\[ \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \xrightarrow{\text{random}} \begin{bmatrix} y_1 \\ \vdots \end{bmatrix} \]

\[ P(y \text{ is odd}) = \frac{2}{3} \]

\[ p = \frac{\alpha}{1+\alpha} \leftarrow \text{odds in favor} \quad \Rightarrow \quad \alpha = \frac{p}{1-p} \]

\[ \begin{cases} 0 & \text{Lindley} \\ \theta^{1000} & \theta \end{cases} \]
sexual orientation

recognizable

hair color

recognizable
1. $0 \leq P(A) \leq 1$
   - impossibility (certainly false)
   - certainty (certainly true)

2. $P(A \text{ and } B) = ?$
   - conditional probability
   - $P(B | A) = P(B \text{ given } A)$
   - $P(B | A) = \frac{P(A \text{ and } B)}{P(A)}$
   - $P(B) = \frac{P(A \text{ and } B)}{P(A)}$
   - $P(A \text{ and } B) = P(A) \cdot P(B | A) = P(B) \cdot P(A | B)$
3. \[ P(A \lor B) = P(A) + P(B) \]

4. \[ P(A) + P(\text{not } A) = 1 \]
   \[ P(A) = 1 - P(\text{not } A) \]

**Bayes' cause & effect**

**Theorem**

\[ P(\text{effect} | \text{cause}) \quad \text{easy} \]
\[ P(\text{cause} | \text{effect}) \quad \text{harder} \]

\[ A = \text{unknown (true HIV status) (cause)} \]
\[ D = \text{data (ELISA result) (effect)} \]

\[ P(A \mid D) \overset{?}{=} P(D \mid A) \]
\[ P(A | D) = \frac{P(A \text{ and } D)}{P(D)} \]

\[ P(D | A) = \frac{P(D \text{ and } A)}{P(A)} \]

\[ P(A \text{ and } D) = P(D) P(A | D) \]

\[ P(D \text{ and } \neg A) = P(A) P(D | A) \]

\[ P(A | D) = \frac{P(A \text{ and } D)}{P(D)} \]

Before

a priori

after

a posteriori

time

prior

information

\[ P(A) \]

\[ P(\text{unknown}) \]

data

posterior

information

\[ P(\text{unknown} | \text{data}) \]
\[ P(A \mid I) = \frac{P(A) P(D \mid A)}{P(D)} \]

- \( P(A) = 0.01 \) (prevalence)
- \( P(D \mid A) = 0.95 \) (sensitivity)
- \( P(\text{not } D \mid \text{not } A) = 0.98 \) (specificity)

\[ P(A \mid I) = P(\text{really is } HIV + \mid \text{ELISA says } +) \sim \frac{95}{293} \approx 0.32 \]
\[ P(D) = P(ELISA+) \]
\[ = P(D \text{ and } A) + P(D \text{ and } \text{not } A) \]
\[ = P(A)P(D | A) + P(\text{not } A)P(D | \text{not } A) \]
\[ = (.01)(.95) + (.99)(1 - .98) \]
\[ = .0293 \]
\[ \therefore \quad 1 - P(D | \text{not } A) \]
\[ P(A | D) = \frac{P(A)P(D | A)}{P(D)} = \frac{(.01)(.95)}{(.0293)} \]
\[ = \frac{.95}{.293} = 0.32 \]
\[ P(A \mid D) = \frac{P(A)P(D \mid A)}{P(D)} \]
\[ P(\text{not} A \mid D) = \frac{P(\text{not} A)P(D \mid \text{not} A)}{P(D)} \]

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

\[ \text{(posterior odds)} = \left( \text{prior odds} \right) \left( \begin{array}{c}
\text{II}
\text{III}
\end{array} \right) \]

Ⅰ likelihood ratio
Ⅱ Bayes factor \leftarrow
Ⅲ "data odds"
population (cause)
whole

sample (effect)
data

probability

easier (deductive reasoning)

statistics (harder)

\text{inductive reasoning} \rightarrow \text{inference}
review of random variables ($rv$) \{ values \} \{ prob. \}

0 one at a time: real-valued rv. ($rv^2$)

Any rv is uniquely characterized by its cumulative distribution function (CDF)

\[
F_Y(y) = P(Y \leq y). \text{ Nicely-behaved } \text{ in } \mathbb{R}
\]

$rvs$ are either discrete or continuous.

($F$ is non-decreasing and $F \leq 1$). A discrete rv is equivalently char. by its probability mass function (PMF) $P(Y = y)$.

ex. $Y \sim \text{Bernoulli}(\theta)$ parameter

\[
P(Y = y) = \begin{cases} \theta & \text{for } y = 1 \\ 1 - \theta & \text{else} \end{cases} = \theta^y(1-\theta)^{1-y}
\]
For a continuous rv $r$ with CDF $F_r(y)$, if you can find a function $f_r(y)$ such that $F_r(y) = \frac{\int_{-\infty}^{y} f_r(t) \, dt}{-\infty}$ then $f_r(y)$ is (a) the density function of $r$; then $f_r(y) = F'_r(y)$. Immediately $F_r(y) = P(r \leq y) = \int_{-\infty}^{y} f_r(t) \, dt$, so probability for cont. rv $r$ is represented by area under the density function.

Ex. $r \sim \text{Uniform}(a, b)$ ($a < b$)

$f_r(y) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq y \leq b \\ 0 & \text{else} \end{cases}$
\[ F_\mathcal{Z}(y) = \begin{cases} \ 0 & \text{for } y \leq a \\ \frac{y - a}{b - a} & \text{for } a \leq y \leq b \\ \ 1 & \text{for } y \geq b \end{cases} \]

\[ \text{ex. } \mathcal{Z} \sim \text{Gaussian}(\mu, \sigma^2) = N(\mu, \sigma^2) \]

\[ f_\mathcal{Z}(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{y - \mu}{\sigma} \right)^2} \]

\[ F_\mathcal{Z}(y) = \left\{ \begin{array}{ll} 0 & \text{closed form} \\ \end{array} \right. \]

In this case densities will typically not be denoted by \( f_\mathcal{Z}(y) \) but by \( p_\mathcal{Z}(y) \), or even more often by \( p(y) \) (inherently from argument) ex. \( p(\Theta) \ldots p(\lambda) \)

\( \Theta \) at a time (\( \geq 2 \) by extension)
The joint CDF of $Z_1$ and $Z_2$ is

$$F_{Z_1, Z_2}(y_1, y_2) = P(Z_1 \leq y_1, Z_2 \leq y_2)$$

If $Z_1, Z_2$ both discrete, make sense to talk about joint probability

$$P_{Z_1, Z_2}(y_1, y_2) = P(Z_1 = y_1, Z_2 = y_2)$$

If continuous, natural to consider joint density $f_{Z_1, Z_2}$ defined by

$$F_{Z_1, Z_2}(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f_{Z_1, Z_2}(t_1, t_2) \, dt_1 \, dt_2$$

The marginal CDFs are given by

$$F_{Z_1}(y_1) = \lim_{y_2 \to \infty} F_{Z_1, Z_2}(y_1, y_2) \quad \text{for } y_1$$

& marginal density is given by
\[ f_{\Xi_1}(\gamma_1) = \int_{-\infty}^{\infty} f_{\Xi_1, \Xi_2}(\gamma_1, \gamma_2) d\gamma_2 \text{ (marginalizing over } \Xi_2) \]

Conditional density of \( \Xi_2 \) given \( \Xi_1 \) is:

\[ f_{\Xi_2|\Xi_1}(\gamma_2|\gamma_1) = \begin{cases} f_{\Xi_1, \Xi_2}(\gamma_1, \gamma_2) & \text{if } f_{\Xi_2}(\gamma_1) > 0 \text{ else } \end{cases} \]

From this it follows that if \( \Xi_1, \Xi_2 \) independent,

\[ f_{\Xi_1, \Xi_2}(\gamma_1, \gamma_2) = f_{\Xi_1}(\gamma_1) f_{\Xi_2}(\gamma_2). \]

Expected value: If a rv \( \Xi \) is discrete then its expectation is
(Weighted average) \[ E(\bar{X}) = \sum \gamma P(X = \gamma) \]

If \( X \sim \Gamma \) then

\[ E(\bar{X}) = \int_{-\infty}^{\infty} \gamma f_\bar{X}(\gamma) d\gamma = \mu \]

The variance or \( \text{spread} \) of the standard deviation (SD) of \( X \) is \( \sigma^2 = \text{Var}(X) \).

\[ \mu - \sigma \quad \mu \quad \mu + \sigma \]

68%
\[
\text{SE}^{2} \left( \hat{p} \right) = \frac{\hat{p}(1-\hat{p})}{n} = \sqrt{\frac{0.18(1-0.18)}{400}} = 0.019
\]

\[
\text{SE} \left( \hat{p} \right) = 1.9\%
\]

\[
\hat{p} = 18\%
\]

\[
\text{SE} \left( \hat{p} \right) = 0.019
\]

95% confidence interval for \( \hat{p} \):

\[
\hat{p} - 2\text{SE} < \hat{p} < \hat{p} + 2\text{SE}
\]

\( p_{L} = 0.14 \) and \( p_{U} = 0.22 \):

\[
\hat{p} \cap (0.14 \leq \hat{p} \leq 0.22) = \frac{0.14}{0.22} \text{ undefined}
\]
\[(1, 0, 1, 0, 1, 0, 1, 0, 1, 0, \ldots)\]

\[s = \sum \gamma_i = \frac{h}{2}\]

\[\theta^7 (1 - \theta)^{1 - \gamma_1} \theta^{\gamma_2} (1 - \theta)^{1 - \gamma_2} \ldots \theta^{\gamma_n} (1 - \theta)^{1 - \gamma_n}
= \theta^{\gamma_1 + \ldots + \gamma_n} (1 - \theta)^{n - (\gamma_1 + \ldots + \gamma_n)}
\]

\[= \theta^s (1 - \theta)^{n-s}\]

\[s = \sum_{i=0}^{n} \gamma_i\]
\[ y = x, \ldots, x_k \]

\[ 1 \text{ or 0} \]

\[ 1 = \text{HIV+}, \quad 0 = \text{HIV-} \]

\[ \text{prior info} = \text{all info about unknown external to data} \]

\[ \text{vs.} \]

\[ \text{data info} \]
\[
P(A | D) = \frac{(0.01)(0.95)}{0.0293} = \frac{0.095}{0.0293} = 3.2639188 \approx 3.26
\]
mixture of (2) normal distributions

\[ p_2(y) = p_1 N(\mu_1, \sigma_1^2) + (1 - p_1) N(\mu_2, \sigma_2^2) \]

\[
\begin{pmatrix}
\text{HIV} \
\text{ELISA}
\end{pmatrix} \leftrightarrow \begin{pmatrix}
\text{spam} \
\text{filter}
\end{pmatrix}
\]

\[
\left[ \begin{array}{c}
p(S|D) \
p(\text{not } S|D)
\end{array} \right] = \left[ \begin{array}{c}
p(S) \
p(\text{not } S)
\end{array} \right] \left[ \begin{array}{c}
p(D|S) \
p(D|\text{not } S)
\end{array} \right]
\]

\[= (1) \cdot (?)\]
P(D|S) = P(w_1, w_2, ..., w_k | S)

= \prod_{i=1}^{k} P(w_i | S) \prod_{i=1}^{k} P(w_{i-1} | w_i, S)

= \prod_{i=1}^{k} P(w_i | S) \cdot P(w_{i-1} | w_i, S)

= \prod_{i=1}^{k} P(w_i | S)

\text{since for } P(D | \text{not S})

BF = \begin{bmatrix}
\frac{P(w_1 | S)}{P(w_1 | \text{not S})}
\frac{P(w_2 | S)}{P(w_2 | \text{not S})}
\vdots
\frac{P(w_k | S)}{P(w_k | \text{not S})}
\end{bmatrix}
predictive distribution for actual result

$$f(x) - 	ext{underlog}$$

ex. $$(21 - 18) = 3$$

$$\mu \pm 2 \sigma$$

$$\mu$$

$$y$$

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

$$F(y) = P(y \leq y) = \int_{-\infty}^{y}$$
If \( Z \sim N(0,1) \) then it is distributed as \( F_Z(y) = \Phi(y) \).

\[
\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}z^2} dz.
\]

Standard units:
\[
t = \frac{y-\mu}{\sigma}
\]

For \( y = -0.5 \), we want \( 1 - \text{pnorm}(-0.5) = 69\%\).

\[
P(\text{Seattle wins}) = 1 - \text{pnorm} \left( \frac{-0.4}{14} \right) = 61\%.
\]
\[ p(\theta | y) \]

\[ \theta = 0, \quad y_i = 1 \]

\[ \gamma \]

\[ \left( \begin{array}{c} 1 \\ \vdots \\ \gamma_i \\ \vdots \\ 1 \end{array} \right) \]

\[ \varphi \]

\[ \text{unknown \ (y_1, \ldots, y_n)} \]

\[ \theta \] scalar (1)

\[ \theta \] vector (2)

\[ \mathcal{Y}_i = \left\{ \begin{array}{c} 1 \text{ \ with \ prob } \theta \\ 0 \end{array} \right\} \quad 1 - \theta \]

\[ \text{r.v. \ (Bernoulli) \ \theta} \]

\[ \begin{align*}
\mathbb{P}(\mathcal{Y}_i = 1) &= \theta \\
\mathbb{P}(\mathcal{Y}_i = 0) &= 1 - \theta
\end{align*} \]

\[ p(\mathcal{Y}_i = y) = \theta^y (1 - \theta)^{1 - y} \]

\[ \text{prob. mass \ fn \ \ (PMF)} \]
\[ p(\bar{y}_1 = y_1, \ldots, \bar{y}_n = y_n) = \]
\[ \theta_1^{y_1} (1-\theta_1)^{1-y_1} \theta_2^{y_2} (1-\theta_2)^{1-y_2} \ldots \]
\[ \theta_n^{y_n} (1-\theta_n)^{1-y_n} \]
\[ = \theta^{\sum_{i=1}^{n} y_i} (1-\theta)^{n - \sum_{i=1}^{n} y_i} \]
\[ = \theta^s (1-\theta)^{n-s}, \quad s = \sum_{i=1}^{n} y_i \]

**Intuition**  \( \hat{\theta} = \frac{s}{n} = \frac{72}{400} = 18\% \)

**Likelihood**  
\[ f(\theta | y_1, \ldots, y_n) = \]
\[ \theta^s (1-\theta)^{n-s}, \quad s = \sum_{i=1}^{n} y_i \]
\[ \ell(\theta|y) \equiv \mathcal{N}(-, -) \]
\[ = c_1, e^{-\frac{c_2 (\theta - c_2)^2}{K}} > 0 \]

\[ \ell(\theta|y) \equiv c_1 - c_2 (\theta - \theta_2)^2 \]

\[ c + c = c, \quad c \cdot c = c \]