11.1 Regular conditional probabilities

We finish with a theorem regarding the existence of regular conditional distributions. Recall the definitions of regular conditional probability and regular conditional distribution.

Definition 11.1 (Regular conditional probability) Assume \((\Omega, \mathcal{F}, P)\) is a probability space. Let \(A \subseteq \mathcal{F}\) be a sub-\(\sigma\)-field. We say that the function \(\Pr(\cdot | C)(\cdot) : A \times \Omega \rightarrow [0,1]\) is a regular conditional probability if

1. \(\forall A \in A, \Pr(A | C)(\cdot)\) is a version of \(\mathbb{E}[1_A | C]\).
2. For \(P\) a.e. \(\omega \in \Omega, \Pr(\cdot | C)(\omega)\) is a probability measure on \((\Omega, A)\).

Definition 11.2 (Regular conditional distribution) Assume \((\Omega, \mathcal{F}, P)\) is a probability space. Let \(X\) be random variable such that \(X : (\Omega, \mathcal{F}) \rightarrow (X, \mathcal{B})\). For each \(B \in \mathcal{B}\), let

\[
\mu_{X|C}(B | C)(\omega) = \Pr(X^{-1}(B) | C)(\omega).
\]

Then the function \(\mu_{X|C}(\cdot | C) : B \times \Omega \rightarrow [0,1]\) is called a regular conditional distribution of \(X\) given \(C\) if

1. \(\forall B \in \mathcal{B}, \mu_{X|C}(B | C)(\cdot)\) is a version of \(\mathbb{E}[1_{X \in B} | C]\).
2. For \(P\) a.e. \(\omega \in \Omega, \mu_{X|C}(\cdot | C)(\omega)\) is a probability measure on \((X, \mathcal{B})\).

Regular conditional distributions are useful as they allow us to compute the conditional expectations of all functions of a random variable \(X\) simultaneously and to generalize the properties of ordinary expectation in a more straightforward way [D10]. See Exercises 5.1.14 and 5.1.15 in [D10].

Theorem 11.3 (Existence of regular conditional distribution) Regular conditional distributions exist if \((X, \mathcal{B})\) is nice, i.e., there is a 1-1 map \(\varphi : X \rightarrow \mathbb{R}\) such that \(\varphi\) and \(\varphi^{-1}\) are measurable.
Some examples of nice spaces are:

1. \((X, \mathcal{B})\) where \(X\) is any topological space and \(\mathcal{B}\) is the Borel \(\sigma\)-field.
2. \((\mathbb{R}^n, \mathcal{B}^n)\)
3. The space \(C[0,1]\) of continuous functions on \([0,1]\) endowed with the sup norm.
4. Polish spaces endowed with the Borel \(\sigma\)-field.

### 11.2 Bayes’ theorem

Assume \((\Omega, \mathcal{F}, P)\) is a probability space. Let \(X\) and \(\Theta\) be random variables such that \(X : (\Omega, \mathcal{F}) \to (X, \mathcal{B})\) and \(\Theta : (\Omega, \mathcal{F}) \to (T, \tau)\). Assume that there exists a regular conditional distribution of \(X\) given \(\Theta\) denoted by \(\mu_{X|\Theta}(\cdot|\theta)\). Assume that there exists a \(\sigma\)-finite measure \(\nu\) on \((X, \mathcal{B})\) such that \(P_{\theta} \ll \nu\). Let \(f_{X|\Theta}(x|\theta) = \frac{d\mu_{X|\Theta}}{d\nu}\).

In fact, let \(P_{\theta}(B) = \mu_{X|\Theta}(B|\theta)\) for all \(\theta \in T\) and \(B \in \mathcal{B}\). Then \(P = \{P_{\theta} : \theta \in T\}\) is called a statistical model where \(T\) is the parameter space, \(f_{X|\Theta}(x|\theta)\) is the likelihood function, and \(\mu_{\Theta}\) (distribution of \(\Theta\)) is the prior.

**Theorem 11.4 (Bayes’ theorem)** Assume the structure above. Let \(\mu_{X|\Theta}\) be the conditional distribution of \(\Theta\) given \(X\). Then

1. \(\mu_{\Theta|X} \ll \mu_{\Theta}\) a.e. with respect to the distribution of \(X\).
2. \(\frac{d\mu_{\Theta|X}}{d\mu_{\Theta}} = \frac{f_{X|\Theta}(x|\theta)}{\int_{\tau} f_{X|\Theta}(x|\theta)d\mu_{\Theta}(\theta)}\) for all \(x\) for which the denominator is neither 0 nor \(\infty\).

Note that Bayes’ theorem is not always applicable. An example where Bayes’ theorem does not apply: (Example 1.36 in [S95])

Consider the case when the conditional distribution of \(X\) given \(\Theta = \theta\) is discrete with \(P_{\theta}(\{\theta - 1\}) = P_{\theta}(\{\theta + 1\}) = \frac{1}{2}\). Suppose that \(\Theta\) has a density \(f_{\Theta}\) with respect to the Lebesgue measure. The distributions \(P_{\theta}\) are not all absolutely continuous with respect to a single \(\sigma\)-finite measure. It is still possible to verify that the posterior distribution of \(\Theta\) given \(X = x\) is the discrete distribution with

\[
P(\Theta = x - 1|X = x) = \frac{f_{\Theta}(x - 1)}{f_{\Theta}(x - 1) + f_{\Theta}(x + 1)}\]
\[
= 1 - P(\Theta = x + 1|X = x)
\]

Note that the posterior is not absolutely continuous with respect to the prior.

### 11.3 Martingales

Martingales are sequences of dependent random variables. Martingales originated from gambling where you can adjust the next bet according to previous outcomes. The two main take aways from this section are:

1. Optional sampling theorem
2. Martingale convergence theorem
Definition 11.5 (Martingale) Let $\mathcal{F}_n$ be a filtration, i.e., an increasing sequence of $\sigma$-fields. A sequence of random variables $X_n$ is said to be adapted to $\mathcal{F}_n$ if $X_n$ is $\mathcal{F}_n$ measurable for all $n$. If $X_n$ is a sequence adapted to $\mathcal{F}_n$ such that for all $n$:

1. $\mathbb{E}[|X_n|] < \infty$
2. $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$

then $X_n$ is said to be a martingale (with respect to $\mathcal{F}_n$). If $=$ in condition 2 is replaced by $\leq$ or $\geq$, then $X_n$ is said to be a submartingale or a supermartingale, respectively.

Note that for martingales $\mathbb{E}[X_n] = \mathbb{E}[X_1]$. We give some examples of martingales below.

1. Sum of independent random variables

Let $Y_n$ be a sequence of independent random variables such that $\mathbb{E}[Y_n] = 0$. Let $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$ and $X_n = \sum_{i=1}^{n} Y_i$. Then $X_n$ is a martingale with respect to $\mathcal{F}_n$. This follows because:

$$
\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}
\left[
\sum_{i=1}^{n+1} Y_i | \mathcal{F}_n
\right]
\quad = \sum_{i=1}^{n+1} \mathbb{E}[Y_i | \mathcal{F}_n]
\quad = \sum_{i=1}^{n} Y_i + \mathbb{E}[Y_{n+1}|\mathcal{F}_n]
\quad = X_n + \mathbb{E}[Y_{n+1}]
\quad = X_n
$$

Note that if $\mathbb{E}[Y_n] \leq 0$ or $\mathbb{E}[Y_n] \geq 0$, then $X_n$ is a submartingale or a supermartingale, respectively.

2. Levy martingale

Let $\mathcal{F}_n$ be a filtration and $X$ be a random variable with finite mean. Define $X_n = \mathbb{E}[X | \mathcal{F}_n]$. Then $X_n$ is martingale, sometime called a Levy martingale. This follows from tower property:

$$
\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X | \mathcal{F}_n] | \mathcal{F}_n]
\quad = \mathbb{E}[X | \mathcal{F}_n]
\quad = X_n
$$

3. Gambler’s ruin

Consider the example 1 of sequence of independent random variables $Y_n$ again. Think of $Y_n$ as the amount a gambler wins on the $n$th round in a sequence of fair games. Let $Y_0$ be the initial fortune. Assume it is a known value for simplicity. Suppose that the gambler comes up with a system for determining how much to bet on the $n$th round denoted by $W_n \geq 0$. Assume that $W_n$ is measurable with respect to $\mathcal{F}_{n-1}$ for each $n$. This condition forces the gambler to choose the bet before knowing the outcome. Define $Z_n = Y_0 + \sum_{i=1}^{n} W_i Y_i$. Then $Z_n$ is a martingale. This is because:

$$
\mathbb{E}[W_{n+1}Y_{n+1}|\mathcal{F}_n] = W_{n+1} \mathbb{E}[Y_{n+1}|\mathcal{F}_n]
\quad = 0
$$

Note that if started with a submartingale or a supermartingale, we will end up with a submartingale or a supermartingale, respectively. In summary, this results says that a gambling system can not change whether a game is favorable, fair, or unfavorable to a gambler.
References
