

Solomonoff Induction Exercises

April 7, 2026

Setup and Definitions

For more setup and definitions, see [the corresponding post on Solomonoff induction](#).

A **semiprobability distribution** over a finite set \mathcal{X} is a function $Q : \mathcal{X} \rightarrow [0, 1]$ with $Q(x) \geq 0$ and $\sum_{x \in \mathcal{X}} Q(x) \leq 1$.

A **semimeasure** over infinite binary sequences is a function $\nu : \{0, 1\}^* \rightarrow [0, 1]$ with $\nu(\epsilon) \leq 1$ and $\nu(x) \geq \nu(x0) + \nu(x1)$ for all $x \in \{0, 1\}^*$. A semimeasure is a **measure** if equality holds for both equations.

The **universal a priori distribution** M is a lower semicomputable semimeasure, which can be written in the form

$$M(x_{\leq n}) = \sum_{\nu \in \mathcal{M}_{sol}} P_{ap}(\nu) \nu(x_{\leq n}) \quad (1)$$

for all finite strings $x_{\leq n}$, where \mathcal{M}_{sol} is the set of all lower semicomputable semimeasures and $P_{ap}(\nu)$ is the **a priori probability** of ν .

With $\nu_i, i = 1, 2, \dots$ an effective enumeration of all lower semicomputable discrete semimeasures and the **Solomonoff prior** $P_{sol}(i) = 2^{-K(i)}$, there also exist two constants $0 < c < C$ such that

$$c \cdot \sum_{i \in \mathbb{N}} P_{sol}(i) \nu_i(x_{\leq n}) < M(x_{\leq n}) < C \cdot \sum_{i \in \mathbb{N}} P_{sol}(i) \nu_i(x_{\leq n}). \quad (2)$$

Problem 1

Let P be a probability distribution over \mathcal{X} and Q be a semiprobability distribution over the same set. The *Kullback–Leibler (KL) divergence* from P to Q is defined as

$$D_{KL}(P \parallel Q) := \sum_{x \in \mathcal{X}} P(x) \ln \frac{P(x)}{Q(x)},$$

where we use the conventions $0 \ln \frac{0}{q} = 0$ for any $q \geq 0$, and $p \ln \frac{p}{0} = \infty$ for $p > 0$.

Now let (X, Y) be a pair of random variables taking values in $\mathcal{X} \times \mathcal{Y}$, and let P and Q be two joint distributions over (X, Y) (with Q only being a *semiprobability* distribution). Define the *conditional KL divergence* as

$$D_{\text{KL}}(P(Y | X) \parallel Q(Y | X)) := \sum_{x \in \mathcal{X}} P(x) \sum_{y \in \mathcal{Y}} P(y | x) \ln \frac{P(y | x)}{Q(y | x)}.$$

Show that the KL divergence satisfies the *chain rule*:

$$D_{\text{KL}}(P(X, Y) \parallel Q(X, Y)) = D_{\text{KL}}(P(X) \parallel Q(X)) + D_{\text{KL}}(P(Y | X) \parallel Q(Y | X)).$$

Problem 2

Let P be a joint probability distribution over random variables X_1, \dots, X_n taking values in $\mathcal{X}_1 \times \dots \times \mathcal{X}_n$, and let Q be a *semiprobability* distribution over the same set of variables. Using the notation $X_{<k} := (X_1, \dots, X_{k-1})$, show by induction on n (using the result of Problem 1) that

$$D_{\text{KL}}(P(X_1, \dots, X_n) \parallel Q(X_1, \dots, X_n)) = \sum_{t=1}^n D_{\text{KL}}(P(X_t | X_{<t}) \parallel Q(X_t | X_{<t})),$$

where $P(X_1 | X_{<1}) := P(X_1)$ and similarly for Q .

Problem 3

Show that for all $z > 0$,

$$\ln(z) \geq 1 - \frac{1}{z}.$$

Hint: Consider the function $f(z) = \ln(z) - 1 + \frac{1}{z}$ and show that it attains its global minimum at $z = 1$.

Problem 4

Let P be a probability distribution over $\mathbb{B} = \{0, 1\}$ and let Q be a semiprobability distribution over \mathbb{B} . Write $p_0 = P(0)$, $p_1 = P(1) = 1 - p_0$, $q_0 = Q(0)$, $q_1 = Q(1)$. We want to show that

$$\sum_{x \in \mathbb{B}} (Q(x) - P(x))^2 \leq \sum_{x \in \mathbb{B}} P(x) \ln \frac{P(x)}{Q(x)}.$$

i) Show that the inequality holds whenever $Q(x) = 0$ for some $x \in \mathbb{B}$.

Hint: Distinguish the cases $P(x) > 0$ and $P(x) = 0$, and use Problem 3 for the latter.

ii) Assuming $q_0, q_1 > 0$, show that it suffices to prove the inequality for the case $q_0 + q_1 = 1$, i.e., when Q is a full probability distribution.

Hint: Define $F(q_0, q_1) := p_0 \ln \frac{p_0}{q_0} + p_1 \ln \frac{p_1}{q_1} - (q_0 - p_0)^2 - (q_1 - p_1)^2$ and show that F is decreasing in q_1 .

iii) Now assume $q_0 + q_1 = 1$. Show that

$$2(p_0 - q_0)^2 \leq p_0 \ln \frac{p_0}{q_0} + p_1 \ln \frac{p_1}{1 - q_0}$$

and observe this finishes the proof.

Hint: Define $g(q_0)$ as the right-hand side minus the left-hand side, compute $g'(q_0)$, and use that $q_0(1 - q_0) \leq \frac{1}{4}$.

Problem 5

Let μ be any lower semicomputable *measure* (!) on binary sequences, and assume it generates our actual universe. Let M be the Solomonoff mixture distribution.

Define the cumulative expected prediction error of predicting sequences sampled by μ via M as:

$$S_\infty^\mu := \sum_{t=1}^{\infty} \sum_{x_{<t} \in \mathbb{B}^*} \mu(x_{<t}) \sum_{x_t \in \mathbb{B}} (M(x_t | x_{<t}) - \mu(x_t | x_{<t}))^2.$$

Show that

$$S_\infty^\mu \leq -\ln P_{ap}(\mu).$$

Hint: Apply Problem 4 to each summand to pass from squared error to KL divergence, then make the infinite sum explicit as a limit of finite sums up to n , then use Problem 2 to telescope the KL divergences, and finally use the mixture representation of M .

Problem 5'

Under the same setup as Problem 5, show that there exists a constant $\kappa \in \mathbb{R}$ (independent of μ) such that

$$S_\infty^\mu \leq K(\mu) \ln 2 + \kappa,$$

where $K(\mu) := \min\{K(i) : \nu_i = \mu\}$ is the Kolmogorov complexity of μ (minimized over all indices i in the effective enumeration from the setup that give rise to μ).

Hint: Follow the same strategy as Problem 5, but instead of the mixture representation Equation (1), use the lower bound $M(x_{\leq n}) > c \cdot \sum_i P_{sol}(i) \nu_i(x_{\leq n})$ from Equation (2), together with $P_{sol}(i) = 2^{-K(i)}$.

Interpretation

Problems 5 and 5' show that the *cumulative* expected prediction error S_∞^μ is finite. Since S_∞^μ is a sum of non-negative terms, this immediately implies that the per-step expected prediction error

$$d_t := \sum_{x_{<t} \in \mathbb{B}^*} \mu(x_{<t}) \sum_{x_t \in \mathbb{B}} (M(x_t | x_{<t}) - \mu(x_t | x_{<t}))^2$$

converges to zero as $t \rightarrow \infty$. In other words, M 's predictions become indistinguishable from μ 's predictions on average over histories.