(8.1) Suppose probability measures satisfy $\pi \ll \nu \ll \mu$. Show that

$$\frac{d\pi}{d\mu} = \frac{d\pi}{d\nu} \frac{d\nu}{d\mu}$$

**proof** Let $h_{\pi,\mu}$ be the Radon-Nikodym derivative of $\pi$ with respect to $\mu$, and similarly for the other measures. First, let us prove that for any integrable function $f$

$$\int f\,d\pi = \int h_{\pi,\mu}\,d\mu$$

this is true for indicator functions, since $\int 1_A\,d\pi = \pi(A) = \int_A h_{\pi,\mu}\,d\mu$ by definition of the Radon-Nikodym derivative. Then by linearity of the integral, the claim is also true for simple functions. Then we can show the claim for non-negative measurable functions, by approximating by simple functions. Indeed, if $f \geq 0$ is measurable, then there exists a sequence of simple functions $0 \leq s_n \uparrow f$ such that $\int s_n\,d\pi \uparrow \int f\,d\pi$. For all $n$, we have $\int s_n\,d\pi = \int s_n h_{\pi,\mu} = \int s_n h_{\pi,\mu}\,d\mu$ where $0 \leq s_n h_{\pi,\mu} \uparrow f h_{\pi,\mu}$ since $h_{\pi,\mu} \geq 0$. By the Monotone Convergence theorem, we have $\int f\,d\pi = \int f h_{\pi,\mu}\,d\mu$. Finally, to show the claim for any integrable function $f$, we can write $f = f^+ - f^-$ where each term is non-negative, and apply the previous result to each term.

Now we have for any integrable function $f$

$$\int f\,d\pi = \int f h_{\pi,\mu}\,d\mu$$

(1)

and applying the result to the measures $\pi \ll \nu$, we have

$$\int f\,d\pi = \int f h_{\pi,\nu}\,d\nu$$

where $f h_{\pi,\nu}$ is $\nu$-integrable. Applying the result again to $\nu \ll \mu$, we have

$$\int f h_{\pi,\nu}\,d\nu = \int f h_{\pi,\nu} h_{\nu,\mu}\,d\mu$$

(2)

From equalities (1) and (2), we have for all integrable functions $f$

$$\int f(h_{\pi,\mu} - h_{\pi,\nu} h_{\nu,\mu})\,d\mu = 0$$

in particular, taking $f = h_{\pi,\mu} - h_{\pi,\nu} h_{\nu,\mu}$ (this is an integrable function), we have

$$\int (h_{\pi,\mu} - h_{\pi,\nu} h_{\nu,\mu})^2\,d\mu = 0$$

therefore $h_{\pi,\mu} - h_{\pi,\nu} h_{\nu,\mu} = 0$ a.e.

(Indeed, if a non-negative function $g$ is such that $\int g\,d\mu = 0$, then $\{\omega : g(\omega) \geq 0 = \cup_{n \geq 1} \{\omega : g(\omega) > \frac{1}{n}\}$, and by Markov’s inequality, $\mu(\{\omega : g(\omega) > \frac{1}{n}\}) = P(g > \frac{1}{n}) \leq \frac{1}{n} \int g\,d\mu = 0$, thus the set $\{\omega : g(\omega) > 0\}$ is the countable union of null sets, and is a null set).
In the setting of Theorem 7 [hard part], where $S_2$ is nice, show that $Q$ is unique in the following sense. If $Q^*$ is another conditional probability kernel for $\mu$, then

$$\mu_1\{x : Q^*(x, B) = Q(x, B) \text{ for all } B \in S_2\} = 1$$

**proof**  
First, by definition of a nice measure space, we can reduce the problem to the case where $S_2 = \mathbb{R}$ and $S_2 = B$ the collection of Borel sets.

First, we show the result for a countable generator of $B$, namely the collection $C = \{(-\infty, c), c \in \mathbb{Q}\}$. This is indeed a generator since for any real number $a$, there exists a non-decreasing sequence of rationals $(c_n)$ such that $c_n \uparrow a$, thus $(-\infty, a) = \bigcup_{n \in \mathbb{N}} (\infty, c_n) \in \sigma(C)$.

Now consider the set

$$X = \{x : Q^*(x, (-\infty, c)) = Q(x, (-\infty, c)) \text{ for all } c \in \mathbb{Q}\}$$

where we have by definition of the conditional probability Kernels, for any product-integrable function,

$$\int_{x \in S_1} \int_{y \in S_2} f(x, y)Q(x, dy)\mu_1(dx) = \int_{x \in S_1} \int_{S_2} f(x, y)Q^*(x, dy)\mu_1(dx)$$

Now fix $c$, and take $f(x, y) = [Q(x, (-\infty, c)) - Q^*(x, (-\infty, c))]1_{(-\infty, c)}(y)$. Then we have

$$\int_{x \in S_1} [Q(x, (-\infty, c)) - Q^*(x, (-\infty, c))]^2\mu_1(dx) = 0$$

therefore

$$\mu_1(\{x : Q(x, (-\infty, c)) \neq Q^*(x, (-\infty, c))\}) = 0$$

and $X^c$ is simply the countable union of such null sets, thus $\mu_1(X^c) = 0$, i.e. $\mu_1(X) = 1$.

This shows the result for a countable generator set. To show the result when $B \in S_2$, it suffices to show that

$$X \subseteq \{x : Q(x, B) = Q^*(x, B) \text{ for all } B \in S_2\}$$

let $x \in X$. We have $Q(x, \cdot)$ and $Q^*(x, \cdot)$ are both probability measures on $S_2$, and by definition of $X$, they agree on $C$, which is a generator of the Borel $\sigma$-field $B$, therefore they agree on $B$ (by a familiar $\pi - \lambda$ argument), i.e. $Q(x, B) = Q^*(x, B)$ for all $B \in B$. This proves the desired inclusion, and the result follows since

$$\mu_1\{x : Q^*(x, B) = Q(x, B) \text{ for all } B \in S_2\} \geq \mu_1(X) = 1$$
(8.3) Let $F$ be a distribution function. Let $c > 0$. Find a simple formula for
\[
\int_{-\infty}^{\infty} (F(x + c) - F(x))dx
\]

Claim:
\[
\int_{-\infty}^{\infty} (F(x + c) - F(x))dx = c
\]

**proof** Let $\mu$ be the probability measure associated to the distribution function $F$. Then we have for all $x$, $F(x) = \mu((\infty, x])$. Therefore we can write
\[
\int_{\mathbb{R}} (F(x + c) - F(x))dx = \int_{\mathbb{R}} \int_{\mathbb{R}} (1_{(y \leq x+c)} - 1_{(y \leq x)})d\mu(y)dx
\]
\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{(y \leq x+c \& y > x)}d\mu(y)dx
\]
\[
= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{(y \leq x+c \& y > x)}dxd\mu(y)
\]
\[
= \int_{\mathbb{R}} \int_{[y-c,y)} dxd\mu(y)
\]
\[
= \int_{\mathbb{R}} c d\mu(y)
\]
\[
= c
\]

which proves the claim.
In the proof of Corollary 8, we used the inverse distribution function
\[ f(x, u) = \inf \{ y : u \leq Q(x, (-\infty, y]) \} \]
associated with the kernel \( Q \). Show that \( f \) is product-measurable.

**proof** First, we show that \( f \) coincides with \( \hat{f} \), defined as \( \hat{f}(x, u) = \inf \{ y \in Q : u \leq Q(x, (-\infty, y]) \} \). We clearly have \( f(x, u) \leq \hat{f}(x, u) \) since the infimum is taken on a larger set for \( f \). Thus it suffices to show that for all \( x, u \), \( f(x, u) \geq \hat{f}(x, u) \). Fix \( x, u \). We have by definition of \( f(x, y) \) as the inf, for all \( \epsilon > 0 \), there exists \( y < f(x, u) + \epsilon \) such that \( u \leq Q(x, (-\infty, y]) \). By density of \( Q \) in \( \mathbb{R} \), there exists \( y' \in Q \) such that \( y \leq y' \leq f(x, u) + \epsilon \), and since \( Q(x, \cdot) \) is a probability measure, we have
\[ Q(x, (-\infty, y')) \geq Q(x, (-\infty, y]) \geq u \]
therefore \( y' \geq \hat{f}(x, u) \). Combining the previous inequalities, we have
\[ \hat{f}(x, y) \leq y' \leq f(x, y) + \epsilon \]
thus \( \hat{f}(x, y) \leq f(x, y) + \epsilon \). Since this holds for arbitrary \( \epsilon > 0 \), we have \( \hat{f}(x, u) \leq f(x, u) \). This proves the first claim.

Now to show that \( f \) (or equivalently \( \hat{f} \)) is product-measurable, it suffices to show that for all \( c \in \mathbb{R} \), \( \hat{f}^{-1}(c, +\infty) \) is in the product \( \sigma \)-field. Fix \( c \in \mathbb{R} \). We have
\[ (x, u) \in \hat{f}^{-1}(c, +\infty) \iff \inf \{ y \in Q : u \leq Q(x, (-\infty, y]) \} > c \]
\[ \iff \forall y \in Q \text{ with } y \leq c, \ u > Q(x, (-\infty, y]) \]
\[ \iff \forall y \in Q \cap [c, +\infty), \ g_y(x, u) > 0 \]
where \( g_y(x, u) = u - Q(x, (-\infty, y]) \) is the difference of two product-measurable functions, thus is measurable. Therefore we can write
\[ \hat{f}^{-1}(c, +\infty) = \cap_{y \in Q \cap [c, +\infty)} g_y^{-1}(0, \infty) \]
which proves it is product-measurable as the countable intersection of product measurable sets \( g_y^{-1}(0, +\infty) \). The proof is complete.
Given a triple \((X_1, X_2, X_3)\), we can define 3 probability measures \(\mu_{12}, \mu_{13}, \mu_{23}\) on \(\mathbb{R}^2\) by

\[
\mu_{ij} \text{ is the distribution of } (X_i, X_j)
\]

These probability measures satisfy a consistency condition: the marginal distribution \(\mu_1\) obtained from \(\mu_{12}\) must coincide with the marginal distribution obtained from \(\mu_{13}\), and similarly for the marginal distributions \(\mu_2\) and \(\mu_3\).

Give an example to show that the converse is false. That is, give an example of \(\mu_{12}, \mu_{13}, \mu_{23}\) that satisfy the consistency conditions, but for which there does not exists a triple \((X_1, X_2, X_3)\) satisfying (3)

**answer** Assume \(\Omega\) is finite, and consider the following joint distributions \(\mu_{12} = \mu_{23} = \mu_{12}\) such that \(\mu_{1,2}(x_1, x_2)\) is given by the following table

\[
\begin{array}{ccc}
0 & 1 \\
0 & 0 & \frac{2}{5} \\
1 & \frac{1}{5} & \frac{1}{5}
\end{array}
\]

where columns are values of \(x_1\), and rows are values of \(x_2\).

The consistency conditions are satisfied, since we have for example

\[
\mu_{1,2}((0, 0)) + \mu_{1,2}((0, 1)) = \mu_{1,3}((0, 0)) + \mu_{1,3}((0, 1)) = \frac{2}{5}
\]

and similarly for other conditions. This is by construction of \(\mu_{i,j}\).

The claim is that there exists no random variables \(X_1, X_2, X_3\) such that \(\mu_{i,j}\) is the joint distribution of \((X_i, X_j)\) for all \(i, j\). Indeed, assume by contradiction that such variables exist, and let \(\mu_{1,2,3}\) be their joint distribution. First, since \(\Omega\) is finite, the image set \((X_1, X_2, X_3)(\Omega)\) is finite. Denote it by \(X\). Then we can write

\[
\mu_{1,2,3}(B) = \sum_{\xi \in \mathcal{X}} \mu_{1,2,3}(\xi) 1(\xi \in B)
\]

where \(a_\xi \geq 0\). Next, we argue that \(a_\xi = 0\) for all \(\xi \notin \{0, 1\}^3\). Indeed, if \((x_1, x_2) \notin \{0, 1\}^2\), then \(\mu_{1,2}(x_1, x_2) = 0\) by definition of \(\mu_{1,2}\), thus \(\sum_{x_3} \mu_{1,2,3}(x_1, x_2, x_3) = 0\). Similarly for \((x_2, x_3) \notin \{0, 1\}^2\) and \((x_1, x_3) \notin \{0, 1\}^2\). Now we can write the marginal distribution conditions

\[
\begin{align*}
\mu_{1,2,3}(0, 0, 0) + \mu_{1,2,3}(0, 0, 1) &= \mu_{1,2}(0, 0) = 0 \\
\mu_{1,2,3}(0, 1, 0) + \mu_{1,2,3}(0, 1, 1) &= \mu_{1,2}(0, 1) = \frac{2}{5} \\
\mu_{1,2,3}(1, 0, 0) + \mu_{1,2,3}(1, 0, 1) &= \mu_{1,2}(1, 0) = \frac{1}{5} \\
\mu_{1,2,3}(1, 1, 0) + \mu_{1,2,3}(1, 1, 1) &= \mu_{1,2}(1, 1) = \frac{1}{5}
\end{align*}
\]

\[
\begin{align*}
\mu_{1,2,3}(0, 0, 0) + \mu_{1,2,3}(0, 1, 0) &= \mu_{1,3}(0, 0) = 0 \\
\mu_{1,2,3}(0, 0, 1) + \mu_{1,2,3}(0, 1, 1) &= \mu_{1,3}(0, 1) = \frac{2}{5} \\
\mu_{1,2,3}(1, 0, 0) + \mu_{1,2,3}(1, 1, 0) &= \mu_{1,3}(1, 0) = \frac{2}{5} \\
\mu_{1,2,3}(1, 0, 1) + \mu_{1,2,3}(1, 1, 1) &= \mu_{1,3}(1, 1) = \frac{1}{5}
\end{align*}
\]
\[
\begin{align*}
\mu_{1,2,3}(0,0,0) + \mu_{1,2,3}(1,0,0) &= \mu_{2,3}(0,0) = 0 \\
\mu_{1,2,3}(0,0,1) + \mu_{1,2,3}(1,0,1) &= \mu_{2,3}(0,1) = \frac{2}{5} \\
\mu_{1,2,3}(0,1,0) + \mu_{1,2,3}(1,1,0) &= \mu_{2,3}(1,0) = \frac{2}{5} \\
\mu_{1,2,3}(0,1,1) + \mu_{1,2,3}(1,1,1) &= \mu_{2,3}(1,1) = \frac{1}{5}
\end{align*}
\]

then

\[
\begin{align*}
\mu_{1,2,3}(0,0,0) &= \mu_{1,2,3}(0,0,1) = \mu_{1,2,3}(0,1,0) = \mu_{1,2,3}(1,0,0) = 0 \\
\mu_{1,2,3}(0,1,1) &= \mu_{1,2,3}(1,0,1) = \mu_{1,2,3}(1,1,0) = \frac{2}{5} \\
\mu_{1,2,3}(1,1,0) + \mu_{1,2,3}(1,1,1) &= \mu_{1,2,3}(1,0,1) + \mu_{1,2,3}(1,1,1) = \mu_{1,2,3}(0,1,1) + \mu_{1,2,3}(1,1,1) = \frac{1}{5}
\end{align*}
\]

which implies that \( \mu_{1,2,3}(1,1,1) = -\frac{1}{5} \), contradiction. This provides a counter-example.