handwritten notes
11 Jan

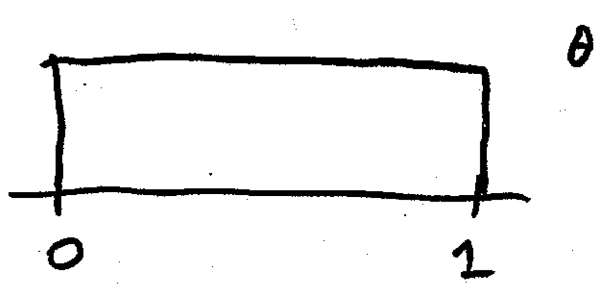
$$\begin{bmatrix}1\\2\\9\end{bmatrix} \xrightarrow[\text{random}]{\text{at}} \begin{bmatrix}Y_1\end{bmatrix}$$

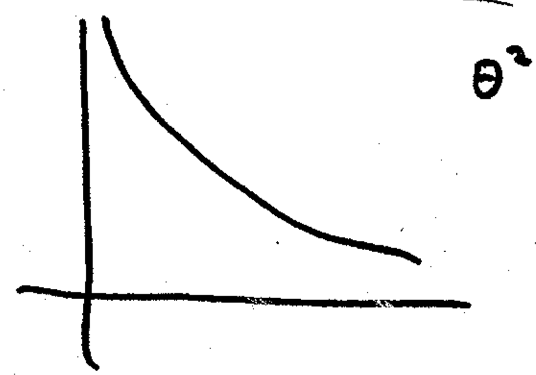
$$P_c(Y_1 \text{ is odd}) = \frac{2}{3}$$

---

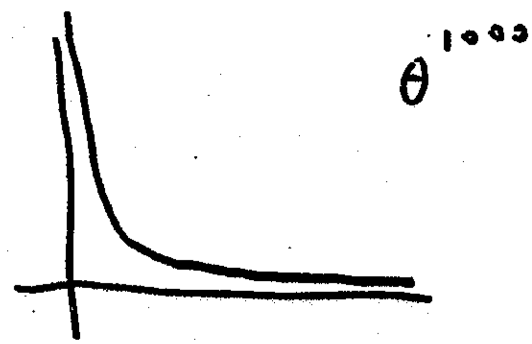
odds in favor

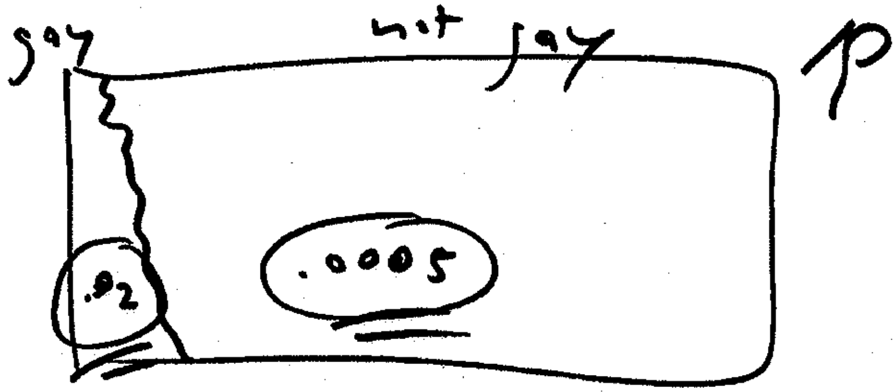
$$p = \frac{o}{1+o} \longleftrightarrow o = \frac{p}{1-p}$$





Lindley

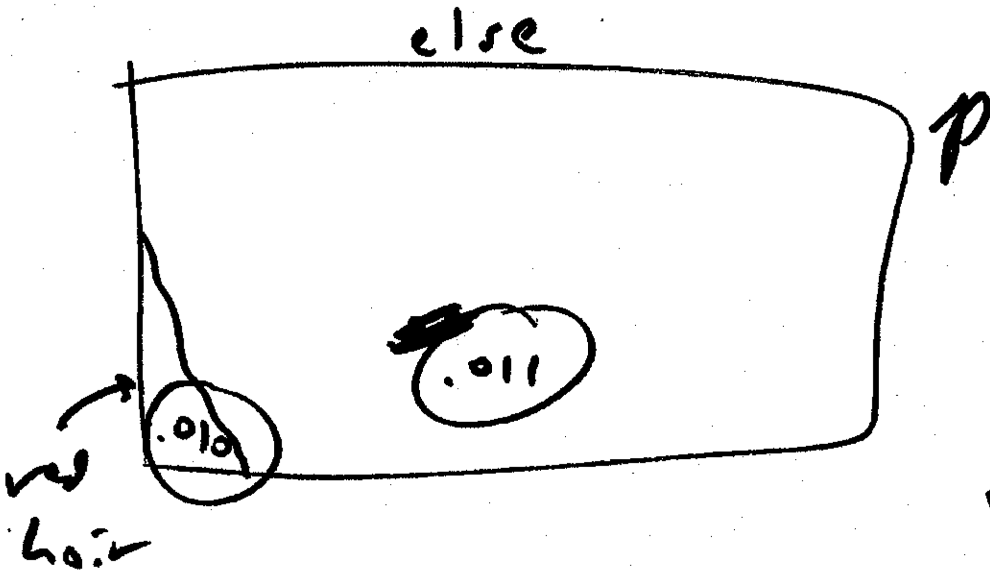




sexual orientation

is

recognizable



hair color not

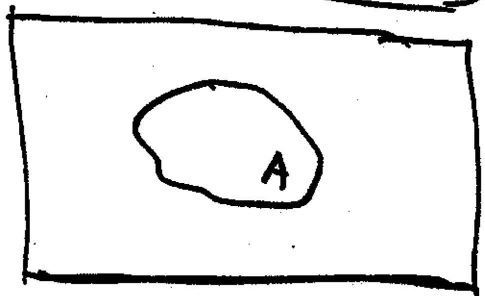
recognizable

basic prob. facts

1. $0 \le P(A) \le 1$

impossibility (certainly false)

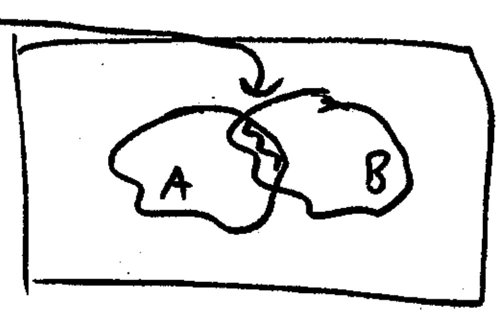
certainty (certainly true)



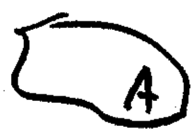
2. and $P(A \text{ and } B) = ?$

conditional probability

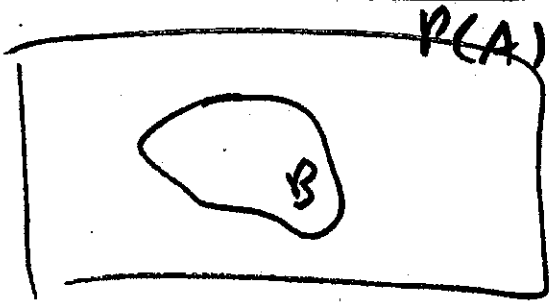
$P(B \mid A) = P(B \text{ given } A)$



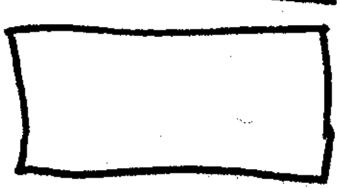
$P(B \mid A) = \dfrac{A \text{ and } B}{\quad}$



$= \dfrac{P(A \text{ and } B)}{P(A)}$



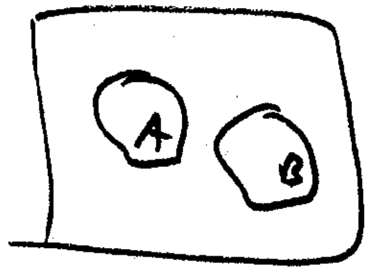
$P(B) = \dfrac{\quad}{\quad}$

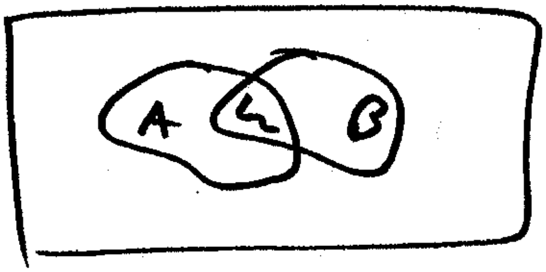


$$\boxed{\begin{aligned} P(A \text{ and } B) &= P(A) \cdot P(B \mid A) \\ &= P(B) \cdot P(A \mid B) \end{aligned}}$$

3 or

$P(A \text{ or } B) = P(A) + P(B)$

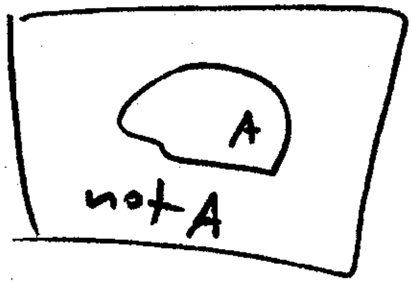




$$\boxed{P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)}$$

4 not

$P(A) + P(\text{not } A) = 1$



$P(A) = 1 - P(\text{not } A)$

Bayes' Theorem

cause & effect

$P(\text{effect} \mid \text{cause})$ easy

$P(\text{cause} \mid \text{effect})$ harder

A = unknown (true HIV status) (cause)

D = data (ELISA result) (effect)

$P(A \mid D) \stackrel{?}{=} P(D \mid A)$

$$P(A \mid D) = \frac{P(A \text{ and } D)}{P(D)}$$

$$P(D \mid A) = \frac{P(D \text{ and } A)}{P(A)}$$

$$P(A \text{ and } D) = \boxed{P(D)\, P(A \mid D)}$$

$$P(D \text{ and } A) = \boxed{P(A)\, P(D \mid A)}$$

$$P(A \mid D) = \frac{P(A)\, P(D \mid A)}{P(D)}$$

before — a priori

after — a posteriori

$\longrightarrow$ time

data

prior information

$P(A)$

$P(\text{unknown})$

posterior info

$P(\text{unknown} \mid \text{data})$

⑥

$$P(A \mid D) = \frac{P(A)\,\boxed{P(D \mid A)}}{P(D)}$$

sensitivity

ELISA +

true HIV+

.01

.95

$P(D)$ ② $\frac{293}{10000}$?

truth

(2x2 contingency table)

① what ELISA says

	HIV+	HIV−	
HIV +	95	198	293
HIV −	5	9702	9707
	100	9900	10000

$P(A) = .01$ (prevalence)

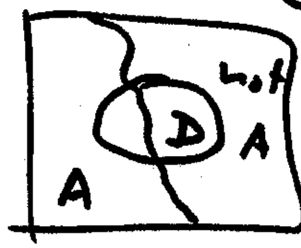
$P(D \mid A) = .95$ (sensitivity)

$P(\text{not } D \mid \text{not } A) = .98$ (specificity)

$$P(A \mid D) = P\left(\begin{array}{c}\text{really is}\\ \text{HIV+}\end{array} \;\middle|\; \begin{array}{c}\text{ELISA}\\ \text{says}\\ +\end{array}\right) = \frac{95}{293}$$

$\doteq 0.32$

7



$P(D) = P(\text{ELISA}+)$

$= P(D \text{ and } A) +$

2

$P(D \text{ and } (\text{not } A))$

$= P(A)\,P(D|A) +$

$P(\text{not } A)\,P(D|\text{not } A)$

$= (.01)(.95) + (.99)(1 - .98)$

$= .0293$

$1 - P(\text{not } D \mid \text{not } A)$

$$P(A|D) = \frac{P(A)\,P(D|A)}{P(D)} = \frac{(.01)(.95)}{(.0293)}$$

$$= \frac{95}{293} = 0.32$$

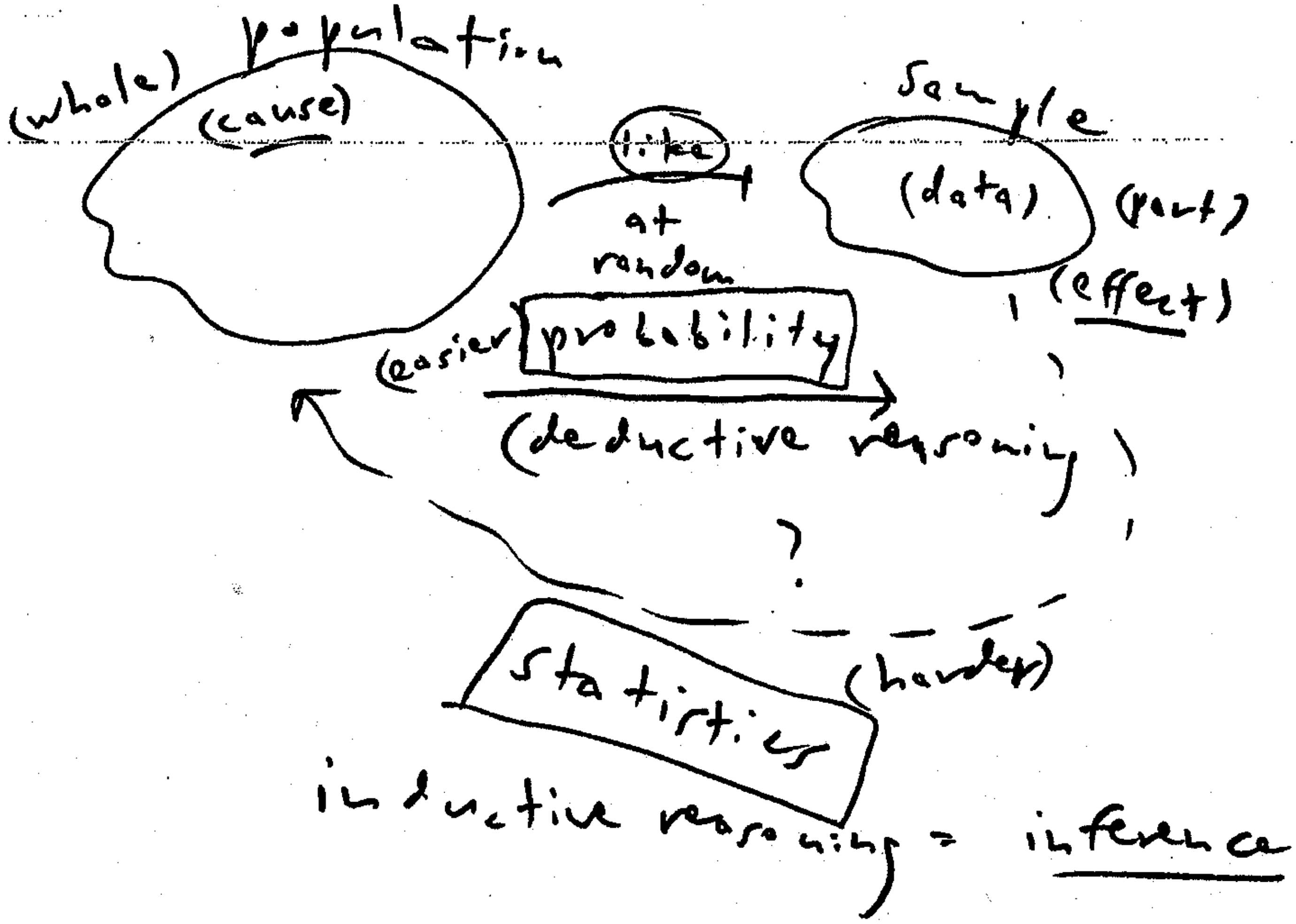
$$P(A \mid D) = \frac{P(A)\,P(D \mid A)}{P(D)}$$

$$P(\text{not } A \mid D) = \frac{P(\text{not } A)\,P(D \mid \text{not } A)}{P(D)}$$

$$\frac{P(A \mid D)}{P(\text{not } A \mid D)} = \left[\frac{P(A)}{P(\text{not } A)}\right]\left[\frac{P(D \mid A)}{P(D \mid \text{not } A)}\right]$$

$$\begin{pmatrix}\text{posterior}\\ \text{odds}\end{pmatrix} = \begin{pmatrix}\text{prior}\\ \text{odds}\end{pmatrix}\begin{pmatrix}\text{I or}\\ \text{II or}\\ \text{III}\end{pmatrix}$$

I likelihood ratio
II Bayes factor
III "data odds"


(whole) population
(cause)
like
at random
sample
(data)
(part)
(effect)
(easier) probability
(deductive reasoning)
?
statistics (harder)
inductive reasoning = inference

review of random variables (r.v.) $\begin{Bmatrix} \text{values} \\ \text{prob.} \end{Bmatrix}$

① one at a time: real-valued r.v. (rv's)

Any rv $\mathbb{Y}$ is uniquely characterized by its cumulative distribution function (CDF)

def. $F_{\mathbb{Y}}(y) = P(\mathbb{Y} \le y)$. Nicely-behaved

value

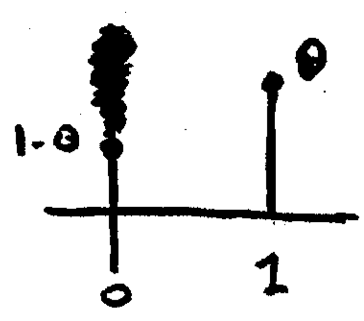
rv →

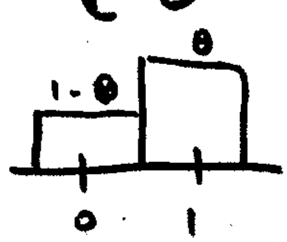
rv's are either discrete or continuous.

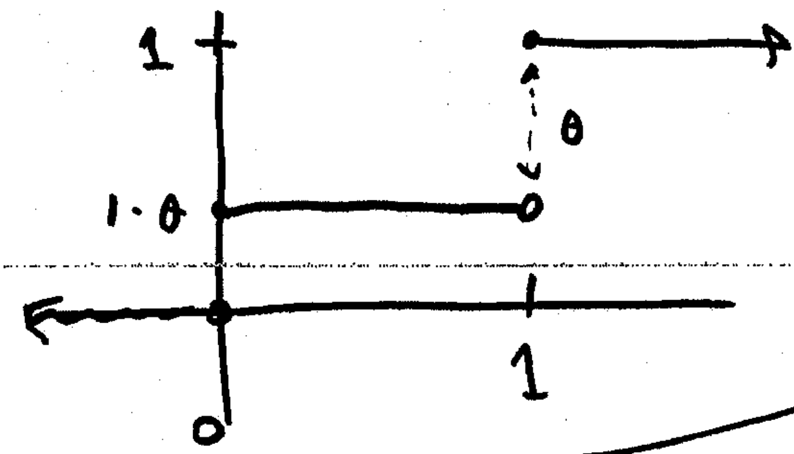
($F_{\mathbb{Y}}$ is nondecreasing & $0 \le F_{\mathbb{Y}} \le 1$). A discrete rv is equivalently char. by its probability mass function (PMF) $P(\mathbb{Y} = y)$.

ex. $\mathbb{Y} \overset{\text{(is distributed as)}}{\sim}$ Bernoulli(θ) parameter

$$P(\mathbb{Y} = \underset{0,1}{y}) = \begin{cases} \theta & \text{for } y = 1 \\ 1-\theta & \phantom{\text{for } y =\ } 0 \\ 0 & \text{else} \end{cases} = \underline{\theta^y (1-\theta)^{1-y}}.$$







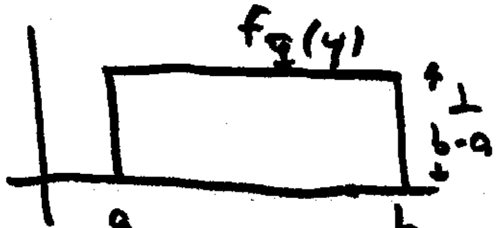
$F_Y(y)$

$P(Y=y)$ for a continuous rv² has to be 0, so subtlety is needed.

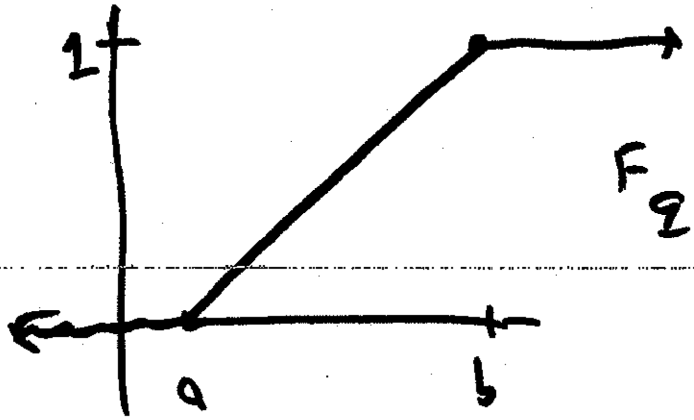
def. For a continuous rv² Y with CDF $F_Y(y)$, if you can find a function $f_Y(y)$ such that $F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt$ then $f_Y(y)$ is (a) the <u>density function</u> of Y; then $f_Y(y) = F'_Y(y)$. Immediately $F_Y(y) = P(Y \le y) = \int_{-\infty}^{y} f_Y(t)\,dt$, so probability for cont. rv² is represented by <u>area</u> under the density function.

---

ex. $Y \sim \text{Uniform}(a, b)$ $(a < b)$

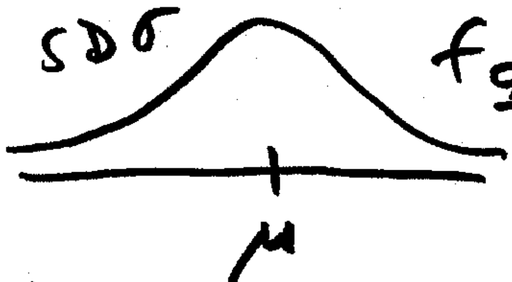


$$f_Y(y) = \begin{cases} \frac{1}{b-a} & \text{for } a \le y \le b \\ 0 & \text{else} \end{cases}$$

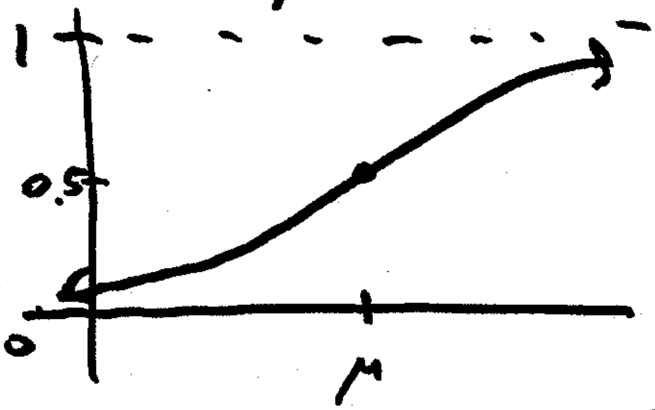


$$F_Y(y) = \begin{cases} 0 & \text{for } y \le a \\ \dfrac{y-a}{b-a} & a \le y \le b \\ 1 & y \ge b \end{cases}$$

ex. $Y \sim \text{Gaussian}(\underline{\mu, \sigma^2}) = N(\underline{\mu, \sigma^2})$ (normal)



$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$



$F_Y(y) =$ (no closed form)

In this class densities will typically not be denoted by $f_Y(y)$ but by $p_Y(y)$, or even more often by $\boxed{p(y)}$ (infer rv from argument) ex. $p(\theta)$ .. $p(\lambda)$

② 2 at a time (>2 by extension)

The joint CDF of Y_1 and Y_2 is

and

$$F_{Y_1,Y_2}(y_1, y_2) = P(Y_1 \le y_1, Y_2 \le y_2)$$

If Y_1, Y_2 both discrete, makes sense to talk about joint PMF

and

$$p_{Y_1,Y_2}(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2);$$

if continuous, natural to consider joint density f'n defined by

$$F_{Y_1,Y_2}(y_1,y_2) = \int_{-\infty}^{y_1}\int_{-\infty}^{y_2} f_{Y_1,Y_2}(t_1, t_2)\, dt_2\, dt_1$$

The marginal CDFs are given by

$$F_{Y_1}(y_1) = \lim_{y_2 \to \infty} F_{Y_1,Y_2}(y_1, y_2) \quad \text{(\& sim. for } Y_2\text{)}$$

& marginal density is given by

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1,y_2)\,dy_2$$

(<u>marginalizing</u> over Y_2). (Ldic. for Y_2)

Conditional density of Y_2 given Y_1 is

$$f_{Y_2|Y_1}(y_2|y_1) = \begin{cases} \dfrac{f_{Y_1,Y_2}(y_1,y_2)}{f_{Y_1}(y_1)} & \text{if } 0 < f_{Y_1}(y_1) < \infty \\ 0 & \text{else} \end{cases}$$

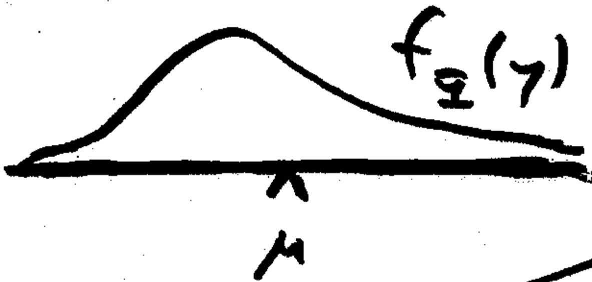
From this it follows that if Y_1, Y_2 <u>independent</u>

$$f_{Y_1,Y_2}(y_1,y_2) = f_{Y_1}(y_1)\,f_{Y_2}(y_2).$$

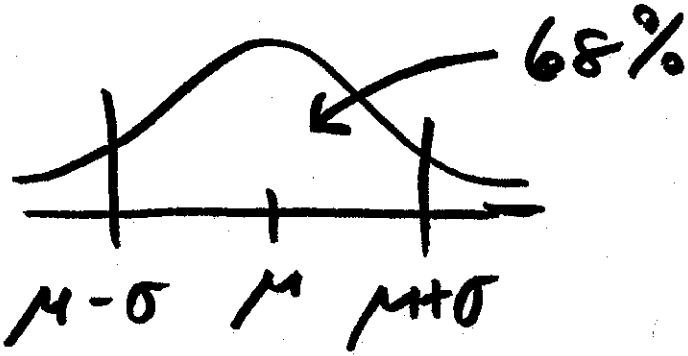
Expected value: If a rv Y is discrete then its <u>expectation</u> is (weighted average) $E(Y) = \sum_{\text{all } y} y \, P(Y=y)$

---

if cont. $E(Y) = \int_{-\infty}^{\infty} y \, f_Y(y) \, dy = \mu$



the variance of Y is then $\sigma^2 = E[Y - E(Y)]^2 = V(Y)$ (spread) & the standard deviation (SD) of Y is $\sigma = \sqrt{V(Y)}$.



$$\boxed{SE_{IID}(\bar{y}) = \frac{\sigma}{\sqrt{n}}} = \sqrt{\frac{p(1-p)}{n}}$$

$$\hat{SE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(.18)(.82)}{400}}$$

$= 0.019$

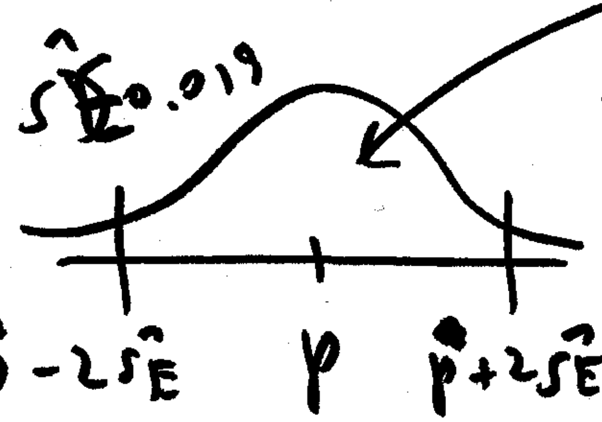
$= 1.9\%$

$\hat{p} = 18\%$

$\hat{SE}(\hat{p}) = 1.9\%$

95%

density in repeated sampling of $\hat{p}$



$$P_F(.14 \le p \le .22) \doteq \cancel{.95}$$

undefined

$$(1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, \ldots)$$

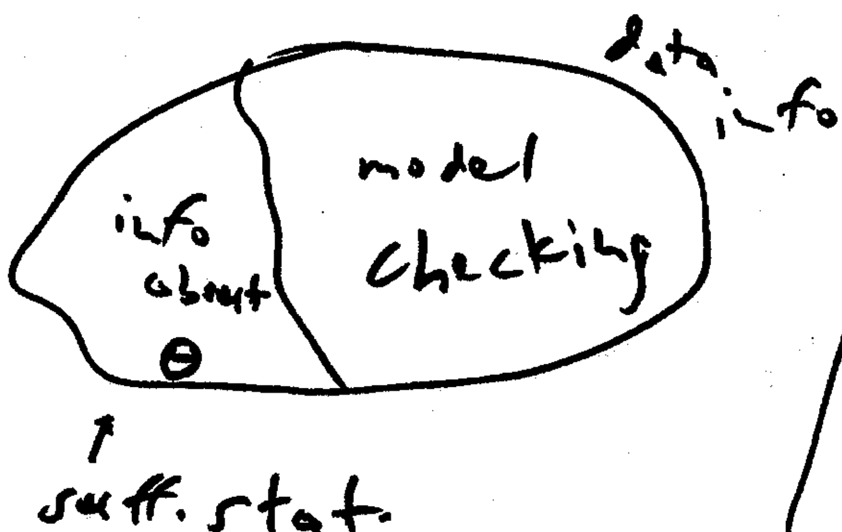
$$s = \Sigma y_i = \frac{n}{2}$$

$$\theta^{y_1}(1-\theta)^{1-y_1}\,\theta^{y_2}(1-\theta)^{1-y_2}\cdots\theta^{y_n}(1-\theta)^{1-y_n}$$

$$= \theta^{y_1+\ldots+y_n}(1-\theta)^{n-(y_1+\ldots+y_n)}$$

$$= \underline{\theta^{s}(1-\theta)^{n-s}}$$

$$s = \sum_{i=1}^{n} y_i$$



c generic positive constant

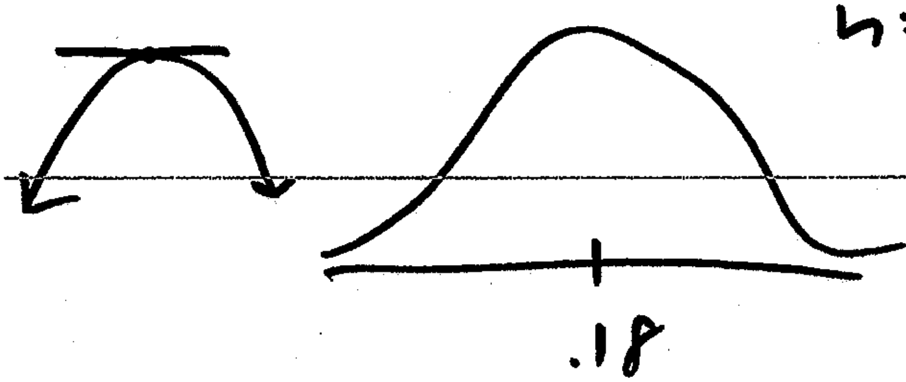
$c + c = c$

$c \cdot c = c$

etc.

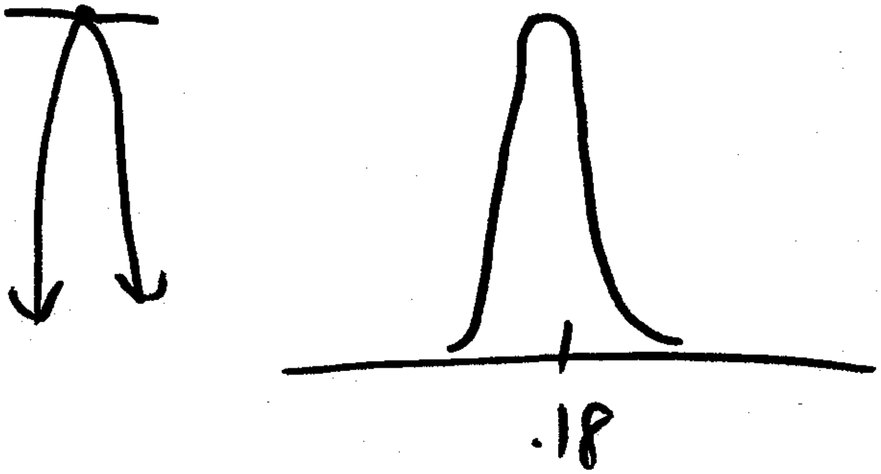
$n = 400$

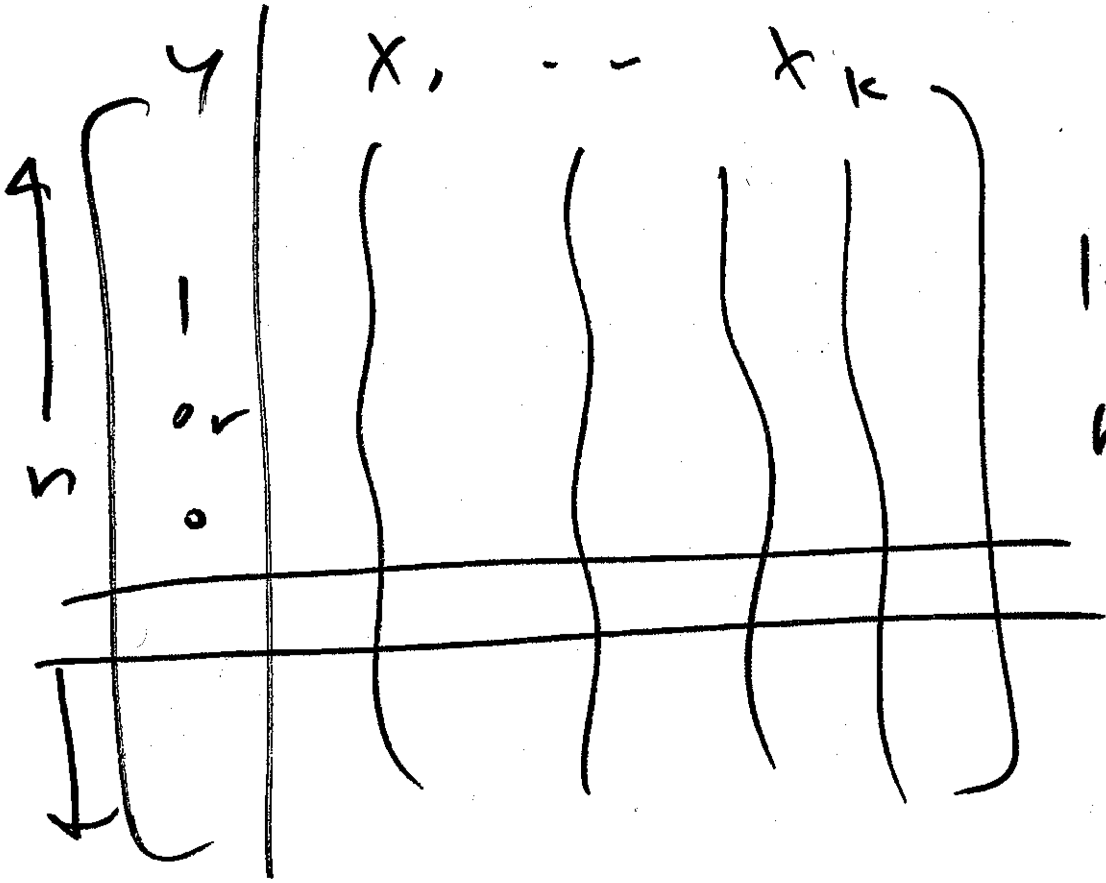
$\ell(\theta | y)$



$n = 1600$

$\ell(\theta | y)$





logistic regression

$1 = \text{HIV}^+$, $0 = \text{not}$

---

prior info = all info about unknown external to data

vs.

data info

truth

①
what
ELISA
says

	Ⓐ HIV +	HIV −	
Ⓓ HIV +	95	198	293
HIV −	5	9702	9707
	100	9900	10000

$$p(D) = \frac{293}{10000}$$

✓ $p(A) = .01$ (prevalence)

✓ $p(D \mid A) = .95$ (sensitivity)

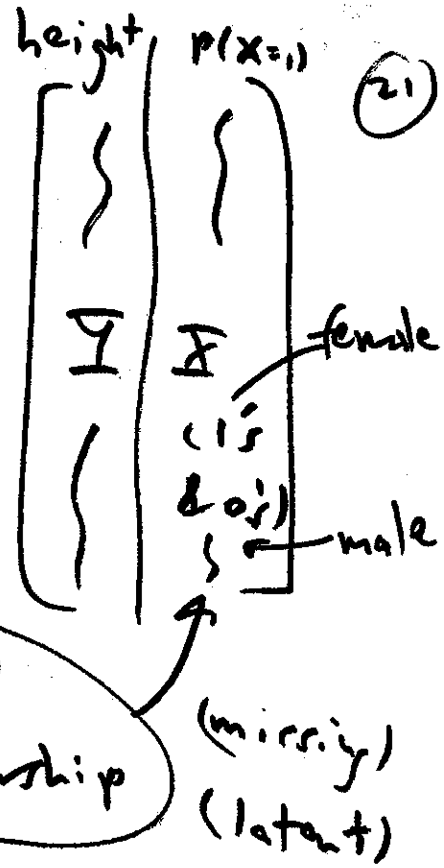
✓ $p(\text{not } D \mid \text{not } A) = .98$ (spec.)

$$p(A \mid D) = \frac{(.01)(.95)}{.0293} = .32\ (!) = \frac{95}{293}$$

	I	ELISA says
HIV−	C	HIV+

Coherence = internal consistency of Bayesian prob.

height of adults in this room



mixture of (2) normal distributions

$$p_Z(y) = p_1 N(\mu_1, \sigma_1^2) + (1-p_1) N(\mu_2, \sigma_2^2)$$

(HIV) ↔ (spam) (unknown) (A)

(ELISA) ↔ (spam filter) (data) (D)

(spam)

$$\left(\frac{P(S \mid D)}{P(\text{not } S \mid D)}\right) = \left[\frac{P(S)}{P(\text{not } S)}\right]\left[\frac{P(D \mid S)}{P(D \mid \text{not } S)}\right]$$

$$= (1) \cdot (\,?\,)$$

$$P(D \mid S) = P(\underline{w}_1 \overset{\text{and}}{,} \underline{w}_2, \ldots, \underline{w}_k \mid S)$$

$$= \boxed{P(w_1 \mid S)}\,\boxed{P(w_2 \mid \cancel{w_1}, \underline{\underline{S}})}\, P(w_3 \mid \cancel{w_1}, \cancel{w_2}, \underline{\underline{S}})$$

$$\cdots\ P(w_k \mid \cancel{w_1}, \ldots, \cancel{w_{k-1}}, S)$$

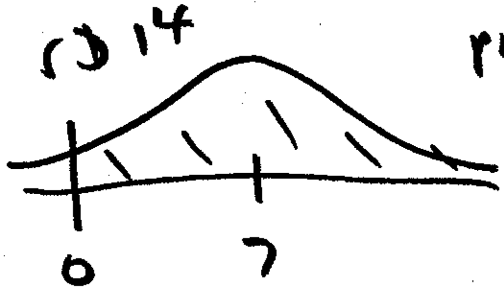
naive Bayes

$$\overset{\text{naive Bayes}}{=} P(w_1 \mid S) \cdot P(w_2 \mid S) \cdots P(w_k \mid S)$$

---

sim for $P(D \mid \text{not } S)$

$$BF = \left[\frac{P(w_1 \mid S)}{P(w_1 \mid \text{not } S)}\right]\left[\frac{P(w_2 \mid S)}{P(w_2 \mid \text{not } S)}\right] \cdots \left[\frac{P(w_k \mid S)}{P(w_k \mid \text{not } S)}\right]$$

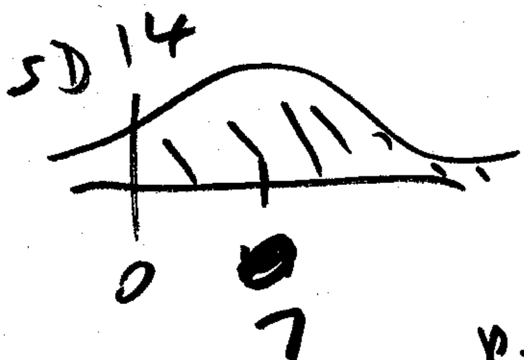


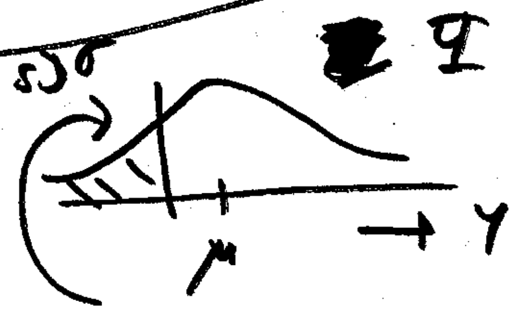
predictive dist for actual result (fav. - underdog)

ex. (21 - 18) = 3

dnorm
pnorm
qnorm
rnorm

d pois

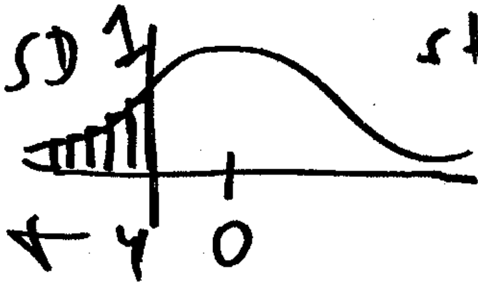




$$p_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

r.v.
value

$$F_Y(y) = P(Y \le y) = \int_{-\infty}^{y}$$



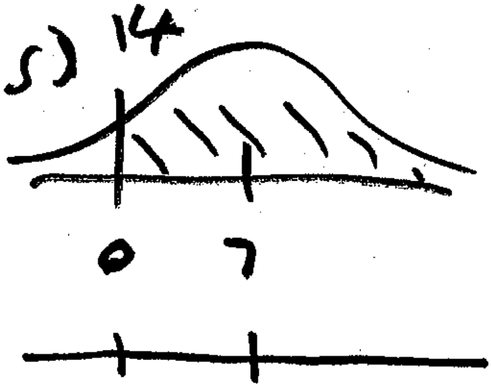
standard normal dist.

if $Y \sim N(0,1)$ then

$\sim$ is dist. as

$F_Y(y) = \Phi(y)$

$\boxed{\text{pnorm}}$ R



raw units y

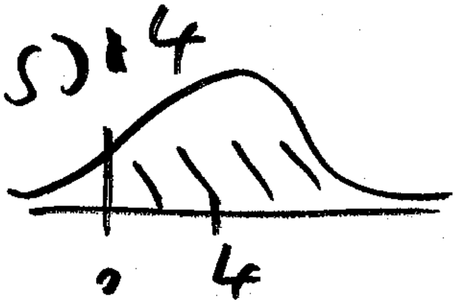
standard units $= z = \frac{y-\mu}{\sigma}$

$\frac{0-7}{14}$ 0

$= -0.5$

we want $1 - \text{pnorm}(-0.5)$

$= 69\%$



P(seattle wins)

$1 - \text{pnorm}\left(\frac{0-4}{14}\right) = 61\%$

$(A \mid B)$

$p(\theta \mid y)$

unknown θ; data $(y_1, \dots, y_n)$

θ scalar (1)

θ vector (2)

$\theta = \begin{matrix} 1 \\ 0 \end{matrix}$ $\quad y_i = \begin{matrix} 1 \\ 0 \end{matrix}$

$$y = \begin{pmatrix} 1 \\ 5 \\ 2 \\ \text{or} \end{pmatrix} \; n$$

(25)

$$Y_i = \begin{cases} 1 & \text{with prob } \theta \\ 0 & 1-\theta \end{cases}$$

r.v. (Bernoulli) (θ) ← parameter

$p(Y_i = 1) = \theta$

$p(Y_i = 0) = 1 - \theta$

(discrete r.v.)

$p(Y_i = y) = \theta^y (1-\theta)^{1-y}$ ← prob. mass fn (PMF)

$P(Y_1 = y_1, \ldots, Y_n = y_n) =$

$$\theta^{y_1}(1-\theta)^{1-y_1}\,\theta^{y_2}(1-\theta)^{1-y_2}\cdots$$

$$\theta^{y_n}(1-\theta)^{1-y_n}$$

$$= \theta^{y_1+\cdots+y_n}(1-\theta)^{n-(y_1+\cdots+y_n)}$$

$$= \theta^{s}(1-\theta)^{n-s}, \quad s = \sum_{i=1}^{n} y_i$$

---

intuition $\quad \hat{\theta} = \dfrac{s}{n} = \dfrac{72}{400} = 18\%$

$$\ell(\theta|y) = \ell(\theta | y_1, \ldots, y_n)$$

$$= \theta^{s}(1-\theta)^{n-s}, \quad s = \sum_{i=1}^{n} y_i$$

$$\ell(\theta \mid y) \doteq N(\sim, \sim)$$

$$= c_1 \, e^{-\boxed{c_2}(\theta - c_3)^2}$$

$c > 0$

$$\ell\ell(\theta \mid y) \doteq c_1^{*} - c_2(\theta - c_3)^2$$

$$c + c = c, \qquad c \cdot c = c$$