Solutions to Homework 2

September 24, 2018

Solution to Exercise 1.4.5: Note that $P_\lambda(\{n\}) > 0$ for all $n \in I\!N$, so if $B \subset I\!N$, then $P_\lambda(B) = \sum_{n\in B} P_\lambda(\{n\}), > 0$ unless $B = \emptyset$. The only P_λ null set is $\emptyset$, for all λ, so $P_\lambda \ll P_1$.

If we let μ be counting measure on $I\!N$, then $P_\lambda \ll \mu$ since for $B \subset I\!N$, $\mu(B) = 0$ implies $B = \emptyset$, which implies $P_\lambda(B) = 0$. The density is given by

$$f_\lambda(n) \;=\; \frac{dP_\lambda}{d\mu}(n) \;=\; e^{-\lambda}\lambda^n/n!, \quad \forall n \in I\!N.$$

To see this, note that for $B \subset I\!N$,

$$P_\lambda(B) \;=\; \sum_{n\in B} P_\lambda(\{n\}) \;=\; \int_B f_\lambda d\mu.$$

(Note: we don't actually need this, but it is probably the most natural way to think of this problem for statistics students.) Therefore,

$$P_\lambda(B) \;=\; = \;\int_B f_\lambda d\mu \;=\; \int_B \frac{f_\lambda}{f_1} f_1 d\mu \;=\; \int_B \frac{f_\lambda}{f_1} dP_1,$$

i.e.

$$\frac{dP_\lambda}{dP_1}(n) \;=\; \frac{f_\lambda(n)}{f_1(n)} \;=\; \frac{e^{-\lambda}\lambda^n/n!}{e^{-1}/n!} \;=\; e^{-\lambda+1}\lambda^n.$$

One could also use Proposition 1.4.2 (c) here.

Solution to Exercise 1.4.9: First of all note that $\ll$ is a transitive relationship, i.e., $\nu \ll \mu$ and $\mu \ll \lambda$ implies $\nu \ll \lambda$. To see this, if we let $\mathcal{N}(\mu)$ denote the collection of μ-null sets for any measure μ, then $\nu \ll \mu$ is the same as $\mathcal{N}(\nu) \supset \mathcal{N}(\mu)$, and $\mu \ll \lambda$ means $\mathcal{N}(\mu) \supset \mathcal{N}(\lambda)$, so the transitivity property of set inclusion gives us that $\mathcal{N}(\nu) \supset \mathcal{N}(\lambda)$ and hence that $\nu \ll \lambda$.

Turning to the more substantive part of the problem, we have that the defining property of $d\nu/d\mu$ says

$$\nu(A) \;=\; \int_A \left(\frac{d\nu}{d\mu}\right) d\mu \quad \forall A \text{ measurable.}$$

Since $\mu \ll \lambda$, we can apply part (a) of Proposition 1.4.2 to the r.h.s. and we obtain

$$\nu(A) = \int_A \left(\frac{d\nu}{d\mu}\right)\left(\frac{d\mu}{d\lambda}\right) d\lambda \quad \forall A \text{ measurable.}$$

However, this just state that $(d\nu/d\mu)(d\mu/d\lambda)$ satisfies the defining property of $d\nu/d\lambda$, which we know exists and is essentially unique by the Radon-Nikodym theorem since we are assuming μ and λ are σ-finite and we know from the previous paragraph that $\nu \ll \lambda$. Thus, we conclude

$$\frac{d\nu}{d\lambda} = \left(\frac{d\nu}{d\mu}\right)\left(\frac{d\mu}{d\lambda}\right), \quad \lambda - a.e.$$

Finally, assuming $\mu \simeq \nu$, i.e., that both $\nu \ll \mu$ and $\mu \ll \nu$, we can substitute ν for λ in the above and get

$$\begin{aligned} 1 &= \frac{d\nu}{d\nu}, \quad \nu - a.e. \\ &= \left(\frac{d\nu}{d\mu}\right)\left(\frac{d\mu}{d\nu}\right), \quad \nu - a.e. \end{aligned}$$

where the first equality follows since the constant function 1 satisfies the defining property to be the Radon-Nikodym derivative (namely, $\int_A 1 d\nu = \nu(A)$). Now the equality above shows that both factors in the last expression must be nonzero, ν-a.e. So the reciprocal $(d\mu/d\nu)^{-1}$ is well defined ν-a.e., and the result follows.

Solution to Exercise 1.4.11: **(a)** Assume of course that the μ_i are σ-finite, and that the joint density factors. Then

$$\begin{aligned} & P[X_1 \in A_1 \,\&\, X_2 \in A_2] \\ &= P[(X_1, X_2) \in A_1 \times A_2] \\ &= \int_{A_1 \times A_2} f(x_1, x_2) d(\mu_1 \times \mu_2)(x_1, x_2) \\ &= \int I_{A_1 \times A_2}(x_1, x_2) f(x_1, x_2) d(\mu_1 \times \mu_2)(x_1, x_2) \\ &= \int I_{A_1}(x_1) I_{A_2}(x_2) f_1(x_1) f_2(x_2) d(\mu_1 \times \mu_2)(x_1, x_2) \\ &= \int \int I_{A_1}(x_1) I_{A_2}(x_2) f_1(x_1) f_2(x_2) d\mu_1(x_1) d\mu_2(x_2) \end{aligned}$$

$$
\begin{aligned}
&= \int I_{A_2}(x_2) f_2(x_2) \int I_{A_1}(x_1) f_1(x_1) d\mu_1(x_1) d\mu_2(x_2) \\
&= \int I_{A_1}(x_1) f_1(x_1) d\mu_1(x_1) \int I_{A_2}(x_2) f_2(x_2) d\mu_2(x_2).
\end{aligned}
$$

In the second to last line, we factored out the functions of x_2 from the $d\mu_1(x_1)$ integral since they are "constants" (because the value of x_2 is held fixed when computing $\int \cdots d\mu_1(x_1)$). In the last line, we factored out the constant from the $\int \cdots d\mu_2(x_2)$ $\int I_{A_1}(x_1) f_1(x_1) d\mu_1(x_1)$ from the $\int \cdots d\mu_2(x_2)$ (since it doesn't involve x_2). We recognize the final expression as $P[X_1 \in A_1]P[X_2 \in A_2]$, and since A_1 and A_2 are arbitrary (measurable) sets, this shows X_1 and X_2 are independent.

(b) For n random variables X_1, ..., X_n such that $\mathrm{Law}[X_1, \ldots, X_n] \ll \mu = \mu_1 \times \cdots \times \mu_n$, where μ_i is on the range of X_i, and all the μ_i are σ-finite, then if the joint density $f_{1\cdots n} = d\mathrm{Law}[X_1, \ldots, X_n]/d\mu$ factors as

$$
f_{1\cdots n}(x_1, \ldots, x_n) \;=\; \prod_{i=1}^{n} f_i(x_i),
$$

(where of course $f_i = d\mathrm{Law}[X_i]/d\mu_i$), then X_1, ..., X_n are independent.

Essentially the same argument as in part (a) will work. We start with

$$
P[X_1 \in A_1 \,\&\cdots\&X_n \in A_n],
$$

write it in terms of a multiple integral using Fubini's theorem, and we can factor out each of the individual integrals of the form $\int_{A_i} f_i d\mu_i$, and express the probability above as $\prod_{i=1}^{n} P[X_i \in A_i]$.

Solution to Exercise 1.5.2: To verify Remark 1.5.4, we simply need to check that $I_{A_i} = I_{\{a_i\}}(Y)$, but this is immediate since the A_is are disjoint and the a_is are distinct, so $Y(\omega) = a_i$ if and only if $\omega \in A_i$.

To verify equation (1.70), if we define

$$
h(y) \;=\; \int_{\Lambda_1} g(x, y)\, f_{X|Y}(x|y)\, d\mu_1(x),
$$

then we observe from the previous equation that

$$
h(Y) \;=\; E[g(X, Y)|Y].
$$

Solution to Exercise 1.5.7: For part (a), let's start with

$$\begin{aligned}
&E[(X - E[X|\mathcal{G}])^2|\mathcal{G}] \\
&\quad = \quad E[X^2|\mathcal{G}] - 2E[E[X|\mathcal{G}]X|\mathcal{G}] + E[E[X|\mathcal{G}]^2|\mathcal{G}] \\
&\quad = \quad E[X^2|\mathcal{G}] - 2E[X|\mathcal{G}]E[X|\mathcal{G}] + E[X|\mathcal{G}]^2.
\end{aligned}$$

In the last equality, we used that $E[X|\mathcal{G}]$ is $\mathcal{G}$ measurable, so can be "factored" out from $E[\cdot|\mathcal{G}]$ (Theorem 1.5.7 (h)), and $E[X|\mathcal{G}]^2$ is $\mathcal{G}$ measurable, so it's conditional expectation w.r.t. $\mathcal{G}$ is itself (Theorem 1.5.7 (f)). Of course, now the $-2E[X|\mathcal{G}]^2$ in the middle term combines with the final term to give the desired result.

For part (b), we start with the definition of $\mathrm{Var}[X]$:

$$\begin{aligned}
&E[(X - E[X])^2] \\
&\quad = \quad E[(X - E[X|\mathcal{G}] + E[X|\mathcal{G}] - E[X])^2] \\
&\qquad\qquad\qquad \text{adding and subtracting } E[X|\mathcal{G}] \\
&\quad = \quad E[(X - E[X|\mathcal{G}])^2] + 2E[(X - E[X|\mathcal{G}])(E[X|\mathcal{G}] - E[X])] + E[(E[X] - E[X|\mathcal{G}])^2] \\
&\qquad\qquad\qquad \text{by algebra and linearity of expectation} \\
&\quad = \quad E[E[(X - E[X|\mathcal{G}])^2|\mathcal{G}]] + 2E[E[(X - E[X|\mathcal{G}])(E[X|\mathcal{G}] - E[X])|\mathcal{G}]] \\
&\qquad\qquad\qquad + E[(E[X|\mathcal{G}] - E[X])^2] \\
&\qquad\qquad\qquad \text{by total expectation (twice)} \\
&\quad = \quad E[\mathrm{Var}[X|\mathcal{G}]] + 2E[(E[X|\mathcal{G}] - E[X])E[(X - E[X|\mathcal{G}])|\mathcal{G}]] + E[(E[X|\mathcal{G}] - E[X])^2] \\
&\qquad\qquad\qquad \text{by the factorization result (Theorem 1.5.7 (h))} \\
&\quad = \quad E[\mathrm{Var}[X|\mathcal{G}]] + 2E[(E[X|\mathcal{G}] - E[X])(E[X|\mathcal{G}] - E[X|\mathcal{G}])] + E[(E[X|\mathcal{G}] - E[X])^2] \\
&\qquad\qquad\qquad \text{by linearity of } E[\cdot|\mathcal{G}] \text{ and Theorem 1.5.7 (f) applied to } E[X|\mathcal{G}] \\
&\quad = \quad E[\mathrm{Var}[X|\mathcal{G}]] + \mathrm{Var}[E[X|\mathcal{G}]] \\
&\qquad\qquad\qquad \text{since } E[E[X|\mathcal{G}]] = E[X] \text{ by total expectation.}
\end{aligned}$$

It follows immediately that $E[\mathrm{Var}[X|\mathcal{G}]] = \mathrm{Var}[X] - \mathrm{Var}[E[X|\mathcal{G}]]$. Also, since $\mathrm{Var}[E[X|\mathcal{G}]] \geq 0$, we have $E[\mathrm{Var}[X|\mathcal{G}]] \leq \mathrm{Var}[X]$.

For part (c), it is clear that if $X = E[X|\mathcal{G}]$ a.s., then $\mathrm{Var}[X|\mathcal{G}] = 0$, a.s. It seems reasonable that this would be necessary as well for $\mathrm{Var}[X|\mathcal{G}] = 0$, a.s. Assume $\mathrm{Var}[X|\mathcal{G}] = 0$, a.s., so

$$\begin{aligned}
0 &= E[\mathrm{Var}[X|\mathcal{G}]] \\
&= E[E[(X - E[X|\mathcal{G}])^2|\mathcal{G}]] \\
&= E[(X - E[X|\mathcal{G}])^2],
\end{aligned}$$

where the last equality follows by total expectation. But $(X - E[X|\mathcal{G}])^2$ is a nonnegative r.v., so its expectation being 0 implies $(X - E[X|\mathcal{G}])^2 = 0$, a.s., i.e., $X = E[X|\mathcal{G}]$ a.s. This shows $X = E[X|\mathcal{G}]$ a.s. is a necessary and sufficient condition for $\mathrm{Var}[X|\mathcal{G}] = 0$, a.s.

Under our supposition, we have from part (a) that

$$\begin{aligned} \mathrm{Var}[X|\mathcal{G}] &= E[X^2|\mathcal{G}] - E[X|\mathcal{G}]^2 \\ &= Y^2 - Y^2 \\ &= 0, \quad a.s. \end{aligned}$$

Thus, by part (c), $X = E[X|\mathcal{G}] = Y$, a.s.

Solution to Exercise 1.5.9: Part (a) is easy: we know $E[X|Y] = k$ if $X = k$, a.s. The function $\phi(y) \equiv k$ is Borel measurable from the range space of Y to $I\!R$, and $E[X|Y] = \phi(Y)$, a.s.

For part (b), if $X_1 \leq X_2$, a.s., then $\phi_1(Y) = E[X_1|Y] \leq E[X_2|Y] = \phi_2(Y)$, a.s., so we claim that $\phi_1(y) = E[X_1|Y = y] \leq E[X_2|Y = y] = \phi_2(y)$, P_Y-a.s. Letting $A = \{y : \phi_1(y) > \phi_2(y)\}$, we have $P_Y(A) = P(Y^{-1}(A)) = P(\{\omega : Y(\omega) \in A\}) = P(\{\omega : \phi_1(Y(\omega)) > \phi_2(Y(\omega))\}) = 0$.

For part (c), the obvious conjecture is that

$$E[a_1X_1 + a_2X_2|Y = y] = a_1E[X_1|Y = y] + a_2E[X_2|Y = y] \quad , P_Y \text{ a.s.}$$

We have from the theorem that

$$E[a_1X_1 + a_2X_2|Y] = a_1E[X_1|Y] + a_2E[X_2|Y] \quad , P \text{ a.s.}$$

Now we know that $E[a_1X_1 + a_2X_2|Y]$, $E[X_1|Y]$, and $E[X_2|Y]$ can each be expressed as a function of Y, say

$$\begin{aligned} E[a_1X_1 + a_2X_2|Y] &= h(Y) \\ E[X_1|Y] &= h_1(Y) \\ E[X_2|Y] &= h_2(Y). \end{aligned}$$

See the discussion beginning in the middle of p. 70. Our equation above then says

$$h(Y) = a_1h_1(Y) + a_1h_1(Y) \quad , P \text{ a.s.}$$

Also, by definition,

$$\begin{aligned} E[a_1X_1 + a_2X_2|Y = y] &= h(y) \\ E[X_1|Y = y] &= h_1(y) \\ E[X_2|Y = y] &= h_2(y). \end{aligned}$$

So, can't we conclude that

$$h(y) \;=\; a_1h_1(y) \;+\; a_1h_1(y) \quad , \; P_Y \text{ a.s.?}$$

Let

$$A \;=\; \{y : h(y) \neq a_1h_1(y) \;+\; a_1h_1(y)\}.$$

This is a subset of the range space of Y. It is measurable in that space since it is the inverse image of the Borel set $I\!R \setminus \{0\}$ under the measurable map $h - (a_1h_1 + a_2h_2)$. We want to show that $P_Y(A) = 0$. Now

$$P_Y(A) \;=\; P\left(Y^{-1}(A)\right) \;=\; P\left(\{\omega : h(Y(\omega)) \neq a_1h_1(Y(\omega)) \;+\; a_1h_1(Y(\omega))\}\right).$$

We already observed this latter event has probability 0 when we noted that $h(Y) = a_1h_1(Y) + a_1h_1(Y)$, P-a.s. So we are done.

The previous paragraph illustrates one way of proving a result for the conditional expectation of the type $E[X|Y = y]$: prove the corresponding result for for conditional expectation of the type $E[X|Y]$ and simply translate it over. This will work in most cases, so one doesn't have to do separate proofs. The P-null sets in Ω where an equality fails will automatically become a P_Y-null set on the range of Y by the same sort of argument as above. However, most students seem to want to derive a result using the defining properties of $E[X|Y = y]$. So, for example, we observe that $a_1E[X_1|Y = y] + a_2E[X_2|Y = y]$ is a Borel measurable function of y (whose domain is the range space of Y), and if A is a measurable set in the range space of Y, then

$$\begin{aligned} \int_{Y^{-1}(A)} (a_1X_1 + a_2X_2)\, dP &= \int_{Y^{-1}(A)} (a_1E[X_1|Y] \;+\; a_2E[X_2|Y])\; dP \\ &= \int_A (a_1E[X_1|Y = y] \;+\; a_2E[X_2|Y = y])\; dP_Y(y), \end{aligned}$$

which shows that $a_1E[X_1|Y = y] + a_2E[X_2|Y = y]$ has satisifies the integral property to be $E[a_1X_1 + a_2X_2|Y = y]$.

Moving on to part (d), it is very tempting to write "$E[E[X|Y = y]]$" but this doesn't make sense. For a r.v. Z, $E[Z] = \int_\Omega Z dP$ is an integral over the underlying probability space, but the domain of the function $E[X|Y = y]$ is the range of Y, not Ω. Of course, the range of Y has the probability measure P_Y, so it makes sense to write

$$\int_\Lambda E[X|Y = y]\, dP_Y(y) \;=\; E[X],$$

where Λ is the range of Y. Since $E[X|Y = y]|_{y=Y}$ is $E[X|Y]$, we have by the law of the unconscious statistician that

$$\int_\Lambda E[X|Y = y]\, dP_Y(y) \;=\; \int_{Y^{-1}(\Lambda)} E[X|Y]\, dP \;=\; E[E[X|Y]] \;=\; E[X]$$

where the last equality follows by part (d) of Theorem 1.5.7(d).

To deal with part (e) of the theorem, we need to translate $E[X|\{\emptyset, \Omega\}]$ into some kind of statement about $E[X|Y = y]$, we need to think of it as $E[X|Y]$ where $\sigma(Y) = \{\emptyset, \Omega\}$. But this happens if and only if Y is a constant r.v. (Check it out!). Hence, we claim

If Y is a constant r.v., $E[X|Y = y] \;=\; E[X], \quad P_Y$-a.s..

Clearly if $Y = c$ where $c \in \Lambda$ is fixed, then for $h(Y) = E[X|Y] = E[X]$, we must have $h(c) = E[X]$, and $h(y)$ can be defined arbitrarily for $y \neq c$. But this means $h(y) = E[X]$, P_Y-a.s., since $P_Y = \delta_c$.

Theorem 1.5.7(f) tells us that if $\sigma(X) \subset \sigma(Y)$, then $E[X|Y] \;=\; X$, a.s. Now, it wouldn't make sense to claim $E[X|Y = y]$ is equal to X, since they are functions with different domains. However, if $\sigma(X) \subset \sigma(Y)$, then we know from Theorem 1.5.1 that $X = \phi(Y)$ for some function ϕ whose domain is Λ, the range of Y. Thus, it would make sense to claim that

If $X = \phi(Y)$for some measurable ϕ, then $E[X|Y = y] \;=\; \phi(y), \quad P_Y - a.s.$

The proof is immediate from

$$\phi(Y) \;=\; X \;=\; E[X|Y], \quad a.s.$$

Part (g) is a little trickier. Suppose Y_1 is some other random element (with range space $(\Lambda_1, \mathcal{G}_1)$, say), and $\sigma(Y_1) \subset \sigma(Y)$. We know (at least if the range of Y_1 is $(I\!\!R, \mathcal{B})$ that then $Y_1 = \psi(Y)$, for some ψ, by Theorem 1.5.1. So let us just assume

$$Y_1 = \psi(Y), \quad \psi : (\Lambda, \mathcal{G}) \longrightarrow (\Lambda_1, \mathcal{G}_1).$$

Now it makes no sense to write "$E[E[X|Y_1 = y_1]|Y = y]$" since the domain of $E[X|Y_1 = y_1]$ is Λ_1, and not Ω. (Recall that when we write $E[Z|Y = y]$, Z must be a r.v., i.e. a mapping from Ω to $I\!\!R$.) Also, how are we to match up the given values y_1 and y? If we are given $Y = y$ and $Y_1 = \psi(Y)$, then it must be that $Y_1 = \psi(y)$. So let's try the following:

$$\textit{If } Y_1 = \psi(Y), \textit{then } E[E[X|Y_1]|Y = y] = E[X|Y_1 = \psi(y)], \quad P_{Y_1} - a.s.$$

Of course, we haven't fully interpreted $E[E[X|Y_1]|Y = y]$ into the requisite kind of conditional expectation, but there is no way to do so. Also, what does the r.h.s. mean? We are taking the function h_1 given by $h_1(y_1) = E[X|Y_1 = y_1]$ and composing it with ψ, i.e. the r.h.s. is $h \circ \psi$ evaluated at y, and $h \circ \psi$ is a map from Λ to $I\!\!R$, so the domains and ranges of the two sides match. Again, the proof is trivial: $E[X|Y_1] = h(Y_1) = h(\psi(Y)) = (h \circ \psi)(Y)$, so we apply our result above on part (f) of the theorem.

Now for the other side of the Law of Successive conditioning, we need to deal with $E[E[X|Y]|Y_1 = y_1]$ when $Y_1 = \psi(Y)$. We could just write

$$\textit{If } Y_1 = \psi(Y), \textit{then } E[E[X|Y]|Y_1 = y_1] = E[X|Y_1 = y_1], \quad P_{Y_1} - a.s.$$

Certainly everything makes sense: the l.h.s. and r.h.s. of the equation are the same kind of objects (both sides are functions with argument y_1 varying over the domain Λ_1 and the ranges are $I\!\!R$). And we know that $E[E[X|Y]|Y_1]$ $= E[X|Y_1]$ a.s. in this case, which proves the result. This one is easy.

Finally, part (h) is fairly straightforward: If X_1 is $\sigma(Y)$-measurable, then $X_1 = \psi(Y)$ for some measurable ψ, and we claim that $E[\psi(Y)X_2|Y = y] = \psi(y)E[X_2|Y = y]$. We see that $\phi(y) = \psi(y)E[X_2|Y = y]$ is a Borel measurable function on the range of Y (since it is the product of two such functions), and $\phi(Y) = \psi(Y)E[X_2|Y] = E[\psi(Y)X_2|Y]$, where the last equality follows from the result given in the theorem.

The translation and proof of Theorem 1.5.8 is straightforward. For part (a), if $0 \leq X_n \uparrow X$, then we claim $E[X_n|Y = y] \to E[X|Y = y]$, P_Y-a.s. We know that $\phi_n(Y) = E[X_n|Y] \to E[X|Y] = \phi(Y)$, a.s. Simply translate the null set to the range space of Y as in part (b) of the previous theorem. The dominated convergence theorem is similar.