Confidence interval on the cross-entropy

→ True distribution F (density $f(\cdot)$)

→ We observed $x = (x_1, \dots, x_n)$ an n-sample from F

→ We considered the parametric model Q_θ (density $q(\cdot|\theta)$) with $\theta \in \mathbb{R}^d$

→ $$\beta = \max_{\theta \in \mathbb{R}^d} \mathbb{E}_F\left[\log(q(Y|\theta))\right] = \max_{\theta \in \mathbb{R}^d} \int \log(q(y|\theta)) f(y)\, dy$$

Goal: construct a confidence interval on β

→ Clearly, $\beta = J(F)$

→ Let us denote by $\hat{F}_n$ the empirical distribution associated to x

→ If $\hat{\theta}_n$ is the MLE of θ, then

$$\hat{\beta}_n = J(\hat{F}_n) = \frac{1}{n} \sum_{i=1}^{n} \log(q(x_i|\hat{\theta}_n))$$

1

We propose to bootstrap the statistics

$$\hat{\beta}_n - \beta = J(\hat{F}_n) - J(F)$$

A gaussian toy example

$F = \mathcal{N}(0,1)$; $q_\theta = \mathcal{N}(\theta,1)$; $\beta = -\frac{1}{2}\log(2\pi) - \frac{1}{2}$

$\hat{\theta}_n = \bar{x}_n$; $\hat{\beta}_n = -\frac{1}{2}\log(2\pi) - \frac{1}{2n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$

The bootstrap version of $S = \hat{\beta}_n - \beta = f(\hat{F}_n) - f(F)$

is $S^* = \frac{1}{2n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 - \frac{1}{2n}\sum_{i=1}^{n}(x_i^* - \bar{x}_n^*)^2$

where $(x_1^*, ---, x_n^*)$ is a bootstrap sample from $(x_1, ---, x_n)$

Question: Is there an asymptotic result?

$$n^\alpha(\hat{\beta}_n - \beta) \longrightarrow \ ?$$

$$\hat{\beta}_n = \frac{1}{n}\sum_{i=1}^{n}\log(q(x_i|\hat{\theta}_n))$$

$$\beta = \max_{\theta\in\mathbb{R}^d}\int \log(q(y|\theta))f(y)dy$$

Let $\theta^* = \arg\max_{\theta\in\mathbb{R}^d}\int \log(q(y|\theta))f(y)\,dy$ and

$$\beta_n^* = \frac{1}{n}\sum_{i=1}^{n}\log(q(x_i|\theta^*))$$

3

$$\hat{\beta}_n - \beta = (\hat{\beta}_n - \beta_n^*) - (\beta_n^* - \beta)$$

$$\sqrt{n}(\beta_n^* - \beta) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \log(q(x_i|\theta^*)) - \mathbb{E}_F[\log(y|\theta^*)]\right)$$

According to the central limit theorem, we have that

$$\sqrt{n}(\beta_n^* - \beta) \xrightarrow[n\to\infty]{} \mathcal{N}(0, \mathbb{V}_F(\log(q(y|\theta^*))))$$

$$\sqrt{n}(\hat{\beta}_n - \beta_n^*) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \log(q(x_i|\hat{\theta}_n)) - \frac{1}{n}\sum_{i=1}^{n}\log(q(x_i|\theta^*))\right)$$

Let $h(\theta^*) = \frac{1}{n}\sum_{i=1}^{n} \log(q(x_i|\theta^*))$

Taylor expand: $h(\theta^*)$ around $\hat{\theta}_n$

$$h(\theta^*) \# h(\hat{\theta}_n) + h'(\hat{\theta}_n)^T(\theta^* - \hat{\theta}_n) + \frac{1}{2}(\theta^* - \hat{\theta}_n)^T h''(\hat{\theta}_n)(\theta^* - \hat{\theta}_n)$$

$\hat{\theta}_n$ is the MLE of θ then $h'(\hat{\theta}_n) = 0$

$$h(\theta^*) \# h(\hat{\theta}_n) + \frac{1}{2}(\theta^* - \hat{\theta}_n) h''(\hat{\theta}_n)(\theta^* - \hat{\theta}_n)$$

We deduce that

$$\sqrt{n}(\beta_n - \beta_n^*) = \sqrt{n}(h(\hat{\theta}_n) - h(\theta^*)) \#$$

$$-\frac{\sqrt{n}}{2}(\hat{\theta}_n - \theta^*) h''(\hat{\theta}_n)(\hat{\theta}_n - \theta^*)$$

4

$$-\frac{\sqrt{n}}{2}(\hat{\theta}_n-\theta^*)\, h''(\hat{\theta}_n)\,(\hat{\theta}_n-\theta^*) \xrightarrow[n\to\infty]{} 0$$

Indeed,
$$\left[\begin{array}{lll} \sqrt{n}(\hat{\theta}_n-\theta^*) & \xrightarrow[n\to\infty]{} & N(0_d, \sigma^2) \\ h''(\hat{\theta}_n) & \xrightarrow[n\to\infty]{} & h(\theta^*) \\ (\hat{\theta}_n-\theta^*) & \xrightarrow[n\to\infty]{} & 0 \end{array}\right]$$

$$\sqrt{n}(\hat{\beta}_n-\beta_n^*) \xrightarrow[n\to\infty]{} 0$$

Then, $\sqrt{n}(\hat{\beta}_n-\beta) \xrightarrow[n\to\infty]{} N(0, V_F(\log(q(\theta|\theta^*))))$

The Gaussian toy example

$$\sqrt{n}(\hat{\beta}_n-\beta) = -\frac{\sqrt{n}}{2n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2+\frac{\sqrt{n}}{2}$$

$$\sqrt{n}(\hat{\beta}_n-\beta) = -\frac{\sqrt{n}}{2n}\sum_{i=1}^{n}x_i^2+\frac{\sqrt{n}}{n}\bar{x}_n\sum_{i=1}^{n}x_i-\frac{\sqrt{n}}{2}\bar{x}_n^2+\frac{\sqrt{n}}{2}$$

$$\sqrt{n}(\hat{\beta}_n-\beta) = -\frac{\sqrt{n}}{2n}\sum_{i=1}^{n}x_i^2+\sqrt{n}\,\bar{x}_n^2-\frac{\sqrt{n}}{2}\bar{x}_n^2+\frac{\sqrt{n}}{2}$$

$$\sqrt{n}(\hat{\beta}_n-\beta) = -\frac{\sqrt{n}}{2}\left[\frac{\sum_{i=1}^{n}x_i^2}{n}-1\right]+\frac{\sqrt{n}}{2}\bar{x}_n^2$$

5

$$\sqrt{n}\,(\hat{\beta}_n - \beta) = -\frac{\sqrt{n}}{2}\left(\frac{\sum_{i=1}^{n} X_i^2}{n} - 1\right) + \frac{\sqrt{n}}{2}\,\bar{X}_n^2$$

$$\frac{\sqrt{n}}{2}\,\bar{X}_n^2 \xrightarrow[n\to\infty]{} 0 \qquad \left[\sqrt{n}\,\bar{X}_n \sim N(0,1)\,;\ n\,\bar{X}_n^2 \sim \chi^2(1)\,;\ \sqrt{n}\,\bar{X}_n^2 = \frac{n\,\bar{X}_n^2}{\sqrt{n}}\right]$$

$$-\frac{\sqrt{n}}{2}\left(\frac{\sum_{i=1}^{n} X_i^2}{n} - 1\right) \xrightarrow[n\to\infty]{} N\left(0, \frac{3}{4}\right)$$

Then, $\sqrt{n}\,(\hat{\beta}_n - \beta) \xrightarrow[n\to\infty]{} N\left(0, \frac{3}{4}\right)$

$$V_F\left(\log\left(q(y|\theta^*)\right)\right) = V_F\left(-\frac{1}{2}\log(2\pi) - \frac{y^2}{2}\right)$$

$$\Leftrightarrow\ V_F\left(\log\left(q(y|\theta^*)\right)\right) = \frac{1}{4}\,V(y^2) = \frac{E(y^4)}{4} = \frac{3}{4}$$

Gaussian example with unknown variance

$$F = N(0,1)\,;\quad \varphi_\theta = N(\mu, \sigma^2)\,;\quad \beta = -\frac{1}{2}\log(2\pi) - \frac{1}{2}$$

$$\theta = (\mu, \sigma^2) \quad \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_*^+$$

$$\hat{\theta}_n = (\bar{X}_n, S_n^2) \quad S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$\hat{\beta}_n = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(S_n^2) - \frac{1}{2n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \times \frac{1}{S_n^2}$$

$$\hat{\beta}_n = -\frac{1}{2}\left(\log(2\pi) + 1\right) - \frac{1}{2}\log(S_n^2)$$

The bootstrap version of S

$$S = \hat{\beta}_n - \beta = g(\hat{F}_n) - g(F) \text{ is}$$

$$S^* = -\frac{1}{2}\log(s_n^{2*}) + \frac{1}{2}\log(s_n^2)$$

where $s_n^{2*} = \frac{1}{n}\sum_{i=1}^{n}(x_i^* - \bar{x}_n^*)$

Gaussian regression example

$F = N(0.1\,x_1 + 0.1\,x_2, 1)$ $\quad x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$

$\boxed{A}$ $\varphi_\theta = N(\theta_1 x_1 + \theta_2 x_2, 1)$ $\quad \theta = (\theta_1, \theta_2)$

$\theta_1 \in \mathbb{R}$ and $\theta_2 \in \mathbb{R}$

$\hat{\theta}_n = (X^T X)^{-1} X^T y$; $y = (y_1, \dots, y_n)$; X design matrix

$$\beta = \int \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}(y - 0.1x_1 - 0.1x_2)^2\right] \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-0.1x_1-0.1x_2)^2} dy$$

$$\beta = -\frac{1}{2}\log(2\pi) - \frac{1}{2}$$

$$\hat{\beta}_n = -\frac{1}{2}\log(2\pi) - \frac{1}{2n}\sum_{i=1}^{n}(y_i - x_{i1}\hat{\theta}_1 - x_{i2}\hat{\theta}_2)^2$$

$\hat{\theta}_n = (\hat{\theta}_1, \hat{\theta}_2)$

The bootstrap version of $S = \widehat{\beta_n} - \beta$

is $S^* = \widehat{\beta_n}^* - \widehat{\beta_n} = \frac{1}{2}\sum_{i=1}^{n}(y_i - x_{i1}\hat{\theta}_1 - x_{i2}\hat{\theta}_2)^2 - \frac{1}{2}\sum_{i=1}^{n}(y_i^* - x_{i1}^*\hat{\theta}_1^* - x_{i2}^*\hat{\theta}_2^*)^2$

$\boxed{B}$ $\mathcal{P}_\theta = N(\theta_1 x_1, 1)$ $\left[F = N(0.1x_1 + 0.1x_2, 1)\right]$

$$\beta = \max_{\theta_1 \in \mathbb{R}} \int \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}(y - \theta_1 x_1)^2\right]\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(y - 0.1x_1 - 0.1x_2)^2\right) dy$$

$$E\left((y - \theta_1 x_1)^2\right) = E(y^2 - 2y\theta_1 x_1 + \theta_1^2 x_1^2)$$
$$= 1 + (0.1x_1 + 0.1x_2)^2 - 2(0.1x_1 + 0.1x_2)\theta_1 x_1 + \theta_1^2 x_1^2$$

$$\frac{d\,E((y - \theta_1 x_1)^2)}{d\theta_1} = -2(0.1x_1 + 0.1x_2)x_1 + 2\theta_1 x_1^2 = 0$$

$$\iff \theta_1^* = \left[0.1 + 0.1\frac{x_2}{x_1}\right]$$

$$\beta = -\frac{1}{2}\log(2\pi) - \frac{1}{2}$$

$$\beta(\theta_1) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}E\left(\sum(y_i - \theta_1 x_i)^2\right)$$

$\uparrow N(0.1x_1 + 0.1x_2, 1)$

$$\beta(\theta_1) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\left[1 + (0.1x_{i1} + 0.1x_{i2})^2 - 2(0.1x_{i1} + 0.1x_{i2})\theta_1 + \theta_1^2 x_i^2\right]$$

$$\theta_1^* = 0.1 + 0.1\left(\frac{\sum x_{i1}x_{i2}}{\sum x_{i1}^2}\right)$$