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.

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

$\leftarrow$ equiv of $\rho_{X,Y} = \rho_X \cdot \rho_{Y|X}$

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

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

I.e. $\mathbb{E}[f(X,Y)] = \int_{S_1} \mu(dx) \int_{S_2} f(x,y)\, Q(x, dy) \quad \forall f \in B(S_1 \times S_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$

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

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

Pf. Suffices to show that

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

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

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

$$= \int_{S_1} \mu_X(dx) \int_{S_2} f(x,y)h(x)\, Q(x,dy) = \int_{S_1} h(x)\,\mu_X(dx) \overbrace{\int_{S_2} f(x,y)\, Q(x,dy)}^{g(x)}$$

$$= \int_{S_1} h(x)\, g(x)\, \mu_X(dx)$$

$$= E[h(X)g(X)] \quad \checkmark$$

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

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

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

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

as just shown (b/c $E[\,\cdot\,|X]$ is L^1-continuous.

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

$$= \int \mu_X(dx) \left| \int (f_n(x,y) - f(x,y)) Q(x,dy) \right| \qquad \text{c.o.v.}$$

$$\le \int \mu_X(dx) \int |f_n(x,y) - f(x,y)| \, Q(x,dy)$$

$$= \int |f_n - f| \, d(\mu_X Q) = \int |f_n - f| \, d\mu_{X,Y} \overset{\text{c.o.v.}}{=} \|f_n(X,Y) - f(X,Y)\|_{L^1(\mathbb{P})} \longrightarrow 0.$$

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

///

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" (if $\mathcal{B}_2$ is countably generated)

$$\mu_X \otimes Q = \mu_X \otimes \tilde{Q} \Rightarrow Q(x,\cdot) = \tilde{Q}(x,\cdot) \text{ for } [\mu_X]\text{-a.e. } x.$$

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.

$$E[\mathbb{1}_{Y\in B_2} | X] =: \mathbb{P}(Y \in B_2 | X) = Q(X, B_2) \quad \forall B_2 \in \mathcal{B}_2$$

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

Note: Some authors prefer to just define it as $\mathbb{P}(B_2|X) = Q(X, B_2)$, 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 (μ_X)-a.s.).

Pf. Exercise.

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

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

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

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.

Like undergrad prob., there are two cases when you can explicitly construct the RCD: "continuous" and "discrete"

See [Driver, §19.5]

(non-constructive proof; similar to proof of Skorohod's theorem.)

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, → required for Radon-Nikodym thm.

$X : (\Omega, \mathcal{F}) \to (S_1, \mathcal{B}_1)$
$Y : (\Omega, \mathcal{F}) \to (S_2, \mathcal{B}_2)$ 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 $\rho_{X,Y} = \dfrac{d\mu_{X,Y}}{d\nu_1 \otimes \nu_2}$, and $\rho_X(x) = \int \rho_{X,Y}(x,y)\, \nu_2(dy)$. $\therefore d\mu_X = \rho_X d\nu_1$ by Tonnelli.

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.

E.g. X, Y discrete.

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

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

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

So $\mathbb{P}(Y \in B \mid X) = \mathbb{P}(Y = y \mid X = x) \big|_{x = X}$

very confusing $\longrightarrow \quad \neq \mathbb{P}(Y = y \mid X = X)$

First compute the 2-var. function $Q(x, y) = \mathbb{P}(Y = y \mid X = x)$

Then evaluate $Q(X, y)$.

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