1.1 The sample space for the roll of two fair dice has 36 equi-probable outcomes:

X = outcome of first die
Y = outcome of second die
$Z = X + Y$ = outcome of experiment of interest.

Z \ Y	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The probs. of the outcome of Z are: $P_Z(2) = \frac{1}{36}$, $P_Z(3) = \frac{1}{18}$, $P_Z(4) = \frac{1}{12}$, $P_Z(5) = \frac{1}{9}$, $P_Z(6) = \frac{5}{36}$, $P_Z(7) = \frac{1}{6}$, $P_Z(8) = \frac{5}{36}$, $P_Z(9) = \frac{1}{9}$, $P_Z(10) = \frac{1}{12}$, $P_Z(11) = \frac{1}{18}$, $P_Z(12) = \frac{1}{36}$

The entropy $H(Z)$ is

$$H(Z) = -\sum_{k=2}^{12} P_Z(k) \log P_Z(k)$$

$$= 3.2744 \text{ bits}$$

1.2

X \ Y	B	W	
1	$\frac{1}{3}$	0	$\frac{1}{3}$
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
3	0	$\frac{1}{3}$	$\frac{1}{3}$
	$\frac{1}{2}$	$\frac{1}{2}$	

The joint and marginal probabilities of X and Y are tabulated to the left.

$$H(X) = \sum_{i=1}^{3} \left(\tfrac{1}{3}\right) \log 3 = \log 3 = 1.5850 \text{ bits}$$

$$H(Y) = \tfrac{1}{2} \log 2 + \tfrac{1}{2} \log 2 = \log 2 = 1 \text{ bit}$$

$$H(X,Y) = 2\left[\tfrac{1}{3} \log 3 + \tfrac{1}{6} \log 6\right] = 1.9183 \text{ bits}$$

$$H(X|Y) = \tfrac{1}{3} \log \tfrac{3}{2} + \tfrac{1}{6} \log 3 + \tfrac{1}{6} \log 3 + \tfrac{1}{3} \log \tfrac{3}{2} = 0.9183 \text{ bits}$$

$$H(Y|X) = \tfrac{1}{3} \log 1 + \tfrac{1}{6} \log 2 + \tfrac{1}{6} \log 2 + \tfrac{1}{3} \log 1 = \tfrac{1}{3} \log 2 = \tfrac{1}{3} \text{ bits}$$

$$I(X;Y) = H(Y) - H(Y|X) = \log 2 - \tfrac{1}{3} \log 2 = \tfrac{2}{3} \log 2 = \tfrac{2}{3} \text{ bits}$$

1.3 $X = \{b_1, b_2, b_3, b_4, g_5, g_6, g_7\}$ with $p_i = \frac{1}{7}$, $i = 1, \ldots, 7$.

$Y = \{b_1, b_2, b_3, b_4\}$ with $q_i = \dfrac{1/7}{P(Y)} = \dfrac{1/7}{4/7} = \frac{1}{4}$, $i = 1, \ldots, 4$, where $P(Y) = \frac{4}{7}$

$Z = \{g_5, g_6, g_7\}$ with $r_i = \dfrac{1/7}{P(Z)} = \dfrac{1/7}{3/7} = \frac{1}{3}$, $i = 5, 6, 7$, where $P(Z) = 3/7$

B is a binary event, with $P(B) = P(Y) = 4/7$ and $P(\bar{B}) = P(Z) = 3/7$.

Calculating entropies: $H(X) = \sum_{i=1}^{7} \left(\frac{1}{7}\right) \log 7 = \log 7 = 2.80735$ bits

$H(Y) = \sum_{i=1}^{4} \left(\frac{1}{4}\right) \log 4 = \log 4 = 2$ bits, $H(Z) = \sum_{i=5}^{7} \left(\frac{1}{3}\right) \log 3 = \log 3 = 1.58496$ bits

$H(B) = P(B) \log 1/P(B) + P(\bar{B}) \log 1/P(\bar{B}) = \frac{4}{7} \log \frac{7}{4} + \frac{3}{7} \log \frac{7}{3} = 0.985228$ bits

Now

$$H(B) + P(B)H(Y) + P(\bar{B})H(Z) = 0.985228 + \tfrac{4}{7}(2) + \tfrac{3}{7}(1.58496) = 2.80735 \text{ bits}$$

n.b. This the same value as $H(X)$. This is an example of the grouping property.

1.4 We must find the $\{p_i\} = \{p_1, p_2, p_3, \ldots\}$ that maximizes $H(X)$ given the constraint that the mean is equal to M. This is most easily done with Lagrange multipliers.

So we wish to maximize $-\sum_{n=0}^{\infty} p_n \log p_n$

Subject to the two constraints

$$\sum_{n=0}^{\infty} p_n = 1 \quad \text{and} \quad \sum_{n=0}^{\infty} n p_n = M.$$

Form the Lagrangian $L(\underline{p})$ as (here λ and μ are Lagrange multipliers.)

$$L(\underline{p}) = -\sum_{n=0}^{\infty} p_n \ln p_n + \lambda\left[\sum_{n=0}^{\infty} p_n - 1\right] + \mu\left[\sum_{n=0}^{\infty} n p_n - M\right]$$

To find the extremum, set $\frac{\partial L(\underline{p})}{\partial p_k} = 0$, $k = 0, 1, 2, \ldots$

$$\frac{\partial L(\underline{p})}{\partial p_k} = \frac{\partial}{\partial p_k}\left[-p_k \ln p_k + \lambda p_k + \mu k p_k\right]$$

$$= -\ln p_k - \frac{p_k}{p_k} + \lambda + \mu k = -\ln p_k - 1 + \lambda + \mu k = 0, \quad k = 0, 1, 2, \ldots$$

Thus we have

$$\ln p_k = \lambda - 1 + \mu k \Rightarrow p_k = e^{\lambda - 1 + \mu k}$$

or $p_n = K e^{+\mu n}$, where $K = e^{\lambda - 1} = \text{constant}$.

Note that $1 = \sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} K e^{\mu n} = K \sum_{n=0}^{\infty} (e^{\mu})^n = \frac{K}{1 - e^{\mu}}$, since $|e^{\mu}| < 1$

$$\Rightarrow K = 1 - e^{\mu}$$

Now

$$M = \sum_{n=0}^{\infty} n p_n = K \sum_{n=0}^{\infty} n e^{\mu n} = K \sum_{n=0}^{\infty} \frac{d e^{\mu n}}{d\mu} = K \frac{d}{d\mu}\left\{\sum_{n=0}^{\infty} e^{\mu n}\right\}$$

$$= K \frac{d}{d\mu}\left\{\frac{1}{1 - e^{\mu}}\right\} = \frac{K e^{\mu}}{(1 - e^{\mu})^2} = \frac{e^{\mu}}{1 - e^{\mu}} \Rightarrow \mu = \log\left(\frac{M}{M+1}\right) *$$

$$\therefore p_n = K e^{\mu n} = \left(1 - e^{\log\left(\frac{M}{M+1}\right)}\right) e^{n \log\left(\frac{M}{M+1}\right)} = \boxed{\frac{1}{M+1}\left(\frac{M}{M+1}\right)^n, \; n = 0, 1, 2, \ldots}$$

The associated maximum entropy H is

$$H = \sum_{n=0}^{\infty} p_n \log \frac{1}{p_n} = \sum_{n=0}^{\infty} p_n \log\left[(M+1)\left(\frac{M+1}{M}\right)^n\right] = \log(M+1) \sum_{n=0}^{\infty} p_n + \log\left(\frac{M+1}{M}\right) \sum_{n=0}^{\infty} n p_n$$

$$= \log(M+1) + M \log\left(\frac{M+1}{M}\right) = \boxed{(M+1)\log(M+1) - M \log M}$$

* n.b. $j = n+1$ is geometrically distributed.

1.5. Proof of Theorem: Most easily done by induction

for $N=2$: Inequality is true by defn. of convexity

for $N>2$: Assume valid for $N-1$, we must prove it valid for N.

$$f\left(\sum_{n=1}^{N}\lambda_n P_n\right) = f\left(\lambda_N P_N + (1-\lambda_N)\sum_{n=1}^{N-1}\frac{\lambda_n}{1-\lambda_N}P_n\right)$$

$$\geq \lambda_N f(P_N) + (1-\lambda_N) f\left(\sum_{n=1}^{N-1}\frac{\lambda_n}{1-\lambda_N}P_n\right), \text{ convexity of } f$$

$$\geq \lambda_N f(P_N) + (1-\lambda_N)\sum_{n=1}^{N-1}\frac{\lambda_n}{1-\lambda_N} f(P_n), \text{ by hypothesis (true for } N-1).$$

$$= \lambda_N f(P_N) + \sum_{n=1}^{N-1}\lambda_n f(P_n) = \sum_{n=1}^{N}\lambda_n f(P_n)$$

$$\therefore\ f\left(\sum_{n=1}^{N}\lambda_n P_n\right) \geq \sum_{n=1}^{N}\lambda_n f(P_n)$$

In order to prove Jensen's Inequality for a finite r.v., we take the λ_n's in the Theorem above and replace them with the probabilities p_n, and we take the points P_n in the theorem above and take them to be the X_n's (the possible r.v. outcomes)

It then follows that for a convex $\cap$ fcn. $f(\cdot)$

$$f\left(\sum_{n=1}^{N} p_n X_n\right) \geq \sum_{n=1}^{N} p_n f(X_n)$$

or

$$f(E(X)) \geq E(f(X))$$

We have yet to show that equality holds in the case of a strictly convex fcn. f iff the r.v. X takes on a single value with certainty:

$\Leftarrow$: Assume X takes on the value X_k with certainty. Then

$$p_n = \begin{cases} 1, & n=k \\ 0, & n\neq k \end{cases} \Rightarrow f(E(X)) = f(X_k), \text{ and } E(f(X)) = 1\cdot f(X_k) = f(X_k)$$

$\Rightarrow$: If $f(\cdot)$ is strictly convex, then for non-negative $\lambda_1, \ldots, \lambda_N$ that sum (N>1) to unity, a simple extension of the above Theorem yields $f\left(\sum_n \lambda_n P_n\right) > \sum_n \lambda_n f(P_n)$

again substituting probs. p_n for λ_n and r.v range values X_n for P_n, we have, for $N>1$

$f\left(\sum_{n=1}^{N} p_n X_n\right) > \sum_{n=1}^{N} p_n f(X_n)$. However, if $p_n = \begin{cases} 1, & n=k \\ 0, & n\neq k \end{cases}$ (effectively, $N=1$)

$f(E(X)) = f(X_k) = E(f(X_k))$

1.6 This is easily shown by applying Jensen's Inequality to show that

$$H(X,Y) - H(X) - H(Y) \leq 0 :$$

$$H(XY) - H(X) - H(Y) = \sum_{x,y} p(x,y) \log \frac{1}{p(x,y)} - \sum_x p(x) \log \frac{1}{p(x)} - \sum_y p(y) \log \frac{1}{p(y)}$$

$$= \sum_{x,y} p(x,y) \log \frac{p(x)p(y)}{p(x,y)} \leq \log \left(\sum_{x,y} p(x,y) \frac{p(x)p(y)}{p(x,y)} \right)$$

$$= \log \left(\sum_{x,y} p(x)p(y) \right) = \log 1 = 0$$

$\therefore H(XY) \leq H(X) + H(Y)$

Furthermore, since $\log(\cdot)$ is strictly convex $\cap$, it follows that equality holds $\iff \frac{p(x)p(y)}{p(x,y)} = \text{constant}, \forall (x,y)$

$\iff p(x,y) = p(x)p(y) \iff$ Marginal ensemble outcomes X and Y are statistically independent.

1.7

$$I(X;Y) = H(X) - H(X|Y) = \sum_x p(x) \log \frac{1}{p(x)} - \sum_{x,y} p(x,y) \log \frac{1}{p(x|y)}$$

$$= \sum_{x,y} p(x,y) \log \frac{1}{p(x)} + \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(y)}$$

$$= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \quad \text{we use this sum to prove the equalities} \quad (*)$$

(i) $$I(Y;X) = H(Y) - H(Y|X) = \sum_y p(y) \log \frac{1}{p(y)} - \sum_{x,y} p(x,y) \log \frac{1}{p(y|x)}$$

$$= \sum_{x,y} p(x,y) \log \frac{1}{p(y)} + \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)} = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \text{ same as } (*).$$

(ii) proved in (i), since $I(X;Y) = I(Y;X)$

(iii) from $(*)$,

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \sum_{x,y} p(x,y) \left[\log \frac{1}{p(x)} + \log \frac{1}{p(y)} - \log \frac{1}{p(x,y)} \right]$$

$$= \sum_x p(x) \log \frac{1}{p(x)} + \sum_y p(y) \log \frac{1}{p(y)} - \sum_{x,y} p(x,y) \log \frac{1}{p(x,y)}$$

$$= H(X) + H(Y) - H(XY).$$

1.8 $p_n = \frac{1}{An\log^2 n}$, $n = 2, 3, 4, \ldots$, $A = \sum_{n=2}^{\infty} \frac{1}{n\log^2 n} = \begin{cases} 2.10974, \text{ base-}e \text{ log.} \\ 1.01363, \text{ base-}2 \text{ log} \end{cases}$

$$H(X) = \sum_{n=2}^{\infty} p_n \log \frac{1}{p_n} = \sum_{n=2}^{\infty} \frac{1}{An\log^2 n} \cdot \log[An\log^2 n]$$

$$= \sum_{n=2}^{\infty} \frac{\log(A)}{An\log^2 n} + \sum_{n=2}^{\infty} \frac{\log n}{An\log^2 n} + \sum_{n=2}^{\infty} \frac{\log[\log^2 n]}{An\log^2 n}$$

$$= \underbrace{\log A}_{\text{Term 1}} + \underbrace{\frac{1}{A}\sum_{n=2}^{\infty} \frac{1}{n\log n}}_{\text{Term 2}} + \underbrace{\frac{1}{A}\sum_{n=2}^{\infty} \frac{\log[\log^2 n]}{An\log^2 n}}_{\text{Term 3}}$$

We examine the 3 terms

Term 1: A is a non finite constant $> 1 \Rightarrow \log A$ is finite

Term 2: From the integral test of series convergence, we know that

$$\sum_{n=2}^{\infty} \frac{1}{n\log n} \geq \int_2^{\infty} \frac{dx}{x\log x} = K \lim_{r\to\infty} \int_2^r \frac{dx}{x\ln x} = K \cdot \lim_{r\to\infty} \left[\ln(\ln r) - \ln(\ln 2)\right]$$

(K: pos. constant dep. on log base)

$= +\infty$

So the seconds term diverges toward $+\infty$

(We need only check to make sure term 3 does not go to $-\infty$ to show divergence to $+\infty$)

Term 3: $\frac{1}{A}\sum_{n=2}^{\infty} \frac{\log[\log^2 n]}{An\log^2 n} > \frac{1}{A}\sum_{n=2}^{\infty} \frac{1}{An\log^2 n} = \frac{1}{A} =$ positive constant.

$\therefore H(X) = +\infty$ (Diverges)

1.9 $P_X(0) = P_X(1) = \frac{1}{2}$, note also by symmetry that $P_Y(0) = P_Y(1) = \frac{1}{2}$

$$H(Y) = \tfrac{1}{2}\log 2 + \tfrac{1}{2}\log 2 = \log 2$$

$$H(Y|X=0) = H(Y|X=1) = -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon) = \mathcal{H}(\varepsilon)$$

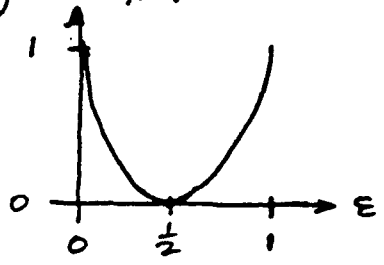
$$H(Y|X) = P_X(0)\,H(Y|X=0) + P_X(1)\,H(Y|X=1) = \mathcal{H}(\varepsilon)$$

$\therefore I(X;Y) = H(Y) - H(Y|X) = \log 2 - \mathcal{H}(\varepsilon)$

Taking logs to be base-2:

$I(X;Y) = 1 - \mathcal{H}_2(\varepsilon)$ bits.

note behavior as a fcn. of ε



10. Cover & Thomas 2.7: The goal of this problem is to investigate the problem of generating a sequence of fair coin flips from a sequence of biased coin flips.

First we must justify the following equalities and inequalities showing that no more than $n\mathcal{H}_2(p)$ fair coin flips can be obtained from n tosses with a probability p of heads: (on average)

$$\begin{aligned} n\mathcal{H}_2(p) &\overset{(a)}{=} H(X_1,\ldots,X_n) \\ &\overset{(b)}{\geq} H(Z_1,\ldots,Z_K,K) \\ &\overset{(c)}{=} H(K)+H(Z_1,\ldots,Z_K|K) \\ &\overset{(d)}{=} H(K)+E\{K\} \\ &\overset{(e)}{\geq} E\{K\} \end{aligned}$$

(a): Because $X_1,\ldots,X_n$ are i.i.d. RVs with

$$P_{X_i}(x) = \begin{cases} p, & x=1 \\ 1-p, & x=0 \end{cases},$$

we have

$$\begin{aligned} H(X_1,\ldots,X_n) &= H(X_1)+H(X_2)+\cdots+H(X_n) \\ &= \mathcal{H}_2(p)+\mathcal{H}_2(p)+\cdots+\mathcal{H}_2(p) \\ &= n\mathcal{H}_2(p). \end{aligned}$$

(b): Because $Z_1,\ldots,Z_K$ is defined by a mapping

$f: \underbrace{\{0,1\}^n}_{\text{binary } n\text{-tuples}} \longrightarrow \underbrace{\{0,1\}^*}_{\text{binary strings}}$, where K may depend on $(X_1,\ldots,X_n)$

we have that $Z_1,\ldots,Z_K,K$ are a function of the RVs $X_1,\ldots,X_n$. Because the entropy of a fcn of a R.V must be less than or equal to the entropy of the RV, it follows that

$$H(X_1,\ldots,X_n) \geq H(Z_1,\ldots,Z_K,K)$$

(c): By the chain rule for entropy (see Cover & Thomas p. 21, Eq. (2.48)), we have

$$H(Z_1,\ldots,Z_K,K) = H(Z_1,\ldots,Z_K \mid K) + H(K).$$

(d): Because the Z_i are independent Bernoulli R.Vs with $p_{Z_i}(0) = p_{Z_i}(1) = 1/2$, it follows that

$$H(Z_1,\ldots,Z_K \mid K=k) = \underbrace{\mathcal{H}_2(\tfrac{1}{2}) + \cdots + \mathcal{H}_2(1/2)}_{k\text{-terms}}$$

$$= 1 + \cdots + 1$$

$$= k$$

But in general K is a R.V. (a ftn. of the RVs $X_1,\ldots,X_n$), so we have

$$H(Z_1,\ldots,Z_K \mid K) = \sum_k P(\{K=k\})\, H(Z_1,\ldots,Z_k \mid K=k)$$

$$= \sum_k p_K(k)\cdot k = E\{K\}$$

$$\therefore\ H(Z_1\ldots,Z_K \mid K) + H(K) = H(K) + E\{K\}$$

(e): Because $H(K) \geq 0 \Rightarrow H(K) + E\{K\} \geq E\{K\}$.

(f): Since we do not know p, the only way to generate random bits is to use the following key fact that we do know:

> "Regardless of the value of p, all sequences $(X_1,\ldots,X_n)$ with the same number of "1"s are equally likely."

So, for example, for $n=4$, we have that 0001, 0010, 0100, and 1000 are equally likely.

So we can use these four sequences to generate two pure random bits each time one of them occurs. An example of a mapping to generate these bits is

$0000 \longmapsto \Lambda$ (empty string)

$0001 \longmapsto 00$, $0010 \longmapsto 01$, $0100 \longmapsto 10$, $1000 \longmapsto 11$

$0011 \longmapsto 00$, $0110 \longmapsto 01$, $1100 \longmapsto 10$, $1001 \longmapsto 11$

$1011 \longmapsto 0$, $0101 \longmapsto 1$

$1110 \longmapsto 00$, $1101 \longmapsto 01$, $1011 \longmapsto 10$, $0111 \longmapsto 11$

$1111 \longmapsto \Lambda$.

The expected number of bits $E\{K\}$ that can be generated from the $n=4$ bent coin outcomes is

$$E\{K\} = 4 \cdot p(1-p)^3 \cdot 2 \text{ bits} + 4p^2(1-p)^2 \cdot 2 \text{ bits} + 2 \cdot p^2(1-p)^2 \cdot 1 \text{ bit} + 4p^3(1-p) \cdot 2 \text{ bits}$$

$$= 8p(1-p)^3 + 10p^2(1-p)^2 + 8p^3(1-p)$$

n.b. $E\{K\}\Big|_{p=1/2} \cong 1.625 = \frac{26}{16} = \frac{13}{8} = 1\,5/8$ bits

So it is not very efficient when $p \simeq 1/2$; however, it does not require knowledge of p.