not. However, there is the following norm inequality for prediction errors.

10.28 Theorem If Y is $\mathcal{F}$-measurable and $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$,

$$\|Y - E(Y|\mathcal{G}_2)\|_p \le 2\|Y - E(Y|\mathcal{G}_1)\|_p,\ p \ge 1. \tag{10.63}$$

Proof Let $\eta = Y - E(Y|\mathcal{G}_1)$. Then by **10.26**(ii),

$$\begin{aligned}\eta - E(\eta|\mathcal{G}_2) &= Y - E(Y|\mathcal{G}_1) - E(Y|\mathcal{G}_2) + E(E(Y|\mathcal{G}_1)|\mathcal{G}_2) \\ &= Y - E(Y|\mathcal{G}_2).\end{aligned} \tag{10.64}$$

The theorem now follows, since

$$\|\eta - E(\eta|\mathcal{G}_2)\|_p \le \|\eta\|_p + \|E(\eta|\mathcal{G}_2)\|_p \le 2\|\eta\|_p \tag{10.65}$$

by, respectively, the Minkowski and conditional Jensen inequalities, and the LIE. ■

10.6 Conditional Distributions

The conditional probability of an event $A \in \mathcal{F}$ can evidently be defined as $P(A|\mathcal{G}) = E(1_A|\mathcal{G})$, where $1_A(\omega)$ is the indicator function of A. But is it therefore meaningful to speak of a conditional distribution on $(\Omega,\mathcal{F})$, which assigns probabilities $P(A|\mathcal{G})$ to each $A \in \mathcal{F}$? There are two ways to approach this question.

First, we can observe straightforwardly that conditional probabilities satisfy the axioms of probability except on sets of probability 0 and, in this sense, satisfactorily mimic the properties of true probabilities, just as was found for the expectations. Thus, we have the following.

10.29 Theorem

(i) $P(A|\mathcal{G}) \ge 0$ a.s., all $A \in \mathcal{F}$.

(ii) $P(\Omega|\mathcal{G}) = 1$ a.s.

(iii) For a countable collection of disjoint sets $A_j \in \mathcal{F}$,

$$P\left(\bigcup_j A_j|\mathcal{G}\right) = \sum_j P(A_j|\mathcal{G}) \text{ a.s.} \tag{10.66}$$

Proof To prove (i), suppose $\exists\, G \in \mathcal{G}$ with $P(G) > 0$, and $P(A|\mathcal{G})(\omega) < 0$ for all $\omega \in G$. Then, by (10.18),

$$\int_{G\cap A} dP = \int_G P(A|\mathcal{G})dP < 0, \tag{10.67}$$

which is a contradiction, since the left-hand member is a probability. To prove (ii), note that $P(\Omega|\mathcal{G})$ is $\mathcal{G}$-measurable and let $G^+ \in \mathcal{G}$ denote the set of ω such that $P(\Omega|\mathcal{G})(\omega) > 1$. Suppose $P(G^+) > 0$. Then since $G^+ \cap \Omega = G^+$,

$$P(G^+) = \int_{G^+} dP < \int_{G^+} P(\Omega|\mathcal{G})dP = \int_{G^+\cap\Omega} dP = P(G^+), \tag{10.68}$$