

Proof of the Halmos-Savage Factorization Theorem (Theorem 50).

We'll take Lemma 51 (Lemma 1.1 of Shao) and Lemma 52 (Lemma 2.1 of Shao) as given, prove Lemma 53, and then prove Theorem 50 given Lemmas 51, 52 and 53.

Consider the proof of Lemma 53 (assuming the truth of Lemmas 51 and 52).

Suppose that T is sufficient. Let $B \in \mathcal{B}$ and note that by sufficiency \exists a \mathcal{B}_0 -measurable function $P(B|\mathcal{B}_0)$ such that for all θ

$$E_\theta(I_B|\mathcal{B}_0) = P_\theta(B|\mathcal{B}_0) = P(B|\mathcal{B}_0) \text{ a.s. } P_\theta$$

Then for $B_0 \in \mathcal{B}_0$

$$\int_{B_0} P(B|\mathcal{B}_0) dP_\theta = P_\theta(B \cap B_0) \text{ for all } \theta \in \Theta \quad (1)$$

by the definition of conditional expectation. Then since $\lambda = \sum c_i P_{\theta_i}$,

$$\int_{B_0} P(B|\mathcal{B}_0) d\lambda = \lambda(B \cap B_0)$$

and so (again by the definition of conditional expectation)

$$E_\lambda(I_B|\mathcal{B}_0) = P(B|\mathcal{B}_0) \text{ a.s. } \lambda$$

and thus

$$E_\lambda(I_B|\mathcal{B}_0) = P(B|\mathcal{B}_0) \text{ a.s. } P_\theta \text{ for all } \theta \quad (2)$$

(that is, this object serves as the conditional expectation wrt each P_θ and wrt λ).

Now consider P_θ^0 and λ^0 , respectively the restrictions of P_θ and λ to \mathcal{B}_0 and the R-N derivative $\frac{dP_\theta^0}{d\lambda^0}$ on $(\mathcal{X}, \mathcal{B}_0)$. By Lemma 51 Theorem, \exists non-negative \mathcal{F} -measurable functions g_θ such that

$$\frac{dP_\theta^0}{d\lambda^0} = g_\theta \circ T \quad (3)$$

We need to show that $g_\theta \circ T$ also serves as an R-N derivative of P_θ wrt λ . So

let $B \in \mathcal{B}$.

$$\begin{aligned}
P_\theta(B) &= \int_{\mathcal{X}} P(B|\mathcal{B}_0) dP_\theta && \text{(by fact (1))} \\
&= \int_{\mathcal{X}} E_\lambda(I_B|\mathcal{B}_0) dP_\theta && \text{(by fact (2))} \\
&= \int_{\mathcal{X}} E_\lambda(I_B|\mathcal{B}_0) dP_\theta^0 && \text{(since } E_\lambda(I_B|\mathcal{B}_0) \text{ is } \mathcal{B}_0\text{-measurable)} \\
&= \int_{\mathcal{X}} E_\lambda(I_B|\mathcal{B}_0) g_\theta \circ T d\lambda^0 && \text{(by relationship (3))} \\
&= \int_{\mathcal{X}} E_\lambda(I_B|\mathcal{B}_0) g_\theta \circ T d\lambda \\
&= \int_{\mathcal{X}} E_\lambda(I_B \cdot g_\theta \circ T|\mathcal{B}_0) d\lambda && \text{(since } g_\theta \circ T \text{ is } \mathcal{B}_0\text{-measurable)} \\
&= \int_{\mathcal{X}} I_B \cdot g_\theta \circ T d\lambda && (\mathcal{X} \in \mathcal{B}_0 \text{ and the definition of conditional expectation)} \\
&= \int_B g_\theta \circ T d\lambda
\end{aligned}$$

That is, $g_\theta \circ T$ works as the R-N derivative of P_θ wrt λ and the first half of the theorem is proved.

So conversely, suppose that \exists non-negative F -measurable functions g_θ such that

$$\frac{dP_\theta}{d\lambda} = g_\theta \circ T \quad \text{a.s. } \lambda$$

Since $g_\theta \circ T$ is \mathcal{B}_0 -measurable,

$$\frac{dP_\theta^0}{d\lambda^0} = g_\theta \circ T \quad \text{a.s. } \lambda^0 \tag{4}$$

To show that T is sufficient, it will be enough to show that for any $B \in \mathcal{B}$

$$E_\theta(I_B|\mathcal{B}_0) = E_\lambda(I_B|\mathcal{B}_0) \quad \text{a.s. } P_\theta \text{ for all } \theta$$

So for fixed θ and $B \in \mathcal{B}$, define a measure ν on $(\mathcal{X}, \mathcal{B})$ by

$$\nu(C) = \int_C I_B dP_\theta = P_\theta(C \cap B)$$

Now $\nu \ll P_\theta$ and

$$\frac{d\nu}{dP_\theta} = I_B \quad \text{a.s. } P_\theta \tag{5}$$

For all $B_0 \in \mathcal{B}_0$,

$$\begin{aligned}
\nu(B_0) &= \int_{B_0} I_B dP_\theta \\
&= \int_{B_0} E_\theta(I_B|\mathcal{B}_0) dP_\theta && \text{(by the definition of conditional expectation)}
\end{aligned}$$

So for ν^0 the restriction of ν to \mathcal{B}_0

$$\frac{d\nu^0}{dP_\theta^0} = \mathbb{E}_\theta (I_B | \mathcal{B}_0) \quad \text{a.s. } P_\theta^0 \quad (6)$$

(since $\mathbb{E}_\theta (I_B | \mathcal{B}_0)$ is \mathcal{B}_0 -measurable and has the right integrals). Then since $\nu^0 \ll P_\theta^0 \ll \lambda^0$

$$\frac{d\nu^0}{d\lambda^0} = \frac{d\nu^0}{dP_\theta^0} \cdot \frac{dP_\theta^0}{d\lambda^0} \quad \text{a.s. } \lambda^0$$

So since $P_\theta^0 \ll \lambda^0$ and facts (4) and (6) hold

$$\frac{d\nu^0}{d\lambda^0} = \mathbb{E}_\theta (I_B | \mathcal{B}_0) \cdot g_\theta \circ T \quad \text{a.s. } P_\theta^0 \quad (7)$$

On the other hand, from (5), we still have $\frac{d\nu}{dP_\theta} = I_B$ a.s. P_θ , so that

$$\frac{d\nu}{d\lambda} = I_B \cdot \frac{dP_\theta}{d\lambda} = I_B \cdot g_\theta \circ T \quad \text{a.s. } P_\theta$$

This then implies by the definition of conditional expectation that

$$\frac{d\nu^0}{d\lambda^0} = \mathbb{E}_\lambda (I_B \cdot g_\theta \circ T | \mathcal{B}_0) \quad \text{a.s. } P_\theta^0$$

(The object on the right is obviously \mathcal{B}_0 -measurable, and if we multiply by I_{B_0} and integrate $d\lambda^0$, we get $\int I_{B \cap B_0} g_\theta \circ T d\lambda = \nu(B_0)$.) So

$$\frac{d\nu^0}{d\lambda^0} = \mathbb{E}_\lambda (I_B | \mathcal{B}_0) \cdot g_\theta \circ T \quad \text{a.s. } P_\theta^0 \quad (8)$$

So from (7) and (8)

$$\mathbb{E}_\theta (I_B | \mathcal{B}_0) \cdot g_\theta \circ T = \mathbb{E}_\lambda (I_B | \mathcal{B}_0) \cdot g_\theta \circ T \quad \text{a.s. } P_\theta^0$$

This is enough to imply that

$$\mathbb{E}_\theta (I_B | \mathcal{B}_0) = \mathbb{E}_\lambda (I_B | \mathcal{B}_0) \quad \text{a.s. } P_\theta^0 \quad (9)$$

Why? Let $A = \{x | g_\theta(T(x)) = 0\}$ and note that

$$P_\theta(A) = \int_A g_\theta \circ T d\lambda = 0$$

So $\mathcal{X} - A$ has P_θ probability 1 and so does $\{x | \mathbb{E}_\theta (I_B | \mathcal{B}_0) \cdot g_\theta \circ T = \mathbb{E}_\lambda (I_B | \mathcal{B}_0) \cdot g_\theta \circ T\} \cap (\mathcal{X} - A)$. And on this set, $\mathbb{E}_\theta (I_B | \mathcal{B}_0) = \mathbb{E}_\lambda (I_B | \mathcal{B}_0)$.

So (9) holds and Lemma 53 has been proved (conditional on Lemmas 51 and 52).

Finally, assume that Lemmas 51, 52, and 53 have been established, and consider the proof of Theorem 50 (the H-S Factorization Theorem).

Suppose that T is sufficient. $P_\theta \ll \lambda \ll \mu$ implies that

$$\frac{dP_\theta}{d\mu} = \frac{dP_\theta}{d\lambda} \cdot \frac{d\lambda}{d\mu} \quad \text{a.e. } \mu$$

(and both functions on the right of this equality are non-negative). By Lemma 53, \exists a nonnegative function g_θ such that

$$\frac{dP_\theta}{d\lambda} = g_\theta \circ T \quad \text{a.s. } \lambda$$

Let $\mathcal{X}^* = \{x \mid \frac{d\lambda}{d\mu}(x) > 0\}$. I claim that as measures on \mathcal{X}^* , λ and μ are equivalent. Clearly, $\lambda \ll \mu$ on \mathcal{X}^* . So suppose that B is a \mathcal{B} -measurable subset of \mathcal{X}^* that has $\lambda(B) = 0$. But

$$\lambda(B) = \int_B \frac{d\lambda}{d\mu} d\mu$$

while since $B \subset \mathcal{X}^*$ the integrand here is positive. So it must be that $\mu(B) = 0$. Thus λ and μ are equivalent on \mathcal{X}^* .

So then consider

$$\Delta = \frac{dP_\theta}{d\mu} - (g_\theta \circ T) \cdot \frac{d\lambda}{d\mu} = \frac{d\lambda}{d\mu} \left[\frac{dP_\theta}{d\lambda} - g_\theta \circ T \right]$$

Letting

$$D = \{x \mid \Delta(x) \neq 0\} = \mathcal{X}^* \cap \{x \mid \frac{dP_\theta}{d\lambda}(x) \neq g_\theta(T(x))\}$$

since the second term in this intersection has λ probability 0, $\mu(D) = 0$ and thus

$$\frac{dP_\theta}{d\mu} = (g_\theta \circ T) \cdot \frac{d\lambda}{d\mu} \quad \text{a.e. } \mu$$

and taking $h = \frac{d\lambda}{d\mu}$ the forward implication is proved.

Conversely, suppose that the factorization holds, that is that $f_\theta = \frac{dP_\theta}{d\mu} = (g_\theta \circ T) \cdot h$ a.e. μ .

Now

$$\begin{aligned} \frac{d\lambda}{d\mu} &= \sum_{i=1}^{\infty} c_i \frac{dP_{\theta_i}}{d\mu} \\ &= \sum_{i=1}^{\infty} c_i h \cdot (g_{\theta_i} \circ T) \\ &= h \cdot \sum_{i=1}^{\infty} c_i (g_{\theta_i} \circ T) \\ &= h \cdot (k \circ T) \end{aligned}$$

for $k = \sum_{i=1}^{\infty} c_i g_{\theta_i}$. Then the fact that $P_{\theta} \ll \lambda \ll \mu$ again implies that

$$\frac{dP_{\theta}}{d\mu} = \frac{dP_{\theta}}{d\lambda} \cdot \frac{d\lambda}{d\mu} \quad \text{a.e. } \mu$$

so that

$$(g_{\theta} \circ T) \cdot h = \frac{dP_{\theta}}{d\lambda} \cdot h \cdot (k \circ T) \quad \text{a.e. } \mu.$$

Then define

$$g_{\theta}^* \circ T = \begin{cases} \frac{g_{\theta} \circ T}{k \circ T} & \text{if } \frac{d\lambda}{d\mu} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

On \mathcal{X}^*

$$\frac{dP_{\theta}}{d\lambda} = g_{\theta}^* \circ T$$

and this set has λ probability 1, i.e.

$$\frac{dP_{\theta}}{d\lambda} = g_{\theta}^* \circ T \quad \text{a.e. } \lambda$$

so that Lemma 53 then implies that T is sufficient.

So Theorem 50 has been proved (conditional on Lemmas 51-53).