Chapter 5

Conditioning

SECTION 1 considers the elementary case of conditioning on a map that takes only finitely many different values, as motivation for the general definition.

SECTION 2 defines conditional probability distributions for conditioning on the value of a general measurable map.

SECTION 3 discusses existence of conditional distributions by means of a slightly more general concept, disintegration, which is essential for the understanding of general conditional densities.

SECTION 4 defines conditional densities. It develops the general analog of the elementary formula for a conditional density: (joint density)/(marginal density).

*SECTION *5 illustrates how conditional distributions can be identified by symmetry considerations. The classical Borel paradox is presented as a warning against the misuse of symmetry.*

SECTION 6 discusses the abstract Kolmogorov conditional expectation, explaining why it is natural to take the conditioning information to be a sub-sigma-field.

*SECTION *7 discusses the statistical concept of sufficiency.*

1. Conditional distributions: the elementary case

In introductory probability courses, conditional probabilities of events are defined as ratios, $\mathbb{P}(A \mid B) = \mathbb{P}AB/\mathbb{P}B$, provided $\mathbb{P}B \neq 0$. The division by $\mathbb{P}B$ ensures that $\mathbb{P}(\cdot \mid B)$ is also a probability measure, which puts zero mass outside the set B, that is, $\mathbb{P}(B^c \mid B) = 0$. The conditional expectation of a random variable X is defined as its expectation with respect to $\mathbb{P}(\cdot \mid B)$, or, more succinctly, $\mathbb{P}(X \mid B) = \mathbb{P}(XB)/\mathbb{P}B$. If $\mathbb{P}B = 0$, the conditional probabilities and conditional expectations are either left undefined or are extracted by some heuristic limiting argument. For example, if Y is a random variable with $\mathbb{P}\{Y = y\} = 0$ for each possible value y, one hopes that something like $\mathbb{P}(A \mid Y = y) = \lim_{\delta \to 0} \mathbb{P}(A \mid y \leq Y \leq y + \delta)$ exists and is a probability measure for each fixed y. Rigorous proofs lie well beyond the scope of the typical introductory course.

In applications of conditioning, the definitions get turned around, to derive probabilities and expectations from conditional distributions constructed by appeals to symmetry or modelling assumptions. The typical calculation starts from a