KL-bounded Natural-Gradient Updates (closed forms and step size)

State space. Open simplex $\Delta_+^{n-1} = \{\, p \in \mathbb{R}^n \mid p_i > 0,\ \sum_i p_i = 1 \,\}$.

KPIs (unit bet).

$$\mathrm{RTP}(p) = \sum_i p_i r_i, \qquad \mathrm{Hit}(p) = \sum_{r_i > 0} p_i, \qquad \mathrm{Var}(p) = \sum_i p_i r_i^2 - \mathrm{RTP}(p)^2. \quad (1$$

1) One natural-gradient step (information-geometry form)

Gradient of Var w.r.t. p:

$$g_i := \frac{\partial\,\mathrm{Var}}{\partial p_i} = r_i^2 - 2\,\mathrm{RTP}(p)\, r_i. \tag{2}$$

Center it to lie in the simplex tangent space:

$$\tilde g_i := g_i - \langle g \rangle_p, \qquad \langle g \rangle_p = \sum_i p_i g_i. \tag{3}$$

A KL/mirror-descent (natural-gradient flavor) update is the **exponential tilt**:

$$q_i(\eta) \propto p_i\ \exp\!\big(\eta\, \tilde g_i\big), \qquad q_i(\eta) = \frac{p_i\, e^{\eta \tilde g_i}}{Z(\eta)}, \quad Z(\eta) = \sum_j p_j e^{\eta \tilde g_j}. \tag{4}$$

Properties: $q_i > 0$ (interior point), $\sum_i q_i = 1$ (mass conserved).

2) KL trust-region step size (1-D monotone solve)

We bound the move by $D_{\mathrm{KL}}(q(\eta) \,\|\, p) \le \varepsilon$.
For $q(\eta)$ above,

$$D_{\mathrm{KL}}(q(\eta)\,\|\,p) = \sum_i q_i \log \frac{q_i}{p_i} = \eta\,\mathbb{E}_{q(\eta)}[\tilde g] - \log Z(\eta). \tag{5}$$

Let $A(\eta) := \log Z(\eta)$. Then

$$\mathbb{E}_{q(\eta)}[\tilde g] = A'(\eta), \qquad D_{\mathrm{KL}}(\eta) = \eta\,A'(\eta) - A(\eta). \tag{6}$$

Facts: $D_{\mathrm{KL}}(0) = 0$, $D'_{\mathrm{KL}}(\eta) = \eta\,\mathrm{Var}_{q(\eta)}(\tilde g) \geq 0$.
Hence there is a unique $\eta^\star > 0$ with $D_{\mathrm{KL}}(\eta^\star) = \varepsilon$.

Practical solve (1-D).

$$\text{find } \eta^\star \text{ s.t. } f(\eta) := \eta\,A'(\eta) - A(\eta) - \varepsilon = 0. \tag{7}$$

Use bisection (robust) or Newton:

$$A'(\eta) = \frac{\sum_i p_i \tilde g_i e^{\eta \tilde g_i}}{Z(\eta)}, \quad A''(\eta) = \mathrm{Var}_{q(\eta)}(\tilde g), \quad f'(\eta) = \eta\,A''(\eta). \tag{8}$$

Newton step: $\eta \leftarrow \eta - \dfrac{f(\eta)}{f'(\eta)}$ with safeguarding.

3) Iterate and accept

One iteration.
(i) From $p^{(t)}$ compute RTP and $\tilde g$.
(ii) Solve $D_{\mathrm{KL}}(\eta) = \varepsilon$ to get $q(\eta)$.
(iii) If guardrails are violated, project q back (see `constraints.md`).
(iv) **Accept if**

$$\mathrm{Var}(p^{(t+1)}) > \mathrm{Var}(p^{(t)}) \;\wedge\; D_{\mathrm{KL}}(p^{(t+1)} \| p^{(t)}) \leq \varepsilon \;\wedge\; \text{GuardrailOK}. \tag{9}$$

Else shrink ε (or η) and retry.

4) Expected improvement (2nd-order audit note)

For small η,

$$\Delta\mathrm{Var} \;\approx\; \eta\,\langle \tilde g, \tilde g\rangle_p - \tfrac{1}{2}\,\eta^2\,\tilde g^\top H\,\tilde g, \tag{10}$$

where H is the Hessian of Var restricted to the tangent space.
This supports an **accept rule** and can be logged as an improvement certificate.