1 Lecture: Measure Theory (solutions)

1.

(a) $\Longrightarrow$) Let $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ be an increasing sequence and let $A := \bigcup_{n=1}^{\infty} A_n$. Then

$$\begin{aligned}\mu(A) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) &\overset{(1)}{=} \mu\left(\biguplus_{n=1}^{\infty} (A_n - A_{n-1})\right) \overset{(2)}{=} \sum_{n=1}^{\infty} \mu(A_n - A_{n-1}) \\ &\overset{(3)}{=} \sum_{n=1}^{\infty} (\mu(A_n) - \mu(A_{n-1})) \overset{(4)}{=} \lim_{n\to\infty} \mu(A_n) - \mu(A_0) = \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

(1) $\biguplus$ denotes the disjoint union of sets. We define $A_0 = \emptyset$.
(2) We use the σ-additivity of μ.
(3) We use the finite additivity of μ. If $\mu(A_n) - \mu(A_{n-1}) = \infty - \infty$ for some $n \in \mathbb{N}$, then the sequence $\{\mu(A_k)\}_{k\geq n}$ would be identically ∞. The monotony property of μ would then imply that $\mu(A) = \infty$ and the result would be also true.
(4) Telescopic series.

$\Longleftarrow$) Let $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ be a disjoint increasing sequence of sets and let $A := \bigcup_{n=1}^{\infty} A_n$. Then

$$\mu(A) \overset{(1)}{=} \mu\left(\bigcup_{n=1}^{\infty}\left(\bigcup_{k=1}^{n} A_k\right)\right) \overset{(2)}{=} \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} A_k\right) \overset{(3)}{=} \lim_{n\to\infty} \sum_{k=1}^{n} \mu(A_k) = \sum_{k=1}^{\infty} \mu(A_k).$$

(1) Standard trick to transform a countable union into a limit of an increasing sequence of sets. $\{\bigcup_{k=1}^{n} A_k\}_{n\in\mathbb{N}}$ is increasing.
(2) We apply the hypothesis.
(3) Because of the finite additivity.

(b) Let $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ be an decreasing sequence such that $\mu(A_1) < \infty$ and let $A := \bigcap_{n=1}^{\infty} A_n$. Then

$$\begin{aligned}\mu(A_1) - \mu(A) = \mu(A_1 - A) = \mu\left(A_1 - \bigcap_{n=1}^{\infty} A_n\right) &\overset{(1)}{=} \mu\left(\bigcup_{n=1}^{\infty} (A_1 - A_n)\right) \\ &\overset{(2)}{=} \lim_{n\to\infty} \mu(A_1 - A_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

(1) We use that the complement of an intersection of sets is the union of the corresponding complement sets.
(2) We apply (a) since $\{A_1 - A_n\}_{n\in\mathbb{N}}$ is an increasing sequence.

(c) Let $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ be a disjoint countable sequence of sets and let $A := \bigcup_{n=1}^{\infty} A_n$. Observe that

$$\mu(A) = \mu\left(\bigcup_{k=1}^{n} A_k\right) + \mu\left(A - \bigcup_{k=1}^{n} A_k\right), \ \forall n. \tag{1}$$

Therefore

$$\mu(A) \overset{(1)}{=} \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} A_k\right) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(A_k) = \sum_{k=1}^{\infty} \mu(A_k).$$

(1) We take limits: $\bigcap_{n=1}^{\infty}\left(A-\bigcup_{k=1}^{n} A_k\right)=\emptyset$ so the second term of (1) is zero because by hypothesis.

4. Let $\mathcal{A} := \left\{A \subseteq \mathbb{R} : f^{-1}(A) \in \mathcal{F}\right\}$. It is easy to prove that $\mathcal{A}$ is a σ-algebra.

   Since $\mathcal{C} \subseteq \mathcal{A}$ and $\mathcal{A}$ is a σ-algebra, $\sigma(\mathcal{C}) \subseteq \mathcal{A}$, because the σ-algebra generated by $\mathcal{C}$ is the smallest σ-algebra containing $\mathcal{C}$. But $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$, therefore $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{A}$ and f is measurable.

5. (a) $\{\sup_{n\in\mathbb{N}} f_n < a\} = \bigcap_{n=1}^{\infty}\{f_n < a\} = \left(\bigcup_{n=1}^{\infty}\{f_n < a\}^c\right)^c \in \mathcal{F}$, because any $\{f_n < a\} \in \mathcal{F}$. We have used that the complement of an intersection of sets is the union of the complement sets.

   (b) $\inf_{n\in\mathbb{N}} f_n = -\sup_{n\in\mathbb{N}}(-f_n)$. The result follows from (a).

   (c) $\overline{\lim\limits_{n\to\infty}} f_n = \inf_{n\in\mathbb{N}}\sup_{m\geq n} f_m$ by definition. We apply (a) and (b).

   (d) $\underline{\lim\limits_{n\to\infty}} f_n = \sup_{n\in\mathbb{N}}\inf_{m\geq n} f_m$. Again, the results follows from (a) and (b).

7. First of all, let $Y := \inf_{n\in\mathbb{N}} f_n$. Since $\int_X Y\,d\mu$ exists, it is a finite number. The sequence $\{f_n - Y\}_{n\in\mathbb{N}}$ is a sequence of positive functions so that

$$\int_X \underline{\lim}_{n\to\infty}(f_n - Y)\,d\mu \leq \underline{\lim}_{n\to\infty}\int_X (f_n - Y)\,d\mu$$

   by hypothesis. Since Y does not depend on n and using the linearity properties of the Lebesgue integral, we have

$$\int_X \underline{\lim}_{n\to\infty} f_n d\mu - \int_X Y\,d\mu \leq \underline{\lim}_{n\to\infty}\int_X f_n d\mu - \int_X Y\,d\mu.$$

   Finally, we can subtract $\int_X Y\,d\mu$.

8. Since $|f| \leq g$ and f is measurable, $f \in L^1(X,\mu)$. Since $|f_n - f| \leq 2g$, Fatou's Lemma applies to the functions $2g - |f_n - f|$ and yields

$$\int_X 2g d\mu \leq \underline{\lim}_{n\to\infty}\int_X (2g - |f_n - f|)\,d\mu$$

   because $\underline{\lim}_{n\to\infty}(2g - |f_n - f|) = 2g$ due to the pointwise convergence $\lim_{n\to\infty} f_n = f$. Now,

$$\begin{aligned}\underline{\lim}_{n\to\infty}\int_X (2g - |f_n - f|)\,d\mu &= \int_X 2g d\mu + \underline{\lim}_{n\to\infty}\left(-\int_X |f_n - f|\,d\mu\right)\\ &= \int_X 2g d\mu - \overline{\lim}_{n\to\infty}\left(\int_X |f_n - f|\,d\mu\right).\end{aligned}$$

   Since $\int_X 2g d\mu$ is finite, we may subtract it and obtain

$$\overline{\lim}_{n\to\infty}\left(\int_X |f_n - f|\,d\mu\right) \leq 0$$

   which clearly implies $\lim_{n\to\infty}\int |f_n - f|\,d\mu = 0$. Finally, since

$$\left|\int_X f_n d\mu - \int_X f d\mu\right| \leq \int_X |f_n - f|\,d\mu,$$

   the result follows.

2 Lecture: Invariant Measures

1. Let $U \subseteq \mathbb{R}^n$ be an open set and let $F(t,x)$ a differentiable function of $I \times U$, where $I \subseteq \mathbb{R}$ is an open neighbourhood of 0. Let us try to compute the derivative

$$\left.\frac{d}{dt}\right|_{t=s} \int_{\varphi_t(U)} F(t,x)\,dx.$$

Applying the change of variables formula, we have

$$\int_{\varphi_t(U)} F(t,x)\,dx = \int_U \varphi_t^*\left(F(t,x)\,dx\right)$$

where

$$\varphi_t^*\left(F(t,x)\,dx\right) = F(t,\varphi_t(x))\,\varphi_t^*(dx) = F(t,\varphi_t(x))\,J_t(x)\,dx$$

and

$$J_t(x) = \det\left|\frac{\partial\varphi_t}{\partial x^i}(x)\right|$$

is the Jacobian of the diffeomorphism φ_t. Therefore,

$$\begin{aligned}
\left.\frac{d}{dt}\right|_{t=s} \int_{\varphi_t(U)} F(t,x)\,dx &= \left.\frac{d}{dt}\right|_{t=s} \int_U F(t,\varphi_t(x))\,J_t(x)\,dx \\
&= \int_U \left(\left.\frac{d}{dt}\right|_{t=s} F(t,\varphi_t(x))\,J_t(x)\right)dx. \qquad (2)
\end{aligned}$$

Now, if $\operatorname{div} v = 0$ then, by Liouville's formula,

$$J_t(x) = \exp\left(\int_0^t \operatorname{div} v\left(\varphi_u(x)\right)du\right) = 1.$$

On the other hand, the volume of $\varphi_t(U)$ is

$$\int_{\varphi_t(U)} 1dx$$

so $F(t,x) \equiv 1$ and the integrand of (2)reduces to

$$\left.\frac{d}{dt}\right|_{t=s} F(t,\varphi_t(x))\,J_t(x)\,dx = 0.$$

Consequently, $\left.\frac{d}{dt}\right|_{t=s} \int_{\varphi_t(U)} 1dx = 0$ and the volume of $\varphi_t(U)$ is constant.

3. Let $X := (0,1)$ and suppose that μ is an invariant Borel probability. Since $T^{-1}((0,1/2)) = (0,1)$ we conclude that

$$\mu((0,1/2)) = (\mu(0,1)) = 1$$

and μ is supported on $(0,1/2)$. In general,

$$T^{-1}\left(\left(0,\frac{1}{2^{n+1}}\right)\right) = \left(0,\frac{1}{2^n}\right), \quad n \geq 1,$$

and, by induction,

$$\mu\left(\left(0,\frac{1}{2^{n+1}}\right)\right)=\mu\left(\left(0,\frac{1}{2}\right)\right)=1$$

Therefore, the sequence

$$A_n:=\left(0,\frac{1}{2^{n+1}}\right)^c=\left[\frac{1}{2^{n+1}},1\right)$$

is increasing and such that $\mu(A_n)=0$ for any $n\in\mathbb{N}$. Moreover,

$$\bigcup_{n\geq 1}A_n=(0,1)$$

and since μ is σ-additive,

$$1=\mu((0,1))=\mu\left(\bigcup_{n\geq}A_n\right)=\lim_{n\to\infty}\mu(A_n)=\lim_{n\to\infty}0=0$$

by Lecture 1 Problem 1 (a), which is a contradiction.

3 Lecture: Birkhoff's ergodic theorem

1. Let $U \in \mathcal{F}$ such that $\mu(U) > 0$ and $B := \bigcup_{n\geq 0} T^{-n}(U)$. Then

$$T^{-1}(B) = \bigcup_{n\geq 1} T^{-n}(U) \subseteq B$$

and $\mu\left(T^{-1}(B)\right) = \mu(B)$ (T is measure preserving). Consequently, $\mathbf{1}_B$ is T-invariant a.s., which implies that $\mathbf{1}_B$ is constant a.s. because T is ergodic. In other words, since $\mu(B) > 0$, $\mathbf{1}_B = \mathbf{1}_X$ and $\mu(B) = 1$.

3. Let $f : X \to \mathbb{R}$ be a $\mathcal{G}$-measurable. Since f is $\mathcal{G}$-measurable, there exists a sequence $\{\zeta_n\}_{n\in\mathbb{N}}$ of elementary functions such that $\zeta_n \to f$ as $n \to \infty$. Suppose that

$$\zeta_n = \sum_{i=1}^{k_n} c_i^{(n)} \mathbf{1}_{A_i^{(n)}}, \text{ where } c_i^{(n)} \in \mathbb{R} \text{ and } A_i^{(n)} \in \mathcal{G}.$$

Then

$$\zeta_n \circ T = \sum_{i=1}^{k_n} c_i^{(n)} \mathbf{1}_{A_i^{(n)}} \circ T = \sum_{i=1}^{k_n} c_i^{(n)} \mathbf{1}_{T^{-1}\left(A_i^{(n)}\right)} = \sum_{i=1}^{k_n} c_i^{(n)} \mathbf{1}_{A_i^{(n)}} = \zeta_n$$

because $T^{-1}(A_i^{(n)}) = A_i^{(n)}$ for any $n \in \mathbb{N}$ and any $i = 1, ..., k_n$. That is, ζ_n are T-invariant. However, $\zeta_n \circ T \to f \circ T$ as $n \to \infty$ (because $\zeta_n \to f$). Therefore,

$$\begin{array}{ccc} \zeta_n \circ T & \underset{n\to\infty}{\longrightarrow} & f \circ T \\ \| & & \\ \zeta_n & \underset{n\to\infty}{\longrightarrow} & f. \end{array}$$

Since the limit is unique, this means that $f \circ T = f$, so f is T-invariant.

4.

(a) By definition, given an arbitrary σ-algebra $\mathcal{F}$, $L^2_{\mathbb{C}}(X, \mathcal{F}, \mu)$ is a Hilbert space. In particular a Banach space: any Cauchy sequence $\{f_n\}_{n\in\mathbb{N}} \subset L^2_{\mathbb{C}}(X, \mathcal{F}, \mu)$ converges to an element $f \in L^2_{\mathbb{C}}(X, \mathcal{F}, \mu)$. This means that $L^2_{\mathbb{C}}(X, \mathcal{F}, \mu)$ is closed. Finally, take $\mathcal{F} = \mathcal{G}$.

(b) If H is a Hilbert space with Hermitian product $\langle \cdot, \cdot \rangle$ and $E \subseteq H$ is a closed subspace, for any $y \in H$,

$$x^* = \min_{x\in E} \|y - x\| = \min_{x\in E} \sqrt{\langle y - x, y - x \rangle}$$

exists and belongs to E. x^* is called the orthogonal projection of y onto E. Moreover, x^* has is characterised by the following property:

$$\langle y - x^*, x \rangle = 0 \text{ for any } x \in E. \tag{3}$$

Indeed, if (3) holds for some $x^* \in E$,

$$\begin{aligned} \|y - x\|^2 &= \langle y - x, y - x \rangle = \langle (y - x^*) + (x^* - x), (y - x^*) + (x^* - x) \rangle \\ &= \langle y - x^*, y - x^* \rangle + \langle y - x^*, x^* - x \rangle + \langle x^* - x, y - x^* \rangle + \langle x^* - x, x^* - x \rangle \\ &= \|y - x^*\|^2 + \|x - x^*\|^2 \end{aligned}$$

where we have used that $x^* - x \in E$. Consequently, since $\|x - x^*\|^2 \geq 0$,

$$\|y - x\|^2 \geq \|y - x^*\|^2 \text{ for any } x \in E$$

and x^* is the orthogonal projection.

We can now prove that $\mathrm{E}[f|\mathcal{G}]$ is the orthogonal projection of f onto $L^2_{\mathbb{C}}(X, \mathcal{G}, \mu)$. Obviously $\mathrm{E}[f|\mathcal{G}] \in L^2_{\mathbb{C}}(X, \mathcal{G}, \mu)$ and if $g \in L^2_{\mathbb{C}}(X, \mathcal{G}, \mu)$ is a characteristic function, $g = \mathbf{1}_A$, $A \in \mathcal{G}$,

$$\begin{aligned}
\langle g, f - \mathrm{E}[f|\mathcal{G}] \rangle &= \int \overline{g}(f - \mathrm{E}[f|\mathcal{G}])\, d\mu = \int \overline{g} f d\mu - \int \overline{g}\, \mathrm{E}[f|\mathcal{G}]\, d\mu \\
&= \int_A f d\mu - \int_A \mathrm{E}[f|\mathcal{G}]\, d\mu = 0
\end{aligned}$$

by the definition of the conditional expectation. The result is also true for elementary functions. For an arbitrary $g \in L^2_{\mathbb{C}}(X, \mathcal{G}, \mu)$, we take an increasing sequence $\{\zeta_n\}_{n \in \mathbb{N}}$ of elementary functions such that $\zeta_n \to g$. Then, for any $n \in \mathbb{N}$,

$$0 = \int \overline{\zeta}_n (f - \mathrm{E}[f|\mathcal{G}])\, d\mu.$$

But $\int \overline{\zeta}_n (f - \mathrm{E}[f|\mathcal{G}])\, d\mu \to \int \overline{g}(f - \mathrm{E}[f|\mathcal{G}])\, d\mu$ because the integrand is dominated. Therefore

$$0 = \int \overline{g}(f - \mathrm{E}[f|\mathcal{G}])\, d\mu.$$

5.

(a) By induction. If $\nu \ll \mu$, then there exists $f \in L^1(X, \mu)$ such that $\nu(A) = \int_A f d\mu$ for any $A \in \mathcal{F}$. On the other hand,

$$T_* \nu(A) = \nu\left(T^{-1}(A)\right) = \int_{T^{-1}(A)} f d\mu = \int_A P(f) d\mu$$

where we have applied the definition of the Frobenius-Perron operator. Therefore, $T_*\nu$ is absolutely continuous with Radon-Nikodym derivative $P(f)$.

$T_*\nu = \nu \Longleftrightarrow P(f) = f$.

$\Longrightarrow$) $T_*\nu = \nu$ means that, for any $A \in \mathcal{F}$, $\int_A P(f)\, d\mu = \int_A f d\mu$ or, what is the same,

$$\int_A (P(f) - f)\, d\mu = 0.$$

In particular, take $A_+ := \operatorname{supp}(P(f) - f)^+ = \{x : P(f)(x) - f(x) > 0\}$ and $A_- := \operatorname{supp}(P(f) - f)^- = \{P(f)(x) - f(x) < 0\}$. Then

$$0 = \int_{A_+} (P(f) - f)\, d\mu = \int_{A_+} (P(f) - f)^+\, d\mu$$

which implies $(P(f) - f)^+ \equiv 0$ because $(P(f) - f)^+$ is a nonnegative function. One can equally prove that $(P(f) - f)^- \equiv 0$ so $P(f) = f$.

$\Longleftarrow$) Obvious.

(b) From the definition of the Frobenius-Perron operator,

$$\int_X \mathbf{1}_{T^{-1}(A)} f d\mu = \int_X \mathbf{1}_A P(f) d\mu,$$

It is clear that $P(f)(x) = 0$ on A implies $f(x) = 0$ on $T^{-1}(A)$ and vice versa. Let $A := X \backslash \operatorname{supp} P(f)$. Then $f(x) = 0$ for any $x \in T^{-1}(A)$ so $\operatorname{supp} f \subseteq X \backslash T^{-1}(A) = T^{-1}(X \backslash A)$. But $X \backslash A = \operatorname{supp} P(f)$ and the result follows.

(c) Let f_1 and f_2 be two different stationary densities. That is, $P(f_1) = f_1$, $P(f_2) = f_2$ and $\|f_1\|_{L^1(X,\mu)} = \|f_2\|_{L^1(X,\mu)} = 1$. Let $g = f_1 - f_2$. Then $P(g) = g$ and $P(g^+) = g^+$ and $P(g^-) = g^-$ as well. Since $f_1 \neq f_2$ and both are densities, g^+ and g^- are not identically zero a.s.. Define $A = \operatorname{supp} g^+$ and $B = \operatorname{supp} g^-$, which are strictly positive disjoint sets. By (b), $A \subseteq T^{-1}(A)$ and $B \subseteq T^{-1}(B)$, where $T^{-1}(A)$ and $T^{-1}(B)$ are also disjoint. We have

$$\begin{aligned} A &\subseteq T^{-1}(A) \subseteq T^{-2}(A) \subseteq ... \\ B &\subseteq T^{-1}(B) \subseteq T^{-2}(B) \subseteq ... \end{aligned}$$

where $T^{-n}(A)$ and $T^{-n}(B)$ are disjoint for any $n \in \mathbb{N}$. Define

$$\overline{A} = \bigcup_{n=0}^{\infty} T^{-n}(A) \text{ and } \overline{B} = \bigcup_{n=0}^{\infty} T^{-n}(B).$$

These sets are invariant. Indeed, consider for instance $\overline{A}$,

$$T^{-1}\left(\overline{A}\right) = \bigcup_{n=1}^{\infty} T^{-n}(A) = \bigcup_{n=0}^{\infty} T^{-n}(A) = \overline{A}$$

because $A \subseteq T^{-1}(A)$. Equivalently, $T^{-1}\left(\overline{B}\right) = \overline{B}$. Since $\overline{A}$ and $\overline{B}$ are invariant disjoint sets with strictly positive measure, the map is not ergodic. Contradiction.

(d) Suppose that T is not ergodic. Then $\exists A \in \mathcal{F}$ such that $T^{-1}(A) = A$ and $0 < \mu(A) < 1$. $B := X \backslash A$ is also invariant with strictly positive measure. Observe that

$$\mathbf{1}_A + \mathbf{1}_B = \mathbf{1}_X = P(\mathbf{1}_X) = P(\mathbf{1}_A + \mathbf{1}_B) = P(\mathbf{1}_A) + P(\mathbf{1}_B)$$

where we have used the linearity of P and the fact that $P(\mathbf{1}_X) = \mathbf{1}_X$. Since $\mathbf{1}_A$ is zero on B (resp. $\mathbf{1}_B$ is zero on A), by (b), $P(\mathbf{1}_A)$ is zero on B as well (resp. $P(\mathbf{1}_B)$ is zero on A). Therefore,

$$P(\mathbf{1}_A) = \mathbf{1}_A \text{ and } P(\mathbf{1}_B) = \mathbf{1}_B$$

which is a contradiction because neither $\mathbf{1}_A$ nor $\mathbf{1}_B$ are constants.

(e) Suppose that T is ergodic.
From previous results, we have

$$\int_X g P^n(f) d\mu = \int_X g d(T_*^n \nu) = \int_X g \circ T^n d\nu.$$

Now,

$$\frac{1}{N} \sum_{n=0}^{N-1} \int_X g P^n(f) d\mu = \int_X \left(\frac{1}{N} \sum\nolimits_{n=0}^{N-1} g \circ T^n\right) d\nu = \int_X \left(\frac{1}{N} \sum\nolimits_{n=0}^{N-1} g \circ T^n\right) f d\mu.$$

Since $g \in L^{\infty}(X,\mu)$ and $\nu \ll \mu$, $g \in L^{\infty}(X,\nu)$ and

$$\left\| \frac{1}{N} \sum\nolimits_{n=0}^{N-1} g \circ T^n \right\|_{L^{\infty}(X,\nu)} \leq \frac{1}{N} \sum\nolimits_{n=0}^{N-1} \|g\|_{L^{\infty}(X,\nu)} = \|g\|_{L^{\infty}(X,\nu)}$$

so

$$\left| \frac{f}{N} \sum\nolimits_{n=0}^{N-1} g \circ T^n \right| \leq |f| \, \|g\|_{L^{\infty}(X,\nu)} \quad \text{for any } N \in \mathbb{N}.$$

On the other hand, since T is ergodic,

$$\frac{1}{N} \sum\nolimits_{n=0}^{N-1} g \circ T^n \longrightarrow \int_X g d\mu \quad \mu\text{-a.s.}.$$

By the Dominated Convergence Theorem,

$$\int_X \left(\frac{1}{N} \sum\nolimits_{n=0}^{N-1} g \circ T^n \right) f d\mu \underset{n \to \infty}{\longrightarrow} \int_X \left(\int_X g d\mu \right) f d\mu = \int_X g d\mu \int_X f d\mu.$$

Now suppose that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \int P^n(f) g d\mu = \int f d\mu \int g d\mu \tag{4}$$

for any $f \in L^1(X,\mu)$ and any $g \in L^{\infty}(X,\mu)$. We want to prove that T is ergodic. So take $A \in \mathcal{F}$ and invariant set, $T^{-1}(A) = A$. Since A is invariant, $P(\mathbf{1}_A) = \mathbf{1}_A$. Indeed, for any $B \in \mathcal{F}$,

$$\begin{aligned} \int_B P(\mathbf{1}_A) d\mu &= \int_{T^{-1}(B)} \mathbf{1}_A d\mu = \int_X \mathbf{1}_{A \cap T^{-1}(B)} d\mu \\ &= \int_X \mathbf{1}_{T^{-1}(A \cap B)} d\mu = \int_X \mathbf{1}_{A \cap B} \circ T d\mu \\ &= \int_X \mathbf{1}_{A \cap B} d(T_* \mu) = \int_X \mathbf{1}_{A \cap B} d\mu = \int_B \mathbf{1}_A d\mu, \end{aligned}$$

which implies $P(\mathbf{1}_A) = \mathbf{1}_A$. Taking $f = \mathbf{1}_A$ in Eq. (4),

$$\int \mathbf{1}_A g d\mu = \int \mathbf{1}_A d\mu \int g d\mu = \mu(A) \int g d\mu. \tag{5}$$

Now, taking $g = \mathbf{1}_A$, (5) we have $\mu(A) = \mu(A)^2$ which can only be satisfied either if $\mu(A) = 0$ or $\mu(A) = 1$. So T is ergodic.