Chapter 1

Eg 1.1 N tosses, prob f. How many heads?

$$\Rightarrow \quad \mathrm{Bin}(r \mid f, N) = f^r (1-f)^{N-r} \binom{N}{r}$$

Stirling from Laplace:

$$\frac{\lambda^\lambda e^{-\lambda}}{\lambda!} \approx \frac{1}{\sqrt{2\pi\lambda}} \Rightarrow \lambda! \approx \left(\frac{\lambda}{e}\right)^\lambda \sqrt{2\pi\lambda}$$

$$\Rightarrow \log_2 \binom{N}{r} \approx r \log_2 \frac{N}{r} + (N-r) \log_2 \frac{N}{N-r}$$

$$\approx N\, H_2\left(\frac{r}{N}\right) - \frac{1}{2} \log \left[2\pi N \, \frac{N-r}{N} \, \frac{r}{N}\right]$$

1.2

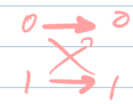
s	0	0	1	0	1
t	000	000	111	000	111
n	000	001	000	000	101
r	000	001	111	000	010

$$P(s \mid r) = \frac{P(r_1 r_2 r_3 \mid s)\, P(s)}{P(r_1 r_2 r_3)}$$

Assume prior probabilities equal

$\Rightarrow$ ML yields s^*

Assume BSC



$$P(\vec{r} \mid s) = \prod p(r_n \mid t_n(s))$$

$$p(r_n \mid t_n) = \begin{cases} 1-f & r_n = t_n \\ f & r_n \neq t_n \end{cases}$$

$$\Rightarrow \text{likelihood ratio} = \frac{P(r \mid s=1)}{P(r \mid s=0)} = \prod_n \frac{p(r_n \mid t_n(1))}{p(r_n \mid t_n(0))}$$

$$= \begin{cases} \dfrac{1-f}{f} & r_n = 1 \\ \dfrac{f}{1-f} & r_n = 0 \end{cases}$$

$\frac{1-f}{f} > 1$ for $f < \frac{1}{2}$

Ex 1.2 $\quad p_b = p_B = 3\underbrace{f^2(1-f)}_{\text{2 flips}} + \underbrace{f^3}_{\text{3 flips}} = 3f^2 - 2f^3 \ll f$

Ex 1.3 For R_N code

a) $p_b = \sum_{n=\frac{N+1}{2}}^{N} \underbrace{\binom{N}{n} f^n (1-f)^{N-n}}_{P(n \text{ flips})}$

error probability

b) leading term is largest $\approx 2^{N H_2(\frac{1}{2})} f^{\frac{N+1}{2}} (1-f)^{\frac{N-1}{2}}$

$$\approx 2^N [f(1-f)]^{N/2}$$

$$\approx [4f(1-f)]^{N/2} \approx p_b$$

Second term is $\sim \frac{f}{1-f}$ times smaller

$$\Rightarrow N = 2\,\frac{\log 10^{-15}}{\log 4f(1-f)} \approx 68$$ ← a bit too big

At next order $\binom{N}{N/2} \approx \frac{2^N}{\sqrt{2\pi N/4}}$ ← can get this from norm: $\sum_k \binom{N}{k} 2^{-N} = 1$ ($\mathrm{Bin}(k|\frac{1}{2},N)$)

$$= 2^{-N}\binom{N}{N/2}\int e^{-r^2/2\sigma^2}$$

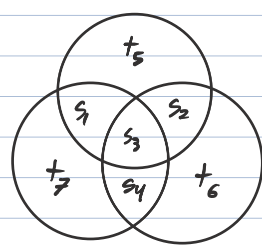
$$= 2^{-N}\binom{N}{N/2}\sqrt{2\pi\sigma^2} \qquad \sigma^2 = \frac{N}{4}$$

$$\Rightarrow \frac{2}{\sqrt{\pi N/4}}\, f\,(4f(1-f))^{\frac{N-1}{2}} = 10^{-15}$$

$$\Rightarrow N \approx 60.9$$

Block code: Add redundancy to blocks of data rather than one bit at a time

7,4 Hamming:



$s_1 \dots s_4$ are source

$t_{5\,6\,7}$ are parity

$\boxed{\Rightarrow \text{parity of each circle is even}}$

eg $0111 \to 0111010$

$t = G^T s$

$$G^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

16 codewords lie in a 7-d space of 2^7 options

Decoding: Assume BSC

We could decode by finding message s whose encoding $t(s)$ is closest to r
(16 choices)

Easier way: Find bit inside violated parity checks
& outside unviolated ones

The violated circles give the syndrome

2-bit errors can't be fixed, and flipping them
gives 3-bit errors

$$G^T = \begin{pmatrix} I \\ P \end{pmatrix} \Rightarrow H = \begin{pmatrix} P & I_3 \end{pmatrix} \qquad z = H \cdot r$$

$=$

Ex 1.4
Codewords (in G) has

$$HG^T = \begin{pmatrix} P & I_3 \end{pmatrix} \begin{pmatrix} I \\ P \end{pmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$\Rightarrow r = G^T s + n$

⇒ find most probable n s.t. $Hn = z$

"Maximum Likelihood Decoder"

Every vector in $\mathbb{F}_2^7$ is either a codeword or 1 flip away from one

Block error:

One or more bits fails to match the source:

$$p_B := P[s \neq \hat{s}]$$

Bit error:

Average probability that a decoded bit doesn't match its source bit:

$$p_b := \frac{1}{K} \sum_{k=1}^{K} P(s_k \neq \hat{s}_k)$$

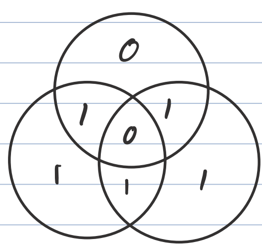
$p_B \propto P[2 \text{ or more flipped}]$ for Hamming 7 4

$O(f^2)$

But rate is $4/7 > 1/3$ from before

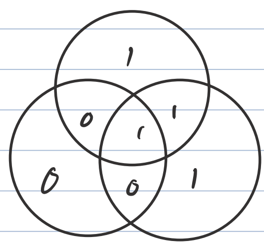
Ex 1.5 Decode

a) 1101011



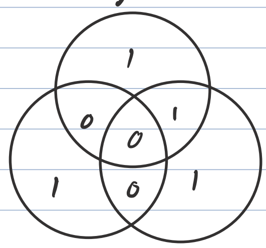
⇒ $z = 011$ ⇒ 1100011

b) 0110110



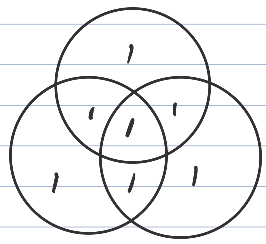
⇒ $z = 111$ ⇒ 0100110

c) 0100111



$\Rightarrow z = 001$ 0100110

d) 1111111



$\Rightarrow z = 000$ 1111111

1.6 a) p_B for Hamming is $\binom{7}{2} f^2 (1-f)^{7-2} + O(f^3)$

$\approx 21 f^2$

b) $p_b \lt p_B$ for 2 flips is $\propto f^2$ since we need 2 bits to flip

$\underbrace{21 f^2}_{P(2\text{ flip})} \cdot \underbrace{\tfrac{3}{7}}_{\#\text{ of wrong bits in that case}} = 9 f^2 + O(f^3)$

because the Hamming code is symmetric any bit (including the source bits) has an equal $\frac{3}{7}$ chance of flipping.

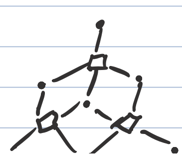
1.7 How many n give the zero syndrome?

The n w/ zero syndrome are exactly the 16 codewords by linearity

1.8 Can't have all 3 incorrectly flipped bits be parity, since flipping any 2 would never lead to the "corrected" 3rd one being the 3rd parity

$\Rightarrow$ 2 noise flips implies block error

1.9

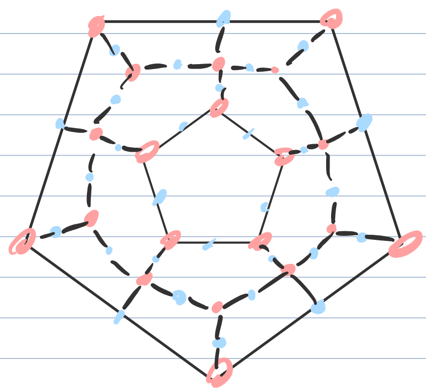


7,4 Hamming

Bipartite graph $\Rightarrow$ 3×7 matrix w/ ones whenever there's an edge

Bipartite graph $\Rightarrow$ error correcting code

30,11 Dodecahedron code



Any • bit flip on pentagonal side still gives 0 syndrome

$\Rightarrow$ lowest weight codewords have weight 5

Distance 5 $\Rightarrow$ can correct 2x bit flip errors

$\Rightarrow$ block error probability goes as $12 \cdot \binom{5}{3} \cdot f^3(1-f)^{27}$

faces — 3 bits from 5 on face — odds

Generically no need to be planar

1.10 For N transmitted bits with ≤ 2 flips

$\binom{N}{2} + \binom{N}{1} + \binom{N}{0}$ patterns

14, 8 code

$\Rightarrow$ 6 syndrome bits

$N = 14 \Rightarrow 91 + 14 + 1 = 106$

$M = 6 \Rightarrow 2^6 = 64$

$K = N - M$

$8 = 14 - 6$

$106 > 64 \Rightarrow$ can't be 2-error correcting

source = tot − syndrome

Linear or nonlinear $2^K \cdot \left[\binom{N}{2} + \binom{N}{1} + \binom{N}{0}\right] \le 2^N$

necessary for error correction

$S_t \cdot S_n \le 2^N$

1.11 see (30,11) code before

1.12 $p_b[R_3^2] \approx p[R_3] \cdot (3\, p[R_3] + \dots)$

$\approx 3(3f^2)^2 = 27 f^4$

same message as R_9 but different decoder

$p_b[R_9] \approx \binom{9}{5} f^5 \sim 126 f^5$

better

Chapter 2

Eg 2.1 Joint ensemble of bigrams

$P(x,y)$

Marginalizing over either var gives same dist

Ex 2.2 X, Y not indep in $P(x,y)$

Eg 2.3 $a = \begin{cases} 1 & \text{Jo has disease} \\ 0 & \text{Jo doesn't} \end{cases}$

$b = \begin{cases} 1 & \text{pos} \\ 0 & \text{neg} \end{cases}$

$p(a=1) = 0.01$

$p(b=1|a=1) = .95$

$p(b=0|a=0) = .95$

$\Rightarrow P(a=1|b=1) = \dfrac{P(b=1|a=1)\,P(a=1)}{P(b=1|a=1)\,P(a=1) + P(b=1|a=0)\,P(a=0)}$

$= \dfrac{.95 \cdot .01}{.95 \cdot .01 + .05 \cdot .99}$

$$= \frac{95}{95 + 99.5} \approx 0.16$$

You cannot do inference w/o making assumptions

2.3 Forward & inverse probability

Fwd:

Ex 2.4 K balls B black $W = K - B$ white

draw w/ replacement N times

a) $n_B = \mathrm{Bin}(f_B, N)$ $f_B = B/K$

b) $\mathbb{E}[n_B] = \frac{B}{K} \cdot N$

$\mathrm{Var}[n_B] = N \frac{B}{K}\left(1 - \frac{B}{K}\right)$

$B = 2$, $K = 10$ $\begin{cases} N = 5: & 1, \ 4/5 \to \sigma = \frac{2}{\sqrt{5}} \\ N = 400: & 80, \ 64 \to \sigma = 8 \end{cases}$

Fwd:

Ex 2.5 $z = \frac{(n_B - f_B N)^2}{N f_B (1 - f_B)} \sim \chi^2$ (approx)

$\mathbb{E}[z] = 1$

$N = 5$ $f_B = 1/5 \Rightarrow$ $\mathrm{Var} = 4/5$ $\mu = 1$

only $n_B = 1$ gives $z < 1$

$P[n_B = 1] = \binom{5}{1} 1/5 \cdot (4/5)^4 \sim \frac{256}{625} - 0.41$

Inv:

Ex 2.6 11 urns

Urn u has u black, $10-u$ white

n_B blacks $N - n_B$ whites

$N = 10$

$n_B = 3$

$P(u, n_B | N) = P(n_B | N, u)\, P(u)$

$P(u | n_B, N) = \frac{P(n_B | u, N)\, P(u)}{P(\ldots)}$

$P(n_B | N)$

$\binom{N}{n_B} f_u^{n_B} (1-f_u)^{N-n_B}$

$\rightarrow P(u | n_B, N) = \frac{1}{P(n_B | N)} \frac{1}{11} \binom{N}{n_B} \left(\frac{u}{10}\right)^{n_B} \left(\frac{10-u}{10}\right)^{N-n_B}$

can eval & see it peaks near $u=3$

Assuming now a new ball is drawn

$$P(N{+}1 \text{ is black} | n_B, N) = \sum_u \underbrace{P(N{+}1 \text{ is black} | n_B, u, N)}_{\frac{u}{10}} P(u | n_B, N)$$

$$\approx 0.333$$

MAP would just give 3/10

Eg 2.7 Observe n_H in N tosses

$\mathbb{P}[N{+}1 \text{ is } H | n_H, N]$

$\rightarrow$ Need assumption (prior)

Ex 2.8 uniform prior

$$\mathbb{P}[f_H | n_H, N] = \frac{P[n_H | f_H, N] \underbrace{P[f_H]}_{\text{prior} = 1}}{P[n_H | N]}$$

$$= f_H^{n_H} (1-f_H)^{N-n_H} \frac{(N+1)!}{n_H! (N-n_H)!}$$

$$\mathbb{P}[N{+}1 \text{ is heads} | n_H, N] = \int_0^1 df_H \underbrace{P[N{+}1 \text{ is heads} | f_H, n_H, N]}_{f_H} P[f_H | n_H, N]$$

$$= \frac{(N+1)!}{n_H! (N-n_H)!} \int_0^1 df_H \, f_H^{n_H+1} (1-f_H)^{N-n_H}$$

$$= \frac{(N+1)!}{(N+2)!} \frac{(n_H+1)!\,(N-n_H)!}{n_H!\,(N-n_H)!}$$

$$\boxed{= \frac{n_H+1}{N+2}}$$

$N=3 \quad n_H=0 \rightarrow \frac{1}{5}$

$N=3 \quad n_H=2 \rightarrow 3/5$

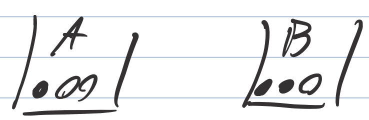
$N=10 \quad n_H=3 \rightarrow 4/12 = 1/3$

$N=300 \quad n_H=29 \rightarrow \frac{30}{302}$

2.9 Compress binary files → Chapter 6

→ estimate $p(1)$ empirically

2.10



$$P(A \mid \text{black}) = \frac{P(\text{black} \mid A)\, P(A)}{P(\text{black})} = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{2}} = 1/3$$

likelihood principle: $\theta \rightarrow D$ generative

$p(D \mid \theta)$

After observing a particular D, all inferences and predictions should depend only on $p(D \mid \theta)$

2.4 Entropy & Related

$H(X) \leq \log |A_X|$

redundancy : $1 - \dfrac{H(X)}{\log |A_X|}$

2.5 Decomposability of entropy

$$H(\vec{p}) = \mathcal{H}_2(p_1) + (1-p_1)\, H\left(\frac{p_2}{1-p_1}, \frac{p_3}{1-p_1}, \cdots\right)$$

keep flipping

Generally $$H(\vec{p}) = \mathcal{H}_2(p_1 + \cdots + p_m, p_{m+1} + \cdots + p_I)$$
$$+ (p_1 + \cdots + p_m)\, H\left(\frac{p_1}{p_1 + \cdots + p_m}, \cdots, \frac{p_m}{p_1 + \cdots + p_m}\right)$$
$$+ (p_{m+1} + \cdots + p_I)\, H\left(\frac{p_{m+1}}{p_{m+1} + \cdots + p_I}, \cdots, \frac{p_I}{p_{m+1} + \cdots + p_I}\right)$$

Ex 2.13 : $\log 3 + \frac{1}{3}\left[\log 10 + \log 5 + \log 21\right]$

$\sim \log 30$ bits

2.6 Gibbs' inequality:

$$D_{KL}(p \| q) \geq 0$$

2.7 Jensen's inequality:

f convex

$$f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$

$$\Rightarrow \mathbb{E}[f] \geq f(\mathbb{E}[x])$$

Ex 2.14: Prove by induction

$$f\left(\sum p_i x_i\right) \leq p_1 f(x_1) + \sum_{i \geq 2} p_i\, f\left(\frac{\sum_{i \geq 2} p_i x_i}{\sum_{i \geq 2} p_i}\right)$$

$$= p_1 f(x_1) + p_2 f(x_2) + \sum_{i \geq 3} p_i f\left(\frac{\sum_{i\geq 3} p_i x_i}{\sum_{i \geq 3} p_i} \right)$$

$\vdots$

$$\leq \sum p_i f(x_i)$$

Ex 2.15: 3 squares w/ $\bar{A} = 100$, $\bar{l} = 10$

$(\bar{l})^2 = \bar{A} \Rightarrow$ all squares the same

Ex 2.16: a) $\left.\begin{matrix} 2, 3, 4, 5 \dots 10, 11, 12 \\ \frac{1}{36}\ \frac{2}{36}\ \frac{3}{36}\ \frac{4}{36} \dots \frac{3}{36}\ \frac{2}{36}\ \frac{1}{36} \end{matrix}\right\}$ sum of 2 dice

0	1	2	3	4	5
$\frac{1}{6}$	$\frac{4}{6}\cdot\frac{1}{3} + \frac{1}{3}\cdot\frac{1}{6}$			$\frac{4}{36}$	$\frac{2}{36}$
$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$		

b) mean $3.5 \cdot 100 = 350$ $\Leftarrow$ Gaussian

var $\frac{35}{12} \cdot 100 \sim 292$

$\frac{1}{3} \cdot \left(\frac{1}{4} + \frac{9}{4} + \frac{25}{4} \right) = \frac{35}{12}$

c) One ordinary 1 2 3 4 5 6

One spiky 6 6 6 0 0 0

~~AA~~ return d) Label r'th dice by $\{0, 1, 2, 3, 4, 5\} \cdot 6^r$

2.17 $\frac{p}{1-p} = \exp(a) \Rightarrow 1 - p = p \exp(-a)$

$1 = p(1 + \exp(-a))$

$\Rightarrow p = \frac{1}{1 + e^{-}}$

$\frac{1 + \tanh(a/2)}{2}$

2.18 $\log \frac{P(x=1|y)}{P(x=0|y)} = \log \frac{P(y|x=1)\,p(x=1)}{P(y|x=0)\,p(x=0)}$ $\square$

2.19 $d_1 \perp d_2 \mid x$

$$\frac{P(x=1|\{d_i\})}{P(x=0|\{d_i\})} = \frac{P(d_1|x=1)\,P(d_2|x=1)\,P(x=1)}{P(d_1|x=0)\,P(d_2|x=0)\,P(x=0)}$$

2.20 $$\frac{S_N \int_{r-\epsilon}^{r} dr'\, r'^{N-1}}{S_N \int_0^r dr'\, r'^{N-1}} = \frac{S_N\left[r^N - (r-\epsilon)^N\right]}{S_N\, r^N} = 1 - \left(1 - \frac{\epsilon}{r}\right)^N$$

as $N \to \infty$
this frac $\to 1$

2.21 $\mathbb{E}\, f(x) = 0.1 \cdot 10 + 0.2 \cdot 5 + 0.7 \cdot \frac{10}{7} = 3$

$\mathbb{E} \frac{1}{p} = 3$

2.22 $\mathbb{E}\left[\frac{1}{p}\right] = |A|$ always

2.23 0.2

2.24 $P\left[P(x) \in [0.15, 0.5]\right] = 0.2$

$$P\left[\left|\log \frac{P(x)}{0.2}\right| > 0.05\right] = p_a + p_c = 0.8$$

2.25 $H(X) = \mathbb{E} \log \frac{1}{p} \leq \log \mathbb{E} \frac{1}{p} = \log |A_X|$

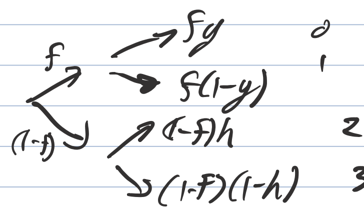
2.26 $D_{KL}(P||Q) = -\mathbb{E}_p\left(\log\frac{q}{p}\right) \geq -\log \mathbb{E}\, 1 = 0$

2.27 $-\sum_{\underline{x}} p(\underline{x}) \log p(\underline{x}) = -\sum_{x_1} p(x_1) \sum p(x_2 \ldots | x_1) \log p(x_1) p(x_2 \cdots | x_1)$

$= -\sum_{x_1} p(x_1) \log p(x_1) - \sum_{x_1} \sum_{x_2 \cdots} p(x_2 \cdots | x_1) \log p(x_2 \cdots | x_1)$

$H(X,Y) = H(X) + H(Y|X)$

2.28 $H(X) = H_2(f) + f H_2(g) + (1-f) H_2(h)$



$\frac{\partial H(X)}{\partial f} = \log \frac{1-f}{f} + H_2(g) - H_2(h)$

2.29 Directly :

$p(1) = 1 - 1/2$
$p(2) = 1/2\,(1-1/2)$
$\vdots$

$1/2$
$1/4 \Rightarrow H(X) = \sum_{n=0}^{\infty} 2^{-n} \cdot n$
$1/8$

$= \frac{\partial}{\partial a} \frac{1}{1-a}\Big|_{1/2}$

$= \frac{1}{1-1/2}$

$= 2$

$H(X) = H(1) + \frac{1}{2} H(2) + \cdots + \frac{1}{2} H(3)$

1 1

$= 2$

2.30 $\mathbb{P}[B_2 = W] = \mathbb{P}[B_1 = W] \cdot \mathbb{P}[B_2 = W | B_1 = W]$ -

$+ \mathbb{P}[B_1 = B] \cdot \mathbb{P}[B_2 = W \mid B_1 = B]$

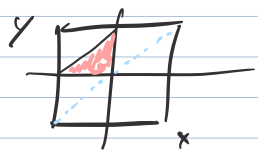
$$= \frac{W}{N}\,\frac{W-1}{N-1} + \frac{N-W}{N}\,\frac{W}{N-1} = \frac{W(N-1)}{N(N-1)} = \frac{W}{N}$$

2.31 $\frac{a}{b}$ chance of intersecting a line horizontally or vertically

$\Rightarrow (1-a/b)^2$ chance of being in a square

2.32 $$\int_0^{2\pi} \frac{d\theta}{2\pi}\,\frac{a|\cos\theta|}{b} = \frac{a}{b}\int_{-\pi/2}^{\pi/2} \frac{d\theta}{\pi}\cos\theta = \frac{2a}{b\pi}$$

2.33



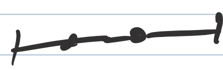
$x,\ y-x,\ 1-y$

$x + y - x > 1-y \Rightarrow y > 1-y$

$x + 1-y > y-x \Rightarrow 1-y+x > y-x$

$1-y + y-x > x \Rightarrow 1-x > x$

$y < x + 1/2$



$y - x < 1/2$

$x < 1/2$

$y > 1/2$

$$\boxed{P = 1/4}$$

2.34 $$\mathbb{E}\,x = \sum_{n=0}^{\infty} n 2^{-n} = 2$$

$$\langle f\rangle = \left\langle \frac{h}{h+t}\right\rangle = \left\langle \frac{1}{h+t}\right\rangle \Rightarrow \mathbb{E}\,\frac{1}{n} = \sum_{n=1}^{\infty} \frac{f(1-f)^{n-1}}{n} = \frac{f}{f-1}\log f$$

2.35 a) Exponential dist

$$P(r) = \left(\frac{5}{6}\right)^{r-1}\frac{1}{6}$$

memoryless

$\mathbb{E}[r] = 6$

e^{-ar} $\quad a = \log 6/5$

Z

b) still 6 (memoryless)

c) still 6

d) $6+6-1 = 11$

e) Yes. More likely to arrive in bigger open spot

2.36 a) $1/2$

b) $2/3$

FAB
FBA
AFB
BFA
ABF
BAF

2.37 $$P(1_T \mid 2_T) = \frac{P(2_T \mid 1_T)\,P(1_T)}{P(2_T \mid 1_T)P(1_T) + P(2_T \mid 1_F)\,P(1_F)}$$

$$= \frac{1/3 \cdot 1/3}{1/3 \cdot 1/3 + 2/3 \cdot 2/3} = \frac{1}{5}$$

2.38 1) $p_B = 3f^2(1-f) + f^3$

2) $$P(r=000) = \overset{P(s=0)}{\tfrac{1}{2}}(1-f)^3 + \overset{P(s=1)}{\tfrac{1}{2}} f^3$$

$$P(r=001) = \tfrac{1}{2} f(1-f)^2 + \tfrac{1}{2} f^2(1-f) = \tfrac{1}{2} f(1-f)$$

p(errors|...)

$$p(s=1 \mid 000) = \frac{f^3 \cdot 1/2}{\frac{1}{2}[(1-f)^3 + f^3]} = \frac{f^3}{(1-f)^3 + f^3}$$

$$p(s=1 \mid 001) = \frac{(1-f)f^2 \cdot 1/2}{\frac{1}{2} f(1-f)} = f$$

$$\Rightarrow p(\text{error}) = \sum_r p(r)\,p(\text{error} \mid r) = 2 \cdot \tfrac{1}{2}[(1-f)^3 + f^3]\,P(\text{err} \mid 000) + 6 \cdot \tfrac{1}{2} f(1-f)\,P(\text{err} \mid 001)$$

$$= f^3 + 3f^2(1-f)$$

2.39 $-\sum \frac{0.1}{n} \log_2 \frac{0.1}{n} \sim 9.72$ bits/word

Chapter 3

Ex 3.1 $\frac{P(A|D)}{P(B|D)} = \frac{1}{2}\frac{3}{2}\frac{1}{1}\frac{3}{2}\frac{1}{2}\frac{2}{2}\frac{1}{2} = \frac{9}{32} \Rightarrow B$

$$P(A|D) = \frac{9}{9+32} = \frac{9}{41}$$

Ex 3.2
$$20^7 \cdot P(A|D) = 3\cdot 1\cdot 2\cdot 1\cdot 3\cdot 1\cdot 1 = 18$$
$$20^7 P(B|D) = 2\cdot 2\cdot 2\cdot 2\cdot 2\cdot 1\cdot 2 = 2^6 = 64$$
$$20^7 P(C|D) = 1 = 1$$

$$\frac{18}{83} \quad \frac{64}{83} \quad \frac{1}{83}$$

Ex 3.3

$$P(x|\lambda) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} Z(\lambda) & 1<\lambda<20 \\ 0 & \text{else} \end{cases}$$

$$P(\lambda|\{x\}) = \frac{1}{(\lambda Z(\lambda))^N} \exp\left[-\sum_n x_n/\lambda\right] P(\lambda)$$

$$Z(\lambda) = \int_1^{20} dx \frac{e^{-x/\lambda}}{\lambda} = e^{-1/\lambda} - e^{-20/\lambda}$$

What you know about λ after the data is what you knew before $P(\lambda)$ & what the data told you $P(\{x\}|\lambda)$

λ is not stochastic. Rather we have a degree of belief

Ex 3.4 $P(O) = .6$
$P(AB) = .01$

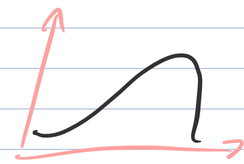
$$P(\text{crime}|O) = \frac{P(O|\text{crime})\,P(\text{crime})}{\quad}$$

$P(\theta)$

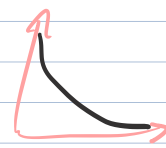
$p(D|S) = p_{AB}$ $\left.\begin{matrix} \\ \end{matrix}\right\} \frac{p(D|S)}{p(D|\bar{S})} = \frac{1}{2p_0}$ $= \frac{1/2}{.6} P(\text{crime}) \leq P(\text{crime})$

$p(D|\bar{S}) = 2p_0 p_{AB}$

Ex 3.5 $P(p_a | aba) = p_a^2 (1-p_a)^1$



$P(p_a | bbb) = (1-p_a)^3$



$P(a|s,F) = \frac{F_a + 1}{F_a + F_b + 2}$ } Laplace

Ex 3.6 $\mathcal{H}_1$ is uniform
$\mathcal{H}_0$ is $p = 1/6$

$$\frac{P(\mathcal{H}_1 | s, F)}{P(\mathcal{H}_0 | s, F)} = \frac{\frac{F_a! \, F_b!}{(F_a + F_b + 1)!}}{p_0^{F_a} (1-p_0)^{F_b}}$$

$$\Rightarrow \log \frac{P(\mathcal{H}_1 | s)}{P(\mathcal{H}_0 | s)} = -\log \binom{F}{F_a} - \log(F+1) - \log p_0^{F_a} (1-p_0)^{F_b}$$

$$= F_a \log p_a F + F_b \log p_b F - F \log F + \frac{1}{2} \log 2\pi \frac{F_a F_b}{F}$$

$$- F_a \log p_0 - F_b \log (1-p_0) - \log(F+1)$$

$$= F(p_a \log p_a + p_b \log p_b) + \frac{1}{2} \log(2\pi p_a p_b F)$$

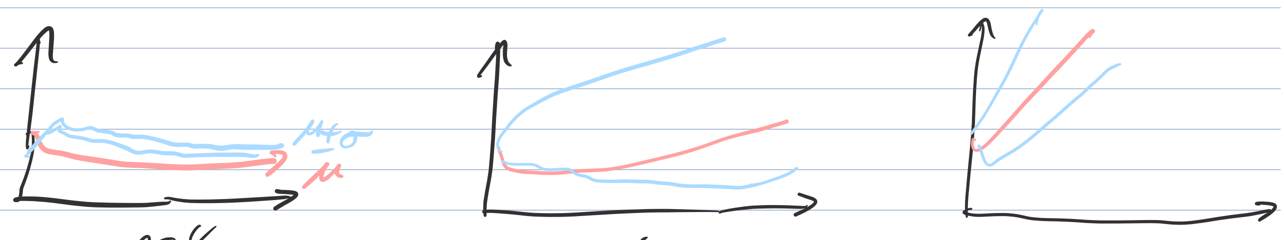
$$= F\left(p_a \log \frac{p_a}{p_0} + \log \frac{p_b}{1-p_0}\right) + \frac{1}{2} \log(2\pi p_a p_b F)$$

$$= \underbrace{F \, D_{KL}(p_a \| p_0)}_{\text{increases linearly unless}} - \underbrace{\frac{1}{2} \log\left(\sqrt{F} + \frac{1}{\sqrt{F}}\right)}_{\text{increases logarithmically}}$$

$p_a \to p_0$ in KL

3.7 $F_a = p_a F \pm \sqrt{F p_A (1-p_A)}$

→ Make plots of evidence



$p = 1/6$ $p = 1/4$ $p = 1/2$

3.8 Monty Hall: switch

3.9 Earthquake MH: no need

3.10 l, m, n are the sexes

$$l = 1 \Rightarrow \begin{cases} m=0 & n=0 \\ m=1 & n=0 \\ m=0 & n=1 \end{cases}$$

⇒ $\frac{2}{3}$ chance of 2 girls 1 boy

$\frac{1}{3}$ chance of 1 girl 2 boys

3.11

$P(h \mid m=1) = 0.28$ $P(b \mid h=0) = 0.02$

husband did it murdered

$P(b \mid h=1) = 0.9$

$$P(h \mid b=1, m=1) = \frac{P(b=1 \mid h=1) \cdot P(h=1 \mid m=1)}{P(b=1 \mid h=1) P(h=1 \mid m=1) + P(b=1 \mid h=0) P(h=0 \mid m=1)}$$

$$= \frac{0.9 \cdot 0.28}{\ldots} \approx 95\%$$

$0.9 \cdot 0.28 + 0.02 \cdot 0.72$

3.12 $P(D \mid H=0) = 1$ $\quad P(D \mid H=1) = 1/2$

draw original was white

$\Rightarrow P(H=0 \mid D) = \dfrac{1 \cdot 1/2}{3/4} = 2/3$

$P(H=1 \mid D) = 1/3$

3.13 $P(D \mid \mathcal{H}_0) = 1$

$P(D \mid \mathcal{H}_1) = \alpha \cdot \beta$

P(valid #) P(busy | valid)

$\approx \dfrac{75000}{10^6} \cdot 0.01$ (fewer if 4AM)

$\approx 0.1 \cdot 1\%$

$\Rightarrow \dfrac{P(D \mid \mathcal{H}_0)}{P(D \mid \mathcal{H}_1)} \approx 10^3$

if prior was 50/50 then

$P(\mathcal{H}_0 \mid D) = \dfrac{1}{1 - \frac{P(D \mid \mathcal{H}_1)}{P(D \mid \mathcal{H}_0)}} \approx 0.99$

3.14 1/3 prob

3.15 $\mathcal{H}_0 \Rightarrow p_0 = 1/2$
$\mathcal{H}_1 \Rightarrow$ biased, w uniform prior on bias

$$\frac{P(D \mid \mathcal{H}_1)}{P(D \mid \mathcal{H}_0)} = \frac{\frac{140!\,110!}{251!}}{(1/2)^{250}} \approx 0.48$$

Tweaking to Beta prior:

$$\frac{P(D|\mathcal{H}_\alpha)}{P(D|\mathcal{H}_0)} = 2^{250} \frac{\Gamma(140+\alpha)\Gamma(110+\alpha)}{\Gamma(250+2\alpha)} \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2}$$

ratio is ≤ 1 for $\alpha < 3$
and never > 2

If 141, 109, get p-val of 0.05
but $\mathcal{H}_1$ is even less likely
under uniform prior

$\Rightarrow$ Don't be a frequentist!