The Principles of Probability: From Formal Logic to Measure Theory to the Principle of Indifference

Jason Swanson

Department of Mathematics
University of Central Florida

Logic Seminar
Indiana University, Bloomington, IN, October 30, 2024

With the exception of mathematics, everything we know is conjecture.

—George Polya, 1954 (paraphrased)

Examples

- Laws of physics
- Guilt of a defendant
- Historical facts
- Economic principles
- The sun will rise tomorrow.

None of these can be shown with complete certainty.

But some are more certain than others.

Deductive reasoning: Used to establish mathematical facts.
Inductive reasoning: Used to establish everything else.

Rules of deductive reasoning

- Formalized in mathematical logic
- Detailed and precise
- Well-understood
- Universally accepted
- Highly successful

Rules of inductive reasoning

- There *are* rules. Example:

  The Fundamental Inductive Pattern (Polya, 1954):

  $$\frac{A \text{ implies } B \qquad B \text{ is true}}{A \text{ becomes more plausible}}$$

  This is the foundation of empirical science.
- There is no successful, agreed-upon formalization
- Obvious tool to use: probability

Mathematicians, philosophers, and physicists have made efforts toward formalizing inductive reasoning with probability.

Incomplete list:

- Leibniz (1670)
- Jacob Bernoulli (1713)
- Bayes (1763)
- Laplace (1774)
- Bolzano (1837)
- De Morgan (1837)
- Boole (1854)
- Keynes (1921)
- Wittgenstein (1922)
- Reichenbach (1949)
- Carnap (1950)
- Scott and Krauss (1966)
- Nilsson (1986)
- Jaynes (2003)

$\mathcal{L} = \mathcal{L}_{\omega_1,\omega}$: a predicate language that allows countable conjunctions and disjunctions

$\mathcal{L}_{\text{fin}} \subseteq \mathcal{L}$: usual first-order language

$\mathcal{L}^0 \subseteq \mathcal{L}$: the set of sentences in $\mathcal{L}$

An *inductive statement* is a triple, (X, φ, p), where $X \subseteq \mathcal{L}^0$, $\varphi \in \mathcal{L}^0$, and $p \in [0, 1]$.

$\mathcal{L}^{\text{IS}}$: the set of inductive statements

$X \vdash \varphi$: natural deduction with obvious generalizations to countable conjunctions

$X \vdash_{\text{fin}} \varphi$: ($X \subseteq \mathcal{L}_{\text{fin}}$, $\varphi \in \mathcal{L}_{\text{fin}}$) usual first-order derivability

Theorem (Prop 5.3.15)

If $X \subseteq \mathcal{L}_{\text{fin}}$ and $\varphi \in \mathcal{L}_{\text{fin}}$, then $X \vdash \varphi \Leftrightarrow X \vdash_{\text{fin}} \varphi$.

$\vdash' \varphi$: (Karp 1959, 1964; Keisler, 1971) Hilbert-type calculus

$X \mid\!\sim \varphi$: Hilbert-type calculus, extension of $\vdash'$

Theorem (Thm 5.2.24)

$$X \vdash \varphi \Leftrightarrow X \mid\!\sim \varphi$$

$Q \vdash (X, \varphi, p)$: ($Q \subseteq \mathcal{L}^{\text{IS}}$) natural deduction based on 9 rules; defined indirectly in terms of inductive theories

$\omega \models \varphi$: (ω an $\mathcal{L}$-structure, $\varphi \in \mathcal{L}^0$) φ is true in ω, we say ω *strictly satisfies* φ

σ-compactness fails for $\models$. There exists $X \subseteq \mathcal{L}^0$ such that every countable subset of X is strictly satisfiable, but X is not strictly satisfiable.

Karp's Completeness Theorem (Karp 1959, 1964)

$$\vdash' \varphi \quad \text{iff} \quad \omega \models \varphi \text{ for all } \omega$$

An *(inductive) model* is a probability space $\mathscr{P} = (\Omega, \Sigma, \mathbb{P})$ where Ω is a set of $\mathcal{L}$-structures.

$\varphi_\Omega = \{\omega \in \Omega \mid \omega \models \varphi\}$

$\mathscr{P} \vDash \varphi \Leftrightarrow \overline{\mathbb{P}}\, \varphi_\Omega = 1$: we say $\mathscr{P}$ *satisfies* φ

Theorem (σ-compactness, Thm 5.3.19)

X is satisfiable iff every countable subset of X is satisfiable.

$$\mathscr{P} \vDash (X, \varphi, p) \Leftrightarrow X \equiv Y \cup \{\psi\}, \text{ where } \mathscr{P} \vDash Y \ \& \ \frac{\overline{\mathbb{P}}\, \varphi_\Omega \cap \psi_\Omega}{\overline{\mathbb{P}}\, \psi_\Omega} = p.$$

Can use this to define $Q \vDash (X, \varphi, p)$ for $Q \subseteq \mathcal{L}^{\mathrm{IS}}$.

Theorem (Soundness and completeness)

- $X \vdash \varphi \Leftrightarrow X \vDash \varphi :\Leftrightarrow (\mathscr{P} \vDash X \Rightarrow \mathscr{P} \vDash \varphi)$
- $Q \vdash (X, \varphi, p) \Leftrightarrow Q \vDash (X, \varphi, p)$

Note: $\omega \models \varphi$ iff $\mathscr{P} = (\{\omega\}, \{\emptyset, \{\omega\}\}, \delta_\omega) \vDash \varphi$.
Therefore, $X \vDash \varphi \Rightarrow (\omega \models X \Rightarrow \omega \models \varphi)$.

Proof sketch of σ-compactness:

Show $\vdash$ is σ-compact (Thm 5.2.11). Use Karp to show
$X \vdash \varphi \Rightarrow X \vDash \varphi$ (Thm 5.3.16) and
X ctble & consistent $\Rightarrow$ X strictly satisfiable (Cor 5.3.17).

Assume every ctble subset of X is satisfiable.
The above shows X is consistent.

$S := \{X_0 \subseteq \mathcal{L}^0 \mid X_0$ ctble & consistent$\}$. For $X_0 \in S$, choose $\omega^{X_0} \Vvdash X_0$. $\Omega := \{\omega^{X_0} \mid X_0 \in S\}$. Then $\varphi_\Omega = \psi_\Omega \Rightarrow \varphi \equiv \psi$.

Then $\Sigma = \{\varphi_\Omega \mid X \vdash \varphi$ or $X \vdash \neg\varphi\}$ is a σ-algebra on Ω. Define

$$\mathbb{P}\,\varphi_\Omega = \begin{cases} 1 & \text{if } X \vdash \varphi, \\ 0 & \text{if } X \nvdash \varphi. \end{cases}$$

Then $\mathbb{P}$ is well-defined, is a probability measure on (Ω, Σ), and $\mathscr{P} = (\Omega, \Sigma, \mathbb{P}) \vDash X$.

$Th\,\mathscr{P} = \{\varphi \in \mathcal{L}^0 \mid \mathscr{P} \vDash \varphi\}$ (a deductive theory)
$[T_0, Th\,\mathscr{P}]$: set of deductive theories T with $T_0 \subseteq T \subseteq Th\,\mathscr{P}$

$X \hookrightarrow Y$: $X \equiv Y \cup \Phi$ for some countable $\Phi \subseteq \mathcal{L}^0$
$X \hookrightarrow [T_0, Th\,\mathscr{P}]$: $X \hookrightarrow T$ for some $T \in [T_0, Th\,\mathscr{P}]$

$\mathbf{Th}\,\mathscr{P} = \{(X, \varphi, p) \in \mathcal{L}^{\mathrm{IS}} \mid \mathscr{P} \vDash (X, \varphi, p)\}$
$\mathbf{Th}\,\mathscr{P}\!\downharpoonleft_{[T_0, Th\,\mathscr{P}]} = \{(X, \varphi, p) \in \mathbf{Th}\,\mathscr{P} \mid X \hookrightarrow [T_0, Th\,\mathscr{P}]\}$

Models Determine Theories (Thm 4.2.4 + Prop 3.5.10)

If $\mathscr{P}$ is a model with $\mathscr{P} \vDash T_0$, then $P = \mathbf{Th}\,\mathscr{P}\!\downharpoonleft_{[T_0, Th\,\mathscr{P}]}$ is a complete inductive theory with root T_0.

Theories Determine Models (Thm 4.2.6)

If P is a complete inductive theory with root T_0, then there exists a model $\mathscr{P}$ such that $P = \mathbf{Th}\,\mathscr{P}\!\downharpoonleft_{[T_0, Th\,\mathscr{P}]}$.

(S, Γ, ν): probability space

$\langle X_i \mid i \in I \rangle$: family of r.v.s

X_i takes values in (R_i, Γ_i); Γ is generated by the X_i's

We call $(S, \Gamma, \nu, \langle X_i \mid i \in I \rangle)$ a **modern probability model**

$\mathcal{L}_R$: predicate language with constant symbols $\{\underline{r} \mid r \in \bigcup_i R_i\}$ and unary relation symbols $\{\underline{V_i} \mid i \in I, V_i \in \Gamma_i\}$

$x \in \underline{V_i}$ is shorthand for $\underline{V_i}\, x$

$\mathcal{R}$: $\mathcal{L}_R$-structure with domain $R = \bigcup_i R_i$ and $\underline{r}^{\mathcal{R}} = r$, $\underline{V}^{\mathcal{R}} = V$

$T_R = \{\varphi \in \mathcal{L}_R^0 \mid \mathcal{R} \models \varphi\}$

$\mathcal{L}$: language obtained from $\mathcal{L}_R$ by adding a constant symbol $\underline{X}_i$ for each $i \in I$

Embedding Theorem I (Thm 5.4.2)

There exists an $\mathcal{L}$-model $\mathscr{P} = (\Omega, \Sigma, \mathbb{P})$ with $\mathscr{P} \vDash T_R$ and a function $h : S \to \Omega$ mapping $x \in S$ to $\omega \in \Omega$ such that

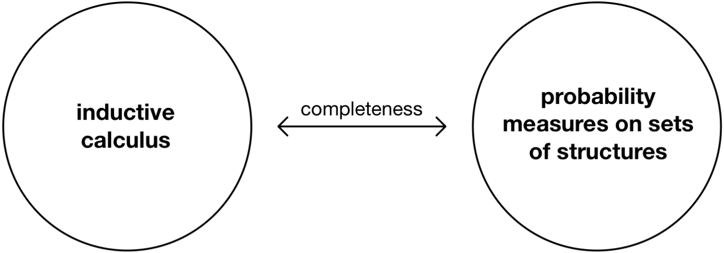
- $x \in \{X_i \in V\} \Leftrightarrow \omega \Vdash \underline{X}_i \in \underline{V}_i$,
- $U \in \Gamma \Rightarrow U = h^{-1}\varphi_\Omega$ for some $\varphi \in \mathcal{L}^0$, and
- h is a measure space isomorphism,

and $P = \mathbf{Th}\, \mathscr{P}\rfloor_{[T_R, Th\, \mathscr{P}]}$ satisfies

$$P(\textstyle\bigwedge_{k=1}^n \underline{X}_{i(k)} \in \underline{V}_k \mid T_R) = \nu \bigcap_{k=1}^n \{X_{i(k)} \in V_k\}.$$

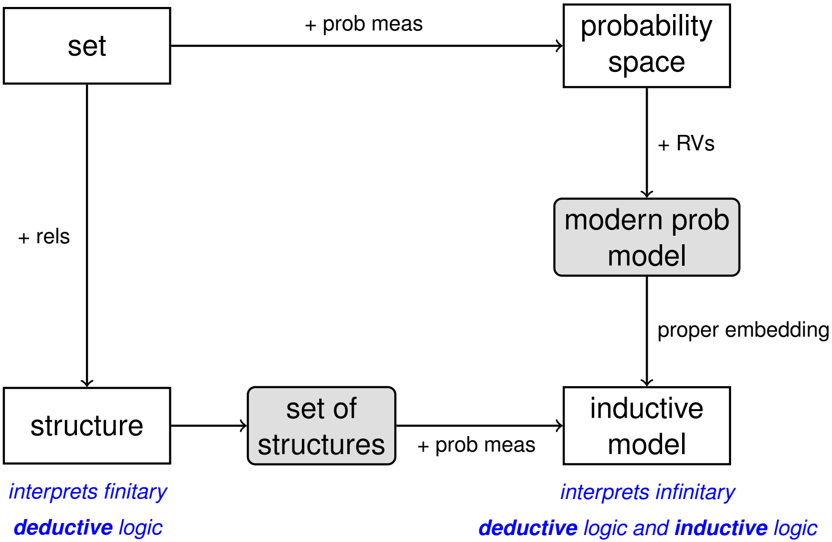
Measure Theory	*Inductive Logic*
outcome	structure
event	sentence
set membership	strict satisfiability
random variable	constant symbol

Inductive Logic



derive probabilities with rules
(no sample spaces,
no measure theory)

all of modern probability is represented here


set
+ prob meas
probability space
+ RVs
modern prob model
+ rels
proper embedding
structure
set of structures
+ prob meas
inductive model
interpretsfinitary
deductive logic
interprets infinitary
deductive logic and **inductive** logic

Example 1

$\mathcal{L}$: a language with constants $\{\underline{X}_n \mid n \in \mathbb{N}\} \cup \{h, t\}$

For $\mathbf{s} = \langle s_n \rangle \in \{h, t\}^{\mathbb{N}}$, let $\psi_{\mathbf{s}} = \neg \bigwedge_{n \in \mathbb{N}} \underline{X}_n = s_n$.

$$X = \{\textstyle\bigwedge_{n \in \mathbb{N}} (\underline{X}_n = h \vee \underline{X}_n = t)\} \cup \{\psi_{\mathbf{s}} \mid \mathbf{s} \in \{h, t\}^{\mathbb{N}}\}$$

Every countable subset of X is satisfiable.

By σ-compactness, X is satisfiable. Therefore, there exists $\mathscr{P}$ such that $\mathscr{P} \vDash X$.

(S, Γ, ν): prob. sp.; $\langle X_n \mid n \in \mathbb{N} \rangle$: i.i.d. coin flips

Build $\mathscr{P}$ as in Embedding Theorem I. Then $\mathscr{P} \vDash X$.

Example 2

I: an uncountable set

$\mathcal{L}$: a language with constants $\{\underline{X}_t \mid t \in I\} \cup \{\underline{n} \mid n \in \mathbb{N}\}$

$$X = \{\underline{m} \neq \underline{n} \mid m, n \in \mathbb{N}, m \neq n\} \\ \cup \{\textstyle\bigvee_{n \in \mathbb{N}} \underline{X}_t = \underline{n} \mid t \in I\} \cup \{\underline{X}_s \neq \underline{X}_t \mid s, t \in I, s \neq t\}$$

Every countable subset of X is satisfiable.

By σ-compactness, X is satisfiable. Choose an inductive model $\mathscr{P}$ such that $\mathscr{P} \vDash T_0 = T(X)$. Let $P = \textbf{Th}\,\mathscr{P}\downarrow_{[T_0, Th\,\mathscr{P}]}$.

Then $P(\underline{X}_s \neq \underline{X}_t \mid T_0) = 1$ for all $s \neq t$.

A “real inductive theory” is an inductive theory capable of talking about the real numbers.

Can construct one with Embedding Theorem I.

A less ad hoc, more flexible way is to talk to about real numbers in set theory.

$\Lambda^{\text{ZFC}}_{-} \subseteq \mathcal{L}_{\text{fin}}$: usual (finitary) axioms of ZFC

Notation for axiom schema of separation:

$$\begin{aligned} &\text{AS}(\varphi) : \exists y \forall z(z \in y \leftrightarrow \varphi \wedge z \in x) \\ &\text{AS} = \{\text{AS}(\varphi) \mid \varphi(x, z, \vec{u}) \in \mathcal{L} \text{ and } y \notin \text{free}\,\varphi\} \\ &\text{AS}_{\text{fin}} = \text{AS} \cap \mathcal{L}_{\text{fin}} \subseteq \Lambda^{\text{ZFC}}_{-} \end{aligned}$$

Λ^{ZFC}: includes all of AS, not just AS_{fin}

$$\mathsf{ZFC}_- = T(\Lambda_-^{\mathsf{ZFC}}), \qquad \mathsf{ZFC}_{\mathrm{fin}} = \mathsf{ZFC}_- \cap \mathcal{L}_{\mathrm{fin}}, \qquad \mathsf{ZFC} = T(\Lambda^{\mathsf{ZFC}})$$

Since $\Lambda_-^{\mathsf{ZFC}} \subseteq \mathcal{L}_{\mathrm{fin}}$, Prop 5.3.15 implies that for $\varphi \in \mathcal{L}_{\mathrm{fin}}$, we have $\Lambda_-^{\mathsf{ZFC}} \vdash \varphi$ iff $\Lambda_-^{\mathsf{ZFC}} \vdash_{\mathrm{fin}} \varphi$.

Therefore, $\mathsf{ZFC}_{\mathrm{fin}} = \{\varphi \in \mathcal{L}_{\mathrm{fin}} \mid \Lambda_-^{\mathsf{ZFC}} \vdash_{\mathrm{fin}} \varphi\}$ is the usual (finitary) version, and ZFC is its minimal infinitary extension.

A *real inductive theory in* ZFC_- is an inductive theory P with root $T_0 \supseteq \mathsf{ZFC}_-$.

If such a P exists, then ZFC_- is consistent.
$(\mathscr{P} \vDash P \Rightarrow \mathscr{P} \vDash T_0 \supseteq \mathsf{ZFC}_-)$

Theorem (Prop 6.2.1)

$\mathscr{P} = (\Omega, \Sigma, \mathbb{P}) \vDash \mathsf{ZFC}_-$ iff $\omega \models \mathsf{ZFC}_{\mathrm{fin}}$ for $\mathbb{P}$-a.e. $\omega \in \Omega$.
$\therefore$ ZFC_- is consistent iff $\mathsf{ZFC}_{\mathrm{fin}}$ is consistent in first-order logic.

Notation in ZFC$_-$:

$$\underline{0} = \emptyset, \quad \underline{n} = \mathrm{S} \cdots \mathrm{S}\emptyset$$

$\underline{q}$: an explicitly defined constant symbol for $q \in \mathbb{Q}$

Embedding Theorem II (Thm 6.3.4)

Assume ZFC$_-$ is consistent. Then there exists a complete inductive theory P with root ZFC$_-$ such that

$$P(\textstyle\bigwedge_{k=1}^n \underline{X}_{i(k)} \leq \underline{q_k} \mid \mathrm{ZFC}_-) = \nu \textstyle\bigcap_{k=1}^n \{X_{i(k)} \leq q_k\}.$$

- Cannot talk about individual real numbers (or Borel sets or measurable functions).
- Loses the connection between outcomes and structures.

A *real inductive theory in* ZFC is an inductive theory P with root $T_0 \supseteq$ ZFC.

If such a P exists, then ZFC is consistent. In fact, we will frequently assume ZFC is strictly satisfiable.

κ: cardinal number
V_κ: set in von Neumann hierarchy
$\boldsymbol{\nu}_\kappa = (V_\kappa, \in^{\nu_\kappa})$, where $\in^{\nu_\kappa} = \in$

Theorem (Thm 6.2.4)

The following are equivalent:

(i) κ *is strongly inaccessible*

(ii) $\boldsymbol{\nu}_\kappa \models \Lambda^{\mathrm{ZFC}}_{-}$

(iii) $\boldsymbol{\nu}_\kappa \models$ ZFC

$\underline{\mathbb{N}_0}$: A constant symbol with an explicit, finitary definition. It is the usual set of natural numbers. That is, it is the smallest set that contains $\underline{0}$ and is closed under the successor operation.

Theorem (Prop 6.3.2)

If ZFC_- *is consistent, then*

$$\mathrm{ZFC}_- \nvdash \exists y \forall x (x \in y \leftrightarrow \bigvee_{n \in \mathbb{N}_0} x = \underline{n}).$$

On the other hand,

$$\mathrm{ZFC} \vdash \forall x (x \in \underline{\mathbb{N}_0} \leftrightarrow \bigvee_{n \in \mathbb{N}_0} x = \underline{n}).$$

Theorem (Prop. 5.3.22)

Suppose $\varphi \in \mathcal{L}^0$ *is strictly satisfied by the standard structure of arithmetic,* $(\mathbb{N}_0, 0, \mathrm{S}, +, \cdot)$. *Then* ZFC $\vdash \varphi$. *Equivalently,* $\mathscr{P} \vDash$ ZFC *implies* $\mathscr{P} \vDash \varphi$ *for all models* $\mathscr{P}$.

If $B \subseteq \mathbb{N}_0$, then

$$\text{ZFC} \vdash \exists! y \, \forall x (x \in y \leftrightarrow \bigvee_{n \in B} x = \underline{n})$$

Can explicitly define each subset of $\mathbb{N}_0$.

Similarly for each subset of $\mathbb{Q}$, and hence, each Dedekind cut.

Can explicitly define each real number, each Borel set, and each measurable function.

ω: a structure with domain A; ${}^\omega t = \{a \in A \mid a \in^\omega t^\omega\}$

The Real Frame of Reference (Thms 6.4.1, 6.4.3, 6.4.5)

If $\mathscr{Q} \vDash$ ZFC, then there exists a model $\mathscr{P} = (\Omega, \Sigma, \mathbb{P})$ such that $\mathscr{Q} \simeq \mathscr{P}$ and

- $\omega \models$ ZFC$_-$ for all $\omega \in \Omega$
- $\underline{q}^\omega = q$ for all $\omega \in \Omega$
- ${}^\omega\underline{\mathbb{R}} \mathrel{\underline{\subseteq}} \mathbb{R}$ for all $\omega \in \Omega$
- if $r \in \mathbb{R}$, then $\underline{r}^\omega = r$ for a.e. $\omega \in \Omega$
- if $V \in \mathcal{B}(\mathbb{R})$, then $\underline{V}^\omega = {}^\omega\underline{V} = V \cap {}^\omega\underline{\mathbb{R}}$ for a.e. $\omega \in \Omega$
- if $h : \mathbb{R} \to \mathbb{R}$ is measurable, then for a.e. $\omega \in \Omega$, $\omega \models (y = \underline{h}(x))[a, b]$ iff $a \in {}^\omega\underline{\mathbb{R}}$, $b \in {}^\omega\underline{\mathbb{R}}$, and $h(a) = b$.

Embedding Theorem III (Thm 6.4.6)

Assume ZFC is strictly satisfiable. Let $(S, \Gamma, \nu, \langle X_i \mid i \in I \rangle)$ be a real-valued modern probability model. Then there exists a model $\mathscr{P} = (\Omega, \Sigma, \mathbb{P})$ with $\mathscr{P} \vDash$ ZFC and a function $h : S \to \Omega$ mapping $x \in S$ to $\omega \in \Omega$ such that

(i) $x \in \{X_i \in V\} \Leftrightarrow \omega \models \underline{X}_i \in \underline{V}$,

(ii) $U \in \Gamma \Rightarrow U = h^{-1}\varphi_\Omega$ for some sentence φ, and

(iii) h is a measure space isomorphism,

and $P = \mathbf{Th}\, \mathscr{P} \rfloor_{[\mathrm{ZFC}_\infty, Th\, \mathscr{P}]}$ satisfies

$$P\Big(\bigwedge_{k=1}^{n} \underline{X}_{i(k)} \in \underline{V}_k \mid \mathrm{ZFC}\Big) = \nu \bigcap_{k=1}^{n} \{X_{i(k)} \in V_k\}.$$

Proof sketch:

Choose ω_0 such that $\omega_0 \models$ ZFC. Add constant symbols $\underline{X}_i$.
For each $x \in S$, let $\omega = \omega^x$ be the ext. of ω_0 with $\underline{X}_i^\omega = \underline{X_i(x)}^{\omega_0}$.

Let $\Omega = \{\omega^x \mid x \in S\}$ and define $h : S \to \Omega$ by $hx = \omega^x$.
Let $\Sigma = \{A \subseteq \Omega \mid h^{-1}A \in \Gamma\}$, $\mathbb{P} = \nu \circ h^{-1}$, and $\mathscr{P} = (\Omega, \Sigma, \mathbb{P})$.

Then $\omega \models$ ZFC for all $\omega \in \Omega$. Therefore, $\mathscr{P} \vDash$ ZFC.

Fix x, i, V. Let $r = X_i(x) \in \mathbb{R}$. Then $\omega^x \models \underline{X}_i \in \underline{V}$ iff $\omega_0 \models \underline{r} \in \underline{V}$.
Using the real frame of reference, we can show

$$r \in V \Rightarrow \text{ZFC} \vdash \underline{r} \in \underline{V}, \text{ and}$$
$$r \notin V \Rightarrow \text{ZFC} \vdash \underline{r} \notin \underline{V}.$$

Therefore, $\omega_0 \models \underline{r} \in \underline{V}$ iff $r \in V$. This proves (i).

Soft arguments gives (ii) and (iii), and (1) follows from the construction of $\mathscr{P}$.