f_n(X)$, by Theorem 3.6, $f(x)$ is $\mathcal{B}^n$-measurable. For general Y, write $Y = Y^+ - Y^-$ which reduces to the preceding case. ∎

Remark 9.1.   *1. An equivalent way to think of Definition 9.3 using the previous theorem is to write*

$$E(Y|X) = \underset{s \in \mathcal{L}^2(\Omega, \sigma(X), P)}{\arg\inf} \|Y - s\| = \arg \inf_{f \in F} \|Y - f(X)\|.$$

where F is the set of Borel measurable functions from $\mathbb{R}^n$ to $\mathbb{R}$.

2. Since $\hat{Y} = E(Y|X)$ is $\sigma(X)$-measurable, by Theorem 9.5, there exists $f : \mathbb{R}^n \to \mathbb{R}$ which is Borel measurable such that $E(Y|X) = f(X)$ and f is unique. Hence, we

can write $E[(Y - f(X))g(X)] = 0$, for all $g : \mathbb{R}^n \to \mathbb{R}$ Borel measurable such that $\int g^2 dP < \infty$.

We can free the concept of conditional expectation from a particular set of random variables (or element) that produces $\sigma(X)$ and speak more generally of conditioning on a σ-algebra $\mathcal{G} \subset \mathcal{F}$, that is a sub-$\sigma$-algebra of $\mathcal{F}$.

Definition 9.4. $Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ be a random variable with $\int Y^2 dP < \infty$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then $E(Y|\mathcal{G})$ is the unique $\hat{Y} \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$ such that

$$E((Y - \hat{Y})s) = E([Y - E(Y|\mathcal{G})]s) = 0,$$

for all measurable $s \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$.

Remark 9.2.   1. The definition gives $E(Ys) = E(sE(Y|\mathcal{G}))$.

     2. Since $s = 1 \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$, $E(Y) = E(E(Y|\mathcal{G}))$.

     3. If $U, V \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$, then $E(U + \alpha V|\mathcal{G})$ satisfies $E((U + \alpha V)s) = E(E(U + \alpha V|\mathcal{G})s)$. But,

$$E((U + \alpha V)s) = E(Us) + \alpha E(Vs)$$
$$= E(E(U|\mathcal{G})s) + \alpha E(E(V|\mathcal{G})s)$$
$$= E([E(U|\mathcal{G}) + \alpha E(V|\mathcal{G})]s).$$

Hence, $E(U + \alpha V|\mathcal{G}) = E(U|\mathcal{G}) + \alpha E(V|\mathcal{G})$. That is $E(\cdot|\mathcal{G})$ is a linear function.

Theorem 9.6. Assume that $Z := \begin{pmatrix} Y \\ X \end{pmatrix}$ is a random vector defined on $(\Omega, \mathcal{F}, P)$ taking values in $\mathbb{R}^2$ and having density f.

     1. Y and X have densities on $(\mathbb{R}, \mathcal{B})$ given by $f_Y(y) = \int_{\mathbb{R}} f(y, x)d\lambda(x)$ and $f_X(x) = \int_{\mathbb{R}} f(y, x)d\lambda(y)$.

2. For every $x \in \mathbb{R}$ such that $f_X(x) \neq 0$ we have that $f_{Y|X=x}(y) = \frac{f(y,x)}{f_X(x)}$ is a density on $\mathbb{R}$.

3. $E(Y|X) = h(X)$ where $h(x) = \int_{\mathbb{R}} y f_{Y|X=x}(y) d\lambda(y)$.

Proof. 1. Let $E \in \mathcal{B}$. Then,

$$P(Y \in E) = P(Z \in E \times \mathbb{R}) = \int_{E \times \mathbb{R}} f(y,x) d\lambda^2(y,x)$$

$$= \int_E \int_{\mathbb{R}} f(y,x) d\lambda(y) d\lambda(x) = \int_E f_Y(y) d\lambda(y)$$

with $f_Y(y) = \int_{\mathbb{R}} f(y,x) d\lambda(x)$. Therefore, $P(Y \in E) = \int_{\mathbb{R}} I_E f_Y(y) d\lambda(y)$ and f_Y is a density for Y.

2. $\int_{\mathbb{R}} f_{Y|X=x}(y) d\lambda(y) = \int_{\mathbb{R}} \frac{f(y,x)}{f_X(x)} d\lambda(y) = 1$.

3. Let $h(x) = \int_{\mathbb{R}} y f_{Y|X=x}(y) d\lambda(y)$ and consider any bounded Borel measurable function $g : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$. Then,

$$E(h(X)g(X)) = \int_{\mathbb{R}} h(x)g(x)f_X(x) d\lambda(x) = \int_{\mathbb{R}} \int_{\mathbb{R}} y f_{Y|X=x}(y) d\lambda(y) g(x) f_X(x) d\lambda(x)$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} y \frac{f(y,x)}{f_X(x)} d\lambda(y) g(x) f_X(x) d\lambda(x) = \int_{\mathbb{R}} \int_{\mathbb{R}} y f(y,x) d\lambda(y) g(x) d\lambda(x)$$

$$= E(Yg(X))$$

Consequently,

$$E(h(X)g(X)) - E(Yg(X)) = E((Y - h(X))g(X)) = 0$$

which gives $E(Y|X) = h(X)$. ∎

Theorem 9.7. *Let Y be a random variable in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ and S be a closed subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, P)$. Then,*

1. *there exists a unique function $P_S : \mathcal{L}^2(\Omega, \mathcal{F}, P) \to S$ such that $(\mathcal{I} - P_S) : \mathcal{L}^2(\Omega, \mathcal{F}, P) \to S^\perp$ where $S^\perp$ is the orthogonal complement of S.[3]*

---

[3]The orthogonal complement of a subset S of an inner-product space is the set of all vectors in the space that are orthogonal to S.

2. $\|Y\|^2 = \|P_S(Y)\|^2 + \|(I - P_S)(Y)\|^2$,

3. $P_S(Y_n) \rightarrow P_S(Y)$ if $\|Y_n - Y\| \rightarrow 0$ as $n \rightarrow \infty$,

4. if S_1, S_2 are closed subspaces of $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ such that $S_1 \subset S_2 \implies P_{S_1}(P_{S_2}(Y)) = P_{S_1}(Y)$.

Proof. 1. By the Projection Theorem, for each $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ there exists a unique $\hat{Y} \in S$. Thus, we write the function $P_S(Y) = \hat{Y}$. In addition $E\{(Y - P_S(Y))s\} = 0$ for all $s \in S$. That is, $Y - P_S(Y)$ is orthogonal to the subspace S. Any $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ can be written as $Y - P_S(Y) + P_S(Y) = Y$ or $Y = (\mathcal{I} - P_S)(Y) + P_S(Y)$ where $\mathcal{I}$ is the identity operator in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ and $\mathcal{I} - P_S$ projects Y onto the orthogonal complement of S.

2. Note that

$$\|Y\|^2 = \|Y - P_S Y + P_S Y\|^2$$
$$= \|Y - P_S(Y)\|^2 + \|P_S(Y)\|^2 \text{ by Pythagoras' theorem}$$
$$= \|(\mathcal{I} - P_S)(Y)\|^2 + \|P_S(Y)\|^2.$$

3. Note that $\|P_S(Y_n) - P_S(Y)\|^2 = \|P_S(Y_n - Y)\|^2$. By the last equality in part 2.,

$$\|Y_n - Y\|^2 = \|(\mathcal{I} - P_S)(Y_n - Y)\|^2 + \|P_S(Y_n - Y)\|^2$$
$$= \|(\mathcal{I} - P_S)(Y_n - Y)\|^2 + \|P_S(Y_n) - P_S(Y)\|^2.$$

Consequently,

$$\|P_S(Y_n) - P_S(Y)\|^2 = \|Y_n - Y\|^2 - \|(\mathcal{I} - P_S)(Y_n - Y)\|^2 \leq \|Y_n - Y\|^2.$$

Hence, if $\|Y_n - Y\| \rightarrow 0$ as $n \rightarrow \infty$, then $\|P_S(Y_n) - P_S(Y)\|^2 \rightarrow 0$ as $n \rightarrow \infty$.

4. $Y = P_{S_2}(Y) + (\mathcal{I} - P_{S_2})(Y)$ and $P_{S_1}(Y) = P_{S_1}(P_{S_2}(Y)) + P_{S_1}((\mathcal{I} - P_{S_2})(Y))$. In the last term, the argument of P_{S_1} is an element of the orthogonal complement of S_2. That is $< (\mathcal{I} - P_{S_2})(Y), s >= 0$ for every $s \in S_2$. But since $S_1 \subset S_2$, it must be that $< (\mathcal{I} - P_{S_2})(Y), s_1 >= 0$ for all $s_1 \in S_1$. Thus, $(\mathcal{I} - P_{S_2})(Y) \in S_1^{\perp}$ and consequently $P_{S_1}((\mathcal{I} - P_{S_2})(Y)) = 0$. ∎

In Theorem 9.7, if we take the closed subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ to be $\mathcal{L}^2(\Omega, \mathcal{G}, P)$ for $\mathcal{G}$ a sub σ-algebra of $\mathcal{F}$, we write $E(Y|\mathcal{G})$ for $P_S(Y)$. In particular, we have:

1. $\|Y\|^2 = \|E(Y|\mathcal{G})\|^2 + \|Y - E(Y|\mathcal{G})\|^2$,

2. $E(Y_n|\mathcal{G}) \to E(Y|\mathcal{G})$ if $Y_n \xrightarrow{\mathcal{L}^2} Y$,

3. if $\mathcal{H} \subset \mathcal{G}$ then $E(E(Y|\mathcal{G})|\mathcal{H}) = E(Y|\mathcal{H})$.

9.3 Conditional expectation for random variables in $\mathcal{L}(\Omega, \mathcal{F}, P)$

It is desirable to extend the concept of conditional expectation to random variables $Y :$ $(\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ such that $Y \in \mathcal{L}$. The word extend is justified, since by the Cauchy-Schwarz Inequality (or Rogers-Hölder Inequality with $p = q = 2$)

$$E(|XY|) \leq \|X\|_2 \|Y\|_2.$$

Taking $Y = 1$ almost everywhere, we have $E(|X|)^2 \leq E(X^2)$. Hence, if $E(X^2) < C$ then $E|X| < C$. Consequently, $\mathcal{L}^2 \subset \mathcal{L}$.

For this purpose, recall that $Y \in \mathcal{L}(\Omega, \mathcal{F}, P)$ if $Y^+ = \max\{Y(\omega), 0\}$ and $Y^- = -\min\{Y(\omega), 0\}$ are such that $E(Y^+)$, $E(Y^-) < \infty$ and, in this case, we define $E(Y) = E(Y^+) - E(Y^-)$. If $Y \geq 0$, then $Y^- = 0$ and $Y = Y^+$. We first consider $Y \in \mathcal{L}_+(\Omega, \mathcal{F}, P)$. As in Definition 4.4 we allow $Y(\omega) = \infty$. The next theorem provides the basis for extending our definition of conditional expectation to random variables in $\mathcal{L}$.

Theorem 9.8. *i) Let $Y \in \mathcal{L}_+(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. There exists a unique element $E(Y|\mathcal{G})$ of $\mathcal{L}_+(\Omega, \mathcal{G}, P)$ such that $E([Y - E(Y|\mathcal{G})]X) = 0$ for all $X \in \mathcal{L}_+(\Omega, \mathcal{G}, P)$.*

ii) If $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ then the conditional expectation $E(Y|\mathcal{G})$ in i) is the same as $E(Y|\mathcal{G})$ in Definition 9.3 with $\sigma(X) = \mathcal{G}$.

iii) If $Y \leq Y'$ then $E(Y|\mathcal{G}) \leq E(Y'|\mathcal{G})$.

Proof. i) We first consider the existence $E(Y|\mathcal{G})$. Let $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ and $Y \geq 0$. In this case, define $E(Y|\mathcal{G})$ as in Definition 9.3. Now, for $X \in \mathcal{L}_+(\Omega, \mathcal{G}, P)$ let

$$X_n(\omega) = \min\{X(\omega), n\} = \begin{cases} X(\omega), & \text{if } X(\omega) \leq n, \\ n, & \text{if } X(\omega) > n, \end{cases}$$

and note that

$$X_n^2(\omega) = \begin{cases} X^2(\omega), & \text{if } X(\omega) \leq n \\ n^2, & \text{if } X(\omega) > n \end{cases}.$$

Hence,

$$\int_\Omega X_n^2 dP = \begin{cases} \int_\Omega X^2 dP \leq n^2 \int_\Omega dP = n^2 < \infty, & \text{if } X(\omega) \leq n \\ n^2 \int_\Omega dP = n^2 < \infty, & \text{if } X(\omega) > n \end{cases}$$

so that $X_n \in \mathcal{L}^2$.

Now, $0 \leq X_1(\omega) \leq X_2(\omega) \leq \cdots \leq X(\omega)$ and $X_n(\omega) \to X(\omega)$ almost everywhere as $n \to \infty$. Then, by Beppo-Levi's Theorem, we have that

$$E\left(\lim_{n\to\infty} Y X_n\right) = E(YX) = \lim_{n\to\infty} E(Y X_n) = \lim_{n\to\infty} E(E(Y|\mathcal{G}) X_n).$$

The last equality follows from the fact that $EY^2 < \infty$, $EX_n^2 < \infty$ and Definition 9.3. Now, again by Beppo-Levi's Theorem, we have

$$E(YX) = \lim_{n\to\infty} E(E(Y|\mathcal{G}) X_n) = E(E(Y|\mathcal{G}) X), \text{ for all } X \in \mathcal{L}_+(\Omega, \mathcal{G}, P).$$

If $Y \in \mathcal{L}_+(\Omega, \mathcal{F}, P)$ then let $Y_m(\omega) = \min\{Y(\omega), m\}$ and from the argument above $Y_m \in \mathcal{L}^2$. Hence,

$$\lim_{n\to\infty} E(Y_m X_n) = \lim_{n\to\infty} E(E(Y_m|\mathcal{G}) X_n) = E(E(Y_m|\mathcal{G}) \lim_{n\to\infty} X_n)$$
$$= E(E(Y_m|\mathcal{G}) X).$$

Now, since $Y_m \geq 0$, then $E(Y_m|\mathcal{G})$ as defined in Definition 9.3 is such that $E(Y_m|\mathcal{G}) \geq 0$. To see this, consider $Z = I_{\{E(Y_m|\mathcal{G})<0\}}$ and note that $E(Z^2) = P(E(Y_m|\mathcal{G}) < 0)$, $E(Y_m Z) = $

157

$E(E(Y_m|\mathcal{G})Z) = E(E(Y_m|\mathcal{G})I_{\{E(Y_m|\mathcal{G})<0\}})$. Now, since $Y_m \geq 0$ and $Z = 1$ or $Z = 0$ we have

that $E(Y_m Z) \geq 0$. But the right-hand side of the last equality is less than 0 if $E(Y_m|\mathcal{G}) < 0$,

so it must be that $E(Y_m|\mathcal{G}) \geq 0$ if $Y_m \geq 0$. Hence, $E(Y_m|\mathcal{G})$ is increasing with m, and by

Beppo-Levi's Theorem we have

$$\lim_{m \to \infty} \lim_{n \to \infty} E(Y_m X_n) = E(YX) = \lim_{m \to \infty} E\left(E(Y_m|\mathcal{G})X\right) = E\left(\lim_{m \to \infty} E(Y_m|\mathcal{G})X\right).$$

Now, since $E(YX) = E\left(\left(\lim_{m \to \infty} E(Y_m|\mathcal{G})\right)X\right)$ or $E\left(\left(Y - \lim_{m \to \infty} E(Y_m|\mathcal{G})\right)X\right) = 0$ for all

$X \in \mathcal{L}_+(\Omega, \mathcal{G}, P)$, we define

$$E(Y|\mathcal{G}) = \lim_{m \to \infty} E(Y_m|\mathcal{G}) \tag{9.1}$$

for $Y \in \mathcal{L}^+(\Omega, \mathcal{F}, P)$.

We now consider uniqueness of $E(Y|\mathcal{G})$. Let U and V be two versions of $E(Y|\mathcal{G})$ and let

$\wedge_n = \{\omega : U < V \leq n\}$. Since U and V are versions of $E(Y|\mathcal{G})$ we know that U and V are

$\mathcal{G}$-measurable. Consequently, $\{\omega : U \leq n\} \in \mathcal{G}$, $\{\omega : V \leq n\} \in \mathcal{G}$ and $\wedge_n = \{\omega : U < V \leq n\} \in \mathcal{G}$.

Note that $E(YI_{\wedge_n}) = E(UI_{\wedge_n}) = E(VI_{\wedge_n})$ since $U = V = E(Y|\mathcal{G})$. Furthermore,

$0 \leq UI_{\wedge_n} \leq VI_{\wedge_n} \leq n$ and if $P(\wedge_n) > 0$ $(\wedge_n \neq \emptyset)$, $UI_{\wedge_n} < VI_{\wedge_n}$ which implies that

$E(UI_{\wedge_n}) < E(VI_{\wedge_n})$, which contradicts $E(UI_{\wedge_n}) = E(VI_{\wedge_n})$. Therefore, $P(\wedge_n) = 0$ for

all n. Now, note that $\wedge_1 \subset \wedge_2 \subset \wedge_3 \subset \cdots \subset \{U < V\}$. Now $\lim_{n \to \infty} \cup_{i=1}^n \wedge_i = \{U < V\}$

and $P\left(\lim_{n \to \infty} \cup_{i=1}^n \wedge_i\right) = \lim_{n \to \infty} P\left(\cup_{i=1}^n \wedge_i\right) \leq \lim_{n \to \infty} \sum_{i=1}^n P(\wedge_i)$. Thus, $P(\{U < V\}) = 0$.

Repeating the argument for $\Gamma_n = \{\omega : V < U \leq n\}$ we conclude that $P(\{V < U\}) = 0$.

Hence, it must be that U and V coincide with probability 1.

ii) The proof follows from the first part of the argument in item i).

iii) If $Y \leq Y'$ then $Y_m \leq Y'_m$ for all m and $E(Y_m|\mathcal{G}) \leq E(Y'_m|\mathcal{G})$ and consequently

$$\lim_{m \to \infty} E(Y_m|\mathcal{G}) \leq \lim_{m \to \infty} E(Y'_m|\mathcal{G}) \iff E(Y|\mathcal{G}) \leq E(Y'|\mathcal{G}).$$

∎

We now consider conditional expectations for random variables in $\mathcal{L}(\Omega, \mathcal{F}, P)$.

Theorem 9.9. *Let $Y \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. There exists a unique element $E(Y|\mathcal{G})$ in $\mathcal{L}(\Omega, \mathcal{G}, P)$ such that*

$$E\left((Y - E(Y|\mathcal{G}))\,X\right) = 0, \text{ for all bounded } \mathcal{G}\text{-measurable } X.$$

$E(Y|\mathcal{G})$ coincides with those in Definition 9.3 and Theorem 9.8 when $Y \in \mathcal{L}^2$ and $Y \in \mathcal{L}_+$. In addition, (i) if $Y \geq 0$, then $E(Y|\mathcal{G}) \geq 0$ and (ii) $E(Y|\mathcal{G})$ is a linear in Y.

Proof. We first consider existence of the conditional expectation. Since $Y \in \mathcal{L}$, we can write $Y = Y^+ - Y^-$ and $Y^+, Y^- \in \mathcal{L}$. Now, Y^+ and Y^- are such that

$$E\left((Y^+ - E(Y^+|\mathcal{G}))X\right) = 0, \text{ for all } X \in \mathcal{L}_+(\Omega, \mathcal{G}, P) \text{ and}$$

$$E\left((Y^- - E(Y^-|\mathcal{G}))X\right) = 0, \text{ for all } X \in \mathcal{L}_+(\Omega, \mathcal{G}, P).$$

Define $E(Y|\mathcal{G}) = E(Y^+|\mathcal{G}) - E(Y^-|\mathcal{G})$ and note that for $X \in \mathcal{L}_+(\Omega, \mathcal{G}, P)$

$$
\begin{aligned}
E(YX) = E((Y^+ - Y^-)X) &= E(Y^+X) - E(Y^-X) \\
&= E(E(Y^+|\mathcal{G})X) - E(E(Y^-|\mathcal{G})X) \text{ by Theorem 9.8} \\
&= E((E(Y^+|\mathcal{G}) - E(Y^-|\mathcal{G})))X) = E(E(Y|\mathcal{G})X).
\end{aligned}
$$

We now establish uniqueness of $E(Y|\mathcal{G})$. Suppose U and V are two versions of $E(Y|\mathcal{G})$ and let $\wedge = \{U < V\}$. Then, since U and V are $\mathcal{G}$-measurable, then $\wedge \in \mathcal{G}$. Therefore $I_\wedge$ is $\mathcal{G}$-measurable.

$$E(YI_\wedge) = E(E(Y|\mathcal{G})I_\wedge) = E(UI_\wedge) = E(VI_\wedge).$$

But, if $P(\wedge) > 0$, then $E(UI_\wedge) < E(VI_\wedge)$, a contradiction. Thus, $P(\wedge) = 0$. A similar reverse argument gives $P(V < U) = 0$.

Now, for any X that is bounded and $\mathcal{G}$-measurable consider

$$E(YX) = E(Y(X^+ - X^-)) = E(YX^+) - E(YX^-)$$

$$= E(X^+ E(Y|\mathcal{G})) - E(X^- E(Y|\mathcal{G}))$$

using the definition of conditional expectation in this proof.

$$= E((X^+ - X^-)E(Y|\mathcal{G})) = E(XE(Y|\mathcal{G})).$$

The proofs of items (i) and (ii) are left as exercises. ∎

Remark 9.3. *Note that if X and Y are independent random variables defined on the same probability space, then by Theorem 6.6, if f is a bounded measurable function $E(Yf(X)) = E(Y)E(f(X))$. Now, $E(Yf(X)) = E(E(Y|\sigma(X))f(X))$ and consequently*

$$E(Y)E(f(X)) = E(E(Y|\sigma(X))f(X)),$$

taking $f(X) = E(Y)$ gives $E(Y) = E(Y|\sigma(X))$.

Lebesgue's monotone and dominated convergence theorems hold for conditional expectations.

Theorem 9.10. *$Y_n(\omega) : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.*

  a) If $Y_n \geq 0$, $Y_1 \leq Y_2 \leq Y_3 \leq \cdots$ with $Y_n \overset{as}{\to} Y$ as $n \to \infty$, then $\lim_{n\to\infty} E(Y_n|\mathcal{G}) = E(Y|\mathcal{G})$ a.s.

  b) If $Y_n \overset{as}{\to} Y$ and $|Y_n| \leq Z$ for some $Z \in \mathcal{L}(\Omega, \mathcal{F}, P)$, then $\lim_{n\to\infty} E(Y_n|\mathcal{G}) = E(Y|\mathcal{G})$ a.s.

Proof. Left as an exercise. ∎

We now give an example where conditional expectation is taken to belong to a specific class of measurable functions.

Example 9.3. *Let $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ and let X be a random vector defined on the same probability space. Assume that for every component of X_k, for $k = 1, \cdots, K$ of X we have $X_k \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$. Now, consider the following class of functions*

$$F = \{f : f(x) = \sum_{k=1}^{K} a_k x_k \text{ where } f \text{ is } \sigma(X)\text{-measurable and } a_k \in \mathbb{R}\}.$$

Using Definition 9.3 or item 1 in Remark 30

$$E(Y|X) = \underset{a_1, \cdots, a_K}{argmin} \int \left(Y - \sum_{k=1}^{k} a_k X_k\right)^2 dP = \underset{a_1, \cdots, a_K}{argmin} O(a_1, \cdots, a_K).$$

Now,

$$O(a_1, \cdots, a_K) = \int (Y^2 - 2Y \sum_{k=1}^{K} a_k X_k + (\sum_{i=1}^{K} a_k X_k)^2) dP$$

$$= \int Y^2 dP - 2 \sum_{k=1}^{K} a_k \int X_k Y dP + \sum_{k=1}^{K} a_k^2 \int X_k^2 dP$$

$$+ \sum_{k=1}^{K} \sum_{k \neq l} a_k a_l \int X_k X_l dP$$

$$= \sigma^2 - 2 \sum_{k=1}^{K} a_k E(X_k Y) + \sum_{k=1}^{K} a_k^2 \int X_k^2 dP + \sum_{k=1}^{K} \sum_{jk \neq l} a_k a_l E(X_k X_l).$$

Now, taking derivatives with respect to a_k we have $\frac{\partial}{\partial a_k} O(a_1, \cdots, a_K) = -2E(X_k Y) + 2a_k E(X_k^2) + 2 \sum_{k \neq l} a_l E(X_k X_l)$ for $k = 1, \cdots, K$. Alternatively, using matrices

$$\frac{\partial}{\partial a} O(a_1, \cdots, a_K) = -2 \begin{bmatrix} E(X_1 Y) \\ \vdots \\ E(X_K Y) \end{bmatrix} + 2 \begin{bmatrix} E(X_1^2) & E(X_1 X_2) & \cdots & E(X_1 X_K) \\ E(X_2 X_1) & E(X_2^2) & \cdots & E(X_2 X_K) \\ \vdots & & & \vdots \\ E(X_K X_1) & E(X_K X_2) & \cdots & E(X_K^2) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_K \end{bmatrix}$$

$$= -2b + 2Aa$$

Choosing $a := \hat{a}$ such that $\frac{\partial}{\partial a} O(\hat{a}_1, \cdots, \hat{a}_K) = 0$ we have $\hat{a} = A^{-1}b$ if A is invertible. Invertibility of A follows positive definiteness of A, which also assures that $\hat{f}(x) = \sum_{k=1}^{K} \hat{a}_k x_k$ corresponds to a minimum.

9.4 Exercises

1. Assess the veracity of the following statement: "Since knowledge of X implies knowledge of $f(X)$, conditioning on X is the same as conditioning on $f(X)$. Hence, $E(Y|f(X)) = E(Y|X)$." Explain using mathematical arguments.

2. Let X and Y be independent random variables defined in the same probability space. Show that if $E(|Y|) < \infty$ then

$$P\left(E(Y|X) = E(Y)\right) = 1.$$

3. Let $(\Omega, \mathcal{F}, P)$ be a probability space. The set of random variables $X : \Omega \to \mathbb{R}$ such that $\int_\Omega X^2 dP < \infty$ is denoted by $L^2(\Omega, \mathcal{F}, P)$. On this set $\|X\| = \left(\int_\Omega X^2 dP\right)^{1/2}$ is a norm and $< X, Y > = \int_\Omega XY dP$ is an inner product. If $\mathcal{G}$ is a σ-algebra and $\mathcal{G} \subset \mathcal{F}$, the conditional expectation of X with respect to $\mathcal{G}$, denoted by $E(X|\mathcal{G})$ is the orthogonal projection of X onto the closed subspace $L^2(\Omega, \mathcal{G}, P)$ of $L^2(\Omega, \mathcal{F}, P)$. Prove the following results:

   (a) For $X, Y \in L^2(\Omega, \mathcal{F}, P)$ we have $< E(X|\mathcal{G}), Y > = < E(Y|\mathcal{G})), X > = < E(X|\mathcal{G}), E(Y|\mathcal{G}) >$.

   (b) If $X = Y$ almost everywhere then $E(X|\mathcal{G}) = E(Y|\mathcal{G})$ almost everywhere.

   (c) For $X \in L^2(\Omega, \mathcal{G}, P)$ we have $E(X|\mathcal{G}) = X$.

   (d) If $\mathcal{H} \subset \mathcal{G}$ is a σ-algebra, then $E(E(X|\mathcal{G})|\mathcal{H}) = E(X|\mathcal{H})$.

   (e) If $Y \in L^2(\Omega, \mathcal{G}, P)$ and there exists a constant $C > 0$ such that $P(|Y| \geq C) = 0$, we have that $E(YX|\mathcal{G}) = YE(X|\mathcal{G})$.

   (f) If $\{Y_n\}_{n \in \mathbb{N}}$, $X \in L^2(\Omega, \mathcal{F}, P)$ and $\|Y_n - X\| \to 0$ as $n \to \infty$, then $E(Y_n|\mathcal{G}) \xrightarrow{p} E(X|\mathcal{G})$ as $n \to \infty$.

4. Let $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ be random variables and assume that $E(Y|X) = aX$ where $a \in \mathbb{R}$.

(a) Show that if $E(X^2) > 0$, $a = E(XY)/E(X^2)$.

(b) If $\{(Y_i\ X_i)^T\}_{i=1}^n$ is a sequence of independent random vectors with components having the same distribution as $(Y\ X)^T$, show that

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \overset{p}{\to} E(X^2) \text{ and } \frac{1}{n}\sum_{i=1}^n Y_iX_i \overset{p}{\to} E(XY).$$

(c) Let $a_n = \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)^{-1}\frac{1}{n}\sum_{i=1}^n Y_iX_i$. Does $a_n \overset{p}{\to} a$? Can a_n be defined for all n? Explain.

5. Prove the following:

(a) If $Y \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra, show that $|E(Y|\mathcal{G})| \leq E(|Y||\mathcal{G})$.

(b) Let c be a scalar constant and suppose $X = c$ almost surely. Show that $E(X|\mathcal{G}) = c$ almost surely.

(c) If $Y \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra, show that for $a > 0$

$$P\left(\{\omega : |Y(\omega)| \geq a\}|\mathcal{G}\right) \leq \frac{1}{a}E(|Y(\omega)||\mathcal{G}).$$

What is the definition of $P\left(\{\omega : |Y(\omega)| \geq a\}|\mathcal{G}\right)$? Is this a legitimate probability measure?

6. Let Y and X be random variables such that $Y, X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ and define $\varepsilon = Y - E(Y|X)$.

(a) Show that $E(\varepsilon|X) = 0$ and $E(\varepsilon) = 0$.

(b) Let $V(Y|X) = E(Y^2|X) - E(Y|X)^2$. Show that $V(Y|X) = V(\varepsilon|X)$, $V(\varepsilon) = E(V(Y|X))$;

(c) $Cov(\varepsilon, h(X)) = 0$ for any function of X whose expectation exists.

163

(d) Assume that $E(Y|X) = \alpha + \beta X$ where $\alpha, \beta \in \mathbb{R}$. Let $E(Y) = \mu_Y$, $E(X) = \mu_X$, $V(Y) = \sigma_Y^2$, $V(X) = \sigma_X^2$ and $\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$. Show that,

$$E(Y|X) = \mu_Y + \rho \sigma_Y \frac{X - \mu_X}{\sigma_X} \text{ and } E(V(Y|X)) = (1 - \rho^2)\sigma_Y^2.$$