STOCHASTIC INTERPOLANTS III

Generative Models, Likelihood Control, Density Estimation, Cross-Entropy, and Proofs

Ziseok Lee

April 30, 2026

DOCUMENT III: MAIN GOAL

Document II proved that the interpolant density satisfies:

$$\partial_t \rho + \nabla \cdot (b\rho) = 0,$$

and:

$$\partial_t \rho + \nabla \cdot ((b + \varepsilon s)\rho) = \varepsilon \Delta \rho.$$

Document III turns these PDEs into:

$$\boxed{\text{generative models}}$$

and proves:

$$\boxed{\text{likelihood control and density estimation formulas.}}$$

Core question:

If $\hat{b}, \hat{s}$ are learned imperfectly, what can we guarantee?

THREE DYNAMICS WITH THE SAME MARGINALS

The same density path $\rho(t)$ can be realized by:

1. Probability-flow ODE

$$\dot{X}_t = b(t, X_t).$$

2. Forward SDE

$$\mathrm{d}X_t^F = b^F(t, X_t^F)\,\mathrm{d}t + \sqrt{2\varepsilon(t)}\;\mathrm{d}W_t, \qquad b^F = b + \varepsilon s.$$

3. Backward SDE

$$\mathrm{d}X_t^B = b^B(t, X_t^B)\,\mathrm{d}t + \sqrt{2\varepsilon(t)}\;\mathrm{d}W_t^B, \qquad b^B = b - \varepsilon s.$$

COROLLARY 18: PROBABILITY-FLOW ODE

If:

$$X_0 \sim \rho_0,$$

and:

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = b(t, X_t),$$

then:

$$X_t \sim \rho(t) \qquad \forall t \in [0,1].$$

In particular:

$$X_1 \sim \rho_1.$$

Likewise, solving backward from:

$$X_1 \sim \rho_1$$

recovers:

$$X_0 \sim \rho_0.$$

PROOF OF PROBABILITY-FLOW ODE CLAIM

Let $\hat{\rho}(t)$ be the law of X_t. The ODE flow implies:

$$\partial_t \hat{\rho} + \nabla \cdot (b\hat{\rho}) = 0.$$

The interpolant density satisfies:

$$\partial_t \rho + \nabla \cdot (b\rho) = 0.$$

Initial data agree:

$$\hat{\rho}(0) = \rho_0 = \rho(0).$$

By uniqueness of the transport equation:

$$\hat{\rho}(t) = \rho(t).$$

Thus:

$$X_1 \sim \rho(1) = \rho_1.$$

COROLLARY 18: FORWARD SDE

Let:

$$X_0^F \sim \rho_0,$$

and solve:

$$\mathrm{d}X_t^F = b^F(t, X_t^F)\,\mathrm{d}t + \sqrt{2\varepsilon(t)}\;\mathrm{d}W_t.$$

The density of X_t^F satisfies:

$$\partial_t \hat{\rho} + \nabla \cdot (b^F \hat{\rho}) = \varepsilon \Delta \hat{\rho}.$$

The interpolant density satisfies the same FPE:

$$\partial_t \rho + \nabla \cdot (b^F \rho) = \varepsilon \Delta \rho.$$

Therefore:

$$X_t^F \sim \rho(t), \qquad X_1^F \sim \rho_1.$$

COROLLARY 18: BACKWARD SDE

Let:

$$X_1^B \sim \rho_1.$$

Solve backward:

$$\mathrm{d}X_t^B = b^B(t, X_t^B)\,\mathrm{d}t + \sqrt{2\varepsilon(t)}\;\mathrm{d}W_t^B.$$

The density satisfies:

$$\partial_t \rho + \nabla \cdot (b^B \rho) = -\varepsilon \Delta \rho,$$

with final condition:

$$\rho(1) = \rho_1.$$

Thus:

$$X_t^B \sim \rho(t), \qquad X_0^B \sim \rho_0.$$

BACKWARD SDE AS FORWARD SDE

Define:

$$Z_t^F = X_{1-t}^B.$$

Then Z_t^F solves a forward SDE:

$$\mathrm{d}Z_t^F = -b^B(1-t, Z_t^F)\,\mathrm{d}t + \sqrt{2\varepsilon(t)}\,\mathrm{d}W_t.$$

Initial condition:

$$Z_0^F = X_1^B \sim \rho_1.$$

Thus backward sampling from ρ_1 to ρ_0 can be implemented in forward time.

IMPORTANT DISTINCTION

The processes:

$$x_t, \quad X_t, \quad X_t^F, \quad X_t^B$$

are not the same stochastic process.

But:

$$\mathcal{L}(x_t) = \mathcal{L}(X_t) = \mathcal{L}(X_t^F) = \mathcal{L}(X_t^B) = \rho(t).$$

Generative modeling only needs the terminal marginal:

$$X_1 \sim \rho_1.$$

Pathwise differences matter for:

numerical error, statistical robustness, likelihood control.

REVERSE ITÔ CALCULUS

For:

$$\mathrm{d}X_t^B = b^B(t, X_t^B)\,\mathrm{d}t + \sqrt{2\varepsilon(t)}\;\mathrm{d}W_t^B,$$

and $f \in C^{1,2}$:

$$\boxed{\mathrm{d}f(t, X_t^B) = \partial_t f(t, X_t^B)\,\mathrm{d}t + \nabla f(t, X_t^B)\cdot\,\mathrm{d}X_t^B - \varepsilon(t)\Delta f(t, X_t^B)\,\mathrm{d}t.}$$

The sign of the Laplacian is negative because the stochastic differential is backward in time.

PROOF OF REVERSE ITÔ FORMULA

Let:

$$X_t^B = Z_{1-t}^F.$$

Set:

$$g(\tau, z) = f(1-\tau, z).$$

Apply forward Itô to:

$$g(\tau, Z_\tau^F).$$

Since:

$$\partial_\tau g(\tau, z) = -\partial_t f(1-\tau, z),$$

and:

$$\mathrm{d}\tau = -\,\mathrm{d}t,$$

the diffusion term changes sign in the backward differential.

This yields:

$$\mathrm{d}f(t, X_t^B) = \partial_t f \ \mathrm{d}t + \nabla f \cdot \mathrm{d}X_t^B - \varepsilon \Delta f \ \mathrm{d}t.$$

BACKWARD ITÔ ISOMETRIES

If g is sufficiently regular, then:

$$\mathbb{E}_B^x \int_t^1 g(\tau, X_\tau^B) \cdot \mathrm{d}W_\tau^B = 0.$$

And:

$$\mathbb{E}_B^x \left| \int_t^1 g(\tau, X_\tau^B) \cdot \mathrm{d}W_\tau^B \right|^2 = \int_t^1 \mathbb{E}_B^x |g(\tau, X_\tau^B)|^2 \,\mathrm{d}\tau.$$

These follow by converting the backward integral into a forward Itô integral under the change:

$$\tau \mapsto 1 - \tau.$$

APPROXIMATE MODELS

In practice, learn:

$$\hat{b} \approx b, \qquad \hat{s} \approx s.$$

Define approximate forward drift:

$$\hat{b}^F = \hat{b} + \varepsilon \hat{s}.$$

The approximate SDE:

$$\mathrm{d}\hat{X}_t^F = \hat{b}^F(t, \hat{X}_t^F)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t.$$

Let $\hat{\rho}(t)$ be its density:

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}^F \hat{\rho}) = \varepsilon \Delta \hat{\rho}, \qquad \hat{\rho}(0) = \rho_0.$$

LIKELIHOOD CONTROL QUESTION

We want to control:

$$\mathrm{KL}(\rho_1 \| \hat{\rho}(1)).$$

This is equivalent to controlling the cross-entropy up to the entropy of ρ_1:

$$\mathrm{H}(\rho_1 \| \hat{\rho}(1)) = \mathrm{KL}(\rho_1 \| \hat{\rho}(1)) + \mathrm{H}(\rho_1).$$

Main result:

$$\mathrm{KL}(\rho_1 \| \hat{\rho}(1)) \leq \frac{1}{2\varepsilon} \Delta L_b + \frac{\varepsilon}{2} \Delta L_s.$$

This holds for stochastic dynamics with $\varepsilon > 0$.

KL DIFFERENTIATION TEMPLATE

Let:

$$K(t) = \mathrm{KL}(\rho(t) \| \hat{\rho}(t)) = \int \rho \log \frac{\rho}{\hat{\rho}}.$$

Differentiate:

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \int \partial_t \rho \log \frac{\rho}{\hat{\rho}} + \int \rho \left(\frac{\partial_t \rho}{\rho} - \frac{\partial_t \hat{\rho}}{\hat{\rho}} \right).$$

Since:

$$\int \partial_t \rho = 0,$$

this becomes:

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \int \partial_t \rho \log \frac{\rho}{\hat{\rho}} - \int \frac{\rho}{\hat{\rho}} \partial_t \hat{\rho}.$$

LEMMA 21: TWO TRANSPORT EQUATIONS

Suppose:

$$\partial_t \rho + \nabla \cdot (b\rho) = 0, \qquad \rho(0) = \rho_0,$$

and:

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}\hat{\rho}) = 0, \qquad \hat{\rho}(0) = \rho_0.$$

Then:

$$\boxed{\mathrm{KL}(\rho(1) \| \hat{\rho}(1)) = \int_0^1 \int (\nabla \log \hat{\rho} - \nabla \log \rho) \cdot (\hat{b} - b)\rho \; \mathrm{d}x \, \mathrm{d}t.}$$

PROOF OF LEMMA 21: START

Let:

$$K(t) = \int \rho \log \frac{\rho}{\hat{\rho}}.$$

Using the KL differentiation template:

$$\dot{K} = \int \partial_t \rho \log \frac{\rho}{\hat{\rho}} - \int \frac{\rho}{\hat{\rho}} \partial_t \hat{\rho}.$$

Insert:

$$\partial_t \rho = -\nabla \cdot (b\rho),$$
$$\partial_t \hat{\rho} = -\nabla \cdot (\hat{b}\hat{\rho}).$$

PROOF OF LEMMA 21: INTEGRATION BY PARTS

First term:

$$-\int \nabla \cdot (b\rho) \log \frac{\rho}{\hat{\rho}} = \int b\rho \cdot \nabla \log \frac{\rho}{\hat{\rho}}.$$

Second term:

$$\int \frac{\rho}{\hat{\rho}} \nabla \cdot (\hat{b}\hat{\rho}) = -\int \hat{b}\hat{\rho} \cdot \nabla \left(\frac{\rho}{\hat{\rho}}\right).$$

Since:

$$\nabla \left(\frac{\rho}{\hat{\rho}}\right) = \frac{\rho}{\hat{\rho}} (\nabla \log \rho - \nabla \log \hat{\rho}),$$

the second term becomes:

$$-\int \hat{b}\rho \cdot (\nabla \log \rho - \nabla \log \hat{\rho}).$$

PROOF OF LEMMA 21: FINAL IDENTITY

Combine:

$$\dot{K} = \int b\rho \cdot (\nabla \log \rho - \nabla \log \hat{\rho}) - \int \hat{b}\rho \cdot (\nabla \log \rho - \nabla \log \hat{\rho}).$$

Thus:

$$\dot{K} = \int (\nabla \log \hat{\rho} - \nabla \log \rho) \cdot (\hat{b} - b)\rho.$$

Since:

$$K(0) = \mathrm{KL}(\rho_0 \| \rho_0) = 0,$$

integrating from 0 to 1 gives Lemma 21.

WHY ODE DRIFT REGRESSION IS NOT ENOUGH

Lemma 21 gives:

$$\mathrm{KL}(\rho(1)\|\hat{\rho}(1)) = \int_0^1 \int (\nabla \log \hat{\rho} - \nabla \log \rho) \cdot (\hat{b} - b)\rho.$$

Even if:

$$\int |\hat{b} - b|^2 \rho$$

is small, the factor:

$$\nabla \log \hat{\rho} - \nabla \log \rho$$

may be large.

Therefore deterministic likelihood control requires Fisher-type control:

$$\mathrm{FI}(\rho\|\hat{\rho}) = \int |\nabla \log \rho - \nabla \log \hat{\rho}|^2 \rho.$$

LEMMA 22: TWO FOKKER–PLANCK EQUATIONS

Suppose:

$$\partial_t \rho + \nabla \cdot (b^F \rho) = \varepsilon \Delta \rho,$$
$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}^F \hat{\rho}) = \varepsilon \Delta \hat{\rho},$$

with:

$$\rho(0) = \hat{\rho}(0) = \rho_0, \qquad \varepsilon > 0.$$

Then:

$$\mathrm{KL}(\rho(1) \| \hat{\rho}(1)) \leq \frac{1}{4\varepsilon} \int_0^1 \int |\hat{b}^F - b^F|^2 \rho \, \mathrm{d}x \, \mathrm{d}t.$$

PROOF OF LEMMA 22: REWRITE AS TRANSPORT

Use:

$$\Delta\rho = \nabla \cdot (\rho \nabla \log \rho).$$

Then:

$$\partial_t \rho + \nabla \cdot (b^F \rho) = \varepsilon \Delta \rho$$

becomes:

$$\partial_t \rho + \nabla \cdot \Big((b^F - \varepsilon \nabla \log \rho) \rho \Big) = 0.$$

Similarly:

$$\partial_t \hat{\rho} + \nabla \cdot \Big((\hat{b}^F - \varepsilon \nabla \log \hat{\rho}) \hat{\rho} \Big) = 0.$$

PROOF OF LEMMA 22: APPLY LEMMA 21

Apply Lemma 21 with effective velocities:

$$u = b^F - \varepsilon \nabla \log \rho,$$

$$\hat{u} = \hat{b}^F - \varepsilon \nabla \log \hat{\rho}.$$

Then:

$$\hat{u} - u = (\hat{b}^F - b^F) - \varepsilon(\nabla \log \hat{\rho} - \nabla \log \rho).$$

Therefore:

$$\dot{K} = \int a \cdot (\hat{b}^F - b^F)\rho - \varepsilon \int |a|^2 \rho,$$

where:

$$a = \nabla \log \hat{\rho} - \nabla \log \rho.$$

PROOF OF LEMMA 22: FISHER DISSIPATION

Thus:

$$\mathrm{KL}(\rho(1)\|\hat{\rho}(1)) = \int_0^1 \int a \cdot (\hat{b}^F - b^F)\rho - \varepsilon |a|^2 \rho.$$

The second term is:

$$-\varepsilon \mathrm{FI}(\rho\|\hat{\rho}).$$

This negative Fisher term is why stochastic dynamics provide likelihood control.

Use Young's inequality:

$$a \cdot c - \varepsilon |a|^2 \leq \frac{1}{4\varepsilon}|c|^2.$$

With:

$$c = \hat{b}^F - b^F.$$

PROOF OF LEMMA 22: FINAL BOUND

Pointwise:

$$a \cdot c - \varepsilon |a|^2 = -\varepsilon \left| a - \frac{c}{2\varepsilon} \right|^2 + \frac{1}{4\varepsilon} |c|^2 \leq \frac{1}{4\varepsilon} |c|^2.$$

Thus:

$$\mathrm{KL}(\rho(1) \| \hat{\rho}(1)) \leq \frac{1}{4\varepsilon} \int_0^1 \int |\hat{b}^F - b^F|^2 \rho.$$

This proves the FPE likelihood bound.

THEOREM 23: SETUP

Let:

$$b^F = b + \varepsilon s, \qquad \hat{b}^F = \hat{b} + \varepsilon \hat{s}.$$

The learned density satisfies:

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}^F \hat{\rho}) = \varepsilon \Delta \hat{\rho}, \qquad \hat{\rho}(0) = \rho_0.$$

The theorem bounds:

$$\mathrm{KL}(\rho_1 \| \hat{\rho}(1)).$$

Recall:

$$\rho(1) = \rho_1.$$

THEOREM 23: LOSS GAPS

From Document II:

$$L_b[\hat{b}] - L_b[b] = \frac{1}{2} \int_0^1 \int |\hat{b} - b|^2 \rho.$$

$$L_s[\hat{s}] - L_s[s] = \frac{1}{2} \int_0^1 \int |\hat{s} - s|^2 \rho.$$

These excess losses are exactly weighted squared errors under the interpolant law.

PROOF OF THEOREM 23

From Lemma 22:

$$\mathrm{KL}(\rho_1 \| \hat{\rho}(1)) \leq \frac{1}{4\varepsilon} \int |\hat{b}^F - b^F|^2 \rho.$$

But:

$$\hat{b}^F - b^F = (\hat{b} - b) + \varepsilon(\hat{s} - s).$$

Use:

$$|u + v|^2 \leq 2|u|^2 + 2|v|^2.$$

Then:

$$|\hat{b}^F - b^F|^2 \leq 2|\hat{b} - b|^2 + 2\varepsilon^2|\hat{s} - s|^2.$$

THEOREM 23: FIRST BOUND

Therefore:

$$\mathrm{KL}(\rho_1 \| \hat{\rho}(1)) \leq \frac{1}{2\varepsilon} \int |\hat{b} - b|^2 \rho + \frac{\varepsilon}{2} \int |\hat{s} - s|^2 \rho.$$

Using excess loss identities:

$$\int |\hat{b} - b|^2 \rho = 2(L_b[\hat{b}] - L_b[b]),$$

$$\int |\hat{s} - s|^2 \rho = 2(L_s[\hat{s}] - L_s[s]).$$

Depending on normalization, the paper’s stated form is:

$$\boxed{\mathrm{KL}(\rho_1 \| \hat{\rho}(1)) \leq \frac{1}{2\varepsilon} \Delta L_b + \frac{\varepsilon}{2} \Delta L_s.}$$

THEOREM 23: v, s VARIANT

Recall:

$$b = v - \gamma\dot{\gamma}s.$$

Define:

$$\hat{v} = \hat{b} + \gamma\dot{\gamma}\hat{s}.$$

Then:

$$\hat{b} - b = (\hat{v} - v) - \gamma\dot{\gamma}(\hat{s} - s).$$

For:

$$\hat{b}^F - b^F = (\hat{b} - b) + \varepsilon(\hat{s} - s),$$

we get:

$$\hat{b}^F - b^F = (\hat{v} - v) + (\varepsilon - \gamma\dot{\gamma})(\hat{s} - s).$$

This yields:

$$\mathrm{KL} \leq \frac{1}{2\varepsilon}\Delta L_v + \frac{\sup_t(\gamma\dot{\gamma} - \varepsilon)^2}{2\varepsilon}\Delta L_s.$$

OPTIMAL CONSTANT DIFFUSION COEFFICIENT

The first likelihood bound has the form:

$$B(\varepsilon) = \frac{A}{2\varepsilon} + \frac{\varepsilon C}{2},$$

where:

$$A = \Delta L_b, \qquad C = \Delta L_s.$$

Differentiate:

$$B'(\varepsilon) = -\frac{A}{2\varepsilon^2} + \frac{C}{2}.$$

Set $B'(\varepsilon) = 0$:

$$\varepsilon^2 = \frac{A}{C}.$$

Thus:

$$\boxed{\varepsilon^* = \left(\frac{\Delta L_b}{\Delta L_s}\right)^{1/2}.}$$

INTERPRETATION OF ε^*

If:

$$\Delta L_b \ll \Delta L_s,$$

then:

$$\varepsilon^* \ll 1.$$

Use mostly deterministic sampling.

If:

$$\Delta L_s \ll \Delta L_b,$$

then:

$$\varepsilon^* \gg 1.$$

More stochasticity helps the KL bound.

But large ε increases numerical cost and sampling variance.

DENSITY ESTIMATION FOR THE ODE

Let $X_{s,t}(x)$ solve:

$$\frac{\mathrm{d}}{\mathrm{d}t} X_{s,t}(x) = \hat{b}(t, X_{s,t}(x)), \qquad X_{s,s}(x) = x.$$

If:

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}\hat{\rho}) = 0, \qquad \hat{\rho}(0) = \rho_0,$$

then:

$$\hat{\rho}(t, x) = \exp\left(- \int_0^t \nabla \cdot \hat{b}(\tau, X_{t,\tau}(x)) \, \mathrm{d}\tau \right) \rho_0(X_{t,0}(x)).$$

PROOF OF ODE DENSITY FORMULA

Along a characteristic:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\rho}(t, X_{s,t}) = \partial_t \hat{\rho} + \hat{b} \cdot \nabla \hat{\rho}.$$

The transport equation:

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}\hat{\rho}) = 0$$

is:

$$\partial_t \hat{\rho} + \hat{b} \cdot \nabla \hat{\rho} = -(\nabla \cdot \hat{b})\hat{\rho}.$$

Therefore:

$$\frac{\mathrm{d}}{\mathrm{d}t} \log \hat{\rho}(t, X_{s,t}) = -\nabla \cdot \hat{b}(t, X_{s,t}).$$

Integrate.

BACKWARD ODE DENSITY FORMULA

If:

$$\hat{\rho}(1) = \rho_1,$$

then:

$$\boxed{\hat{\rho}(t, x) = \exp\left(\int_t^1 \nabla \cdot \hat{b}(\tau, X_{t,\tau}(x))\, \mathrm{d}\tau\right) \rho_1(X_{t,1}(x)).}$$

This follows by integrating the same characteristic identity from t to 1.

ODE CROSS-ENTROPY FORMULA

If $\hat{\rho}(0) = \rho_0$, then:

$$\mathrm{H}(\rho_1 \| \hat{\rho}(1)) = -\mathbb{E}_{x_1 \sim \rho_1} \log \hat{\rho}(1, x_1).$$

Using the ODE density formula:

$$\boxed{\mathrm{H}(\rho_1 \| \hat{\rho}(1)) = \mathbb{E}_1 \int_0^1 \nabla \cdot \hat{b}(\tau, X_{1,\tau}(x_1)) \, \mathrm{d}\tau - \mathbb{E}_1 \log \rho_0(X_{1,0}(x_1)).}$$

REVERSE ODE CROSS-ENTROPY

If $\hat{\rho}(1) = \rho_1$, then:

$$\mathrm{H}(\rho_0 \| \hat{\rho}(0)) = -\mathbb{E}_{x_0 \sim \rho_0} \log \hat{\rho}(0, x_0).$$

Using the backward formula:

$$\boxed{\mathrm{H}(\rho_0 \| \hat{\rho}(0)) = -\mathbb{E}_0 \int_0^1 \nabla \cdot \hat{b}(\tau, X_{0,\tau}(x_0))\, \mathrm{d}\tau - \mathbb{E}_0 \log \rho_1(X_{0,1}(x_0)).}$$

SDE DENSITY ESTIMATION: SETUP

Let:

$$\hat{b}^F = \hat{b} + \varepsilon \hat{s}, \qquad \hat{b}^B = \hat{b} - \varepsilon \hat{s}.$$

Forward model density:

$$\partial_t \hat{\rho}^F + \nabla \cdot (\hat{b}^F \hat{\rho}^F) = \varepsilon \Delta \hat{\rho}^F, \qquad \hat{\rho}^F(0) = \rho_0.$$

Backward model density:

$$\partial_t \hat{\rho}^B + \nabla \cdot (\hat{b}^B \hat{\rho}^B) = -\varepsilon \Delta \hat{\rho}^B, \qquad \hat{\rho}^B(1) = \rho_1.$$

AUXILIARY SDES

To estimate densities, switch the drifts.

Auxiliary forward SDE:

$$\mathrm{d}Y_t^F = \hat{b}^B(t, Y_t^F)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t, \qquad Y_0^F = x.$$

Auxiliary backward SDE:

$$\mathrm{d}Y_t^B = \hat{b}^F(t, Y_t^B)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t^B, \qquad Y_1^B = x.$$

These are not the samplers; they are density-estimation devices.

THEOREM 26: FORWARD FPE DENSITY FORMULA

For:

$$\partial_t \hat{\rho}^F + \nabla \cdot (\hat{b}^F \hat{\rho}^F) = \varepsilon \Delta \hat{\rho}^F, \qquad \hat{\rho}^F(0) = \rho_0,$$

we have:

$$\boxed{\hat{\rho}^F(1,x) = \mathbb{E}_B^x \left[\exp\left(-\int_0^1 \nabla \cdot \hat{b}^F(t, Y_t^B)\, \mathrm{d}t \right) \rho_0(Y_0^B) \right].}$$

Here Y_t^B solves the auxiliary backward SDE with drift $\hat{b}^F$ and terminal condition $Y_1^B = x$.

THEOREM 26: BACKWARD FPE DENSITY FORMULA

For:

$$\partial_t \hat{\rho}^B + \nabla \cdot (\hat{b}^B \hat{\rho}^B) = -\varepsilon \Delta \hat{\rho}^B, \qquad \hat{\rho}^B(1) = \rho_1,$$

we have:

$$\boxed{\hat{\rho}^B(0,x) = \mathbb{E}_F^x \left[\exp\left(\int_0^1 \nabla \cdot \hat{b}^B(t, Y_t^F)\, \mathrm{d}t \right) \rho_1(Y_1^F) \right].}$$

Here Y_t^F solves the auxiliary forward SDE with drift $\hat{b}^B$ and initial condition $Y_0^F = x$.

PROOF OF THEOREM 26: KEY COMPUTATION

Let:

$$\hat{\rho}^F$$

solve the forward FPE.

Apply backward Itô to:

$$\hat{\rho}^F(t, Y_t^B),$$

where:

$$\mathrm{d}Y_t^B = \hat{b}^F(t, Y_t^B)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t^B.$$

Backward Itô:

$$\mathrm{d}\hat{\rho}^F = \partial_t \hat{\rho}^F\,\mathrm{d}t + \nabla \hat{\rho}^F \cdot \mathrm{d}Y_t^B - \varepsilon \Delta \hat{\rho}^F\,\mathrm{d}t.$$

PROOF OF THEOREM 26: USE THE PDE

Substitute:

$$\partial_t \hat{\rho}^F = -\nabla \cdot (\hat{b}^F \hat{\rho}^F) + \varepsilon \Delta \hat{\rho}^F .$$

Then:

$$\mathrm{d}\hat{\rho}^F(t, Y_t^B) = -\nabla \cdot (\hat{b}^F \hat{\rho}^F)\,\mathrm{d}t + \hat{b}^F \cdot \nabla \hat{\rho}^F\,\mathrm{d}t + \sqrt{2\varepsilon}\nabla \hat{\rho}^F \cdot \mathrm{d}W_t^B .$$

Since:

$$\nabla \cdot (\hat{b}^F \hat{\rho}^F) = (\nabla \cdot \hat{b}^F)\hat{\rho}^F + \hat{b}^F \cdot \nabla \hat{\rho}^F ,$$

we get:

$$\mathrm{d}\hat{\rho}^F = -(\nabla \cdot \hat{b}^F)\hat{\rho}^F\,\mathrm{d}t + \text{martingale.}$$

PROOF OF THEOREM 26: EXPONENTIAL MARTINGALE

Define:

$$A_t = \exp\left(-\int_t^1 \nabla \cdot \hat{b}^F(\tau, Y_\tau^B)\,\mathrm{d}\tau\right).$$

Then:

$$\mathrm{d}A_t = A_t\,\nabla \cdot \hat{b}^F(t, Y_t^B)\,\mathrm{d}t.$$

Using product rule:

$$\mathrm{d}(A_t \hat{\rho}^F(t, Y_t^B)) = A_t \sqrt{2\varepsilon} \nabla \hat{\rho}^F(t, Y_t^B) \cdot \,\mathrm{d}W_t^B.$$

Hence:

$$A_t \hat{\rho}^F(t, Y_t^B)$$

is a backward martingale.

PROOF OF THEOREM 26: TAKE EXPECTATIONS

Condition on:

$$Y_1^B = x.$$

At $t = 1$:

$$A_1 = 1, \qquad \hat{\rho}^F(1, Y_1^B) = \hat{\rho}^F(1, x).$$

At $t = 0$:

$$A_0 = \exp\left(-\int_0^1 \nabla \cdot \hat{b}^F(t, Y_t^B)\,\mathrm{d}t\right),$$

$$\hat{\rho}^F(0, Y_0^B) = \rho_0(Y_0^B).$$

Martingale equality gives the formula.

PROOF OF BACKWARD FORMULA

The backward density formula is analogous.

Apply forward Itô to:

$$\hat{\rho}^B(t, Y_t^F),$$

where:

$$\mathrm{d}Y_t^F = \hat{b}^B(t, Y_t^F)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t.$$

Use:

$$\partial_t \hat{\rho}^B + \nabla \cdot (\hat{b}^B \hat{\rho}^B) = -\varepsilon \Delta \hat{\rho}^B.$$

The product:

$$\exp\left(\int_0^t \nabla \cdot \hat{b}^B(\tau, Y_\tau^F)\,\mathrm{d}\tau\right) \hat{\rho}^B(t, Y_t^F)$$

is a forward martingale.

SDE CROSS-ENTROPY: EXACT FORMULA

For the forward FPE model:

$$\mathrm{H}(\rho_1 \| \hat{\rho}^F(1)) = -\mathbb{E}_{x_1 \sim \rho_1} \log \hat{\rho}^F(1, x_1).$$

Using Theorem 26:

$$\boxed{\mathrm{H}(\rho_1 \| \hat{\rho}^F(1)) = -\mathbb{E}_1 \log \mathbb{E}_B^{x_1} \left[e^{-\int_0^1 \nabla \cdot \hat{b}^F(t, Y_t^B)\, \mathrm{d}t} \rho_0(Y_0^B) \right].}$$

BACKWARD SDE CROSS-ENTROPY

For the backward FPE model:

$$\mathrm{H}(\rho_0 \| \hat{\rho}^B(0)) = -\mathbb{E}_{x_0 \sim \rho_0} \log \hat{\rho}^B(0, x_0).$$

Using Theorem 26:

$$\boxed{\mathrm{H}(\rho_0 \| \hat{\rho}^B(0)) = -\mathbb{E}_0 \log \mathbb{E}_F^{x_0} \left[e^{\int_0^1 \nabla \cdot \hat{b}^B(t, Y_t^F)\, \mathrm{d}t} \rho_1(Y_1^F) \right].}$$

JENSEN UPPER BOUND

The exact SDE cross-entropy contains:

$$-\log \mathbb{E}[\cdots].$$

Using Jensen:

$$-\log \mathbb{E}[e^{-A}\rho_0(Y_0)] \le \mathbb{E}[A] - \mathbb{E}[\log \rho_0(Y_0)].$$

Thus:

$$\mathrm{H}(\rho_1 \| \hat{\rho}^F(1)) \le \mathbb{E}_1 \mathbb{E}_B^{x_1} \int_0^1 \nabla \cdot \hat{b}^F(t, Y_t^B)\,\mathrm{d}t - \mathbb{E}_1 \mathbb{E}_B^{x_1} \log \rho_0(Y_0^B).$$

WHAT JENSEN LOSES

The paper also identifies the missing correction:

$$\mathrm{H}(\rho_1 \| \hat{\rho}^F(1)) = \mathbb{E}_1 \mathbb{E}_B^{x_1} \int_0^1 \left[\nabla \cdot \hat{b}^F - \varepsilon |\nabla \log \hat{\rho}^F|^2 \right] \mathrm{d}t - \mathbb{E}_1 \mathbb{E}_B^{x_1} \log \rho_0(Y_0^B).$$

The term:

$$\nabla \log \hat{\rho}^F$$

is generally unavailable.

Using $\hat{s}$ as a proxy is possible but not controlled in general.

ODE VS SDE DENSITY ESTIMATION

ODE density estimation:

$$\text{one path } X_{1,t}(x_1) \quad \Longrightarrow \quad \log \hat{\rho}(1, x_1).$$

SDE density estimation:

$$\text{expectation over auxiliary paths } Y_t^B.$$

ODE:

$$\hat{\rho}(1, x) = e^{-\int \nabla \cdot \hat{b}} \rho_0(X_{1,0}(x)).$$

SDE:

$$\hat{\rho}^F(1, x) = \mathbb{E}_B^x \left[e^{-\int \nabla \cdot \hat{b}^F} \rho_0(Y_0^B) \right].$$

LIKELIHOOD CONTROL VS DENSITY ESTIMATION

Likelihood control theorem:

$$\mathrm{KL}(\rho_1 \| \hat{\rho}(1)) \leq \frac{1}{2\varepsilon}\Delta L_b + \frac{\varepsilon}{2}\Delta L_s.$$

Density estimation formula:

$$\hat{\rho}^F(1, x) = \mathbb{E}_B^x \left[e^{-\int \nabla \cdot \hat{b}^F} \rho_0(Y_0^B) \right].$$

The first is a theoretical guarantee. The second is a numerical estimator for evaluating model likelihood or cross-entropy.

HUTCHINSON TRACE ESTIMATOR

In high dimension:

$$\nabla \cdot b(t,x) = \mathrm{Tr}(\nabla_x b(t,x))$$

is expensive.

If ξ is a random vector with:

$$\mathbb{E}[\xi\xi^\top] = I_d,$$

then:

$$\mathrm{Tr}(A) = \mathbb{E}_\xi[\xi^\top A \xi].$$

Thus:

$$\nabla \cdot b(t,x) = \mathbb{E}_\xi[\xi^\top \nabla_x b(t,x) \xi].$$

This is used in likelihood estimation for continuous normalizing flows and stochastic interpolants.

MODEL SELECTION BY CROSS-ENTROPY

Given validation samples:

$$x_1^j \sim \rho_1,$$

estimate:

$$\mathrm{H}(\rho_1 \| \hat{\rho}(1)) = -\mathbb{E}_1 \log \hat{\rho}(1, x_1).$$

For ODE:

$$-\log \hat{\rho}(1, x_1) = \int_0^1 \nabla \cdot \hat{b}(\tau, X_{1,\tau})\, \mathrm{d}\tau - \log \rho_0(X_{1,0}).$$

For SDE:

$$-\log \hat{\rho}^F(1, x_1) = -\log \mathbb{E}_B^{x_1}[\cdots].$$

CONCEPTUAL LESSON: WHY SDES ARE ROBUST

For ODEs:

$$\dot{K} = \int a \cdot (\hat{b} - b)\rho,$$

where:

$$a = \nabla \log \hat{\rho} - \nabla \log \rho.$$

There is no negative term.

For SDEs:

$$\dot{K} = \int a \cdot (\hat{b}^F - b^F)\rho - \varepsilon \int |a|^2 \rho.$$

The diffusion gives a dissipative Fisher term:

$$-\varepsilon \mathrm{FI}(\rho || \hat{\rho}).$$

This term converts drift error into KL control.

FULL THEOREM CHAIN SO FAR

$$x_t = I + \gamma z$$

$$\Downarrow$$

$$\rho(t) = \mathcal{L}(x_t), \qquad \partial_t \rho + \nabla \cdot (b\rho) = 0$$

$$\Downarrow$$

$$s = \nabla \log \rho, \qquad b^F = b + \varepsilon s$$

$$\Downarrow$$

$$\partial_t \rho + \nabla \cdot (b^F \rho) = \varepsilon \Delta \rho$$

$$\Downarrow$$

$$\mathrm{d}X_t^F = b^F(t, X_t^F)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}W_t$$

$$\Downarrow$$

$$X_1^F \sim \rho_1.$$

FULL STATISTICAL CHAIN

Training objectives:

$$L_b, \qquad L_s.$$

Exact excess-risk identities:

$$\Delta L_b = \frac{1}{2}\|\hat{b} - b\|^2_{L^2(\rho)},$$
$$\Delta L_s = \frac{1}{2}\|\hat{s} - s\|^2_{L^2(\rho)}.$$

SDE drift error:

$$\hat{b}^F - b^F = (\hat{b} - b) + \varepsilon(\hat{s} - s).$$

KL control:

$$\mathrm{KL}(\rho_1 \| \hat{\rho}(1)) \leq \frac{1}{2\varepsilon}\Delta L_b + \frac{\varepsilon}{2}\Delta L_s.$$

PROOF EXERCISE 1

Reprove Lemma 21:

Assume:

$$\partial_t \rho + \nabla \cdot (b\rho) = 0, \qquad \partial_t \hat{\rho} + \nabla \cdot (\hat{b}\hat{\rho}) = 0.$$

Show:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{KL}(\rho(t)\|\hat{\rho}(t)) = \int (\nabla \log \hat{\rho} - \nabla \log \rho) \cdot (\hat{b} - b)\rho.$$

Then integrate from 0 to 1.

PROOF EXERCISE 2

Reprove Lemma 22:

Assume:

$$\partial_t \rho + \nabla \cdot (b^F \rho) = \varepsilon \Delta \rho,$$
$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}^F \hat{\rho}) = \varepsilon \Delta \hat{\rho}.$$

Rewrite each FPE as a transport equation with velocities:

$$b^F - \varepsilon \nabla \log \rho,$$

$$\hat{b}^F - \varepsilon \nabla \log \hat{\rho}.$$

Apply Lemma 21 and Young's inequality.

PROOF EXERCISE 3

Prove Lemma 25:

Given:

$$\partial_t \hat{\rho} + \nabla \cdot (\hat{b}\hat{\rho}) = 0,$$

and characteristics:

$$\dot{X}_{s,t} = \hat{b}(t, X_{s,t}),$$

show:

$$\frac{\mathrm{d}}{\mathrm{d}t} \log \hat{\rho}(t, X_{s,t}) = -\nabla \cdot \hat{b}(t, X_{s,t}).$$

Integrate to obtain both forward and backward density formulas.

PROOF EXERCISE 4

Prove Theorem 26:

Apply backward Itô to:

$$\hat{\rho}^F(t, Y_t^B).$$

Use the PDE:

$$\partial_t \hat{\rho}^F + \nabla \cdot (\hat{b}^F \hat{\rho}^F) = \varepsilon \Delta \hat{\rho}^F.$$

Show:

$$\mathrm{d}\left[e^{-\int_t^1 \nabla \cdot \hat{b}^F} \hat{\rho}^F(t, Y_t^B)\right] = \text{martingale.}$$

Condition on $Y_1^B = x$.

DOCUMENT III SUMMARY

This document proved:

TE $\Rightarrow$ probability-flow ODE sampler.

FPE $\Rightarrow$ forward/backward SDE samplers.

SDE drift error $\Rightarrow$ KL control.

ODE characteristics $\Rightarrow$ exact likelihood formula.

auxiliary SDEs $\Rightarrow$ SDE density and cross-entropy estimators.

Next documents:

$$\boxed{\text{instantiations, extensions, spatially linear interpolants, connections, algorithms, experiments.}}$$