$(\Omega, \mathcal{F}, P)$ $\quad X : \Omega \to \mathbb{R}$ r.v.

$(\mathbb{R}, \mathcal{B})$ $\quad \mathcal{G} \subseteq \mathcal{F}$ sub-σ-algebra

$\uparrow$ Borel

$E[X \mid \mathcal{G}]$: - $\mathcal{G}$-measurable random var.

conditional exp. $\quad$ - $E\left[I_G E[X \mid \mathcal{G}]\right] = E[I_G X]$ $\quad \forall G \in \mathcal{G}$

$$\int_G E[X \mid \mathcal{G}] \, dP = \int_G X \, dP$$

$$E[X+Y \mid \mathcal{G}](\omega) \overset{\text{a.s.}}{=} E[X \mid \mathcal{G}](\omega) + E[Y \mid \mathcal{G}](\omega)$$

a.s. almost surely with respect P

conditional prob.:

$A \in \mathcal{F}$: $\quad P(A \mid \mathcal{G}) \overset{\text{def}}{=} E[I_A \mid \mathcal{G}]$

- $\mathcal{G}$-meas
- $\forall G \in \mathcal{G}$: $\quad \int_G P(A \mid \mathcal{G}) \, dP = \int_G I_A \, dP = P(A \cap G)$.

Prop. : $P(A|\mathcal{G})$

$$0 \leq P(A|\mathcal{G}) \leq 1$$

$$\int_G 0 \, dP \leq \int_G P(A|\mathcal{G}) \, dP \leq \int_G 1 \, dP$$

holds for all $G \in \mathcal{G}$ $\Rightarrow$ $0 \leq P(A|\mathcal{G}) \leq 1$ a.s.

$A_1, A_2, \ldots \in \mathcal{A}$ disjoint sets

$$P\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, \mathcal{G}\right)(\omega) = \sum_{i=1}^{\infty}{}' P(A_i|\mathcal{G})(\omega) \text{ a.s.}$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty}{}' P(A_i) \leftarrow \text{these are numbers}$$

Indeed: – $\mathcal{G}$-meas. ✓

– $\forall G \in \mathcal{G}$: $\displaystyle\int_G \sum_{i=1}^{\infty}{}' P(A_i|\mathcal{G}) \, dP \overset{?}{=}$

$$\overset{?}{=} \int_G \sum{}' I_{A_i} \, dP$$

$$\int_G \sum{}' P(A_i|\mathcal{G}) \, dP = \sum{}' \int_G P(A_i|\mathcal{G}) \, dP \underset{\text{def. cond. prob.}}{=} \sum{}' \int_G I_{A_i} \, dP \; ✓$$

So $A_1, A_2, \ldots$ disjoint then:

$$P\left(\bigcup_i A_i \mid \mathcal{G}\right) = \sum_i P(A_i \mid \mathcal{G}) \quad \text{a.s.}$$

That is $P(\cdot \mid \mathcal{G})$ behaves like a measure

Problem:

$\forall A \in \mathcal{F}: 0 \leq P(A \mid \mathcal{G}) \leq 1$

$\Rightarrow \exists N_A \in \mathcal{F}: P(N_A) = 0$

such that $P(A \mid \mathcal{G})(\omega) \in [0,1]$

for every $\omega \in \Omega \setminus N_A$

Exceptional set N_A

For different A's N_A is different

These sets together can be large.

$\{x\}$ $\bigcup_{A \in \mathcal{F}} N_A$ can be large

4.1.3 Regular Conditional Probabilities*

$(S, \mathcal{S})$ σ-alg. Borel

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ a measurable map, and $\mathcal{G}$ a σ-field $\subset \mathcal{F}$. $\mu : \Omega \times \mathcal{S} \to [0,1]$ is said to be a **regular conditional distribution** for X given $\mathcal{G}$ if

(i) For each A, $\omega \to \mu(\omega, A)$ is a version of $P(X \in A|\mathcal{G})$. $P(X \in A|\mathcal{G})(\omega) = \mu(\omega, A)$

(ii) For a.e. ω, $A \to \mu(\omega, A)$ is a probability measure on $(S, \mathcal{S})$. ← our set up

When $S = \Omega$ and X is the identity map, μ is called a **regular conditional probability**.

Continuation of Example 4.1.6. Suppose X and Y have a joint density $f(x, y) > 0$. If

$$\mu(y, A) = \int_A f(x, y)\, dx \Big/ \int f(x, y)\, dx$$

then $\mu(Y(\omega), A)$ is a r.c.d. for X given $\sigma(Y)$.

(i) in the definition follows by taking $h = 1_A$ in Example 4.1.1. To check (ii) note that the dominated convergence theorem implies that $A \to \mu(y, A)$ is a probability measure.

Regular conditional distributions are useful because they allow us to simultaneously compute the conditional expectation of all functions of X and to generalize properties of ordinary expectation in a more straightforward way.

Theorem 4.1.16. *Let $\mu(\omega, A)$ be a r.c.d. for X given $\mathcal{F}$. If $f : (S, \mathcal{S}) \to (\mathbf{R}, \mathcal{R})$ has $E|f(X)| < \infty$ then*

$$E(f(X)|\mathcal{F}) = \int \mu(\omega, dx) f(x) \quad a.s. \qquad \int f(x)\, \mu(\omega, dx)$$

$f = I_A$

$E[f(X)|\mathcal{G}] = P(X \in A|\mathcal{G})$

$\int f(x) \mu(\omega, dx) = \mu(\omega, A)$

Proof. If $f = 1_A$ this follows from the definition. Linearity extends the result to simple f and monotone convergence to nonnegative f. Finally we get the result in general by writing $f = f^+ - f^-$. □

Unfortunately, r.c.d.'s do not always exist. The first example was due to Dieudonné (1948). See Doob (1953), p. 624, or Faden (1985) for more recent developments. Without going into the details of the example, it is easy to see the source of the problem. If $A_1, A_2, \ldots$ are disjoint, then (4.1.1) and (4.1.3) imply

$$P(X \in \cup_n A_n|\mathcal{G}) = \sum_n P(X \in A_n|\mathcal{G}) \quad \text{a.s.}$$

but if $\mathcal{S}$ contains enough countable collections of disjoint sets, the exceptional sets may pile up. Fortunately,

$(\Omega, \mathcal{F}, P)$

$X : \Omega \to \mathbb{R}$

dist. function

$$F(x) = P(X \leq x)$$

$$E(X) \overset{\text{def.}}{=} \int_\Omega X \, dP = \int_\Omega X(\omega) \, dP(\omega)$$

density function

$$\underset{\text{theorem}}{=} \int_{\mathbb{R}} x \, \underbrace{dF(x)}_{\text{Lebesgue-Stieltjes integral}} = \int_{-\infty}^{\infty} x \cdot f(x) \, dx \quad X \text{ is cont.}$$

$$= \sum_i x_i \cdot P(X = x_i) \quad X \text{ is discrete}$$

$$\mu_F((-\infty, x]) = F(x)$$

Special case: $dF(x) = f(x) dx$ $\ll L$

$$dF(x) = \sum_i P(X = x_i) \cdot \delta_{x_i}$$

$$E(h(X)) = \int_{-\infty}^{\infty} h(x) \, dF(x)$$

$$E[h(X) | \mathcal{G}] = \int h(x) \overbrace{\mu(\omega, dx)}^{dP(X \leq x \,|\, \mathcal{G})}$$

$(\mathbb{R}, \mathcal{B})$ $(S, \mathcal{S})$ is nice if $\exists\, \varphi : S \to \mathbb{R}$ $\varphi^{-1} : \tilde{\mathbb{R}} \to S$

 $(\mathbb{R}^d, \mathcal{B})$ is nice

Theorem 4.1.17. *r.c.d.'s exist if $(S, \mathcal{S})$ is nice.*

Proof. By definition, there is a 1-1 map $\varphi : S \to \mathbf{R}$ (into) so that φ and φ^{-1} are measurable. Using monotonicity (4.1.2) and throwing away a countable collection of null sets, we find there is a set Ω_o with $P(\Omega_o) = 1$ and a family of random variables $G(q, \omega)$, $q \in \mathbf{Q}$ so that $q \to G(q, \omega)$ is nondecreasing and $\omega \to G(q, \omega)$ is a version of $P(\varphi(X) \le q|\mathcal{G})$. Let $F(x, \omega) = \inf\{G(q, \omega) : q > x\}$. The notation may remind the reader of the proof of Theorem 3.2.12. The argument given there shows F is a distribution function. Since $G(q_n, \omega) \downarrow F(x, \omega)$, the remark after Theorem 4.1.9 implies that $F(x, \omega)$ is a version of $P(\varphi(X) \le x|\mathcal{G})$.

If S is a Borel-set of a separable complete metric space then it is nice.

Now, for each $\omega \in \Omega_o$, there is a unique measure $\nu(\omega, \cdot)$ on $(\mathbf{R}, \mathcal{R})$ so that $\nu(\omega, (-\infty, x]) = F(x, \omega)$. To check that for each $B \in \mathcal{R}$, $\nu(\omega, B)$ is a version of $P(\varphi(X) \in B|\mathcal{G})$, we observe that the class of B for which this statement is true (this includes the measurability of $\omega \to \nu(\omega, B)$) is a λ-system that contains all sets of the form $(a_1, b_1] \cup \cdots (a_k, b_k]$ where $-\infty \le a_i < b_i \le \infty$, so the desired result follows from the $\pi - \lambda$ theorem. To extract the desired r.c.d., notice that if $A \in \mathcal{S}$ and $B = \varphi(A)$, then $B = (\varphi^{-1})^{-1}(A) \in \mathcal{R}$, and set $\mu(\omega, A) = \nu(\omega, B)$. □

The following generalization of Theorem 4.1.17 will be needed in Section 6.1.

Theorem 4.1.18. *Suppose X and Y take values in a nice space $(S, \mathcal{S})$ and $\mathcal{G} = \sigma(Y)$. There is a function $\mu : S \times \mathcal{S} \to [0, 1]$ so that*

(i) for each A, $\mu(Y(\omega), A)$ is a version of $P(X \in A|\mathcal{G})$ $= \mu(Y, A)$

(ii) for a.e. ω, $A \to \mu(Y(\omega), A)$ is a probability measure on $(S, \mathcal{S})$.

Proof. As in the proof of Theorem 4.1.17, we find there is a set Ω_o with $P(\Omega_o) = 1$ and a family of random variables $G(q, \omega)$, $q \in \mathbf{Q}$ so that $q \to G(q, \omega)$ is nondecreasing and $\omega \to G(q, \omega)$ is a version of $P(\varphi(X) \le q|\mathcal{G})$. Since $G(q, \omega) \in \sigma(Y)$ we can write $G(q, \omega) = H(q, Y(\omega))$. Let $F(x, y) = \inf\{G(q, y) : q > x\}$. The argument given in the proof of Theorem 4.1.17 shows that there is a set A_0 with $P(Y \in A_0) = 1$ so that when $y \in A_0$, F is a distribution function and that $F(x, Y(\omega))$ is a version of $P(\varphi(X) \le x|Y)$.

For each $y \in A_o$, there is a unique measure $\nu(y, \cdot)$ on $(\mathbf{R}, \mathcal{R})$ so that $\nu(y, (-\infty, x]) = F(x, y))$. To check that for each $B \in \mathcal{R}$, $\nu(Y(\omega), B)$ is a version of $P(\varphi(X) \in B|Y)$, we observe that the class of B for which this statement is true (this includes the measurability of $\omega \to \nu(Y(\omega), B))$ is a λ-system that contains all sets of the form $(a_1, b_1] \cup \cdots (a_k, b_k]$ where $-\infty \le a_i < b_i \le \infty$, so the desired result follows from the $\pi - \lambda$ theorem. To extract the desired r.c.d. notice that if $A \in \mathcal{S}$, and $B = \varphi(A)$ then $B = (\varphi^{-1})^{-1}(A) \in \mathcal{R}$, and set $\mu(y, A) = \nu(y, B)$. □

Proof of Thm. 4.1.17. For $S = \mathbb{R}$

$X : \Omega \to \mathbb{R}$

$P(X \leq x \mid \mathcal{G})$

consider $P(X \leq q \mid \mathcal{G})$ $\quad q \in \mathbb{Q}$ ↗ countable!!

$\forall q \in \mathbb{Q}: \quad 0 \leq P(X \leq q \mid \mathcal{G}) \leq 1$ a.s.

$\exists\, N_q : P(N_q) = 0$ such that

$\forall w \notin N_q : \; P(X \leq q \mid \mathcal{G})(w) \in [0,1]$

$N = \bigcup_{q \in \mathbb{Q}} N_q$ ⟸ countable union of sets of meas. 0.

$\Rightarrow P(N) = 0.$

$\exists N : (\forall \omega \notin N)(\forall q \in \mathbb{Q}) : \; P(X \leq q \mid \mathcal{G})(\omega) \in [0,1]$

If $q < r$: $\quad P(X \leq q \mid \mathcal{G}) \leq P(X \leq r \mid \mathcal{G})$

monotonicity $\Rightarrow$ $E[I(X \leq q) \mid \mathcal{G}] \leq E[I(X \leq r) \mid \mathcal{G}]$

$\exists \; N_{q,r} : P(N_{q,r}) = 0$

$\forall \omega \notin N_{q,r} : P(X \leq q \mid \mathcal{G})(\omega) \leq P(X \leq r \mid \mathcal{G})(\omega)$

$N_2 = \bigcup_{\substack{q<r \\ q,r \in \mathbb{Q}}} N_{q,r}$ countable!! $\Rightarrow P(N_2) = 0$.

$\Rightarrow$ If $\omega \notin N_2 \cup N$ then

$$P(X \leq q \mid \mathcal{G})(\omega) \in [0,1]$$

and is monotone nondecreasing in q.

$x \in \mathbb{R}$

$$G(x,\omega) = \inf\{P(X \leq q \mid \mathcal{G})(\omega) : q > x\}$$

If $\omega \notin N_2 \cup N$: $G(x,\omega)$ is nondecreasing in x.

$G(x,\omega)$ is a d.f. in x.

$$P(X \leq x \mid \mathcal{G})(\omega) = G(x,\omega) \quad \text{a.s.}$$

$q_n \downarrow x : P(X \leq q_n \mid \mathcal{G}) \downarrow P(X \leq x \mid \mathcal{G})$

So $G(x,\omega)$ is a version of $P(X\leq x|Y)$

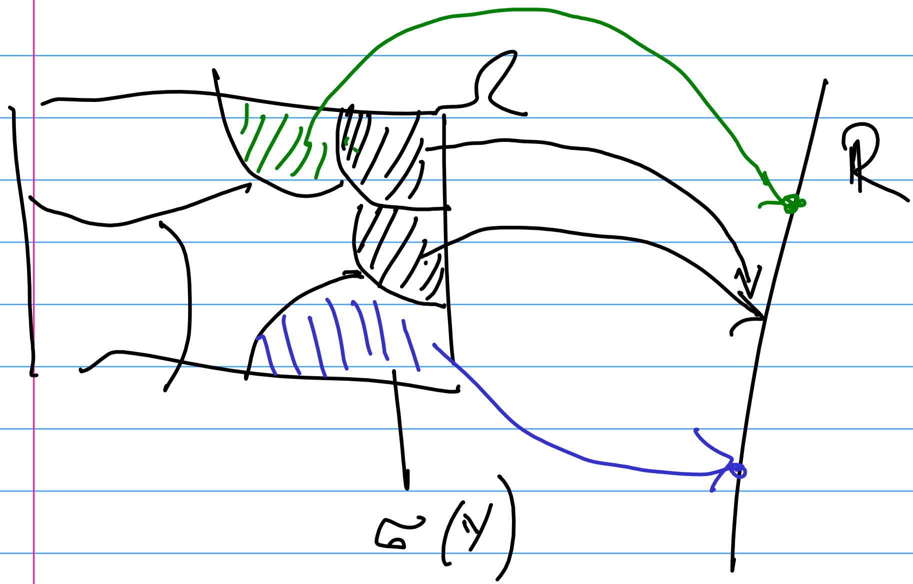
defines a measure on $\mathbb{R}$. $\square$

— o —

$$E[X|Y] \overset{\text{def}}{=} \underbrace{E[X|\sigma(Y)]}_{\text{def. measurable with respect to } \sigma(Y)}$$

$$\Rightarrow E[X|\sigma(Y)] = a(Y)$$

a: measurable



$$P(X \in A \mid Y) = P(X \in A \mid \mathcal{Y})$$

$$\|$$

$$a(Y)$$

Exercises

4.1.1. Bayes' formula. Let $G \in \mathcal{G}$ and show that

$$P(G|A) = \int_G P(A|\mathcal{G})\,dP \Big/ \int_\Omega P(A|\mathcal{G})\,dP$$

When $\mathcal{G}$ is the σ-field generated by a partition, this reduces to the usual Bayes' formula

$$P(G_i|A) = P(A|G_i)P(G_i) \Big/ \sum_j P(A|G_j)P(G_j)$$

4.1.2. Prove **Chebyshev's inequality.** If $a > 0$ then

$$P(|X| \geq a|\mathcal{F}) \leq a^{-2}E(X^2|\mathcal{F})$$

4.1.3. Imitate the proof in the remark after Theorem 1.5.2 to prove the conditional Cauchy-Schwarz inequality.

$$E(XY|\mathcal{G})^2 \leq E(X^2|\mathcal{G})E(Y^2|\mathcal{G})$$

4.1.4. Use regular conditional probability to get the conditional Hölder inequality from the unconditional one, i.e., show that if $p,\ q \in (1,\infty)$ with $1/p + 1/q = 1$ then

$$E(|XY||\mathcal{G}) \leq E(|X|^p|\mathcal{G})^{1/p}E(|Y|^q|\mathcal{G})^{1/q}$$

4.1.5. Give an example on $\Omega = \{a, b, c\}$ in which

$$E(E(X|\mathcal{F}_1)|\mathcal{F}_2) \neq E(E(X|\mathcal{F}_2)|\mathcal{F}_1)$$

4.1.6. Show that if $\mathcal{G} \subset \mathcal{F}$ and $EX^2 < \infty$ then

$$E(\{X - E(X|\mathcal{F})\}^2) + E(\{E(X|\mathcal{F}) - E(X|\mathcal{G})\}^2) = E(\{X - E(X|\mathcal{G})\}^2)$$

Dropping the second term on the left, we get an inequality that says geometrically, the larger the subspace the closer the projection is, or statistically, more information means a smaller mean square error.

4.1.7. An important special case of the previous result occurs when $\mathcal{G} = \{\emptyset, \Omega\}$. Let $\text{var}\,(X|\mathcal{F}) = E(X^2|\mathcal{F}) - E(X|\mathcal{F})^2$. Show that

$$\text{var}\,(X) = E(\,\text{var}\,(X|\mathcal{F})) + \text{var}\,(E(X|\mathcal{F}))$$

4.1.8. Let $Y_1, Y_2, \ldots$ be i.i.d. with mean μ and variance σ^2, N an independent positive integer valued r.v. with $EN^2 < \infty$ and $X = Y_1 + \cdots + Y_N$. Show that $\text{var}\,(X) = \sigma^2\,EN + \mu^2\,\text{var}\,(N)$. To understand and help remember the formula, think about the two special cases in which N or Y is constant.

4.1.9. 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.

4.1.10. 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 $\text{sgn}\,(X) = \text{sgn}\,(E(X|\mathcal{G}))$ a.s., and then take $X = Y - c$ to get the desired result.

conditional prob.

information

4.2 Martingales, Almost Sure Convergence

In this section we will define martingales and their cousins supermartingales and submartingales, and take the first steps in developing their theory. Let $\mathcal{F}_n$ be a **filtration**, i.e., an increasing sequence of σ-fields. A sequence X_n is said to be **adapted** to $\mathcal{F}_n$ if $X_n \in \mathcal{F}_n$ for all n. If X_n is sequence with

$(\Omega, \mathcal{F}, P)$, $(\mathcal{F}_n)_n$

(i) $E|X_n| < \infty$,

(ii) X_n is adapted to $\mathcal{F}_n$,

$\mathcal{F}_0 \subset \dots \subset \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \dots$

(iii) $E(X_{n+1}|\mathcal{F}_n) = X_n$ for all n,

then X is said to be a **martingale** (with respect to $\mathcal{F}_n$). If in the last definition, $=$ is replaced by $\leq$ or $\geq$, then X is said to be a **supermartingale** or **submartingale**, respectively.

increasing sequence of σ-alg.

We begin by describing three examples related to random walk. Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed. Let $S_n = S_0 + \xi_1 + \cdots + \xi_n$ where S_0 is a constant. Let $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ for $n \geq 1$ and take $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Example 4.2.1. Linear martingale. If $\mu = E\xi_i = 0$ then S_n, $n \geq 0$, is a martingale with respect to $\mathcal{F}_n$.

(X_n) is adapted to $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$ meas.

To prove this, we observe that $S_n \in \mathcal{F}_n$, $E|S_n| < \infty$, and ξ_{n+1} is independent of $\mathcal{F}_n$, so using the linearity of conditional expectation, (4.1.1), and Example 4.1.4,

$$E(S_{n+1}|\mathcal{F}_n) = E(S_n|\mathcal{F}_n) + E(\xi_{n+1}|\mathcal{F}_n) = S_n + E\xi_{n+1} = S_n$$

If $\mu \leq 0$ then the computation just completed shows $E(\overset{S}{\cancel{X}}_{n+1}|\mathcal{F}_n) \leq$ $\overset{S}{\cancel{X}}_n$, i.e., $\overset{S}{\cancel{X}}_n$ is a supermartingale. In this case, $\overset{S}{\cancel{X}}_n$ corresponds to betting on an unfavorable game so there is nothing "super" about a supermartingale. The name comes from the fact that if f is superharmonic (i.e., f has continuous derivatives of order ≤ 2 and $\partial^2 f/\partial x_1^2 + \cdots + \partial^2 f/\partial x_d^2 \leq 0$), then

$$f(x) \geq \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\, dy \tag{4.2.1}$$

$E[S_{n+1}|\mathcal{F}_n] = E[S_n + \xi_{n+1}|\mathcal{F}_n] = E[S_n|\mathcal{F}_n] +$

$$+ \; E[\xi_{n+1} \mid \mathcal{F}_n] = S_n + E[\xi_{n+1}] = S_n$$

$$\mathcal{F}_n = G(\xi_1, \dots, \xi_n)$$

where $B(x,r) = \{y : |x-y| \le r\}$ is the ball of radius r, and $|B(x,r)|$ is the volume of the ball.

If $\mu \ge 0$ then S_n is a submartingale. Applying the first result to $\xi_i' = \xi_i - \mu$ we see that $S_n - n\mu$ is a martingale.

Example 4.2.2. Quadratic martingale. Suppose now that $\mu = E\xi_i = 0$ and $\sigma^2 = \text{var}(\xi_i) < \infty$. In this case $S_n^2 - n\sigma^2$ is a martingale.

Since $(S_n + \xi_{n+1})^2 = S_n^2 + 2S_n\xi_{n+1} + \xi_{n+1}^2$ and ξ_{n+1} is independent of $\mathcal{F}_n$, we have

„constant"

$E[S_n \xi_{n+1}|\mathcal{F}_n] = S_n E[\xi_{n+1}|\mathcal{F}_n]$

↓ $\mathcal{F}_n$ meas.

$$E(S_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n) = S_n^2 + 2S_n \overbrace{E(\xi_{n+1}|\mathcal{F}_n)}^{=0} + E(\xi_{n+1}^2|\mathcal{F}_n) - (n+1)\sigma^2$$
$$= S_n^2 + 0 + \sigma^2 - (n+1)\sigma^2 = S_n^2 - n\sigma^2$$

i.i.d.

Example 4.2.3. Exponential martingale. Let $Y_1, Y_2, \ldots$ be nonnegative i.i.d. random variables with $EY_m = 1$. If $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$ then $M_n = \prod_{m \le n} Y_m$ defines a martingale. To prove this note that

$$E(M_{n+1}|\mathcal{F}_n) = M_n E(\overset{Y}{\cancel{X}}_{n+1}|\mathcal{F}_n) = M_n \, .$$

Suppose now that $\cancel{Y_i = e^{\theta\xi_i}}$ and $\phi(\theta) = Ee^{\theta\xi_i} < \infty$. $Y_i = \exp(\theta\xi)/\phi(\theta)$ has mean 1 so $EY_i = 1$ and

$$E(Y_i) = \frac{Ee^{\theta\xi}}{\phi(\theta)} = \frac{\phi(\theta)}{\phi(\theta)} = 1$$

$$M_n = \prod_{i=1}^n Y_i = \exp(\theta S_n)/\phi(\theta)^n \quad \text{is a martingale.}$$

We will see many other examples below, so we turn now to deriving properties of martingales. Our first result is an immediate consequence of the definition of a supermartingale. We could take the conclusion of the result as the definition of supermartingale, but then the definition would be harder to check.

Theorem 4.2.4. *If X_n is a supermartingale then for $n > m$, $E(X_n|\mathcal{F}_m) \le X_m$.*

Proof. The definition gives the result for $n = m+1$. Suppose $n = m+k$ with $k \ge 2$. By Theorem 4.1.2, tower rule

$$E(X_{m+k}|\mathcal{F}_m) = E(E(X_{m+k}|\mathcal{F}_{m+k-1})|\mathcal{F}_m) \le E(X_{m+k-1}|\mathcal{F}_m)$$

def. + monotonicity of $E[\cdot|\mathcal{F}_n]$

by the definition and (4.1.2). The desired result now follows by induction. □

Theorem 4.2.5. *(i) If X_n is a submartingale then for $n > m$, $E(X_n|\mathcal{F}_m) \ge X_m$.*
(ii) If X_n is a martingale then for $n > m$, $E(X_n|\mathcal{F}_m) = X_m$.

Proof. To prove (i), note that $-X_n$ is a supermartingale and use (4.1.1). For (ii), observe that X_n is a supermartingale and a submartingale. □