

Example: Letter Frequencies

i	a_i	p_i		
1	a	0.0575	a	■
2	b	0.0128	b	■
3	c	0.0263	c	■
4	d	0.0285	d	■
5	e	0.0913	e	■
6	f	0.0173	f	■
7	g	0.0133	g	■
8	h	0.0313	h	■
9	i	0.0599	i	■
10	j	0.0006	j	■
11	k	0.0084	k	■
12	l	0.0335	l	■
13	m	0.0235	m	■
14	n	0.0596	n	■
15	o	0.0689	o	■
16	p	0.0192	p	■
17	q	0.0008	q	■
18	r	0.0508	r	■
19	s	0.0567	s	■
20	t	0.0706	t	■
21	u	0.0334	u	■
22	v	0.0069	v	■
23	w	0.0119	w	■
24	x	0.0073	x	■
25	y	0.0164	y	■
26	z	0.0007	z	■
27	–	0.1928	–	■

Figure 2.1. Probability distribution over the 27 outcomes for a randomly selected letter in an English language document (estimated from *The Frequently Asked Questions Manual for Linux*). The picture shows the probabilities by the areas of white squares.

Example: Letter Frequencies

i	a_i	p_i	
1	a	0.0575	a
2	b	0.0128	b
3	c	0.0263	c
4	d	0.0285	d
5	e	0.0913	e
6	f	0.0173	f
7	g	0.0133	g
8	h	0.0313	h
9	i	0.0599	i
10	j	0.0006	j
11	k	0.0084	k
12	l	0.0335	l
13	m	0.0235	m
14	n	0.0596	n
15	o	0.0689	o
16	p	0.0192	p
17	q	0.0008	q
18	r	0.0508	r
19	s	0.0567	s
20	t	0.0706	t
21	u	0.0334	u
22	v	0.0069	v
23	w	0.0119	w
24	x	0.0073	x
25	y	0.0164	y
26	z	0.0007	z
27	-	0.1928	-



Figure 2.1. Probability distribution over the 27 outcomes for a randomly selected letter in an English language document (estimated from *The Frequently Asked Questions Manual for Linux*). The picture shows the probabilities by the areas of white squares.

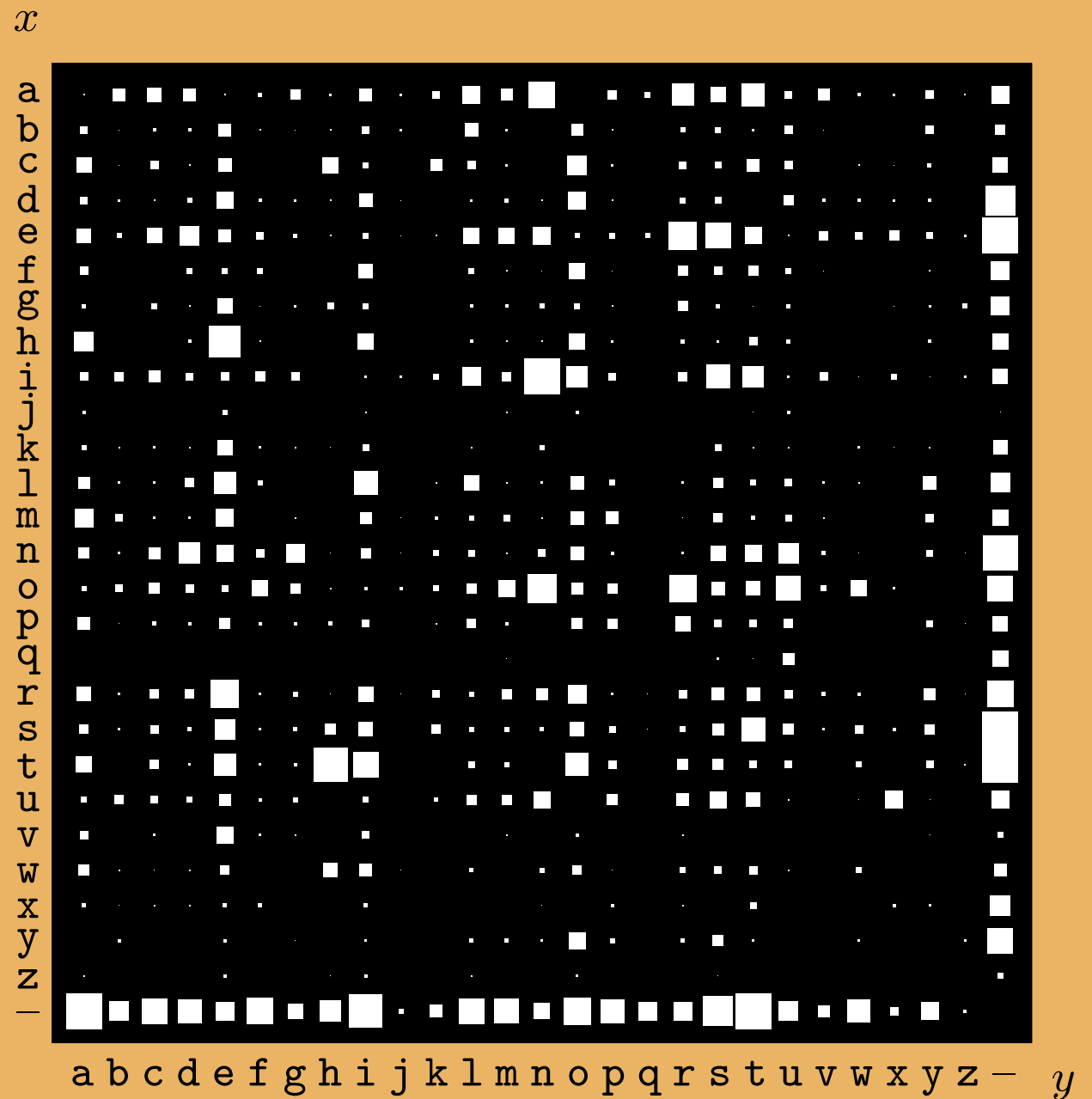


Figure 2.2. The probability distribution over the 27×27 possible bigrams xy in an English language document, *The Frequently Asked Questions Manual for Linux*.

Example: Surprisal Values

from <http://www.umsl.edu/~fraundorfp/egsurpri.html>

situation	probability $p = 1/2^{\text{\#bits}}$	surprisal $\text{\#bits} = \ln_2[1/p]$
one equals one	1	0 bits
wrong guess on a 4-choice question	3/4	$\ln_2[4/3] \sim 0.415$ bits
correct guess on true-false question	1/2	$\ln_2[2] = 1$ bit
correct guess on a 4-choice question	1/4	$\ln_2[4] = 2$ bits
seven on a pair of dice	$6/6^2 = 1/6$	$\ln_2[6] \sim 2.58$ bits
snake-eyes on a pair of dice	$1/6^2 = 1/36$	$\ln_2[36] \sim 5.17$ bits
random character from the 8-bit ASCII set	1/256	$\ln_2[2^8] = 8$ bits = 1 byte
N heads on a toss of N coins	$1/2^N$	$\ln_2[2^N] = N$ bits
harm from a smallpox vaccination	$\sim 1/1,000,000$	$\sim \ln_2[10^6] \sim 19.9$ bits
win the UK Jackpot lottery	1/13,983,816	~ 23.6 bits
RGB monitor choice of one pixel's color	$1/256^3 \sim 5.9 \times 10^{-8}$	$\ln_2[2^{8 \cdot 3}] = 24$ bits
gamma ray burst mass extinction event TODAY!	$< 1/(10^9 \cdot 365) \sim 2.7 \times 10^{-12}$	hopefully > 38 bits
availability to reset 1 gigabyte of random access memory	$1/2^{8E9} \sim 10^{-2.4E9}$	8×10^9 bits $\sim 7.6 \times 10^{-14}$ J/K
choices for 6×10^{23} Argon atoms in a 24.2L box at 295K	$\sim 1/2^{1.61E25} \sim 10^{-4.8E24}$	$\sim 1.61 \times 10^{25}$ bits ~ 155 J/K
one equals two	0	∞ bits

i	a_i	p_i	$h(p_i)$
1	a	.0575	4.1
2	b	.0128	6.3
3	c	.0263	5.2
4	d	.0285	5.1
5	e	.0913	3.5
6	f	.0173	5.9
7	g	.0133	6.2
8	h	.0313	5.0
9	i	.0599	4.1
10	j	.0006	10.7
11	k	.0084	6.9
12	l	.0335	4.9
13	m	.0235	5.4
14	n	.0596	4.1
15	o	.0689	3.9
16	p	.0192	5.7
17	q	.0008	10.3
18	r	.0508	4.3
19	s	.0567	4.1
20	t	.0706	3.8
21	u	.0334	4.9
22	v	.0069	7.2
23	w	.0119	6.4
24	x	.0073	7.1
25	y	.0164	5.9
26	z	.0007	10.4
27	-	.1928	2.4

$$\sum_i p_i \log_2 \frac{1}{p_i} \quad 4.1$$

Table 2.9. Shannon information contents of the outcomes a–z.

MacKay's Mnemonic

convex

concave

MacKay's Mnemonic

convex



concave

MacKay's Mnemonic

convex



concave



MacKay's Mnemonic

convex

convec-smile



concave



MacKay's Mnemonic

convex

convec-smile

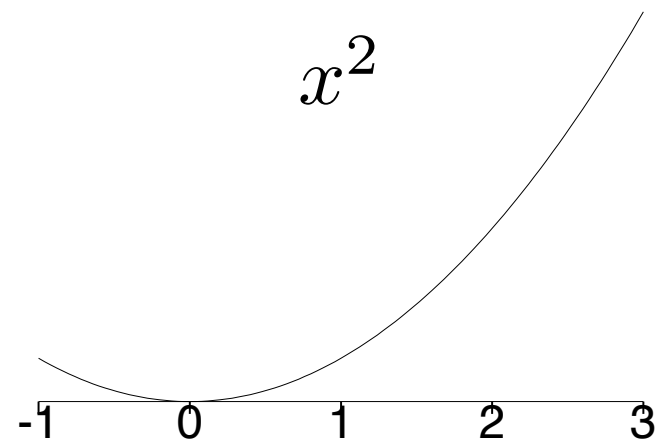


concave

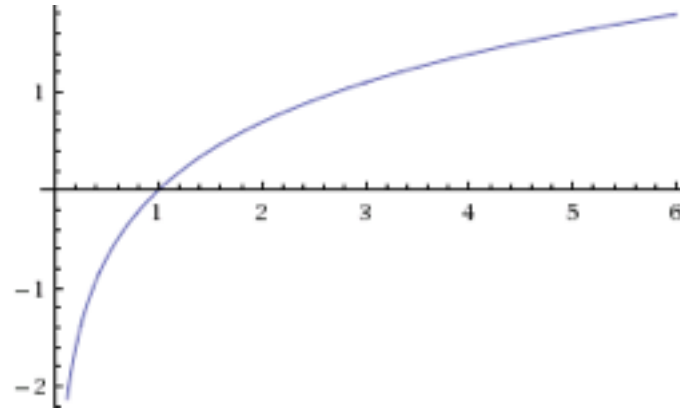
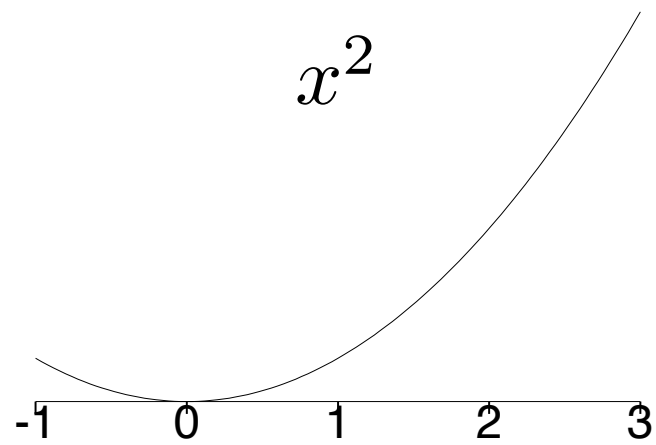
conca-frown



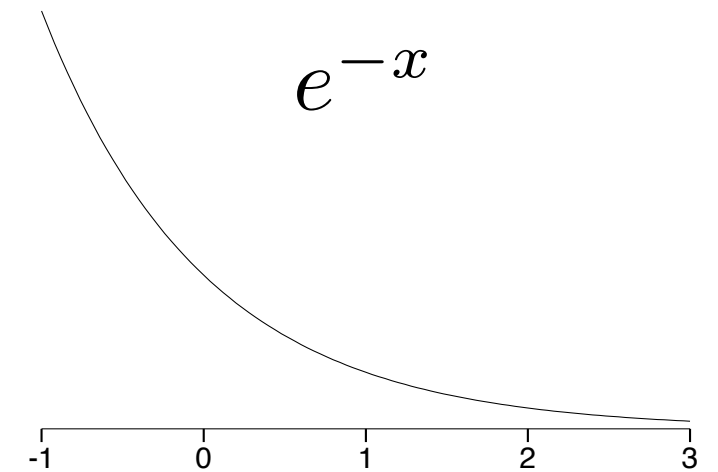
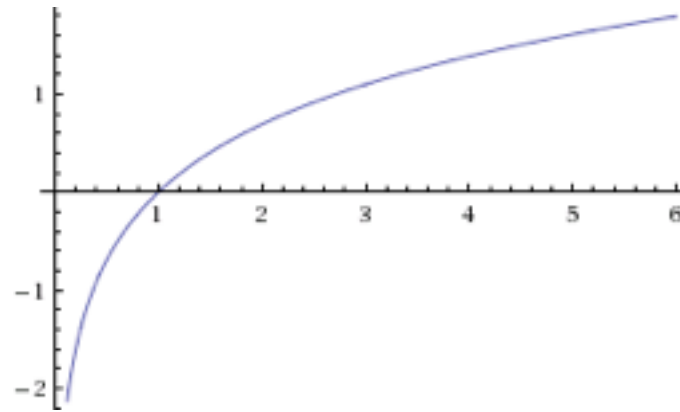
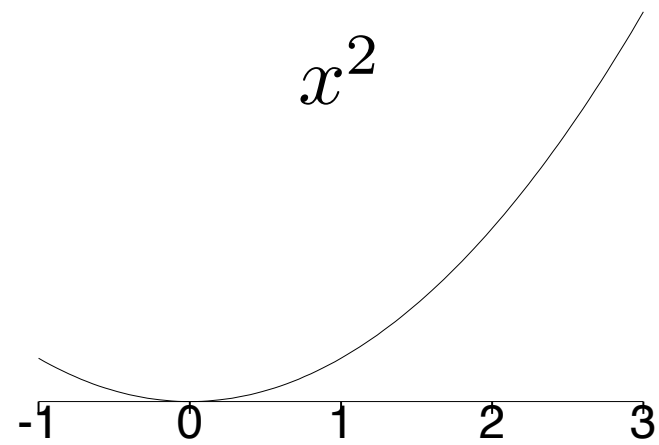
Examples: Convex & Concave Functions



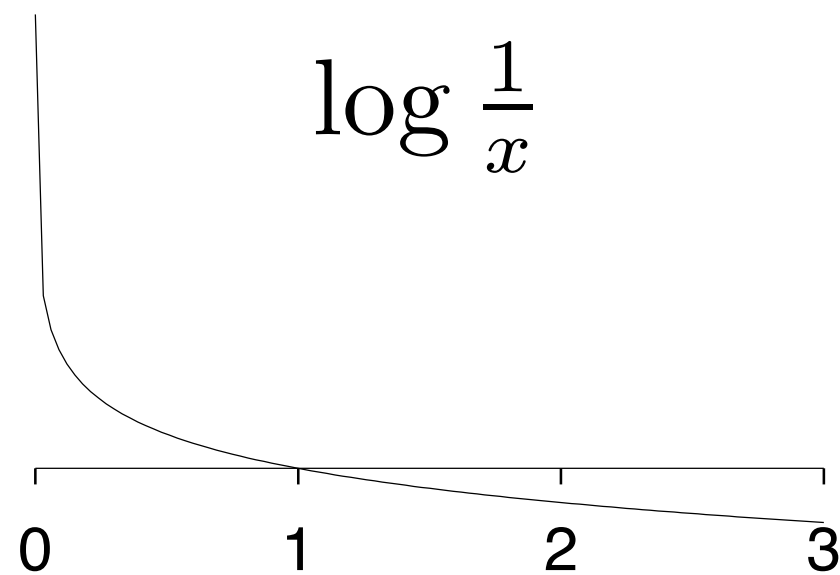
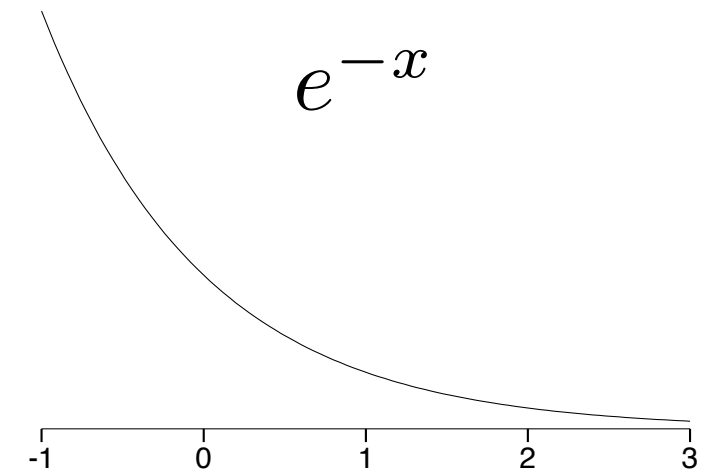
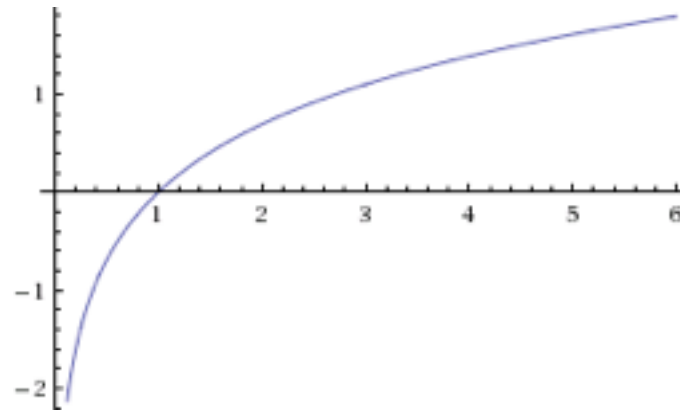
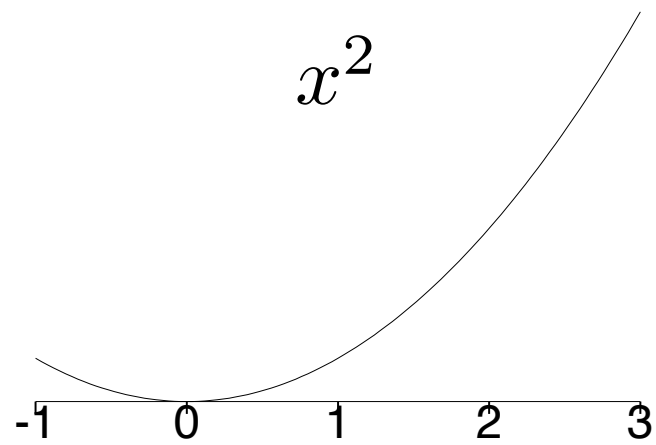
Examples: Convex & Concave Functions



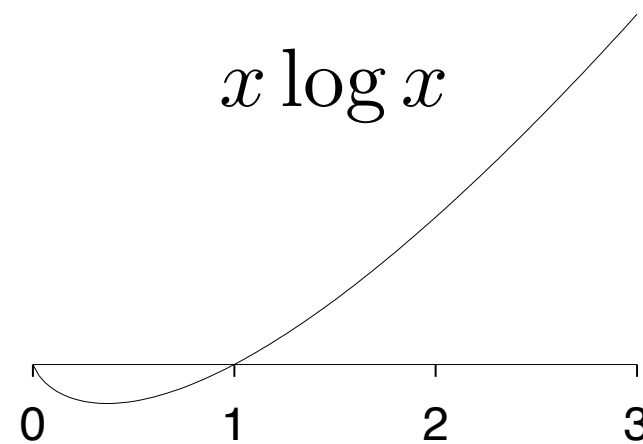
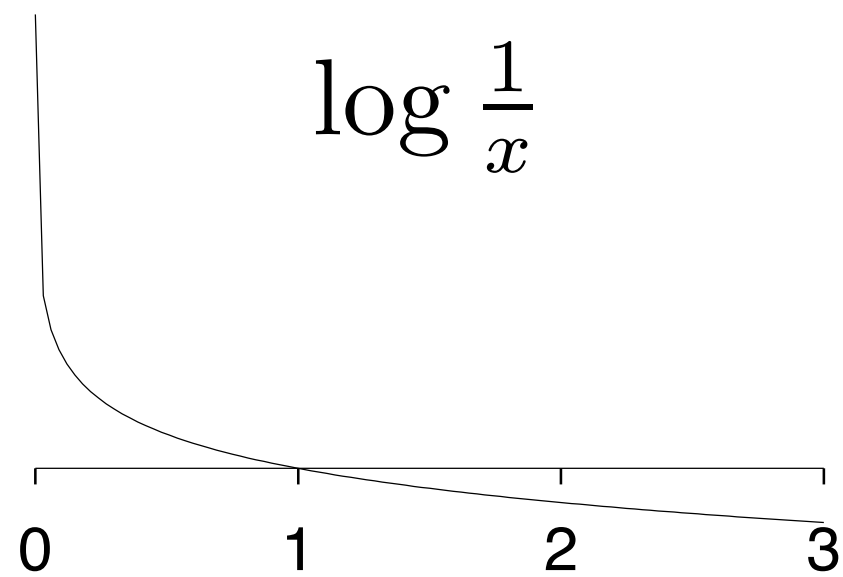
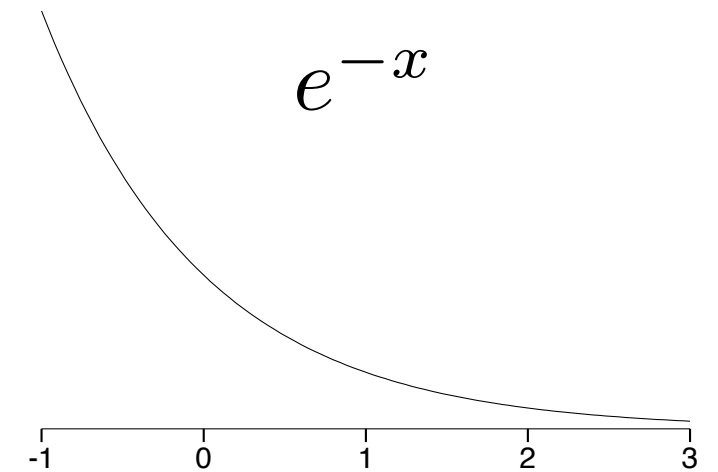
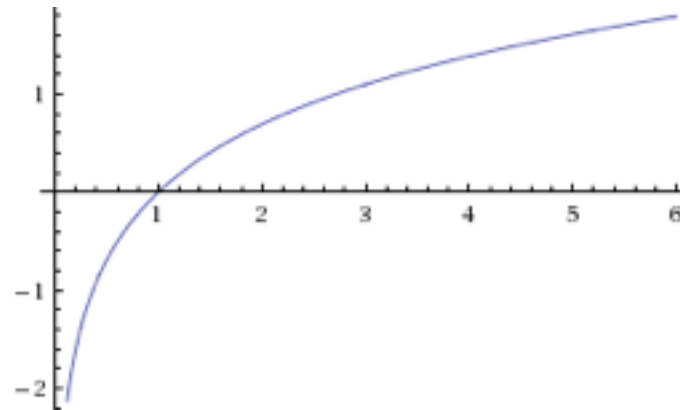
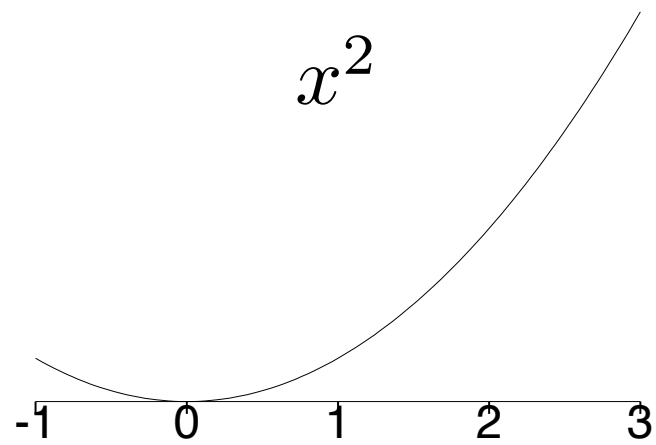
Examples: Convex & Concave Functions



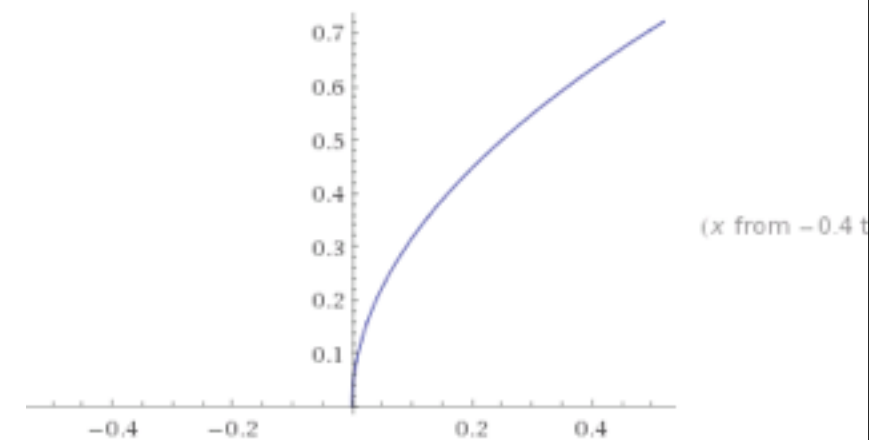
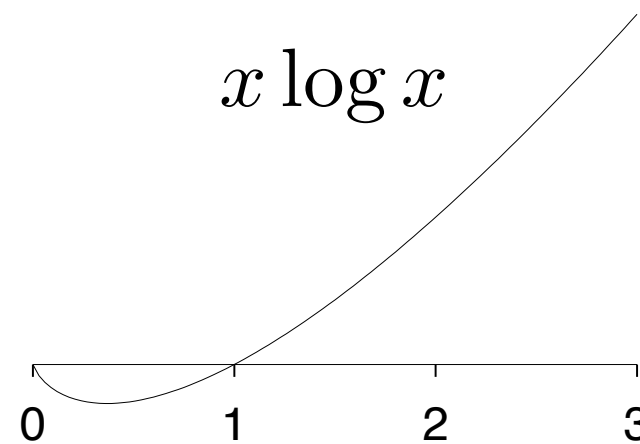
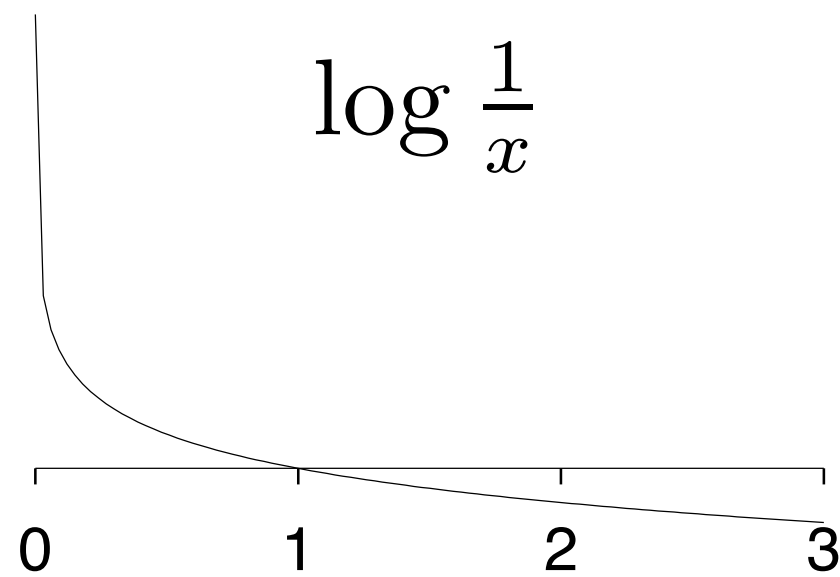
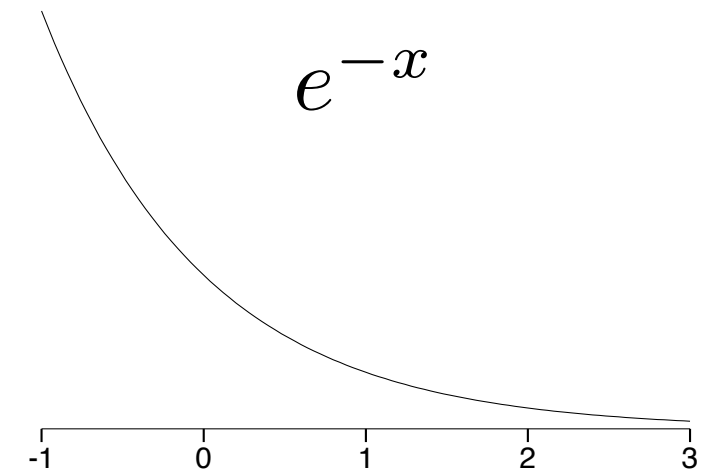
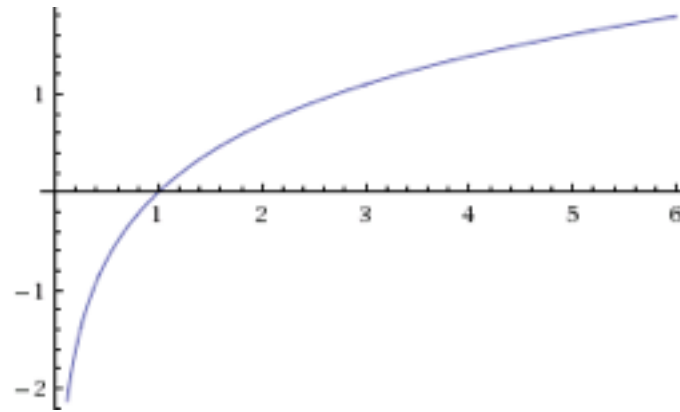
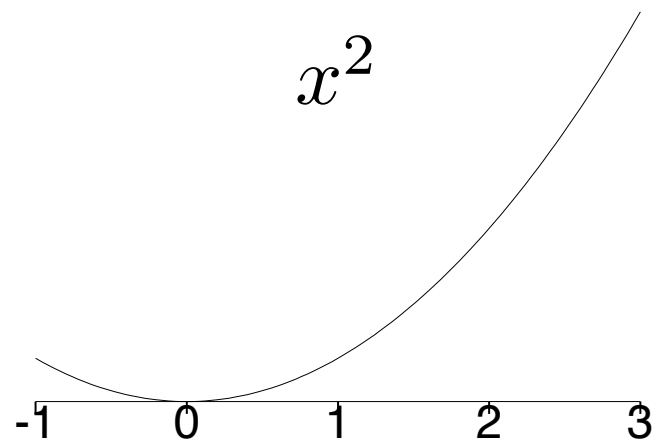
Examples: Convex & Concave Functions



Examples: Convex & Concave Functions



Examples: Convex & Concave Functions



Binary Entropy Function

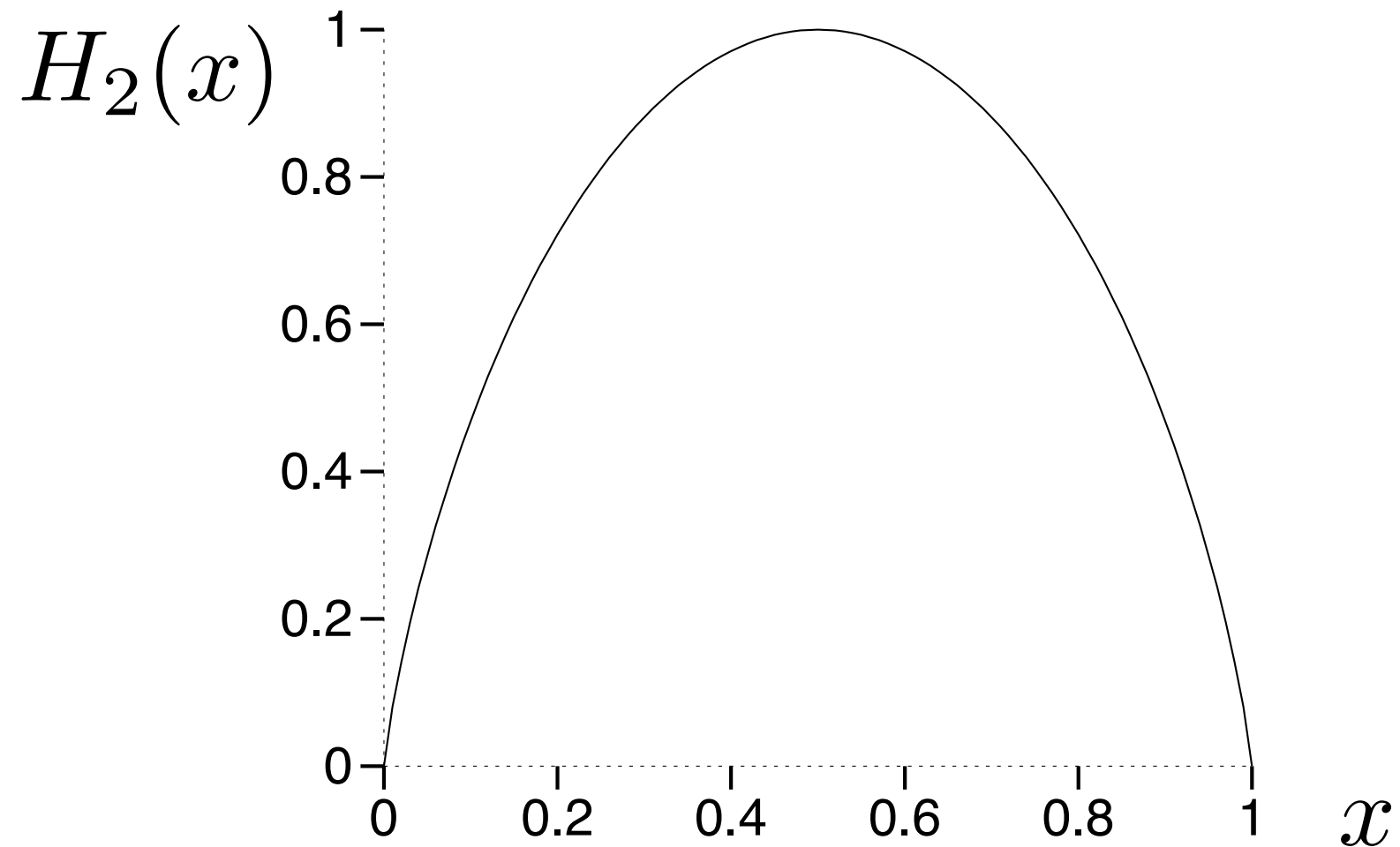
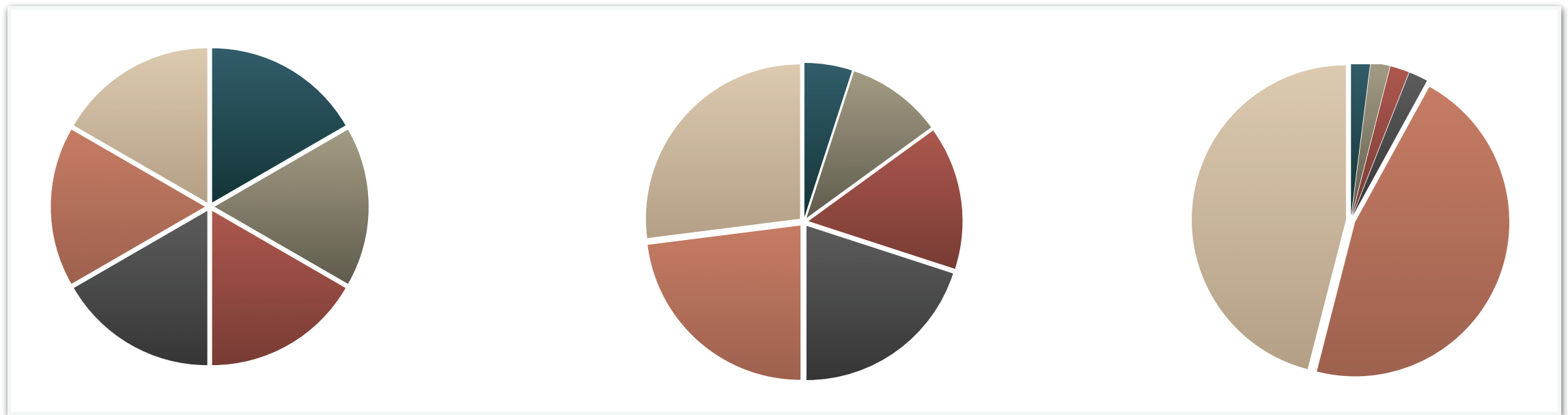


Figure 1.3. The binary entropy function.

Order These in Terms of Entropy



Order These in Terms of Entropy



Mutual Information and Entropy

Theorem: Relationship between mutual information and entropy.

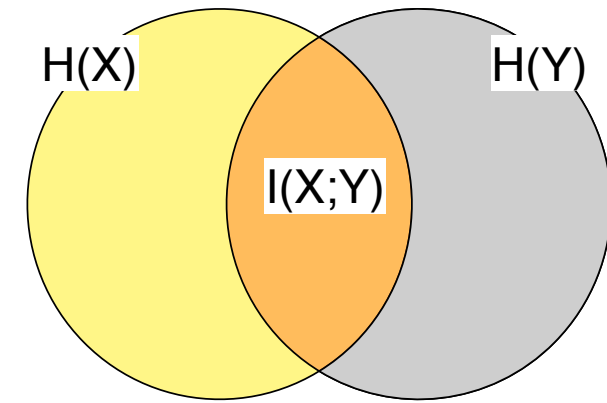
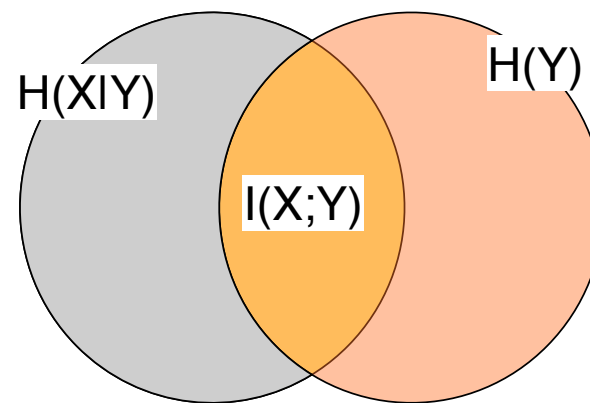
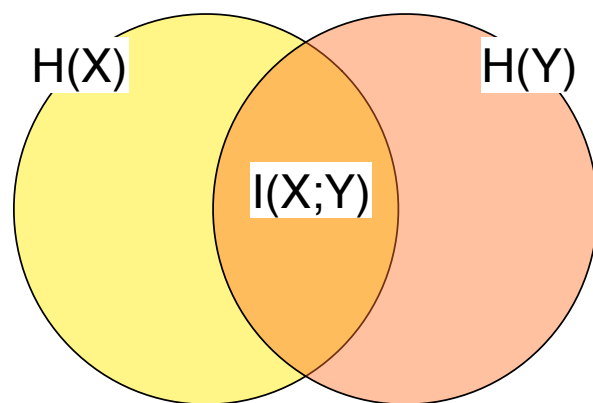
$$I(X; Y) = H(X) - H(X|Y)$$

$$I(X; Y) = H(Y) - H(Y|X)$$

$$I(X; Y) = H(X) + H(Y) - H(X, Y)$$

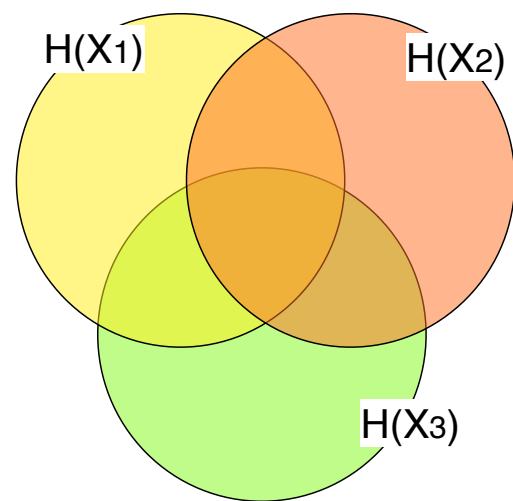
$$I(X; Y) = I(Y; X) \quad (\text{symmetry})$$

$$I(X; X) = H(X) \quad (\text{“self-information”})$$

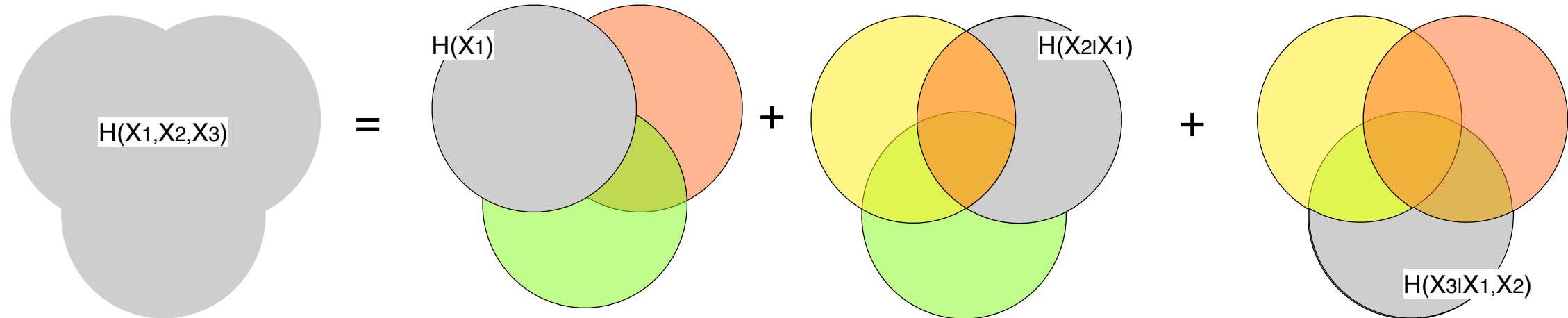


Chain Rule for Entropy

Theorem: (Chain rule for entropy): $(X_1, X_2, \dots, X_n) \sim p(x_1, x_2, \dots, x_n)$



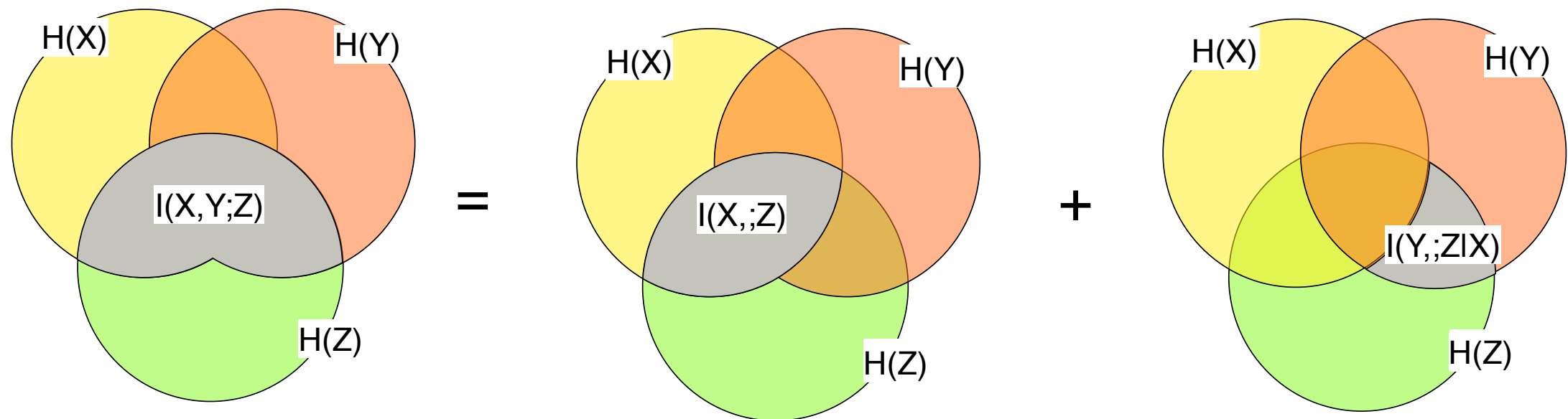
$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i | X_{i-1}, \dots, X_1)$$



Chain Rule for Mutual Information

Theorem: (Chain rule for mutual information)

$$I(X_1, X_2, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y | X_{i-1}, X_{i-2}, \dots, X_1)$$



What are the Grey Regions?

