Probability kernels give us a general tool to identify
$\mathbb{E}[f(X,Y) \mid X]$.

Theorem: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(S_1, \mathcal{B}_1), (S_2, \mathcal{B}_2)$ measurable spaces, and $X : (\Omega, \mathcal{F}) \to (S_1, \mathcal{B}_1)$, $Y : (\Omega, \mathcal{F}) \to (S_2, \mathcal{B}_2)$ random variables.

Let $\mu_{X,Y} = (X \times Y)^* \mathbb{P} \in \mathrm{Prob}(S_1 \times S_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ and $\mu_X = X^* \mathbb{P} \in \mathrm{Prob}(S_1, \mathcal{B}_1)$.

If there exists a probability kernel Q on $(S_1, \mathcal{B}_1), (S_2, \mathcal{B}_2)$ s.t.

$$\mu_{X,Y} = \mu_X \otimes Q$$

then for any $f \in L^1(S_1 \times S_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_{X,Y})$,

$$\mathbb{E}[f(X,Y) \mid X = x] = \int_{S_2} f(x,y)\, Q(x, dy)$$

To be clear, the claim is that $\mathbb{E}[f(X,Y)\,|X] = g(X)$ where

$$g(x) = \int_{S_2} f(x,y)\, Q(x,dy)$$

Pf. Suffices to show that

$$\mathbb{E}[f(X,Y)h(X)] = \mathbb{E}[g(X)h(X)] \quad \forall h \in \mathbb{B}(S_1, \mathcal{B}_1)$$

Start by assuming $f \in \mathbb{B}(S_1 \times S_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$

$$\mathbb{E}[f(X,Y)h(X)] = \int_{S_1 \times S_2} f(x,y)h(x)\, \mu_{X,Y}(dx\,dy)$$

If $f \in L^1(S_1 \times S_2, \mathcal{B}_1 \otimes \mathcal{B}_2, \mu_{X,Y})$, let $f_n = f \mathbb{1}_{|f| \le n}$

Then $f_n \to f$ in $L^1(\mu_{X,Y})$

$\therefore$ $f_n(X,Y) \to f(X,Y)$ in $L^1(\mathbb{P})$.

$\therefore$ $\int f_n(x,y) Q(x,dy) \big|_{x=X} = E[f_n(X,Y) | X] \to E[f(X,Y)|X]$ in $L^1(\mathbb{P})$

But also $E\left[\left| \int f_n(x,y) Q(x,dy) \big|_{x=X} - \int f(x,y) Q(x,dy) \big|_{x=X} \right|\right]$

$\therefore$ $E[f(X,Y)|X] = \int f(x,y) Q(x,dy) \big|_{x=X}$.

This leaves us with the question:
can we always find a probability kernel Q
s.t. $\mu_{X,Y} = \mu_X \otimes Q$?

Note: if so, it is "unique"

$$\mu_X \otimes Q = \mu_X \otimes \tilde{Q}$$

Def: Given random variables $X : (\Omega, \mathcal{F}) \to (S_1, \mathcal{B}_1)$
$Y : (\Omega, \mathcal{F}) \to (S_2, \mathcal{B}_2)$

a regular conditional distribution (RCD) of Y given X is a probability kernel Q on $(S_1, \mathcal{B}_1), (S_2, \mathcal{B}_2)$ s.t.

$$\mathbb{P}(Y \in B_2 \mid X) = Q(X, B_2) \quad \forall B_2 \in \mathcal{B}_2$$

I.e. $\mathbb{E}[h(Y) \mid X] = \int h(y)\, Q(x, dy) \Big|_{x = X} \quad \forall h \in B(S_2, \mathcal{B}_2)$.

Note: Some authors prefer to just define it as $\mathbb{P}(B_2 \mid X) = Q(X, B_2)$,
i.e. Q is the RCD of X given $\mathcal{B}_2$.

RCDs look weaker than what we want:

if Q is an RCD for Y given X,

$$E[h(Y)|X] = \int h(y)\, Q(X,dy)$$

Want $$E[f(X,Y)|X] = \int f(X,y)\, Q(X,dy)$$

Prop: If Q is a RCD for Y given X, then $\mu_{X,Y} = \mu_X \otimes Q$

(and hence Q is unique $[\mu_X]$-a.s.).

Pf. Exercise.

Eg. If $f(x,y) = g(x)h(y)$,

$$E[f(X,Y)|X] =$$

$$\int f(X,y)\, Q(X,dy) =$$

So the proposition is a jazzed up version of the product rule.

Question: Given $X : (\Omega, \mathcal{F}) \to (S_1, \mathcal{B}_1)$
$Y : (\Omega, \mathcal{F}) \to (S_2, \mathcal{B}_2)$

does there always exist a RCD of Y given X?

NO !

But the answer is yes, if $(S_2, \mathcal{B}_2)$ is nice enough

"standard Borel space"
measure isomorphic to a Borel set in $\mathbb{R}^d$
equipped with its Borel σ-field.

See [Driver, §19.5]

Eg. $(S_1, \mathcal{B}_1, \nu_1)$, $(S_2, \mathcal{B}_2, \nu_2)$ σ-finite measure spaces.

$(\Omega, \mathcal{F}, \mathbb{P})$ a probability space,

$$X : (\Omega, \mathcal{F}) \to (S_1, \mathcal{B}_1)$$
$$Y : (\Omega, \mathcal{F}) \to (S_2, \mathcal{B}_2)$$ random variables

and suppose $\mu_{X,Y} = (X,Y)^* \mathbb{P} \in \mathrm{Prob}(S_1 \times S_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ is $\ll \nu_1 \otimes \nu_2$.

Let $\rho_{X,Y} = \dfrac{d\mu_{XY}}{d\,\nu_1 \otimes \nu_2}$, and $\rho_X(x) = \int \rho_{X,Y}(x,y)\, \nu_2(dy)$.

Set $\rho_{Y|X}(y|x) := \mathbb{1}_{0 < \rho_X(x) < \infty} \dfrac{\rho_{X,Y}(x,y)}{\rho_X(x)}$.

Then $Q(x, B) = \int_B \rho_{Y|X}(y|x)\, \nu_2(dy)$ is a RCD for Y given X.

(The proof is essentially identical to the calculation at the beginning of [Lecture 33.1] that motivated the whole discussion of probability kernels. See [Driver, Prop 19.23] for details.

Eg. X, Y discrete.

$$\left.\begin{aligned} X &: \Omega \to S_1 \\ Y &: \Omega \to S_2 \end{aligned}\right\} \text{ countable, } \quad \begin{aligned} &P(X=x) > 0 \ \forall x \in S_1 \\ &P(Y=y) > 0 \ \forall y \in S_2 \end{aligned}$$

Then $\mathbb{P}(Y \in B \mid X = x) = \sum_{y \in B} \mathbb{P}(Y = y \mid X = x)$

$\therefore \ \mathbb{P}(Y \in B \mid X) = Q(X, B)$

where $Q(x, B) = \sum_{y \in B} \mathbb{P}(Y = y \mid X = x)$

Exercise.

I.e. $Q(x, B) = \sum_{y \in B} Q(x, y)$ where $Q(x, y) = \mathbb{P}(Y = y \mid X = x)$