A note on the cross-entropy method

Patrick Laub, Institut de Science Financière et d'Assurances

March 16, 2019

1 When X is a random variable, and we only have one parameter to fit

The goal is to find

$$\ell = \mathbb{P}(X > \gamma).$$

Say that $f(\,\cdot\,)$ is the p.d.f. of X, and $f(\,\cdot\,; v)$ is the p.d.f. inside the chosen family of distributions with parameter v.

1. Choose a starting point v_0, R (e.g. $R = 10^6$), and ρ (e.g. $\rho = 0.10$).
2. For $i = 1, 2, \ldots$:

    1. Simulate $X^r \overset{\text{i.i.d.}}{\sim} f(\,\cdot\,; v_{i-1})$ for $r = 1, \ldots, R$.
    2. Find the quantile $\gamma_i = \text{Quantile}(\{X^1, \ldots, X^R\}, 1 - \rho)$.
    3. If $\gamma_i \geq \gamma$, set $v_* = v_{i-1}$, and quit the loop.
    4. Set
    $$v_i = \arg\max_v \frac{1}{R} \sum_{r=1}^{R} 1\{X^r > \gamma_i\} \frac{f(X^r)}{f(X^r; v_{i-1})} \log\big[f(X^r; v)\big]$$
    by letting v_i be the solution to
    $$\sum_{r=1}^{R} 1\{X^r > \gamma_i\} \frac{f(X^r)}{f(X^r; v_{i-1})} \frac{\mathrm{d}}{\mathrm{d}v} \Big\{ \log\big[f(X^r; v)\big] \Big\} = 0.$$

3. Return the result of IS with $f(\,\cdot\,; v_*)$ proposal.

2 When X is a random variable, and we have p distributional parameters to fit, $p > 1$

The goal is to find

$$\ell = \mathbb{P}(X > \gamma).$$

Say that $f(\,\cdot\,)$ is the p.d.f. of X, and $f(\,\cdot\,; \mathbf{v})$ is the p.d.f. inside the chosen family of distributions with parameter vector $\mathbf{v} = (v_1, \ldots, v_p)$.

1. Choose a starting point $\mathbf{v}_0$, R (e.g. $R = 10^6$), and ρ (e.g. $\rho = 0.10$).
2. For $i = 1, 2, \ldots$:

    1. Simulate $X^r \overset{\text{i.i.d.}}{\sim} f(\,\cdot\,; \mathbf{v}_{i-1})$ for $r = 1, \ldots, R$.

2. Find the quantile $\gamma_i = \text{Quantile}(\{X^1, \ldots, X^R\}, 1-\rho)$.
3. If $\gamma_i \geq \gamma$, set $\mathbf{v}_* = \mathbf{v}_{i-1}$, and quit the loop.
4. Set

$$\mathbf{v}_i = \arg\max_{\mathbf{v}} \frac{1}{R} \sum_{r=1}^{R} 1\{X^r > \gamma_i\} \frac{f(X^r)}{f(X^r; \mathbf{v}_{i-1})} \log\left[f(X^r; \mathbf{v})\right]$$

by letting $\mathbf{v}_i$ be the solution to

$$\sum_{r=1}^{R} 1\{X^r > \gamma_i\} \frac{f(X^r)}{f(X^r; \mathbf{v}_{i-1})} \boldsymbol{\nabla}_{\mathbf{v}} \Big\{ \log\left[f(X^r; \mathbf{v})\right] \Big\} = \mathbf{0}_p$$

where $\boldsymbol{\nabla}_{\mathbf{v}} = (\frac{\mathrm{d}}{\mathrm{d}v_1}, \ldots, \frac{\mathrm{d}}{\mathrm{d}v_p})$ and $\mathbf{0}_p$ is the vector of p zeros.

3. Return the result of IS with $f(\,\cdot\,; \mathbf{v}_*)$ proposal.

3 When X is a random vector, and we have p distributional parameters to fit, $p > 1$

The goal is to find

$$\ell = \mathbb{P}(H(\mathbf{X}) > \gamma)$$

where $\mathbf{X} = (X_1, \ldots, X_n)$ and $H : \mathbb{R}^n \to \mathbb{R}$. Say that $f(\,\cdot\,)$ is the joint p.d.f. of $\mathbf{X}$, and $f(\,\cdot\,; \mathbf{v})$ is the joint p.d.f. inside the chosen family of distributions with parameter vector $\mathbf{v} = (v_1, \ldots, v_p)$.

1. Choose a starting point $\mathbf{v}_0$, R (e.g. $R = 10^6$), and ρ (e.g. $\rho = 0.10$).
2. For $i = 1, 2, \ldots$:

    1. Simulate $\mathbf{X}^r \overset{\text{i.i.d.}}{\sim} f(\,\cdot\,; \mathbf{v}_{i-1})$ for $r = 1, \ldots, R$.
    2. Find the quantile $\gamma_i = \text{Quantile}(\{H(\mathbf{X}^1), \ldots, H(\mathbf{X}^R)\}, 1-\rho)$.
    3. If $\gamma_i \geq \gamma$, set $\mathbf{v}_* = \mathbf{v}_{i-1}$, and quit the loop.
    4. Set

    $$\mathbf{v}_i = \arg\max_{\mathbf{v}} \frac{1}{R} \sum_{r=1}^{R} 1\{H(\mathbf{X}^r) > \gamma_i\} \frac{f(\mathbf{X}^r)}{f(\mathbf{X}^r; \mathbf{v}_{i-1})} \log\left[f(\mathbf{X}^r; \mathbf{v})\right]$$

    by letting $\mathbf{v}_i$ be the solution to

    $$\sum_{r=1}^{R} 1\{H(\mathbf{X}^r) > \gamma_i\} \frac{f(\mathbf{X}^r)}{f(\mathbf{X}^r; \mathbf{v}_{i-1})} \boldsymbol{\nabla}_{\mathbf{v}} \Big\{ \log\left[f(\mathbf{X}^r; \mathbf{v})\right] \Big\} = \mathbf{0}_p$$

    where $\boldsymbol{\nabla}_{\mathbf{v}} = (\frac{\mathrm{d}}{\mathrm{d}v_1}, \ldots, \frac{\mathrm{d}}{\mathrm{d}v_p})$ and $\mathbf{0}_p$ is the vector of p zeros.

3. Return the result of IS with $f(\,\cdot\,; \mathbf{v}_*)$ proposal.

This is the most general form of the algorithm. The most general form of the optimization version is below.

3.1 For optimization, the general form

The goal is to find

$$H^* = \max_{\mathbf{x}} H(\mathbf{x})$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ and $H : \mathbb{R}^n \to \mathbb{R}$. Say that $f(\,\cdot\,; \mathbf{v})$ is the joint p.d.f. inside the chosen family of distributions with parameter vector $\mathbf{v} = (v_1, \ldots, v_p)$.

1. Choose a starting point $\mathbf{v}_0$, R (e.g. $R = 10^6$), and ρ (e.g. $\rho = 0.10$).
2. For $i = 1, 2, \ldots$:
    1. Simulate $\mathbf{X}^r \overset{\text{i.i.d.}}{\sim} f(\,\cdot\,; \mathbf{v}_{i-1})$ for $r = 1, \ldots, R$.
    2. Find the quantile $\gamma_i = \text{Quantile}(\{H(\mathbf{X}^1), \ldots, H(\mathbf{X}^R)\}, 1 - \rho)$.
    3. Set
    $$\mathbf{v}_i = \arg\max_{\mathbf{v}} \frac{1}{R} \sum_{r=1}^{R} 1\{H(\mathbf{X}^r) > \gamma_i\} \log\big[f(\mathbf{X}^r; \mathbf{v})\big]$$
    by letting $\mathbf{v}_i$ be the maximum-likelihood estimate fit to the elite samples.
    4. Calculate $H_i^* = \max\{H(\mathbf{X}^1), \ldots, H(\mathbf{X}^R)\}$.
    5. If H_i^* has not increased much recently (compare it to H_{i-1}^*, H_{i-2}^*, …) then quit now, giving $H^* \approx \max\{H_1^*, \ldots, H_i^*\}$.