Example

	CA	US	MX
Buy = True	20	50	10
Buy = False	300	500	200

$P(CA) = 320 / (320 + 550 + 210) = 0.19$

$P(US) = 550 / (320 + 550 + 210) = 0.51$

$P(MX) = 210 / (320 + 550 + 210) = 0.30$

$$P(\text{country}, \text{bool}) = P(\text{bool} \mid \text{country}) P(\text{country})$$

$P(\text{Buy} = \text{True}, \text{Country} = CA) = 20 / \underline{1080} = 0.019$

→ sum of all values

→ very small numbers

32-bit number means 32 binary digits

↳ log probability

Independence

$A \perp B$ iff $p(A, B) = p(A) p(B)$

$\text{Buy} \perp \text{Country} \Leftrightarrow p(\text{Buy}, \text{Country})$

$= P(\text{Buy}) P(\text{Country})$

$$P(\text{Buy} \mid \text{country}) = \frac{P(\text{Buy}, \text{Country})}{P(\text{country})}$$

$$= \frac{P(\text{Buy})\,P(\text{country})}{P(\text{country})}$$

$$= P(\text{Buy})$$

IId (independent and identically distributed)

- the outcome is independent and identically distributed

The Gamblers Fallacy (誤謬)

- It is a false believe that things will "balance out" in the end

eg) gambler who has just lost a bunch of times before it is more likely to they will win next

Monty Hall

assume you choose #1

C: which door the car is behind ($C = 1, 2, 3$)

H: which door Monty Hall opens (assume $H = 2$)

$P(H=2 \mid C=1) = 0.5$

$P(H=2 \mid C=2) = 0$

$P(H=2 \mid C=3) = 1$

Want $P(C=3 \mid H=2)$ $P(C=1 \mid H=2)$

$$P(C=3 \mid H=2) = \frac{P(H=2, C=3)}{P(H=2)}$$

$$= \frac{P(H=2 \mid C=3)\, P(C=3)}{\sum P(H=2 \mid C=c)\, P(C=c)}$$

$$= \frac{P(H=2 \mid C=3)\, P(C=3)}{P(H=2 \mid C=1)\, P(C=1) + P(H=2 \mid C=2)\, P(C=2) + P(H=2 \mid C=3)\, P(C=3)}$$

$$= \frac{1 \cdot (1/3)}{(1/2)(1/3) + 0\,(1/3) + 1\,(1/3)}$$

$$= 2/3$$

$$P(C=1 \mid H=2) = \frac{(1/2)(1/3)}{(1/2)(1/3) + 0\,(1/3) + 1\,(1/3)}$$

$$= 1/3$$

→ should always switch door

Bernoulli Distribution

$$P(x) = \theta^{x}(1-\theta)^{1-x}$$

ex) $P(1) = \theta^{1}(1-\theta)^{1-1} = \theta$

$P(0) = \theta^{0}(1-\theta)^{1-0} = 1-\theta$

$P(x) = P(0) + P(1) = 1$

Maximum Likelihood Estimation

data $= \{x_1, x_2 \cdots x_N\}$.

$$L(\theta) = p(\text{data} \mid \theta)$$

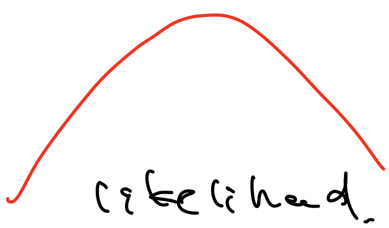
$$= \prod_{i=1}^{N} p(x_i \mid \theta)$$

$$= \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{1-x_i}$$

why is it called maximum likelihood?

- what value of θ makes the data we collected most probable?

- what value of θ maximizes the likelihood?



Log - Likelihood

log( ) is monotonically increasing

whatever θ maximizes L also maximizes $\log L$

$$l(\theta) = \log L(\theta) = \log \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{1-x_i}$$

$$= \sum^{N} \{x_i \log\theta + (1-x_i)\log(1-\theta)\}$$

$$\frac{dL}{d\theta} = \sum^{N}\left(x_i \frac{1}{\theta} - (1-x_i)\frac{1}{(1-\theta)}\right)$$

$$= \frac{1}{\theta}\underbrace{\sum x_i}_{\text{1が出る回数}} - \frac{1}{1-\theta}\sum \underbrace{(1-x_i)}_{\text{0が出る回数}}$$

CLT (central limit theorem)

- observation may due to sum of many random sources

ubiquitous (普遍的な)

Log-Likelihood of Normal distribution

$$l(\mu, \sigma^2) = \log L(\mu, \sigma^2) = \log \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

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

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

$$= -\frac{N}{2}\log 2\pi\sigma^2 - \frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2 \longrightarrow l$$

$$\frac{\partial l}{\partial \mu} = \sum\left(\frac{x_i-\mu}{\sigma}\right)\frac{1}{\sigma} \qquad \left(\left(\frac{x_i-\mu}{\sigma}\right)' = \left(\frac{x_i}{\sigma} - \frac{\mu}{\sigma}\right)' = -\frac{1}{\sigma}\right.$$

$$= 0$$

$$\mu = \frac{1}{N}\sum^{N} x_i$$

$$l(\mu, \sigma^2) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$v = \sigma^2$ とする.

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

$$\frac{\partial l}{\partial v} = -\frac{N}{2}\cdot\frac{(2\pi v)'}{2\pi v} - \frac{1}{2}\cdot(-1)\frac{1}{v^2}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$= -\frac{N}{2}\cdot\frac{1}{v} + \frac{1}{2}\frac{1}{v^2}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$= 0$$

$$-\frac{N}{2}\cdot\frac{1}{v} + \frac{1}{2}\frac{1}{v^2}\sum_{i=1}^{N}(x_i-\mu) = 0$$

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

Functions of random variables are random variable

$Z = X_1 + X_2$

$X_1 \sim Be(p_1)$ $X_2 \sim Be(p_2)$

$Z = 0, 1, 2$ - random

MLE estimates are also random

$\hat{\mu} = \frac{1}{N}\sum x_i = f(x_1, x_2, \cdots x_n)$

$\hat{\sigma^2} = \frac{1}{N}\sum(x_i-\mu)^2 = g(x_i, x_2 \cdots x_n)$

$\downarrow$

if it's Random variable, we can ask,

- what is their distribution
- what is their expectation

$\hat{\mu} \sim Normal(\mu, \frac{\sigma^2}{N})$

$E(\hat{\mu}) = \mu$

$E(\hat{\sigma^2}) \neq \sigma^2$ $E(\hat{\sigma^2}) = \frac{N-1}{N}\sigma^2$

expected value of the unbiased estimate is equal to the true variance

unbiased Estimate : $\hat{\sigma^2} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2$

(in ML, data is really large that $\frac{1}{N}$ and $\frac{1}{N-1}$ doesn't make difference.

CDF (culmulative distribution function)

- CDF : $F(x) = P(X \leq x)$
- CDF calculated from PMF

$$P(X \leq x) = \sum_{k=-\infty}^{x} p(k), \text{ where } p(k) = Prob(X=k)$$

- ex. $F(3) = p(3) + p(2) + p(1)$

CDF & PMF is both probabilities

continuous.

- CDF : $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(\tau) d\tau$
- PDF $f(x) = \frac{dF(x)}{dx}$

CDF is probability but pdf is not.

- cdf is always 0 at $-\infty$, and 1 at ∞

point estimate

- MLE - single value
- How confidence ?

  cf) two estimates (CTR A vs CTR B)

  if we are not confident about our estimates , can we be confident that A is better than B ?