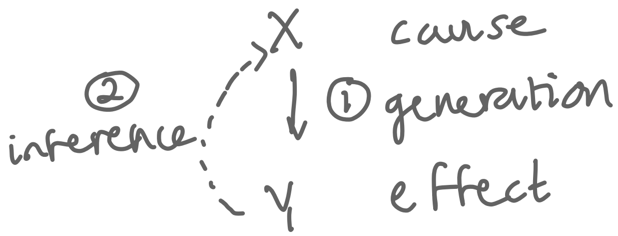
11/10/22

Baye's rule



prior: $p(x)$

likelihood $p(y|x)$

(function of x)

posterior:

$$p(x|y) \propto \underbrace{p(x)}_{\text{prior}}\underbrace{p(y|x)}_{\text{likelihood}}$$

↓ as a function of x

Normalization:

$$p(x|y) = \frac{p(x)\,p(y|x)}{p(y)} = \frac{p(x)\,p(y|x)}{\sum_x p(x)\,p(y|x)}$$

$$\sum_x p(x|y) = 1$$

ex.)

$$\pi(x) \propto \exp(F(x))$$ — Gibbs distribution

$$\pi(x) = \frac{\exp(F(x))}{\sum_x \exp(F(x))}$$

(soft max)

to make a pdf, need to sum over x

→ Z: normalizing constant or "partition function"

(2) X continuous, Y continuous

$X \sim N(\mu, \tau^2)$

σ^2 : measurement error

$[Y | X = x] \sim N(x, \sigma^2)$

posterior $[X | Y = y]$

to make density, sum over x

$p(x|y) \propto p(x) p(y|x)$

$\propto \exp\left(-\frac{(x-\mu)^2}{2\tau^2} - \frac{(y-x)^2}{2\sigma^2}\right)$

expand & only keep terms related to x

$\propto \exp\left(-\frac{1}{2}\left[\left(\frac{1}{\tau^2} + \frac{1}{\sigma^2}\right) - 2x\left(-\frac{\mu}{\tau^2} + \frac{y}{\sigma^2}\right)\right]\right)$

$\propto \exp\left(-\frac{1}{2} \frac{\left(x - \frac{\mu\frac{1}{\tau^2} + y\frac{1}{\sigma^2}}{\frac{1}{\tau^2} + \frac{1}{\sigma^2}}\right)^2}{\frac{1}{\tau^2 + \sigma^2}}\right)$

$[X | Y = y] \sim N\left(\frac{\mu\frac{1}{\tau^2} + y\frac{1}{\sigma^2}}{\frac{1}{\tau^2} + \frac{1}{\sigma^2}}, \frac{1}{\tau^2 + \sigma^2}\right)$

precision

interpolation between prior & posterior
weight is τ^2 and σ^2

compare to the form

$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

$\tau^2 \to \infty$

$N(y, \sigma^2)$

$\sigma^2 \to \infty$

$N(\mu, \tau^2)$

Bayesian

↳ using data & Baye's law to determine unknown parameters by assuming a prior distribution

ex.) X : Speed of light

$X \sim p(x)$ prior

$[Y|X] \sim N(X, \sigma^2)$

↑ measurement data

Empirical Bayes

$X_i \sim N(\mu, \tau^2)$ → estimate $\hat{\mu}$, $\hat{\tau}$ from data

$[Y_i | X_i] \sim N(x_i, \sigma^2)$

ex.) poll result, election result in state i

like website "538"

joint normal

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(\_\_, \_\_) \longrightarrow [X | Y]$$

↑ use Adam/Eve formulas to estimate

$Y = X + \varepsilon$

$\varepsilon \sim N(0, \sigma^2)$

$\varepsilon \perp X$

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \quad \end{bmatrix} \begin{bmatrix} X \\ \varepsilon \end{bmatrix}$$

(3) X discrete Y continuous

"mixture model/ classification"

prior distribution

	0	1
$p(x)$	$1-\lambda$	λ

$X \in \{0, 1\}$

$[Y \mid X = 0] \sim f_0(y) = p(Y|0)$

$[Y \mid X = 1] \sim f_1(y) = p(Y|1)$

$Y \sim f(y) =$

$p(y) = \sum_x p(x, y)$ → marginalization

$= \sum_x p(x)\, p(y|x)$ ↪ factorization $= p(X=0)\,p(Y|X=0) + p(X=1)\,p(Y|X=1)$

$= (1-\lambda) f_0(y) + \lambda f_1(y)$ "mixture distribution"

$p(x \mid Y) = \dfrac{p(x, y)}{p(y)}$ → normalization/conditioning

$= \dfrac{p(x)\, p(y|x)}{\sum_x p(x)\, p(y|x)}$

$p(X=1 \mid y) = \dfrac{\lambda f_1(y)}{\lambda f_1(y) + (1-\lambda) f_0(y)}$

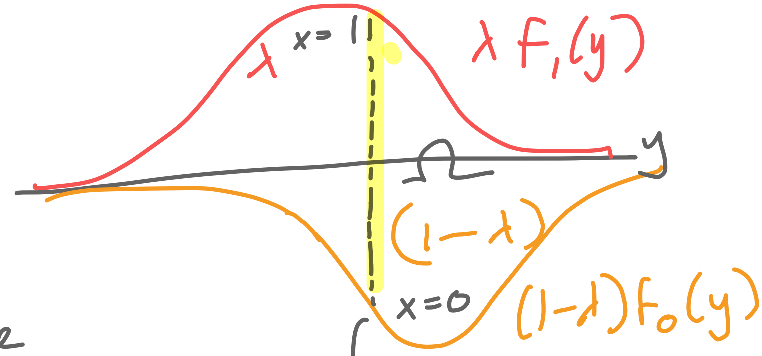
Baye's classification

$p(X=0 \mid y) = \dfrac{(1-\lambda) f_0(y)}{\lambda f_1(y) + (1-\lambda) f_0(y)}$

$\dfrac{p(X=1 \mid y)}{p(X=0 \mid y)} = \dfrac{\lambda}{1-\lambda} \; \dfrac{f_1(y)}{f_0(y)}$

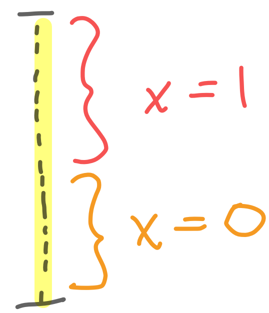
log ↓ logit( )

prior ↓ likelihood ↓



like sampling a lot of fish and λ = male fish $1-\lambda$ = female fish

to see conditional distribution, fix y (dotted line)



(4) X continuous Y discrete

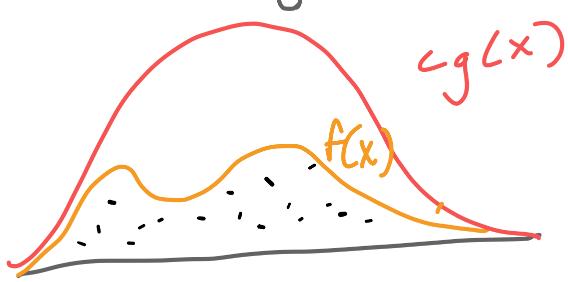
$X \sim g(x)$

given $X = x$ $\quad P(Y=1 \mid X=x) = \dfrac{f(x)}{c\,g(x)}$

↓

accept

accept-reject sampling



$c\,g(x) \geq f(x) \quad \forall x$

↓

envelope

f(x) is a weird distribution that is difficult to sample from

cg(x) is an easy distribution to sample from - like Normal

(1) $X \sim g(x)$

(2) accept X with prob $\dfrac{f(x)}{c\,g(x)}$

accept → return x

reject

distribution of returned X:

$[X \mid Y=1]$

$$p(x|y) = \frac{p(x,y)}{p(y)}$$

conditioning

$\downarrow$

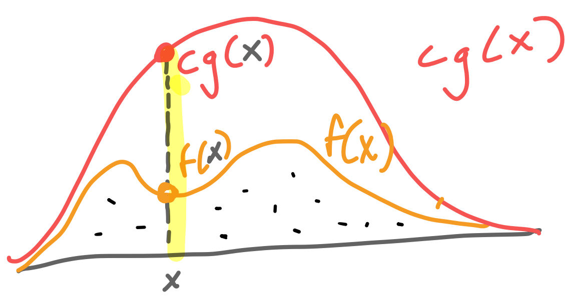
1

$$= \frac{p(x)\,p(y|x)}{\int p(x)\,p(y|x)\,dx}$$

$$= \frac{g(x)\,\frac{f(x)}{c\,g(x)}}{\int \quad dx}$$

$$= \frac{\frac{f(x)}{c}}{\int \frac{f(x)}{c}\,dx} = \frac{\frac{f(x)}{c}}{\frac{1}{c}} = f(x)$$

$\frac{1}{c} = p(y) = P(Y=1)$ accept $\rightarrow$ how often do we accept?



For conditioning, fix x

ex.) send survey to 1000 people

X: income

Y: $\begin{cases} 1 & \text{if response} \\ 0 & \text{if non-response} \end{cases}$

$[X \mid Y=1]$

missing not at random

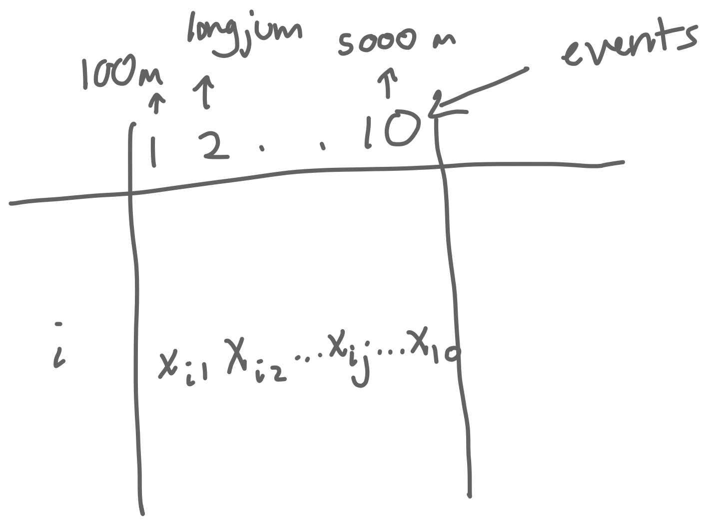
non-ignorable non-response

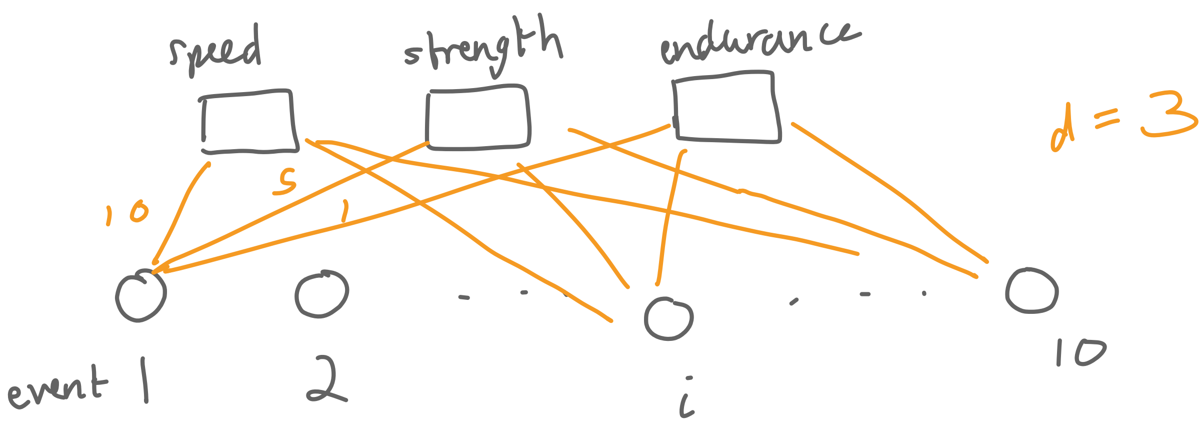
Back to (2): Z continuous X continuous

(using Z & X instead of X & Y)

ex.)

Factor analysis: decathalon

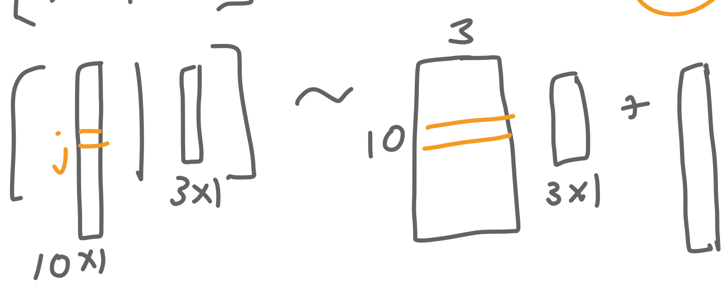




$$Z \sim N(0, I_d)$$

$$[X \mid Z] \sim WZ + \varepsilon$$

ε: $N \sim (0, \sigma^2 I_p)$



$[Z \mid X] \Rightarrow$ given scores of players, find ex.) Netflix

$$\text{rating}_{user,\, movie} = \langle z_{user}, w_{movie} \rangle$$

generator model

$$z \sim N(0, I_d)$$

$$X = g(z) + \varepsilon$$