

Economics of Information

LECTURE 2

Jacopo Staccioli^{†‡}

[†] *Università Cattolica del Sacro Cuore, Milan*

[‡] *Scuola Superiore Sant'Anna, Pisa*

31st March 2020



Information and production

- information is an economic good: it is costly and valuable
- a firm can purchase information, or it can expend resources in research and development to wrest it from nature

BUT

- the way information enters the production function differs from other goods
- the same piece of information can be used over and over again, by the same or a different producer, regardless of the scale of production

Arrow (1996)

- increasing returns can occur for other reasons than information
- but with information, constant returns are impossible



Information and production

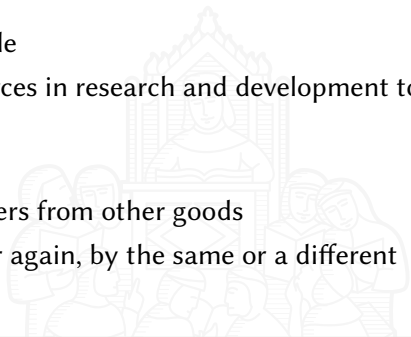
- information is an economic good: it is costly and valuable
- a firm can purchase information, or it can expend resources in research and development to wrest it from nature

BUT

- the way information enters the production function differs from other goods
- the same piece of information can be used over and over again, by the same or a different producer, regardless of the scale of production

Arrow (1996)

- increasing returns can occur for other reasons than information
- but with information, constant returns are impossible



Information and production

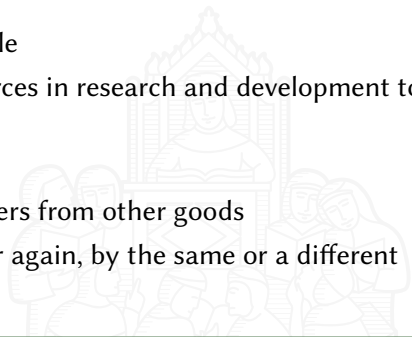
- information is an economic good: it is costly and valuable
- a firm can purchase information, or it can expend resources in research and development to wrest it from nature

BUT

- the way information enters the production function differs from other goods
- the same piece of information can be used over and over again, by the same or a different producer, regardless of the scale of production

Arrow (1996)

- increasing returns can occur for other reasons than information
- but with information, constant returns are impossible



Information and production (cont'd)

information as a choice variable

information is a *signal*, i.e. an observed random variable which may be of no economic interest itself but which is not independent of unobserved variables which affect benefits or costs

- the cost of information is related to its accuracy
- information is often about *rates* (e.g. returns on securities, unit costs...)
- then, the cost is independent of absolute magnitudes

BUT

- the *value* depends on the scale of its utilisation
- e.g. gross income $\propto$ scale of utilisation $\perp$ cost of information
- reasonably, demand for information increases with scale
- then uses on high scale are more accurate than uses on low scale

THEN

- also the expected rate of return increases with scale !



Information and production (cont'd)

information as a choice variable

information is a *signal*, i.e. an observed random variable which may be of no economic interest itself but which is not independent of unobserved variables which affect benefits or costs

- the cost of information is related to its accuracy
- information is often about *rates* (e.g. returns on securities, unit costs...)
- then, the cost is independent of absolute magnitudes

BUT

- the *value* depends on the scale of its utilisation
- e.g. gross income $\propto$ scale of utilisation $\perp$ cost of information
- reasonably, demand for information increases with scale
- then uses on high scale are more accurate than uses on low scale

THEN

- also the expected rate of return increases with scale !



Information and production (cont'd)

information as a choice variable

information is a *signal*, i.e. an observed random variable which may be of no economic interest itself but which is not independent of unobserved variables which affect benefits or costs

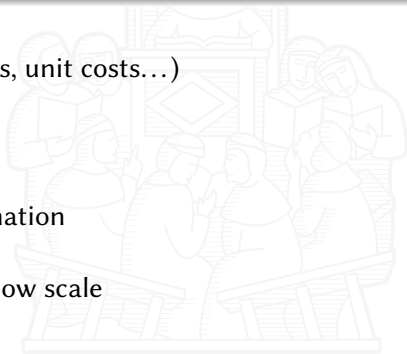
- the cost of information is related to its accuracy
- information is often about *rates* (e.g. returns on securities, unit costs...)
- then, the cost is independent of absolute magnitudes

BUT

- the *value* depends on the scale of its utilisation
- e.g. gross income $\propto$ scale of utilisation $\perp$ cost of information
- reasonably, demand for information increases with scale
- then uses on high scale are more accurate than uses on low scale

THEN

- also the expected rate of return increases with scale !



Information and production (cont'd)

information as a choice variable

information is a *signal*, i.e. an observed random variable which may be of no economic interest itself but which is not independent of unobserved variables which affect benefits or costs

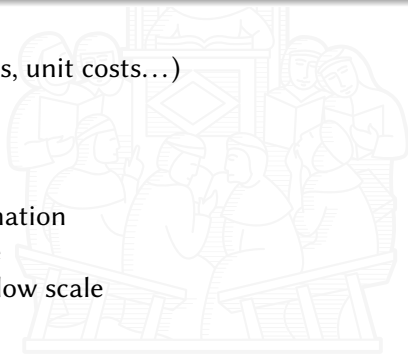
- the cost of information is related to its accuracy
- information is often about *rates* (e.g. returns on securities, unit costs...)
- then, the cost is independent of absolute magnitudes

BUT

- the *value* depends on the scale of its utilisation
- e.g. gross income $\propto$ scale of utilisation $\perp$ cost of information
- reasonably, demand for information increases with scale
- then uses on high scale are more accurate than uses on low scale

THEN

- also the expected rate of return increases with scale !



Preliminary assumptions

- choice of actions $a \in A =$ opportunity set
- payoff $w(a, x)$ depends on state of the world $x \in X \sim$ known *pmf*
- without additional information the problem is

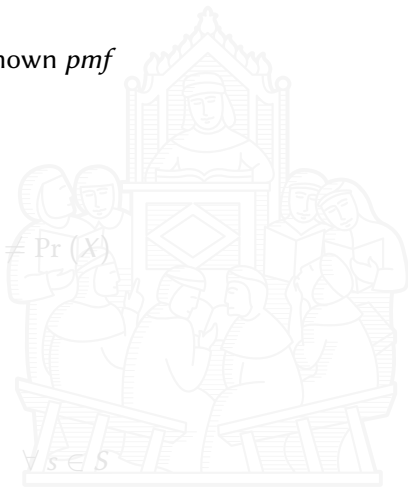
$$\max_a E_X[w(a, X)]$$

- now suppose we can observe *signal* $S \not\perp X \implies \Pr(X|S) \neq \Pr(X)$
- the new problem is

$$\max_a E_{X|S}[w(a, X)]$$

- this defines a *decision function*

$$a(s) = \arg \max_a E_{X|S}[w(a, X)], \quad \forall s \in S$$



Preliminary assumptions

- choice of actions $a \in A =$ opportunity set
- payoff $w(a, x)$ depends on state of the world $x \in X \sim$ known *pmf*
- without additional information the problem is

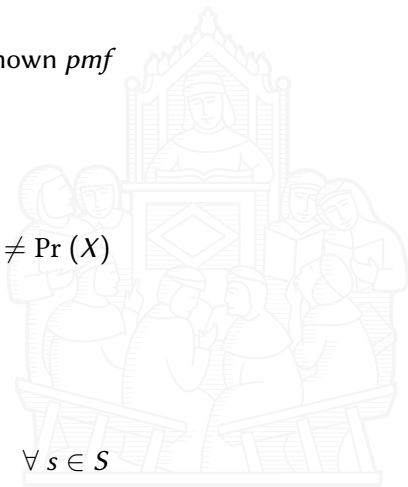
$$\max_a E_X[w(a, X)]$$

- now suppose we can observe *signal* $S \not\perp X \implies \Pr(X|S) \neq \Pr(X)$
- the new problem is

$$\max_a E_{X|S}[w(a, X)]$$

- this defines a *decision function*

$$a(s) = \arg \max_a E_{X|S}[w(a, X)], \quad \forall s \in S$$



Preliminary assumptions (cont'd)

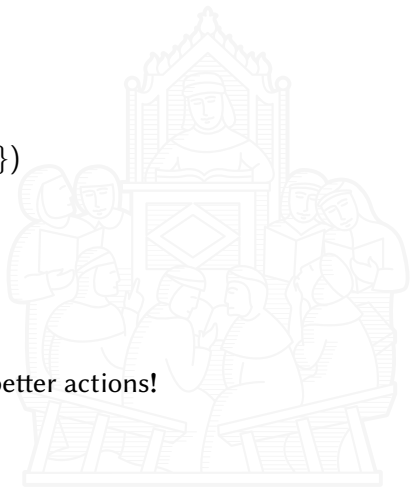
- the signal *costs* and is an *object of choice*

$$w(a, x) \implies w(a, x, \{S\})$$

- for fixed a and x the payoff decreases with cost of signal

$$\frac{\partial w(\cdot)}{\partial C(\{S\})} < 0$$

- but we may still choose a more costly signal if it yields better actions!



Preliminary assumptions (cont'd)

- define income as gross income minus cost of signal

$$Y = v(a, x) - C(\{S\})$$

- define payoff as utility of income

$$w(a, x, \{S\}) = U(Y)$$

- define cost proportional to information content

$$C(\{S\}) = cl(X, S), \quad c > 0$$

- what is the information content of the signal $I(X, S)$?



Measure of information

- the keyword here is **entropy**
- in statistical mechanics, entropy is a property of a thermodynamic system
- expresses the number Ω of different microstate configurations that a system defined by macroscopic variables (volume, pressure and temperature) could assume

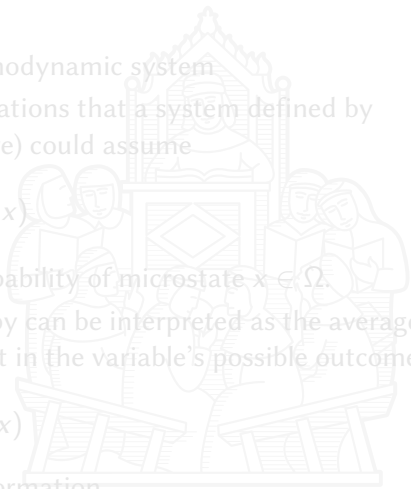
$$S = -k_B \sum p(x) \log p(x)$$

where k_B is the Boltzmann constant and $p(x)$ is the probability of microstate $x \in \Omega$.

- in information theory, given a random variable X , entropy can be interpreted as the average level of 'information', 'surprise', or 'uncertainty' inherent in the variable's possible outcomes

$$H = - \sum p(x) \log_2 p(x)$$

- H is called Shannon entropy or Shannon measure of information



Measure of information

- the keyword here is *entropy*
- in statistical mechanics, entropy is a property of a thermodynamic system
- expresses the number Ω of different microstate configurations that a system defined by macroscopic variables (volume, pressure and temperature) could assume

$$S = -k_B \sum p(x) \log p(x)$$

where k_B is the Boltzmann constant and $p(x)$ is the probability of microstate $x \in \Omega$.

- in information theory, given a random variable X , entropy can be interpreted as the average level of 'information', 'surprise', or 'uncertainty' inherent in the variable's possible outcomes

$$H = - \sum p(x) \log_2 p(x)$$

- H is called Shannon entropy or Shannon measure of information



Measure of information

- the keyword here is *entropy*
- in statistical mechanics, entropy is a property of a thermodynamic system
- expresses the number Ω of different microstate configurations that a system defined by macroscopic variables (volume, pressure and temperature) could assume

$$S = -k_B \sum p(x) \log p(x)$$

where k_B is the Boltzmann constant and $p(x)$ is the probability of microstate $x \in \Omega$.

- in information theory, given a random variable X , entropy can be interpreted as the average level of ‘information’, ‘surprise’, or ‘uncertainty’ inherent in the variable’s possible outcomes

$$H = - \sum p(x) \log_2 p(x)$$

- H is called Shannon entropy or Shannon measure of information



Measure of information (cont'd)

general intuition

- the 'informational value' of a signal depends on the degree to which the content of the message is surprising
- if an event is very probable, it is no surprise (and generally uninteresting) when that event happens as expected; hence transmission of such a message carries very little new information
- however, if an event is unlikely to occur, it is much more informative to learn that the event happened or will happen
- e.g. the knowledge that some particular number will not be the winning number of a lottery provides very little information, because any particular chosen number will almost certainly not win
- however, knowledge that a particular number will win a lottery has high value because it communicates the outcome of a very low probability event



Measure of information (cont'd)

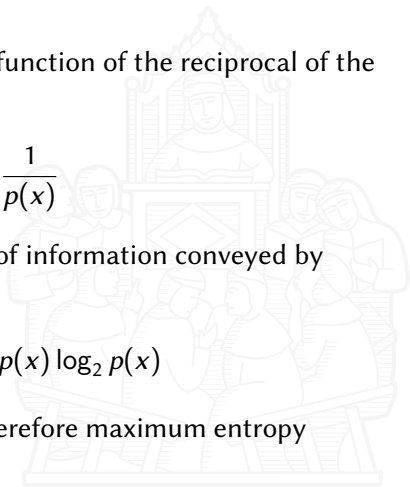
- the information content I of an event x is an increasing function of the reciprocal of the probability $p(x)$ of the event

$$I(x) = -\log_2 p(x) = \log_2 \frac{1}{p(x)}$$

- entropy H measures the expected (i.e. average) amount of information conveyed by identifying the outcome of a random trial

$$H = \mathbb{E}[I] = \sum p(x)I(x) = -\sum p(x) \log_2 p(x)$$

- *uniform* probability yields maximum uncertainty and therefore maximum entropy



Entropy – example I

consider two numbers:

1000000000

2147483647

which one has the highest entropy?



Entropy – example I

consider two numbers:

$$1000000000 = 1 \cdot 10^9$$

2147483647

which one has the highest entropy?



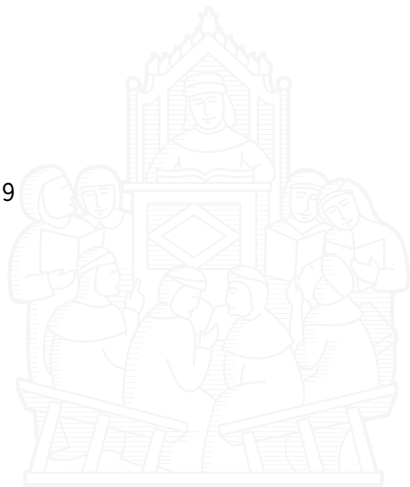
Entropy – example I

consider two numbers:

$$1000000000 = 1 \cdot 10^9 = 1E9$$

2147483647

which one has the highest entropy?



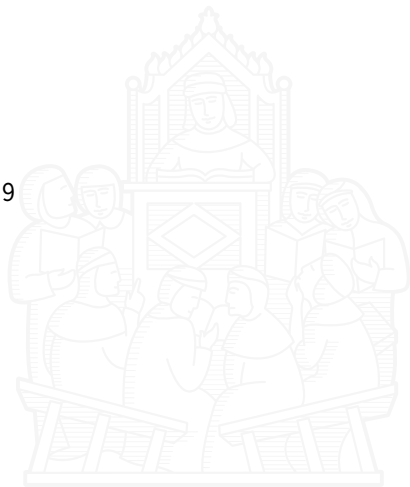
Entropy – example I

consider two numbers:

$$1000000000 = 1 \cdot 10^9 = 1\text{E}9$$

$$2147483647 = 2^{31} - 1$$

which one has the highest entropy?



Entropy – example I

consider two numbers:

$$1000000000 = 1 \cdot 10^9 = 1\text{E}9$$

$$2147483647 = 2^{31} - 1 = \mathcal{M}_{31}$$

which one has the highest entropy?



Entropy – example I

consider two numbers:

$$1000000000 = 1 \cdot 10^9 = 1\text{E}9$$

$$2147483647 = 2^{31} - 1 = \mathcal{M}_{31} = \mathcal{M}_{\mathcal{M}_5}$$

which one has the highest entropy?



Entropy – example I

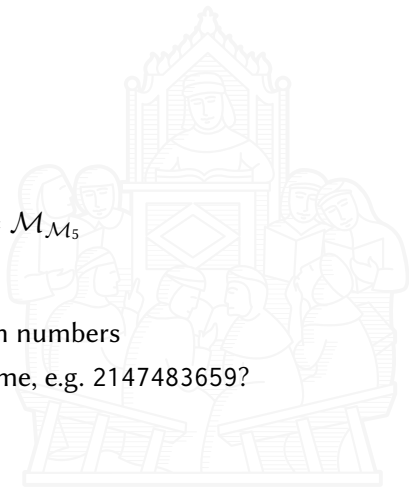
consider two numbers:

$$1000000000 = 1 \cdot 10^9 = 1\text{E}9$$

$$2147483647 = 2^{31} - 1 = \mathcal{M}_{31} = \mathcal{M}_{\mathcal{M}_5}$$

which one has the highest entropy?

- in this case we found a way to (efficiently) *compress* both numbers
- but what if the second number was a non-Mersenne prime, e.g. 2147483659?



Entropy – example II

consider two numbers:

$1.00000\dots$ or $0.99999\dots$

$3.14159\dots$

which one has the highest entropy?



Entropy – example II

consider two numbers:

$1.00000 \dots$ or $0.99999 \dots = 1$

$3.14159 \dots$

which one has the highest entropy?



Entropy – example II

consider two numbers:

$$1.00000 \dots \text{ or } 0.99999 \dots = 1$$

$$3.14159 \dots = \pi$$

which one has the highest entropy?



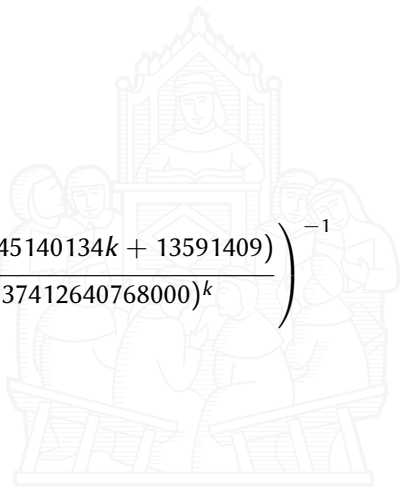
Entropy – example II

consider two numbers:

$$1.00000\dots \text{ or } 0.99999\dots = 1$$

$$3.14159\dots = \pi = 426880\sqrt{10005} \cdot \left(\sum_{k=0}^{\infty} \frac{\frac{(6k)!}{(3k)!(k!)^3} \cdot (545140134k + 13591409)}{(-262537412640768000)^k} \right)^{-1}$$

which one has the highest entropy?



Entropy – example II

consider two numbers:

$$1.00000\dots \text{ or } 0.99999\dots = 1$$

$$3.14159\dots = \pi = 426880\sqrt{10005} \cdot \left(\sum_{k=0}^{\infty} \frac{\frac{(6k)!}{(3k)!(k!)^3} \cdot (545140134k + 13591409)}{(-262537412640768000)^k} \right)^{-1}$$

which one has the highest entropy?

Fact

all finite subsequences belong to the decimal trail of π almost surely



Entropy – example III

consider two strings:

‘X CXN STXLL RXXD THXS’

‘C U L8R’

which one has the highest entropy?



Entropy – example III

consider two strings:

‘X CXN STXLL RXXD THXS’ $\Rightarrow$ ‘I CAN STILL READ THIS’

‘C U L8R’

which one has the highest entropy?



Entropy – example III

consider two strings:

‘X CXN STXLL RXXD THXS’ $\Rightarrow$ ‘I CAN STILL READ THIS’

‘C U L8R’ $\Rightarrow$ ‘SEE YOU LATER’

which one has the highest entropy?



Entropy – example III

consider two strings:

‘X CXN STXLL RXXD THXS’ $\implies$ ‘I CAN STILL READ THIS’

‘C U L8R’ $\implies$ ‘SEE YOU LATER’

which one has the highest entropy?

- English text, treated as a string of characters, has fairly low entropy, i.e. is fairly predictable
- e.g. ‘e’ will be far more common than ‘z’, the combination ‘qu’ will be much more common than any other combination with a ‘q’ in it...
- English text has between 0.6 and 1.3 bits of entropy per character of the message
- less than the theoretical limit of a uniform distribution – $\sum_1^{26} \frac{1}{26} \log_2 \frac{1}{26} = \underbrace{\log_2 26}_{\text{Hartley formula}} \approx 4.7 \text{ bits}$



Preliminary assumptions (cont'd)

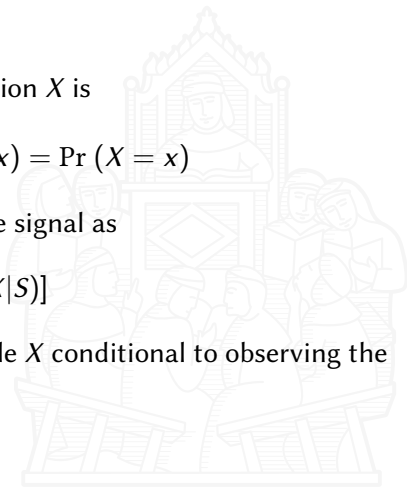
- Shannon measure of information of probability distribution X is

$$H(X) = - \sum p(x) \log_2 p(x), \quad p(x) = \Pr(X = x)$$

- therefore we can measure the information content of the signal as

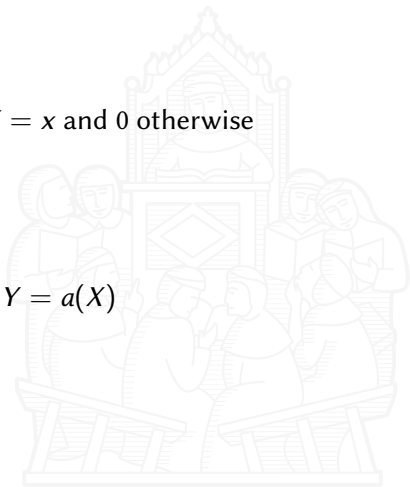
$$I(X, S) = H(X) - E_S[H(X|S)]$$

- i.e. we subtract the (expected) entropy of random variable X conditional to observing the signal S from the unconditional entropy of X



Model assumptions

- let there be unit contingent securities x which pay 1 if $X = x$ and 0 otherwise
- let $a(x)$ be the amount invested in security x
- let A be the initial wealth of the individual
- a portfolio is such that $\sum a(x) = A$
- when the actual state is realised, the individual receives $Y = a(X)$
- suppose utility is logarithmic $U(Y) = \log_2(Y)$



Use of information in portfolio choice I

- in the *absence of information* the problem is

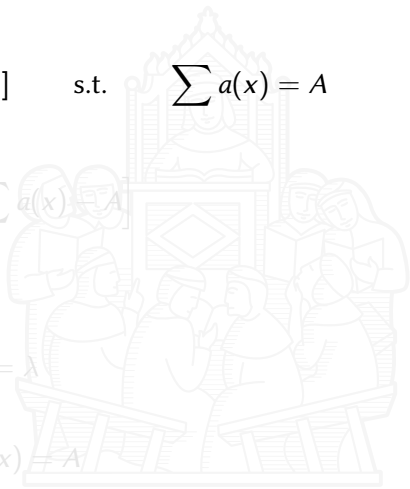
$$\max_a E_X[U(Y)] \implies \max_a \sum p(x) \log_2[a(x)] \quad \text{s.t.} \quad \sum a(x) = A$$

$$\mathcal{L} = \sum p(x) \log_2[a(x)] - \lambda \left[\sum a(x) - A \right]$$

- FOCs:

$$\frac{\partial \mathcal{L}}{\partial a(x)} = 0 \implies \frac{p(x)}{a(x)} = \lambda$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \implies \sum a(x) = A$$



Use of information in portfolio choice I

- in the *absence of information* the problem is

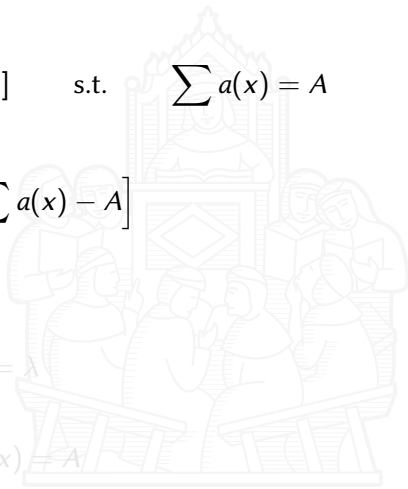
$$\max_a E_X [U(Y)] \implies \max_a \sum p(x) \log_2 [a(x)] \quad \text{s.t.} \quad \sum a(x) = A$$

$$\mathcal{L} = \sum p(x) \log_2 [a(x)] - \lambda \left[\sum a(x) - A \right]$$

- FOCs:

$$\frac{\partial \mathcal{L}}{\partial a(x)} = 0 \implies \frac{p(x)}{a(x)} = \lambda$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \implies \sum a(x) = A$$



Use of information in portfolio choice I

- in the *absence of information* the problem is

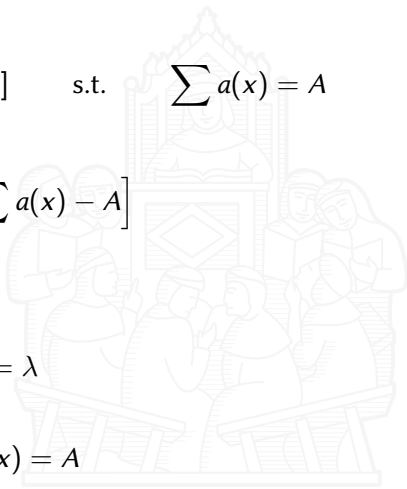
$$\max_a E_X [U(Y)] \implies \max_a \sum p(x) \log_2 [a(x)] \quad \text{s.t.} \quad \sum a(x) = A$$

$$\mathcal{L} = \sum p(x) \log_2 [a(x)] - \lambda \left[\sum a(x) - A \right]$$

- FOCs:

$$\frac{\partial \mathcal{L}}{\partial a(x)} = 0 \implies \frac{p(x)}{a(x)} = \lambda$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \implies \sum a(x) = A$$



Use of information in portfolio choice I (cont'd)

- FOCs:

$$\frac{p(x)}{a(x)} = \lambda \qquad \sum a(x) = A$$

- from 1st FOC

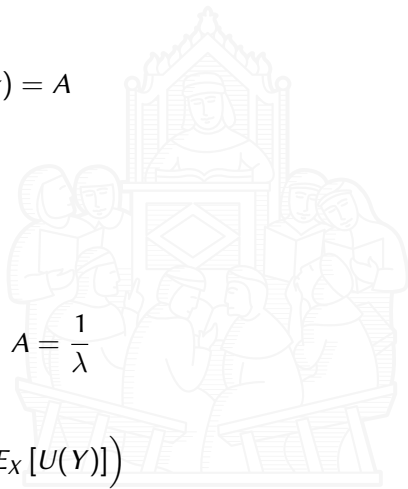
$$a(x) = \frac{p(x)}{\lambda}$$

- take sum and substitute in 2nd FOC

$$\sum a(x) = \frac{\sum p(x)}{\lambda} \implies A = \frac{1}{\lambda}$$

- substitute back in 1st FOC

$$a(x) = Ap(x) \qquad \left(= \arg \max_a E_X [U(Y)] \right)$$



Use of information in portfolio choice I (cont'd)

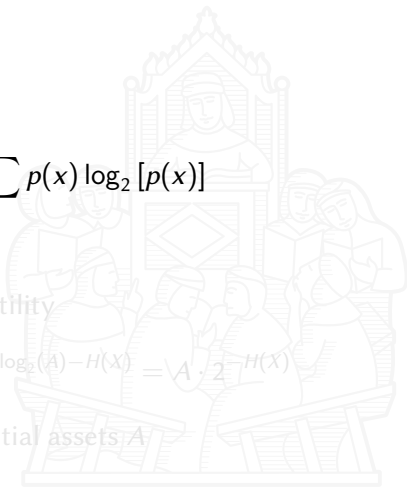
- *optimal expected utility* is

$$\begin{aligned}\max_a E_X [U(Y)] &= \sum p(x) \log_2 [Ap(x)] \\ &= \log_2(A) \sum p(x) + \sum p(x) \log_2 [p(x)] \\ &= \log_2(A) - H(X)\end{aligned}$$

- *certainty equivalent*: W_c such that $\log_2(W_c) = \text{optimal utility}$

$$\log_2(W_c) = \log_2(A) - H(X) \implies W_c = 2^{\log_2(A) - H(X)} = A \cdot 2^{-H(X)}$$

- certainty equivalent of final income is *proportional* to initial assets A
- *constant* returns to scale



Use of information in portfolio choice I (cont'd)

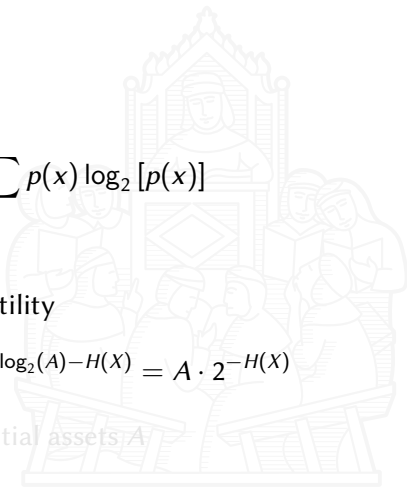
- *optimal expected utility* is

$$\begin{aligned}\max_a E_X [U(Y)] &= \sum p(x) \log_2 [Ap(x)] \\ &= \log_2(A) \sum p(x) + \sum p(x) \log_2 [p(x)] \\ &= \log_2(A) - H(X)\end{aligned}$$

- *certainty equivalent*: W_c such that $\log_2(W_c) = \text{optimal utility}$

$$\log_2(W_c) = \log_2(A) - H(X) \implies W_c = 2^{\log_2(A) - H(X)} = A \cdot 2^{-H(X)}$$

- certainty equivalent of final income is *proportional* to initial assets A
- *constant* returns to scale



Use of information in portfolio choice I (cont'd)

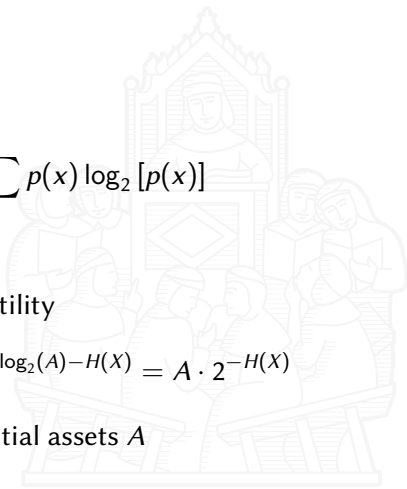
- *optimal expected utility* is

$$\begin{aligned}\max_a E_X [U(Y)] &= \sum p(x) \log_2 [Ap(x)] \\ &= \log_2(A) \sum p(x) + \sum p(x) \log_2 [p(x)] \\ &= \log_2(A) - H(X)\end{aligned}$$

- *certainty equivalent*: W_c such that $\log_2(W_c) = \text{optimal utility}$

$$\log_2(W_c) = \log_2(A) - H(X) \implies W_c = 2^{\log_2(A) - H(X)} = A \cdot 2^{-H(X)}$$

- certainty equivalent of final income is *proportional* to initial assets A
- *constant* returns to scale



Use of information in portfolio choice I (cont'd)

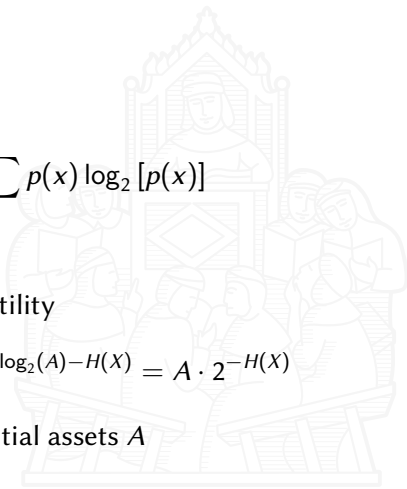
- *optimal expected utility* is

$$\begin{aligned}\max_a E_X [U(Y)] &= \sum p(x) \log_2 [Ap(x)] \\ &= \log_2(A) \sum p(x) + \sum p(x) \log_2 [p(x)] \\ &= \log_2(A) - H(X)\end{aligned}$$

- *certainty equivalent*: W_c such that $\log_2(W_c) = \text{optimal utility}$

$$\log_2(W_c) = \log_2(A) - H(X) \implies W_c = 2^{\log_2(A) - H(X)} = A \cdot 2^{-H(X)}$$

- certainty equivalent of final income is *proportional* to initial assets A
- *constant* returns to scale



Use of information in portfolio choice II

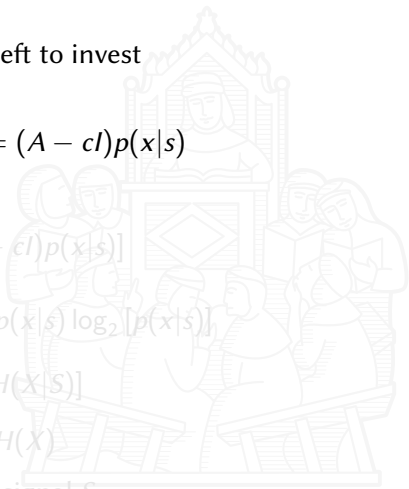
- now let's introduce the *costly signal*
- if S has mutual information I , the individual has $A - cI$ left to invest
- we can rephrase the optimal condition

$$a(x) = Ap(x) \quad \Rightarrow \quad a(x, s) = (A - cI)p(x|s)$$

- new optimal expected utility is

$$\begin{aligned} \max_a E_{X|S} [U(Y)] &= \sum p(x|s) \log_2 [(A - cI)p(x|s)] \\ &= \log_2 (A - cI) + \sum p(x|s) \log_2 [p(x|s)] \\ &= \log_2 (A - cI) - E_S [H(X|S)] \\ &= \log_2 (A - cI) + I - H(X) \end{aligned}$$

- this is the payoff of the optimal investment policy given signal S



Use of information in portfolio choice II

- now let's introduce the *costly signal*
- if S has mutual information I , the individual has $A - cI$ left to invest
- we can rephrase the optimal condition

$$a(x) = Ap(x) \quad \implies \quad a(x, s) = (A - cI)p(x|s)$$

- new optimal expected utility is

$$\begin{aligned} \max_a E_{X|S} [U(Y)] &= \sum p(x|s) \log_2 [(A - cI)p(x|s)] \\ &= \log_2 (A - cI) + \sum p(x|s) \log_2 [p(x|s)] \\ &= \log_2 (A - cI) - E_S [H(X|S)] \\ &= \log_2 (A - cI) + I - H(X) \end{aligned}$$

- this is the payoff of the optimal investment policy given signal S



Use of information in portfolio choice II (cont'd)

- now we want to choose the *optimal signal* (i.e. the amount of information I)

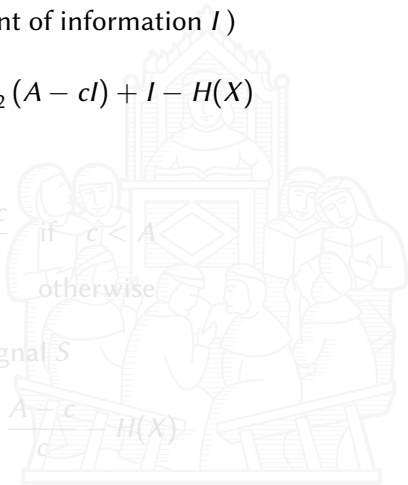
$$\arg \max_I \left[\max_a E_{X|S} [U(Y)] \right] = \arg \max_I \log_2 (A - cI) + I - H(X)$$

- FOC:

$$\frac{-c}{A - cI} + 1 = 0 \quad \Rightarrow \quad I = \begin{cases} \frac{A - c}{c} & \text{if } c < A \\ 0 & \text{otherwise} \end{cases}$$

- payoff of the optimal investment policy given optimal signal S

$$\max_I \left[\max_a E_{X|S} [U(Y)] \right] = \log_2(c) + \frac{A - c}{c} - H(X)$$



Use of information in portfolio choice II (cont'd)

- now we want to choose the *optimal signal* (i.e. the amount of information I)

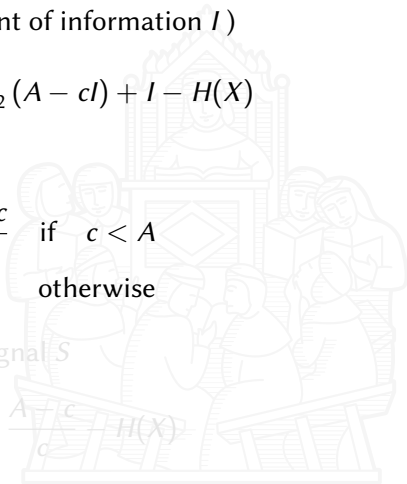
$$\arg \max_I \left[\max_a E_{X|S} [U(Y)] \right] = \arg \max_I \log_2 (A - cI) + I - H(X)$$

- FOC:

$$\frac{-c}{A - cI} + 1 = 0 \quad \Rightarrow \quad I = \begin{cases} \frac{A - c}{c} & \text{if } c < A \\ 0 & \text{otherwise} \end{cases}$$

- payoff of the optimal investment policy given optimal signal S

$$\max_I \left[\max_a E_{X|S} [U(Y)] \right] = \log_2(c) + \frac{A - c}{c} - H(X)$$



Use of information in portfolio choice II (cont'd)

- now we want to choose the *optimal signal* (i.e. the amount of information I)

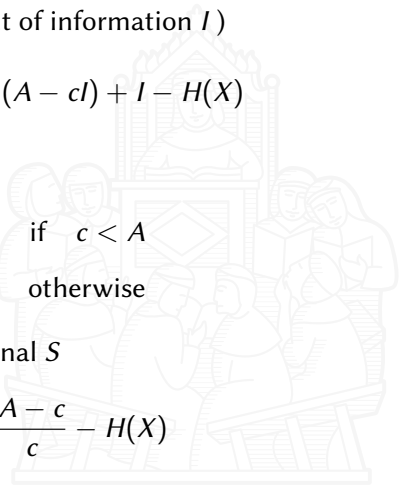
$$\arg \max_I \left[\max_a E_{X|S} [U(Y)] \right] = \arg \max_I \log_2 (A - cI) + I - H(X)$$

- FOC:

$$\frac{-c}{A - cI} + 1 = 0 \quad \Rightarrow \quad I = \begin{cases} \frac{A - c}{c} & \text{if } c < A \\ 0 & \text{otherwise} \end{cases}$$

- payoff of the optimal investment policy given optimal signal S

$$\max_I \left[\max_a E_{X|S} [U(Y)] \right] = \log_2(c) + \frac{A - c}{c} - H(X)$$

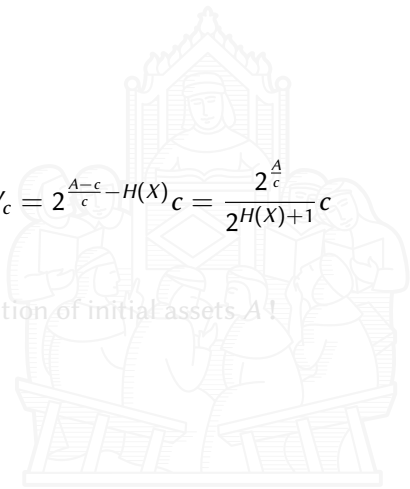


Use of information in portfolio choice II (cont'd)

- certainty equivalent

$$\log_2(W_c) = \log_2(c) + \frac{A-c}{c} - H(X) \implies W_c = 2^{\frac{A-c}{c} - H(X)} c = \frac{2^{\frac{A}{c}}}{2^{H(X)+1}} c$$

- certainty equivalent of final income is an *increasing* function of initial assets A !
- *increasing* returns to scale

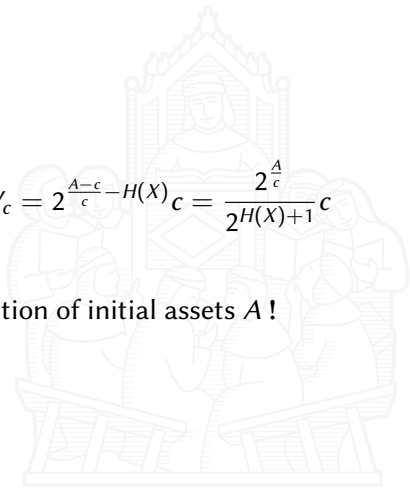


Use of information in portfolio choice II (cont'd)

- certainty equivalent

$$\log_2(W_c) = \log_2(c) + \frac{A-c}{c} - H(X) \implies W_c = 2^{\frac{A-c}{c} - H(X)} c = \frac{2^{\frac{A}{c}}}{2^{H(X)+1}} c$$

- certainty equivalent of final income is an *increasing* function of initial assets A !
- *increasing* returns to scale



Concluding remarks

“the presence of increasing returns leads to the conclusion that the opportunity to acquire information about investments leads to a situation in which the rate of return to investment rises with the amount invested” [ibidem, p.125]

- this mechanism can explain part of the inequality of income
- departures from equality are self-reinforcing !

Example: process innovation

- extensive search for lower unit costs is economically justified only if the new process is applied on a large scale
- the larger the scale, the lower the unit cost because of more R&D
- justifies endogenous theories of economic growth



Concluding remarks

“the presence of increasing returns leads to the conclusion that the opportunity to acquire information about investments leads to a situation in which the rate of return to investment rises with the amount invested” [ibidem, p.125]

- this mechanism can explain part of the inequality of income
- departures from equality are self-reinforcing !

Example: process innovation

- extensive search for lower unit costs is economically justified only if the new process is applied on a large scale
- the larger the scale, the lower the unit cost because of more R&D
- justifies endogenous theories of economic growth



Concluding remarks

“the presence of increasing returns leads to the conclusion that the opportunity to acquire information about investments leads to a situation in which the rate of return to investment rises with the amount invested” [ibidem, p.125]

- this mechanism can explain part of the inequality of income
- departures from equality are self-reinforcing !

Example: process innovation

- extensive search for lower unit costs is economically justified only if the new process is applied on a large scale
- the larger the scale, the lower the unit cost because of more R&D
- justifies endogenous theories of economic growth

