466/566: Sept. 4 – Sept. 12

2019

The definition of independence is relative to the given distribution p, not the true underlying distribution in the world.

<u>Example</u>: I know X and Y are Bernoulli distributed with $\alpha = 0.1$

By definition, $p(x,y) = p_\alpha(x) p_\alpha(y)$

$$= \alpha^x (1-\alpha)^{1-x} \alpha^y (1-\alpha)^y$$

So they are independent.

<u>Example 2</u>: I know X and Y are Bernoulli distributed, <u>but do not know α</u>

But maybe I know $\alpha = 0.1$ with prob 0.5 and $\alpha = 0.8$ with prob 0.5

$$p(x,y) = \sum_{\alpha \in \{0.1, 0.8\}} 1/2 \; p_\alpha(x) p_\alpha(y)$$

$$= \frac{1}{2} 0.1^x 0.9^{1-x} 0.1^y 0.9^y + \frac{1}{2} 0.8^x 0.2^{1-x} 0.8^y 0.2^{1-y}$$

$$\neq p_{\alpha'}(x) p_{\alpha'}(y) \quad \text{for some } \alpha' \in [0,1]$$

p is NOT the product of two Bernoulli's, by definition

$$\mathcal{D} = \{2, 5, 9, 5, 4, 8\} \quad \text{iid Poisson}$$

$$p(x \mid \lambda) = \lambda^x e^{-\lambda} / x! \qquad \lambda \in \mathbb{R}^+$$

$$\lambda = \underbrace{\operatorname*{argmax}_{\lambda \in (0, \infty)}} p(D \mid \lambda)$$

$$\underbrace{\ln p(D \mid \lambda)}_{c(\lambda)} = \ln \prod_{i=1}^{n} p(x_i \mid \lambda)$$

$$= \sum_{i=1}^{n} \ln p(x_i \mid \lambda)$$

$$\ln p(x \mid \lambda) = \ln (\lambda^x e^{-\lambda} / x!)$$

$$= \ln(\lambda^x) + \ln(e^{-\lambda}) - \ln x!$$

$$= \underline{x \ln \lambda} - \lambda - \underline{\ln} x!$$

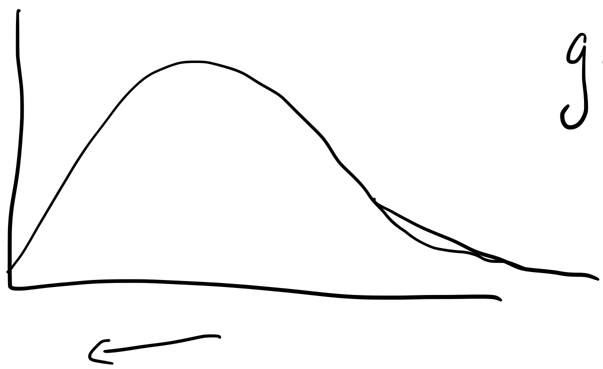
$$\frac{dc(\lambda)}{d\lambda} = \sum_{i=1}^{n} \left(\frac{x_i}{\lambda} - 1 - 0 \right) = 0$$

$$= \frac{1}{\lambda} \sum_{i=1}^{n} x_i = n$$

$$\Rightarrow \lambda = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad = \underline{5.5}$$

Example 10 : Poisson, same $\mathcal{D}$

Prior on λ = gamma distribution



$$\text{gamma}(\lambda \mid k = 3, \theta = 1)$$

$$\lambda > 0$$

$$p(\lambda) = \frac{\lambda^{k-1} e^{-\lambda/\theta}}{\theta^k \Gamma(k)}$$

$$\operatorname*{argmax}_{\lambda \in (0,\infty)} \ln p(\lambda \mid \mathcal{D}) = \operatorname*{argmax}_{\lambda \in (0,\infty)} \ln p(\mathcal{D} \mid \lambda) p(\lambda)$$

$$\ln \underbrace{[p(\mathcal{D} \mid \lambda) p(\lambda)]}_{c(\lambda)} = \ln p(\mathcal{D} \mid \lambda) + \ln p(\lambda)$$

$$\frac{d}{d\lambda} c(\lambda) = 0 \implies \lambda_{MAP} = \frac{k - 1 + \sum_{i=1}^{n} x_i}{n + \frac{1}{\theta}}$$

$$\lambda_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$|\lambda_{MAP} - \lambda_{MLE}| = \left| \underbrace{\frac{k-1+\sum_{i=1}^{n} x_i}{n+1/\theta}} - \underbrace{\frac{\sum_{i=1}^{n} x_n}{n}} \right|$$

$$\leq \underbrace{\frac{|k-1|}{n+1/\theta}}_{\to 0} + \underbrace{\frac{\sum_{i=1}^{n} x_i}{n(n+1/\theta)}}_{\frac{\bar{x}}{n+1/\theta} \to 0} \xrightarrow{n\to\infty} 0$$

$$\lambda_{MAP} \xrightarrow{n\to\infty} \lambda^*$$

$$\lambda_{MLE} \xrightarrow{n\to\infty} \lambda^*$$

$$E[\lambda_{MAP}] \stackrel{?}{=} \lambda^*$$

$$\underline{E\left[\frac{1}{n}\sum_{i=1}^{n} \cancel{X}_i\right] = \mu}$$

① consistency

② biasedness

③ statistical efficiency
(data)

$$p(y|x) = \mathcal{N}(\mu = x, \sigma^2)$$

. σ^2 is unknown

$\mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$

$$\underset{\sigma^2 \in \mathbb{R}}{\text{argmax}} \ \ln p(\mathcal{D}|\sigma)$$

$$= \underset{\sigma^2}{\text{argmax}} \sum_{i=1}^{n} \underbrace{\ln p(y_i | x_i, \sigma^2)}$$

$$\ln p(y_i|x_i, \sigma^2) = \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - x_i)^2}{2\sigma^2}\right)\right)$$

$$= \ln\left(\frac{1}{\sqrt{2\pi}}\right) + \underbrace{\ln\left(\frac{1}{\sigma}\right)}_{-\ln(\sigma)} - \frac{(y_i - x_i)^2}{2\sigma^2}$$

$$\frac{d}{d\sigma} \ln p(y_i|x_i, \sigma^2) = -\frac{1}{\sigma} + \frac{2}{2}\sigma^{-3}(y_i - x_i)^2$$

$$\sum_{i=1}^{n} -\frac{1}{\sigma} + \sigma^{-3}(y_i - x_i)^2 = 0$$

$$-\frac{n}{\sigma} + \sigma^{-3} \sum_{i=1}^{n} (y_i - x_i)^2 = 0$$

$$\Rightarrow -n\sigma^2 + \sum_{i=1}^{n} (y_i - x_i)^2 = 0$$

$$\Rightarrow \underline{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i)^2$$

$$\frac{dz^{-2}}{dz} = -2z^{-3}$$