1. ML estimate of Bernoulli parameter p

$\{X_i\}_{i=1}^n \quad \forall i \; X_i \sim Ber(p)$

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

p?

due to indep n of the samples

$$f_{X_1, X_2 \ldots X_n}(x_1, x_2, \ldots, x_n; p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

$$P(X_i = 1) = p = p^1 (1-p)^0$$
$$P(X_i = 0) = 1-p = p^0 (1-p)^1$$

$$\log f(x_1, x_2, \ldots, x_n; p) = \sum_{i=1}^{n} x_i \log p + (1-x_i)\log(1-p)$$

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

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

$$= \frac{1}{1-p}\left(n - \sum_i x_i\right)$$

$$(1-p)\sum_i x_i = np - p\sum_i x_i$$

$$\sum_i x_i - \cancel{p\sum_i x_i} = np - \cancel{p\sum_i x_i}$$

$$\hat{p} = \frac{1}{n}\sum_i x_i \longrightarrow \text{ML estimate of } p.$$

② ML estimate of Poiss. par λ

$$\{X_i\}_{i=1}^{n} \text{ iid } \forall i, X_i \sim \text{Poisson}(\lambda)$$

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

$$f(\{x_i\}_{i=1}^{n} ; \lambda) = \prod_{i=1}^{n} e^{-\lambda} \lambda^{x_i} / x_i!$$

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

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

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i \longrightarrow \text{ML est of } \lambda.$$

③ ML estimate of μ, σ of a Gaussian

$\{X_i\}_{i=1}^n$ independent $\forall i, X_i \sim N(\mu, \sigma^2)$

$$f\left(\{x_i\}_{i=1}^n ; \mu, \sigma^2\right) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/2\sigma^2}$$

$$\log f = \sum_{i=1}^n \left(-\log\sqrt{2\pi} - \log\sigma - \frac{(x_i-\mu)^2}{2\sigma^2}\right) = 0$$

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

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

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

$$\frac{\partial \log f}{\partial \sigma} = -\frac{n}{\cancel{\sigma}} + \sum_{i=1}^n \frac{(x_i-\mu)^2}{2} \frac{(+2)}{\cancel{\sigma^3}\,\sigma^2} = 0$$

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

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

(4) ML estimate for Uniform Distr.

$\{X_i\}_{i=1}^n$ iid $\forall i$ $X_i \sim \text{Unif}(0, \theta)$

$$f(x_i; \theta) = \begin{cases} 1/\theta & 0 \leq x_i \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

$$f(\{x_i\}_{i=1}^n; \theta) = 1/\theta^n \quad \text{if } 0 \leq x_i \leq \theta \ \forall i$$
$$= 0 \quad \text{otherwise.}$$

$\theta = 0 \longrightarrow$ to max f.

θ must be the smallest possible value such that $\forall i$, $x_i \leq \theta$! θ cannot be smaller than $\max_i x_i$.

$$\hat{\theta} = \max_i x_i \rightarrow \text{smallest possible value} \geq x_i \text{ (all } x_i \text{ vals)}$$

What about $X_i \sim \text{Unif}[v, \theta]$

$$f(x_i; v, \theta) = 1/(\theta - v) \quad v \leq x_i \leq \theta$$
$$= 0 \quad \text{otherwise.}$$

$\theta \longrightarrow$ smallest value $\geq$ all x_i values

$\longrightarrow \tilde{\theta} = \max_i x_i$

$v \longrightarrow$ largest value $\leq$ all x_i values

$\longrightarrow \hat{v} = \min_i x_i$.

5) ML with a twist — Least squares line fitting (linear regression).

n pairs of values (x_i, y_i) such that

exactly known to you

$$y_i = m x_i + c + \varepsilon_i \text{ where } \forall i,\ \varepsilon_i \sim N(0, \sigma^2)$$

noisy

All ε_i values are independent

ε_i values are drawn from iid random vars.

You know $\{x_i\}_{i=1}^n$ accurately, you have noisy values $\{y_i\}_{i=1}^n$

To determine: m, c

$$y_i \sim N(m x_i + c, \sigma^2)$$

not iid (even though indep.)

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

$$p(\{y_i\} ; \{x_i\}, m, c) = \prod_{i=1}^{n} \frac{e^{-(y_i - m x_i - c)^2 / 2\sigma^2}}{\sigma\sqrt{2\pi}}$$

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

Proof that MSE $= \text{Bias}^2 + \text{Variance}$

$\text{MSE} = E\left[(\theta' - \theta)^2\right]$ where θ' is an estimate of θ.

$\text{Bias}^2 = \left(E(\theta') - \theta\right)^2$

$\text{Variance} = E\left[\left(\theta' - E(\theta')\right)^2\right]$

$$\text{LHS} = \text{MSE} = E\left[(\theta' - \theta)^2\right]$$

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

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

$\left(E(\theta') - \theta\right) \rightarrow$ constant

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

$$+ \underline{2E\left[\left(\theta' - E(\theta')\right)\right]\left(E(\theta') - \theta\right)} \rightarrow 0$$

$$\underline{E[\theta'] - E[E(\theta')]} = 0$$

$$E[E(\theta')] = E(\theta')$$

Bias and Variance of ML estimators

ML estimator for μ, σ^2 of a Gaussian

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

$X_i \sim N(\mu, \sigma^2)$

$$E[\hat{\mu}] = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu$$

$\therefore$ ML est of μ for a Gaussian is unbiased

$$Var(\hat{\mu}) = \frac{1}{n^2}\sum_i Var(X_i) \quad \text{due to indep}$$

$$= \frac{1}{n} \times n\sigma^2 = \sigma^2/n$$

$$MSE = \sigma^2/n + 0 = \sigma^2/n \quad (Bias = 0)$$

$$E(\hat{\sigma}^2) = \frac{1}{n}\sum_{i=1}^{n} E[(X_i - \hat{\mu})^2]$$

Suppose we knew μ beforehand → true mean

then

$$E(\hat{\sigma}^2) = \frac{1}{n}\sum_i E[(X_i - \mu)^2]$$

$$= \frac{1}{n}\sum_i \sigma^2 \swarrow var(X_i)$$

$$= n\sigma^2/n = \sigma^2$$

→ unbiased estimate

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

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

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

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

Solve (1) & (2) simultaneously

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

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

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

$$E(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^{n} E\left[(X_i - \hat{\mu})^2\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} E\left[X_i^2 + \hat{\mu}^2 - 2X_i\hat{\mu}\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left\{E[X_i^2] + E[\hat{\mu}^2] - 2E[X_i\hat{\mu}]\right\}$$

$$= \frac{1}{n} \sum_{i=1}^{n} E[X_i^2] + \frac{1}{n} n E[\hat{\mu}^2] - \frac{2}{n} \sum_i E[X_i\hat{\mu}]$$

$$= \frac{1}{n} \sum_i \left\{Var(X_i) + (E(X_i))^2\right\} + E[\hat{\mu}^2] - \frac{2}{n} \sum_i \left\{E[X_i\hat{\mu}]\right\}$$

$$\left(\sum_i E[X_i\hat{\mu}] = E\left(\left(\sum_i X_i\right)\hat{\mu}\right) = E[n\hat{\mu}^2] \right)$$

$$= \frac{1}{n} \sum_i \left(Var(X_i) + (E(X_i))^2\right) - E[\hat{\mu}^2]$$

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

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

$$= \sigma^2(1 - 1/n) \neq \sigma^2 \quad \text{Biased estimator}$$

One more biased estimator.

$\{X_i\}_{i=1}^n \sim \text{Unif}(0, \theta)$.

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

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

ML est. is $d_2(\theta) = \max(\{X_i\}_{i=1}^n)$

$MSE(d_1) = Var(d_1) + Bias^2(d_1)$

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

$\longrightarrow$ unbiased est.

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

$$= \theta^2/3n$$

$$= MSE$$

$$F_{d_2}(x) = P(d_2 \le x) = P(\max_i X_i \le x)$$

$$= P(X_1 \le x, X_2 \le x, \ldots, X_n \le x)$$

$$= \prod_{i=1}^n \underline{P(X_i \le x)} \quad \text{due to indep of } \{X_i\}_{i=1}^n$$

$$= (x/\theta)^n \quad (x \le \theta)$$

$$f_{d_2}(x) = n x^{n-1}/\theta^n \quad x \le \theta$$

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

$$= n \theta^{-n} \int_0^\theta x^n dx = \frac{n}{\theta^n} \frac{\theta^{n+1}}{n+1} = \frac{n\theta}{n+1}$$

biased estimator. $\neq \theta$

$$\text{Bias}^2 = \left(\frac{n\theta}{n+1} - \theta\right)^2 = \theta^2 \left(\frac{n}{n+1} - 1\right)^2$$

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

$$\text{Var}(d_2) = E(d_2^2) - (E(d_2))^2$$

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

$$\text{Var}(d_2) = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} = \frac{n\theta^2}{(n+2)(n+1)^2}$$

$$MSE(d_2) = \frac{\theta^2}{(n+1)^2} + \frac{n}{(n+2)(n+1)^2} \theta^2$$

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

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

ML est. is better $\xrightarrow{}$ MSE than the other one
(biased) (unbiased)

$\{X_i\}_{i=1}^{n}$ random variables iid

$E(X_i) = \mu$.

Consider estimator for μ : $\hat{\mu} = 3$

$\rightarrow$ high bias.

$\rightarrow Var(\hat{\mu}) = 0$.

Consider estimator: $\hat{\mu} = X_j$ for some j

Unbiased but not consistent

Bias of this est $= (E(\hat{\mu}) - \mu)^2$

$= (\mu - \mu)^2 = 0$

$Var(\hat{\mu}) = Var(X_j)$ (high var)

$\rightarrow$ does not decrease with n.