ML Estimate: Bernoulli parameter

$$X_i = \begin{cases} 1 & \longrightarrow \text{trial } i \text{ is a success} \\ 0 & \longrightarrow \text{otherwise} \end{cases}$$

$$f(X_1, X_2, \ldots, X_n \mid p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

$$= p^{\sum_i x_i} (1-p)^{n - \sum_i x_i}$$

$$\log f = \sum_{i=1}^{n} x_i \log p + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-p)$$

$$\frac{\partial \log f}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} + \frac{n - \sum_{i=1}^{n} x_i}{1-p}(-1) = 0$$

$$\therefore \hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

— X —

ML estimate: Poisson parameter

$$f(x_i \mid \lambda) = e^{-\lambda} \lambda^{x_i} / x_i!$$

$$\log f = \sum_{i=1}^{n} (-\lambda + x_i \log \lambda - \log x_i!)$$

$$= -n\lambda + \log \lambda \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \log x_i!$$

$$\frac{\partial \log f}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$$

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

— x —

ML estimate for Gaussian with unknown μ and σ.

$$f(x_1, x_2, \ldots, x_n | \mu, \sigma)$$

$$= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\log f = \sum_{i=1}^{n} -\log\frac{2\pi}{2} - \log\sigma - \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$= -\frac{n}{2}\log 2\pi - n\log\sigma - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \log f}{\partial \mu} = \sum_{i=1}^{n}\frac{2(x_i-\mu)}{2\sigma^2} = 0 \longrightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial \log f}{\partial \sigma} = \frac{-n}{\sigma} - \frac{1(-2)}{2\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\therefore \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \hat{\mu})^2}{n}}$$

— X —

ML estimate of parameter of a $(0, \theta)$ uniform distribution.

$$f(x_i \mid \theta) = \begin{cases} \frac{1}{\theta} & 0 \le x_i \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore f(\{x_i\}_{i=1}^{n} \mid \theta) = \begin{cases} \frac{1}{\theta^n} & 0 \le x_i \le \theta \\ 0 & \text{Otherwise} \end{cases}$$

f is maximized by choosing θ to be as small as possible. But θ clearly must be at least as large as the largest observed value of $\{x_i\}_{i=1}^{n}$,

we set $\theta = \max(\{x_i\}_{i=1}^{n})$

What about a Uniform distribution over $[v, \theta]$?

$$f(\{x_i\}_{i=1}^{n} \mid \theta, v) = \begin{cases} \frac{1}{(\theta - v)^n} & v \le x_i \le \theta \\ 0 & \text{otherwise.} \end{cases}$$

$\theta = \max\left(\{x_i\}_{i=1}^{n}\right)$

$v = \min\left(\{x_i\}_{i=1}^{n}\right)$

— X —

Least squares line fitting

Consider $y_i = mx_i + c + \varepsilon_i$

where $\varepsilon_i \sim N(0, \sigma)$

Known: $m, c, \{x_i\}_{i=1}^{n}$ (accurately)

The $\{y_i\}_{i=1}^{n}$ estimates are inaccurate.

Now: $y_i \sim N(mx_i + c, \sigma)$

$$p(y_i \mid x_i, m, c) = \frac{e^{-\frac{(y_i - (mx_i + c))^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

$$\log p(\{y_i\} \mid \{x_i\}, m, c) = -\sum_{i=1}^{n}\left[\frac{y_i - (mx_i + c)}{\sqrt{2}\sigma}\right]^2 - n\log\sqrt{2\pi} - n\log\sigma$$

$$\frac{\partial \log p}{\partial m} = +\sum_{i=1}^{n} \frac{\cancel{2}[y_i - mx_i - c](+x_i)}{\cancel{2}\sigma^2} = 0$$

$$\therefore \sum_{i=1}^{n} x_i y_i = m \sum_{i=1}^{n} x_i^2 + c \sum_{i=1}^{n} x_i \longrightarrow (1)$$

Also $\dfrac{\partial \log p}{\partial c} = 0$ gives

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

$$\therefore \sum_{i=1}^{n} y_i = m \sum_{i=1}^{n} x_i + c(n) \longrightarrow (2)$$

Solving (1) and (2) simultaneously yields $\hat{m}$ and $\hat{c}$.

$$\hat{c} = \sum_{i=1}^{n} \frac{y_i}{n} - \hat{m} \sum_{i=1}^{n} \frac{x_i}{n}$$

where

$$\hat{m} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$, $\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$

Note that we did <u>not</u> assume knowledge of σ so far. In fact, σ can be estimated as

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{m}x_i - \hat{c})^2$$

Estimator bias or lack thereof.

① ML estimator for μ of a Gaussian is

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

where $x_i \sim N(\mu, \sigma^2)$

$$E(\hat{\mu}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n} \times n\mu = \mu.$$

Hence this is an unbiased estimator

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

Suppose μ were known. Then

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$E(\hat{\sigma}^2) = \frac{1}{n} n \sigma^2 = \sigma^2 \longrightarrow \text{unbiased.}$$

otherwise $E(\hat{\sigma}) = \frac{1}{n} \sum_{i=1}^{n} E(x_i - \hat{\mu})^2$

$$= \frac{1}{n} \sum_{i=1}^{n} E(x_i^2 + \hat{\mu}^2 - 2 x_i \hat{\mu})$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(x_i^2) + E(\hat{\mu})^2 - 2 E(x_i \hat{\mu})$$

(Now $\sum_i E(x_i \hat{\mu}) =$ ~~[illegible]~~ $E\left(\sum_i x_i \hat{\mu}\right)$

$$= E(n \hat{\mu}^2) \;)$$

$$= \frac{1}{n} \sum_{i=1}^{n} E(x_i^2) - E(\hat{\mu}^2)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left[\mathrm{Var}(x_i) + \left(E(x_i)\right)^2 \right] - E\left[\mathrm{Var}(\hat{\mu}) + \left(E(\hat{\mu})\right)^2 \right]$$

$$= \frac{1}{n} \left[n\sigma^2 + n\mu^2 \right] - \left(\frac{\sigma^2}{n^2} + \mu^2 \right) n = \sigma^2 \left(1 - \frac{1}{n} \right)$$

$\neq \sigma^2 \therefore$ biased

Let $X_1, X_2 \ldots X_n \sim \text{unif}(0,\theta)$.

Since $E(X_i) = \theta/2$,

$d_1(\theta) = 2\sum_{i=1}^{n} X_i/n$ is one estimator of θ

ML estimator is $d_2(\theta) = \max\left(\{X_i\}_{i=1}^{n}\right)$

$MSE(d_1) = Var(d_1) + Bias(d_1)$

$$E(d_1) = 2\sum_{i=1}^{n} E(X_i)/n = \frac{2}{n}\frac{\theta}{2} \times n = \theta$$

$$Var(d_1) = \frac{4}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{4}{n^2} \times n\frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

Note that $P(d_2(X) \le x) = P(\max_i X_i \le x)$

$$= \prod_{i=1}^{n} P(X_i \le x) = \left(\frac{x}{\theta}\right)^n \quad x \le \theta$$

$$\therefore f_{d_2}(x) = \frac{nx^{n-1}}{\theta^n}, \; x \le \theta$$

$$E(d_2) = \int_0^\theta \frac{nx^{n-1}}{\theta^n} x\,dx = \frac{n}{\theta^n}\int_0^\theta x^n\,dx = \frac{n}{\theta^n}\frac{\theta^{n+1}}{n+1}$$

$$= \frac{n\theta}{n+1} \neq \theta$$ $\therefore$ biased estimator

$$E(d_2^2) = \int_0^\theta x^2 \frac{n x^{n-1}}{\theta^n} dx = \frac{n\theta^2}{n+2}$$

$$Var(d_2) = E(d_2^2) - \left(E(d_2)\right)^2$$

$$= \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} = \frac{n\theta^2}{(n+2)(n+1)^2}$$

$$MSE(d_2) = \underbrace{\frac{\theta^2}{(n+1)^2}}_{\left(E(d_2)-\theta\right)^2} + \frac{n\theta^2}{(n+2)(n+1)^2}$$

$$= \frac{\theta^2}{(n+1)^2}\left[1 + \frac{n}{n+2}\right] = \frac{2\theta^2}{(n+1)(n+2)} \rightarrow O(n^{-2})$$

$$\leq \frac{\theta^2}{3n} \rightarrow O(n^{-1})$$

$$MSE = E\left((\hat{\theta} - \theta)^2\right)$$

$$= E\left[\, \underline{\hat{\theta} - E(\hat{\theta})} + \underline{E(\hat{\theta}) - \theta} \,\right]^2$$

$$= E\left[\underbrace{\left(\hat{\theta} - E(\hat{\theta})\right)^2}_{\text{Var}}\right] + \underbrace{E\left(E(\hat{\theta}) - \theta\right)^2}_{\text{bias}^2}$$

$$+ 2E\left[\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)\right]$$

As

$$E\left(\hat{\theta}E(\hat{\theta}) - \hat{\theta}\theta - \left(E(\hat{\theta})\right)^2 + E(\hat{\theta})\theta\right)$$

$$= \left(E(\hat{\theta})\right)^2 - E(\hat{\theta})\theta - \left(E(\hat{\theta})\right)^2 + E(\hat{\theta})\theta$$

$$= 0,$$

we have

$$E\left((\hat{\theta} - \theta)^2\right) = E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2\right] + E\left[\left(E(\hat{\theta}) - \theta\right)^2\right]$$