Prof. Fabio Nobile

Series 1 - September 18, 2024

Exercise 1.

Let $(\Omega, \mathcal{F}, P)$ be a probability space
and $\{F_n\}_{n\geqslant 1}$ be a countable sequence of $\mathcal{F}$. Show that:

i) if $F_n \subseteq F_{n+1}$ for all $n \geqslant 1$ then $P\Big(\bigcup_{n=1}^{\infty} F_n\Big) = \lim_{n\to\infty} P(F_n)$,

ii) if $F_n \supseteq F_{n+1}$ for all $n \geqslant 1$ then $P\Big(\bigcap_{n=1}^{\infty} F_n\Big) = \lim_{n\to\infty} P(F_n)$.

Moreover, show that there exists a sequence $\{F'_n\}_{n\geqslant 1}$ such that $F'_i \cap F'_j = \varnothing$ if $i \neq j$ and

$$\bigcup_{n=1}^{\infty} F'_n = \bigcup_{n=1}^{\infty} F_n.$$

Solution

Consider the sequence of events A_n defined by $A_1 = F_1$ and $A_n = F_n \setminus F_{n-1}$. Then by the axioms of probability

$$P(\bigcup_{n=1}^{\infty} F_n) = P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) = \lim_{N\to\infty} \sum_{n=1}^{N} P(A_n) = \lim_{N\to\infty} P(\bigcup_{n=1}^{N} A_n)$$
$$= \lim_{N\to\infty} P(\bigcup_{n=1}^{N} F_n) = \lim_{N\to\infty} P(F_N).$$

The same result can be obtained for (ii) thanks to

$$P(\bigcap_{n=1}^{\infty} F_n) = 1 - P(\bigcup_{n=1}^{\infty} F_n^c),$$

and then proceeding as in (i). Now let $\{F_n\}_{n\geqslant 1}$ be a sequence of $\mathcal{F}$. Define the sequence $\{F'_n\}_{n\geqslant 1}$ as

$$F'_1 = F_1, \qquad F'_n = F_n \cap \big(\cup_{k=1}^{n-1} F_k\big)^c \quad n \geqslant 2.$$

By induction we show that $\cup_{n=1}^{m} F'_n = \bigcup_{n=1}^{m} F_n$ for all $m \geqslant 1$. The case $m = 1$ follows from the definition of F'_1. Then, due to the induction step we have

$$\cup_{n=1}^{m} F'_n = \big(\cup_{n=1}^{m-1} F'_n\big) \cup F'_m = \big(\cup_{n=1}^{m-1} F_n\big) \cup \big(F_m \cap \big(\cup_{n=1}^{m-1} F_n\big)^c\big) = \cup_{n=1}^{m} F_n,$$

where we employed the associativity

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{1.1}$$

Now, if $i < j$ we can write $F'_j = F_j \cap \big(\cup_{k=1}^{j-1} F'_k\big)^c$ and since clearly $F'_i \subset \cup_{k=1}^{j-1} F'_k$ we have $F'_i \cap F'_j = \varnothing$. The argument is analogous if $i > j$.

Exercise 2.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_i\}_{i\geqslant 1}$ be a sequence of events, i.e., $A_i \in \mathcal{F}$.

i) Show that the event $E = \{\text{"infinitely many } A_i \text{ occur"}\}$ can be written

$$E = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i.$$

Remark. The set E is often called "A_i i.o.", which means "A_i infinitely often".

ii) Describe the event $H = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$.

iii) Show that if $\sum_{i=1}^{\infty} P(A_i) < \infty$ then $P(A_i \text{ i.o.}) = 0$ (Borel–Cantelli lemma).

Solution

i) Set $F_n = \cup_{i=n}^{\infty} A_i$. If F_n occurs, there are some A_i $i \geqslant n$ that occur. If E occurs, it means that the F_n occur for all $n \geqslant 1$, as $E = \cap_{n=1}^{\infty} F_n$. But if F_n occurs for all n, it means that infinitely many A_i occur. Reciprocally, if infinitely many A_i occur, then F_n occur for all $n \geqslant 1$ and then E occurs. We write this as

$$\begin{aligned} \omega \in E &\iff \omega \in F_n \ \forall n \geqslant 1 \\ &\iff \forall n \geqslant 1 \ \exists \, i_n \geqslant n \text{ s.t. } \omega \in A_{i_n} \\ &\iff |\{i \text{ s.t. } \omega \in A_i\}| = \infty. \end{aligned} \tag{2.1}$$

ii) Set $G_n = \cap_{i=n}^{\infty} A_i$. If G_n occurs, all A_i occur for $i \geqslant n$. If $H = \cup_{n=1}^{\infty} G_n$ occurs, there exists a n such that for $i \geqslant n$ A_i occur. Consequently, $H = \{\exists n \geqslant 1 : \text{ all } A_i \text{ occur for } i \geqslant n\}$. We write this as

$$\begin{aligned} \omega \in H &\iff \exists \, n \geqslant 1 \text{ s.t. } \omega \in G_n \\ &\iff \exists \, n \geqslant 1 \text{ s.t. } \omega \in A_i \ \forall i \geqslant n. \end{aligned} \tag{2.2}$$

iii) We have to show that $P(E) = 0$. Note that $F_n = \cup_{i=n}^{\infty} A_i$ satisfies $F_n \supseteq F_{n+1}$. From Exercise , we have

$$P(E) = \lim_{n\to\infty} P(F_n) = \lim_{n\to\infty} P(\cup_{i=n}^{\infty} A_i).$$

Furthermore, $P(\cup_{i=n}^{\infty} A_i) \leqslant \sum_{i=n}^{\infty} P(A_i)$ and as we assume $\sum_{i=1}^{\infty} P(A_i) < \infty$, we conclude that $\lim_{n\to\infty} P(\cup_{i=n}^{\infty} A_i) = 0$ and $P(E) = 0$.

Exercise 3.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X \colon \Omega \to \mathbb{R}^n$ be an n-dimensional random variable. Show that $\mathcal{F}(X) = \{X^{-1}(B) \colon B \in \mathcal{B}\}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}^n$, is a σ-algebra. Observe that $\mathcal{F}(X)$ is the smallest σ-algebra with respect to which X is measurable.

Solution

We check the properties of a σ-algebra.

i) $\Omega = X^{-1}(\mathbb{R}^n)$ and as $\mathbb{R}^n \in \mathcal{B}$ then $\Omega \in \mathcal{F}(X)$.

ii) Let $F \in \mathcal{F}(X)$. Then there exists $B \in \mathcal{B}$ such that $F = X^{-1}(B)$. Hence

$$F^c = \left(X^{-1}(B)\right)^c = \{\omega \in \Omega : X(\omega) \notin B\} = \{\omega \in \Omega : X(\omega) \in B^c\} = X^{-1}(B^c),$$

and as $B^c \in \mathcal{B}$ then $F^c \in \mathcal{F}(X)$.

iii) Let $\{F_i\}_{i\geqslant 1} \subset \mathcal{F}(X)$. Then there exist $\{B_i\}_{i\geqslant 1} \subset \mathcal{B}$ such that $F_i = X^{-1}(B_i)$ for all $i \geqslant 1$. Then

$$\cup_i F_i = \cup_i \{\omega \in \Omega : X(\omega) \in B_i\} = \{\omega \in \Omega : X(\omega) \in \cup_i B_i\} = X^{-1}(\cup_i B_i),$$

and as $\cup_i B_i \in \mathcal{B}$ then $\cup_i F_i \in \mathcal{F}(X)$.

Exercise 4.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\{X_n\}_{n\geqslant 1}$ be a sequence of real-valued random variables $X_n\colon \Omega \to \mathbb{R}$ and $X\colon \Omega \to \mathbb{R}$ be another real-valued random variable. Denote by F_n and F the distribution function of X_n and X, respectively, and recall the following notions of convergence of random variables:

- $X_n \to X$ in L^2 if $\mathbb{E}[|X_n - X|^2] \to 0$ as $n \to \infty$,
- $X_n \to X$ in probability if for all $\epsilon > 0$ it holds $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$,
- $X_n \to X$ in distribution if $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ as $n \to \infty$, for any $g \in C_b^0(\mathbb{R})$.

i) Prove that $X_n \to X$ in L^2 implies $X_n \to X$ in probability.

ii) Show that if $X_n \to X$ in probability, then one can extract a subsequence from X_n a.s converging to X. Prove also the following implication: $X_n \to X$ in probability if and only if for every subsequence X_{n_k} one can extract a subsequence from $X_{n_{k_h}}$ a.s converging to X

iii) Prove that $X_n \to X$ in probability implies $X_n \to X$ in distribution.

iv) Show that if $X_n \to X$ in distribution, then $\lim_{n\to\infty} F_n(x) = F(x)$ for all x such that F is continuous *Hint.* Let $\epsilon > 0$ and show that for all $n \geqslant 1$

$$F(x - \epsilon) - P(|X_n - X| > \epsilon) \leqslant F_n(x) \leqslant F(x + \epsilon) + P(|X_n - X| > \epsilon).$$

v) By providing a counterexample show that $X_n \to X$ in probability does not imply $X_n \to X$ in L^2.

vi) By providing a counterexample show that $X_n \to X$ in distribution does not imply $X_n \to X$ in probability.

Solution

i) Fix $\epsilon > 0$. Using Markov's inequality, we have

$$P(|X_n - X| > \epsilon) = P(|X_n - X|^2 > \epsilon^2) \leqslant \frac{1}{\epsilon^2}\mathbb{E}[|X_n - X|^2] \to 0.$$

ii) Via the property of convergence in probability, one has that for all $k \geqslant 1$, there exists $\bar{n}$ such that for all $n > \bar{n}$ one has $P(|X_n - X| > \frac{1}{k}) \leqslant 2^{-k}$. Therefore, one can extract a subsquence $(X_{n_k})_k$ such that $P(|X_{n_k} - X| > \frac{1}{k}) \leqslant 2^{-k}$ for all k. Applying Borel-Cantelli's lemma, one gets the thesis.

For the second part of the claim, if $X_n \to X$ in probability, then thanks to properties of convergent sequences any arbitrary subsequence X_{n_k} of X_n converges in probability to X and, therefore, one can extract a subsequence $X_{n_{k_h}}$ of X_{n_k} converging a.s. to X. On the contrary, suppose by contradiction that $X_n \nrightarrow X$. Then, by definition there exists ε and $\delta > 0$ and a subsequence X_{n_k}, such that $P(|X_{n_k} - X| > \varepsilon) \geqslant \delta$ for all k. But from X_{n_k} we cannot chose an a.s. convergent subsequence to X, which is impossible by hypothesis.

iii) According to point (ii), we have that if $X_n \to X$ in probability, then for every subsequence X_{n_k} we can subtract an a.s subsequence $X_{n_{k_h}}$ convergent to X. For any arbitrary continuous function g, we have that $g(X_{n_{k_h}})$ converges a.s. to $g(X)$. Therefore, for the same point ii), $g(X_n)$ converges in probability to $g(X)$. Then, by boundness and continuity of any $g \in C_b^0(\mathbb{R})$ and defining $A_n = \{\omega \;:\; |g(X_n) - g(X)| < \frac{\varepsilon}{2}\}$, for large n one has

$$\begin{aligned} |\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| &\leqslant \mathbb{E}[|g(X_n) - g(X)|] \\ &\leqslant \frac{\varepsilon}{2}P(A_n) + \|g\|_\infty P(A_n^C) \quad \leqslant \frac{\varepsilon}{2}\cdot 1 + \|g\|_\infty \frac{\varepsilon}{2\|g\|_\infty} = \varepsilon, \quad \forall g \in C_b^0(\mathbb{R}), \end{aligned} \tag{4.1}$$

where in the last line we use the convergence in probability. Via arbitrariness of ε we obtain the thesis.

iv) Fix $\epsilon > 0$ and let x be a point where F is continuous. Then

$$\begin{aligned} F_n(x) &= P(X_n \leqslant x) \\ &= P(X_n \leqslant x, |X_n - X| > \epsilon) + P(X_n \leqslant x, |X_n - X| \leqslant \epsilon) \\ &\leqslant P(|X_n - X| > \epsilon) + P(X \leqslant x + \epsilon) \\ &= P(|X_n - X| > \epsilon) + F(x + \epsilon), \end{aligned} \tag{4.2}$$

and

$$\begin{aligned} F(x - \epsilon) &= P(X \leqslant x - \epsilon) \\ &= P(X \leqslant x - \epsilon, |X_n - X| > \epsilon) + P(X \leqslant x - \epsilon, |X_n - X| \leqslant \epsilon) \\ &\leqslant P(|X_n - X| > \epsilon) + P(X_n \leqslant x) \\ &= P(|X_n - X| > \epsilon) + F_n(x). \end{aligned}$$

Therefore, for all $n \geqslant 1$

$$F(x - \epsilon) - P(|X_n - X| > \epsilon) \leqslant F_n(x) \leqslant F(x + \epsilon) + P(|X_n - X| > \epsilon). \tag{4.3}$$

Hence, taking the limit as $n \to \infty$ and by convergence in probability we have

$$F(x - \epsilon) \leqslant \liminf_{n\to\infty} F_n(x) \leqslant \limsup_{n\to\infty} F_n(x) \leqslant F(x + \epsilon).$$

Since F is continuous in x, $\lim_{\epsilon\to 0} F(x - \epsilon) = \lim_{\epsilon\to 0} F(x + \epsilon) = F(x)$, so taking the limit $\epsilon \to 0$ in (4.3) shows that $\lim_{n\to\infty} F_n(x) = F(x)$.

v) Consider $U \sim \mathrm{Unif}(0, 1)$ and $X_n = \sqrt{n}\chi_{[0,1/n]}(U)$. We have

$$\begin{aligned} P(|X_n - 0| > \epsilon) = P(\sqrt{n}\chi_{[0,1/n]}(U) > \epsilon) &= P(\chi_{[0,1/n]}(U) > \epsilon/\sqrt{n}) \\ &\leqslant P(\chi_{[0,1/n]}(U) > 0) = P(0 \leqslant U \leqslant 1/n) = \frac{1}{n} \to 0, \end{aligned}$$

hence $X_n \to 0$ in probability but $\mathbb{E}(|X_n - 0|^2) = \int_0^{1/n} n = 1$ for all $n \geqslant 1$, so X_n does not converge to 0 in L^2.

vi) Let $X \sim N(0, 1)$, $X_n = -X$ for all $n \geqslant 1$. Then X_n is $N(0, 1)$ and hence $F_n(x) = F(x)$ for all $n \geqslant 1$, but

$$P(|X_n - X| > \epsilon) = P(2|X| > \epsilon) = \frac{2}{\sqrt{2\pi}} \int_{\epsilon/2}^{\infty} e^{-\frac{s^2}{2}} \neq 0.$$

Exercise 5.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{X_n\}_n$ a sequence of $\mathbb{R}^d$-valued random variables with Gaussian distribution $\mu_{X_n} = \mathcal{N}(m_n, \Sigma_n)$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ for all n.

i) Show that if $m_n \to m$ and $\Sigma_n \to \Sigma$ for $n \to \infty$, then $X_n \to X$ in distribution, where X has distribution $\mathcal{N}(m, \Sigma)$.

ii) Show that if $X_n \to X$ in L^2, then μ_X is Gaussian.

iii) Show that if $X_n \to X$ in distribution, then μ_X is Gaussian.

Hint. Recall that a family of probability measure $\{\mu_n\}_n$ is called tight if for any $\epsilon > 0$, there exists a compact set K_ϵ such that for all μ_n one has $\mu_n(K_\epsilon) > 1 - \epsilon$. Moreover, for $\mathbb{R}^d$-valued random variables, one has that a sequence of weakly convergent probability measure is tight (Prokhorov's Theorem). Use this information to prove by contradiction that both m_n, Σ_n are bounded sequences and contain converging subsequences.

Remark. If $X\colon \Omega \to \mathbb{R}^d$ has Gaussian distribution $\mu_X = \mathcal{N}(m, \Sigma)$, then its characteristic function is

$$\hat{\mu}(t) := \mathbb{E}[e^{iX^\top t}] = \exp\{im^\top t - \frac{1}{2}t^\top \Sigma t\}, \quad \text{for } t \in \mathbb{R}^d. \tag{5.1}$$

Moreover, one has that

- if $\mu_n \rightharpoonup \mu$, (equivalently $X_n \to X$ in distribution), then $\hat{\mu}_n(t) \to \hat{\mu}(t), \quad \forall t \in \mathbb{R}^d$
- if $\hat{\mu}_n(t) \to \psi(t), \quad \forall t \in \mathbb{R}^d$ and $\psi(t)$ is continuous at $t = 0$, then ψ is the characteristic function of a probability distribution μ, and $\mu_n \rightharpoonup \mu$.

Solution

i) $\mu_{X_n} = \mathcal{N}(m_n, \Sigma_n)$ implies $\hat{\mu}_n(t) = e^{im_n^\top t - \frac{1}{2}t^\top \Sigma_n t}$. Clearly, if $n \to +\infty$, then $\hat{\mu}_n(t) \to \hat{\mu}(t) := \mathbb{E}[e^{iX^\top t}] = \exp\{im^\top t - \frac{1}{2}t^\top \Sigma t\}$, for all $t \in \mathbb{R}^d$. Moreover, $\hat{\mu}(t)$ is continuous at zero and corresponds to the characteristic function of a random variable X with distribution $\mathcal{N}(m, \Sigma)$, hence $\mu_{X_n} \rightharpoonup \mathcal{N}(m, \Sigma) \iff X_n \to X$ in distribution.

ii) Since $X_n \to X$ in L^2 we have

$$\begin{aligned} m_n &:= \mathbb{E}[X_n] \to \mathbb{E}[X] =: m \\ \Sigma_n &:= \mathbb{E}[(X_n - \mathbb{E}[X_n])(X_n - \mathbb{E}[X_n])^\top] \to \mathrm{Cov}(X) =: \Sigma \end{aligned} \tag{5.2}$$

Moreover, since $x \to e^{ix}$ is Lipschitz, then

$$\left|\mathbb{E}[e^{iX^\top t} - e^{iX_n^\top t}]\right| \leqslant \mathbb{E}[|(X - X_n)^\top t|] \leqslant |t| \|X - X_n\|_{L^2} \to +\infty, \ n \to +\infty. \tag{5.3}$$

Hence

$$\hat{\mu}(t) := \mathbb{E}[e^{iX^\top t}] = \lim_{n\to+\infty} \mathbb{E}[e^{iX_n^\top t}] = \exp\{im_n^\top t - \frac{1}{2}t^\top \Sigma_n t\} = \exp\{im^\top t - \frac{1}{2}t^\top \Sigma t\}, \tag{5.4}$$

and X has Gaussian distribution $\mathcal{N}(m, \Sigma)$.

iii) By contradiction, suppose $\sup_n |m_n| = +\infty$. Then, for all $M > 0$, there exists $\hat{n}$, such that for all $n \geqslant \hat{n}$ one has $|m_n| \geqslant M$. Since for a Gaussian measure $\mathcal{N}(m_n, \Sigma_n)$ one has $\mathbb{P}(|X_n| > |m_n|) \geqslant \frac{1}{2}$, we conclude that $\mathbb{P}(|X_n| > M) \geqslant \frac{1}{2}$. But this fact contradicts tightness, hence $\sup_n |m_n| < +\infty$. One can prove $\sup_n |\Sigma_n| < +\infty$ in a similar way. As both sequences m_n, Σ_n are bounded, one can extract convergent subsequences m_{n_k}, Σ_{n_k} and called their limit m, Σ, respectively. Since X_n is converging in distribution to X, it follow that for all t

$$\mathbb{E}[e^{iX^\top t}] = \lim_{n\to+\infty} \mathbb{E}[e^{iX_n^\top t}] = \lim_{n_k\to+\infty} \mathbb{E}[e^{iX_{n_k}^\top t}] = \exp\{im^\top t - \frac{1}{2}t^\top \Sigma t\}, \tag{5.5}$$

hence $X \sim \mathcal{N}(m, \Sigma)$.

Exercise 6.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Let $X, Y\colon \Omega \to \mathbb{R}$ be integrable random variables on Ω. Using the definition of conditional expectation of random variables given a σ-algebra, prove the following statements.

i) $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ a.s. for $a, b \in \mathbb{R}$.

ii) If X is $\mathcal{G}$-measurable then $\mathbb{E}(X|\mathcal{G}) = X$ a.s.

iii) If X is $\mathcal{G}$-measurable and XY is integrable then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$ a.s.

iv) If X is independent of $\mathcal{G}$ then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ a.s.

v) If $\mathcal{H}$ is a σ-algebra such that $\mathcal{H} \subseteq \mathcal{G}$ then

$$\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})\big|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{H})\big|\mathcal{G}) \text{ a.s.}$$

vi) If $X \leqslant Y$ a.s. then $\mathbb{E}(X|\mathcal{G}) \leqslant \mathbb{E}(Y|\mathcal{G})$ a.s.

Solution

i) It is clear that $Z = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ is $\mathcal{G}$-measurable. Moreover, using the definition of conditional expectation, Z satisfies for any $A \in \mathcal{G}$

$$\int_A Z\,P = a\int_A X\,P + b\int_A Y P = \int_A aX + bY P.$$

Hence, we conclude that $Z = \mathbb{E}(aX + bY|\mathcal{G})$ by the uniqueness of the conditional expectation.

ii) $Z = X$ satisfies the two conditions of the definition of $\mathbb{E}(X|\mathcal{G})$, hence $Z = \mathbb{E}(X|\mathcal{G})$ by uniqueness.

iii) Note that as X is $\mathcal{G}$-measurable, $Z = X\mathbb{E}(Y|\mathcal{G})$ is $\mathcal{G}$-measurable. We prove the result for a simple function $X = \sum_{i=1}^n g_i\chi_{G_i}$, where $\{G_i\} \subset \mathcal{G}$. Note that for $A \in \mathcal{G}$,

$$\int_A Z\,P = \sum_{i=1}^n g_i \int_{A\cap G_i} \mathbb{E}(Y|\mathcal{G})\,P = \sum_{i=1}^n g_i \int_{A\cap G_i} YP = \int_A XYP,$$

and we conclude that $Z = \mathbb{E}(XY|\mathcal{G})$. The general result is then obtained by approximation with simple functions, applying the monotone convergence theorem and considering the positive and negative parts.

iv) For $Z = \mathbb{E}(X)$ and $A \in \mathcal{G}$, using the independence we have

$$\int_A Z\,P = \mathbb{E}(X)P(A) = \mathbb{E}(X)\mathbb{E}(\chi_A) = \mathbb{E}(X\chi_A) = \int_A X\,P,$$

and therefore $Z = \mathbb{E}(X|\mathcal{G})$.

v) First note that as $\mathcal{H} \subset \mathcal{G}$ $Z = \mathbb{E}(X|\mathcal{H})$ is $\mathcal{G}$-measurable. Using *ii)*, it is clear that $\mathbb{E}(\mathbb{E}(X|\mathcal{H})\big|\mathcal{G}) = \mathbb{E}(X|\mathcal{H})$. Now, for any $A \in \mathcal{H}$ it holds $\int_A Z\,P = \int_A X\,P = \int_A \mathbb{E}(X|\mathcal{G})\,P$, because also $A \in \mathcal{G}$. We conclude that $Z = \mathbb{E}(\mathbb{E}(X|\mathcal{G})\big|\mathcal{H})$.

vi) We prove it by contradiction. Consider the event $A = \{\mathbb{E}(X \mid \mathcal{G}) > \mathbb{E}(Y \mid \mathcal{G})\} \in \mathcal{G}$. Then, by *i)* we rewrite $A = \{\mathbb{E}(X - Y \mid \mathcal{G}) > 0\}$. Assume by contradiction that $P(A) > 0$ and denote $Z = \mathbb{E}(X - Y \mid \mathcal{G})$. Then, clearly

$$\int_A ZP > 0.$$

On the other hand, since $X \leqslant Y$ a.s. and by definition of conditional expectation,

$$\int_A ZP = \int_A X - YP \leqslant 0,$$

which gives a contradiction. Hence $P(A) = 0$ and $\mathbb{E}(X \mid \mathcal{G}) \leqslant \mathbb{E}(Y \mid \mathcal{G})$ a.s.