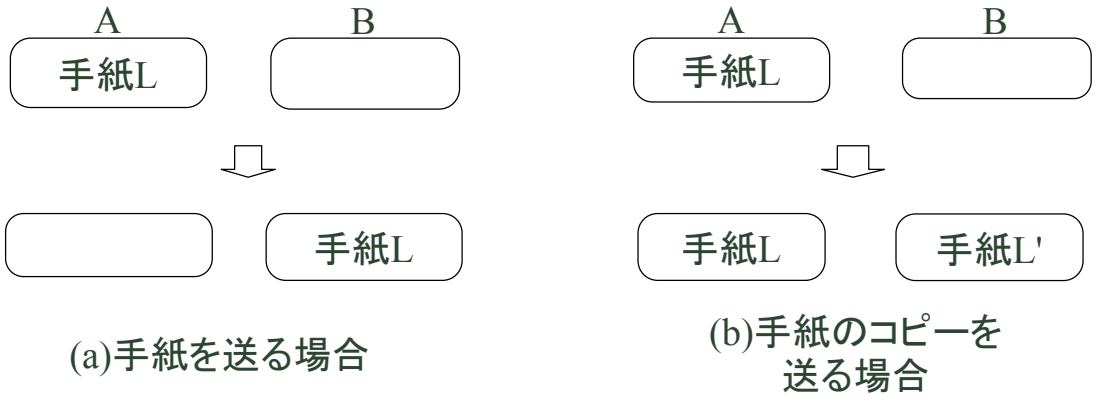
第3章
情報の伝達と通信

3.1 情報の伝達と情報量

- **情報の伝達 (3.1.1) とは:**
  受取側の状態の変化が本質
  - 様々な伝え方で同じ「情報」(メッセージ)が伝わる
    - 「手紙」を送る / 「手紙のコピー」を送る
    - 電子メールを送る
  - 手紙の物理的な移動は本質でない
- **情報量 (3.1.2): 情報を受け取った効果を測る**
  - メッセージ:「今回は日本史から出題する」
  - このメッセージの効果を量で表現すると?

情報の伝達と伝達手段



(a)手紙を送る場合

(b)手紙のコピーを送る場合

- **Bさんが受け取る情報はどちらの場合も同じ**
- **物理的な手紙の移動は無関係**
  - Aさんの手元から手紙が消えることは本質でない

情報を受け取った効果とは?

- **直感的な説明**
  - 情報を受け取った場合
    - 自分に影響がある、これまで知らなかった事実を知った
    - なんらかの判断の材料にできる事実を知った
  - 情報を受け取ったと言い難い場合
    - 関心のない手紙を受け取った (e.g. 迷惑メイル)
- **情報を受け取る効果は、受け取る人の「状態」と関係がある**
- **メッセージの効果を「情報量」として表現したい**

情報の量は何で測るべきか

- **文字数では直感に反する**
  - 1000文字のダイレクトメール vs. 100文字の「情報」試験問題
  - 明日の東京の天気 vs. 明日のブエノスアイレスの天気
    → 受け手にとっての役立ち度が違う
  - 前回のルーレットの目 vs.次回のルーレットの目
    → 受け手がすでに知っているかどうかが違う
  - サイコロの次の目 vs. ルーレットの次の目
    → ルーレットの方が価値が高い?
- **直感に合う情報の量のあるべき性質**
  - 受け手が知っているかどうかを反映したもの
  - 受け手の役立ち度/価値を反映したもの

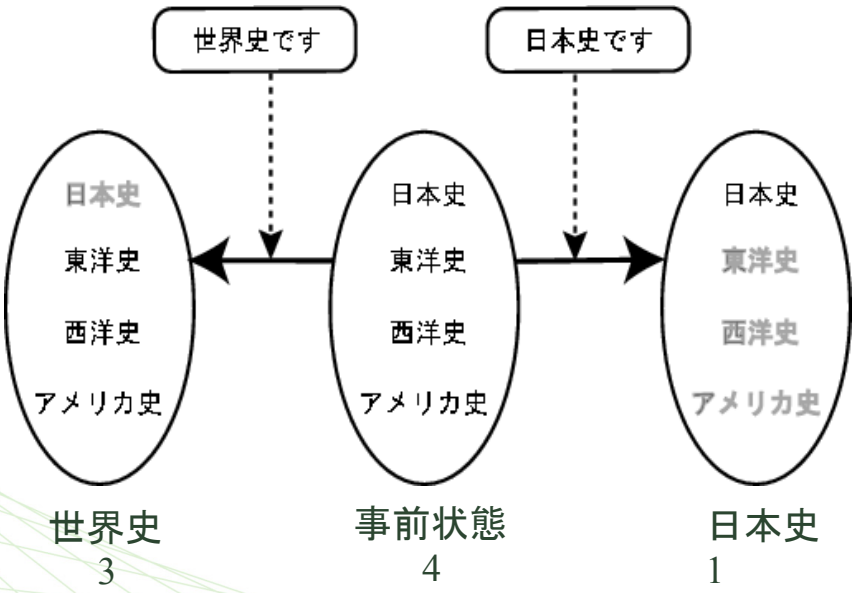
情報の価値 = 場合の数の減少量

- **おおまかな定義: 情報の価値は、受け手の選択肢をどれだけ減らすかで測る**
  - 1000文字のダイレクトメール vs. 100文字の「情報」試験問題
    → 選択肢: 教科書のどのページを勉強するか
  - 明日の東京の天気 vs. 明日のブエノスアイレスの天気
    → 選択肢: どの服を着るか
  - 前回のルーレットの目 vs.次回のルーレットの目
    → すでに知っている情報は選択肢を減らさない
  - サイコロの次の目 vs. ルーレットの次の目
    → 選択肢の数: 6→1 vs. 100→1
- **以降、「選択肢の数」で単純化**

試験に関する情報の価値

- 科目「歴史」の試験
  - 日本史、東洋史、西洋史、アメリカ史のどれか一つが出題
  - 事前にはどれが出題されるかは分からない
- メッセージ:「今回は日本史から出題する」
- 状況の変化
  - 事前: 日本史からアメリカ史の4種類全部の試験勉強が必要である
  - 事後: 日本史の勉強だけですむ

メッセージによる場合の数の変化



場合の数に基づく情報量(の候補)

- **案1: 差**
  - 定義: 事前の場合の数 − 事後の場合の数
  - 問題点: 100 → 97 の場合と 4 → 1 が同じ価値?
- **案2: 商**
  - 定義: 事前の場合の数 / 事後の場合の数
  - 問題点: 情報量の加法性(後述)を満たさない
- **案3: 商の対数**
  - 定義: log(事前の場合の数 / 事後の場合の数)

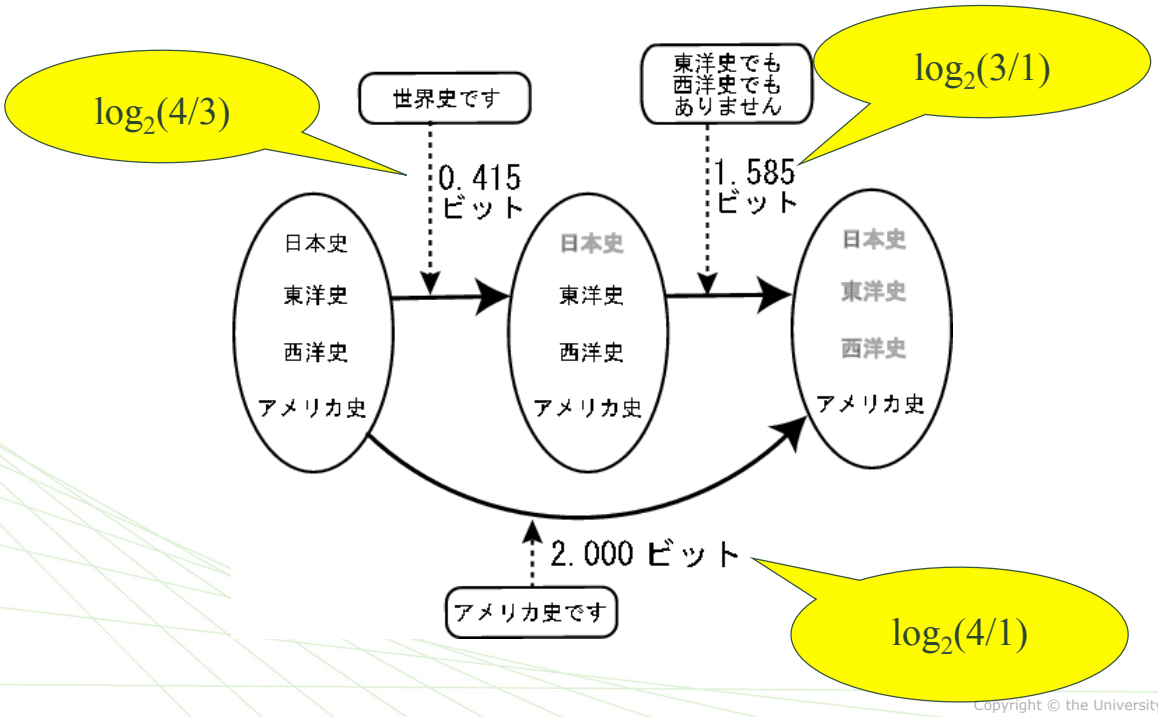
情報量の加法性

- **情報を一度に受け取った場合 (A)**
  - メッセージA:「アメリカ史を出題する」
    場合の数 4 → 1
- **分割して受け取った場合 (B+C)**
  - メッセージB:「世界史を出題する」
    場合の数 4 → 3
  - メッセージC:「東洋史と西洋史は出題しない」
    場合の数 3 → 1
- **情報量(A) = 情報量(B)+情報量(C) としたい**

場合の数に基づいた情報量の定義

- 定義:
  $\log_2$(事前の場合の数 / 事後の場合の数)
- 単位: ビット (bit)
- 性質
  - 場合の数が大きく減る程数が大きい
  - 底が2なので
    二者択一(場合の数が2から1になる場合)に 1.0
  - 情報量の加法性を満たす

情報量の加法性の確認



情報量の定義

↘ 定義: m通りの選択肢をn通りに減らす情報の量

$\log_2(m/n)$ (ビット)

- サイコロの次の目 — $\log_2(6/1)$ ≒ 2.59 ビット
- ルーレットの次の目 — $\log_2(100/1)$ ≒ 6.64 ビット
- 3桁の宝籤の下2桁 — $\log_2(1000/10)$ ≒ 6.64 ビット
- 前回のルーレットの目 — $\log_2(1/1) = 0$ ビット
- コイン投げの裏表 — $\log_2(2/1) = 1$ ビット

情報量の加法性

↘ 2つの情報をまとめた情報の情報量 ＝ 個々の情報量の和

20ビット　30ビット　50ビット

- 例: 3桁の宝籤

$\log_2(1000/10)$ ≒ 6.64 ビット ＋ $\log_2(1000/100)$ ≒ 3.32 ビット ＝ $\log_2(1000/1)$ ≒ 9.97 ビット

配布不可

場合の数から確率へ

- 例題: 科目「歴史」の試験 (改)
  - 日本史、世界史のどれか一つが出題させる
  - 事前にはどれが出題されるかは分からない
  - 世界史が75%、日本史 は25%で出題される
- どちらが価値が高いメッセージか?
  - 「世界史が出題される」
  - 「日本史が出題される」
- 場合の数の変化はどちらも同じだが...

確率に基づく情報量の定義

- 定義: $-\log_2$(確率)
- 単位: ビット (bit)
- 性質
  - 確率が低いことを伝えるメッセージほど大きい
    確率1.0 → 情報量 0, 確率0.5 → 情報量 1.0
    確率 0.25 → 情報量 2.0, 確率0 → 情報量無限大
  - c.f. 犬が人間を噛んだ v.s. 人間が犬を噛んだ
  - 場合の数に基づく定義の一般化:
    全てが等確率で起こる時は、場合の数の定義と同じ

平均情報量

➘ 平均情報量

- **p1*log2(p1) - p2*log2(p2) - …**
- **日本史: 25%,世界史: 75%の場合**
  - −0.25 log2(0.25) − 0.75 log2(0.75)

  = 0.25*2 +0.75*0.415

  = 0.811 < 1
  - 1−0.811=0.189はヤマの分

クイズ

天秤を使って9枚のコインから偽物を見つけるための最小回数は?

- 見掛けは全く同じ / 1枚が偽物 /
  本物の重さは全て同じ / 偽物は本物より軽い

- 方法1) A vs B, A vs C, A vs D, …, A vs I (最悪8回)
  - 実は8回目は不要(A〜Hが同じ→Iが偽物)
- 方法2) A vs B, C vs D, E vs F, G vs H (最悪4回)
- 方法3) ABC vs DEF
  → ABCが軽かったら A vs B (2回)

情報量からの考察

- 計測1回の情報量: 一方が重い/軽い/同じの3通り→ $\log_2(3)$ ≒ 1.58 ビット
- 偽物の存在する場合の数
  - 9枚中1枚だけ軽い: $\log_2(9)$ ≒ 3.17 ビット

    → 3.17/1.58 = 2回で分かる
  - 9枚中1枚だけ軽いか重い: $\log_2(9x2)$ ≒ 4.17 ビット

    → 3 > 4.17/1.58 > 2 → 3回以上
  - 13枚中: $\log_2(13x2)$ ≒ 4.7 ビット

    → 3 > 4.7/1.58 > 2 → 3回以上

情報量の差異の応用:符号化と情報量

- 情報は0,1の符号で表され，伝達される
- 伝送速度が一定ならば，小さいデータほど早く伝送できる
  - データは復元可能なように圧縮して伝送する
- 例:二年分の試験出題情報を符号化する場合，珍しい情報には長い符号を，珍しくない符号には短い符号を割り当てる.
  - 平均符号長＝Σ（符号長i×確率p）

符号化と情報量(例)

出題	確率	符号	符号長
日本史＋日本史	1／16	111	3
日本史＋世界史	3／16	110	3
世界史＋日本史	3／16	10	2
世界史＋世界史	9／16	0	1

- **平均符号長＝0. 844**
  - 1年分の試験の情報を表す符号長(＝1)より平均符号長が短くなっている

3.2 情報通信

- **プロトコル (3.2.2) (a)**
  - 通信の際の決めごと
- **通信の秘密と相手の認証 (3.2.3) (a)**
  - 暗号 盗聴を防ぐ
  - (認証 通信参加者の身元の保証)
  - (署名 通信内容の改竄の防止、否認の防止)

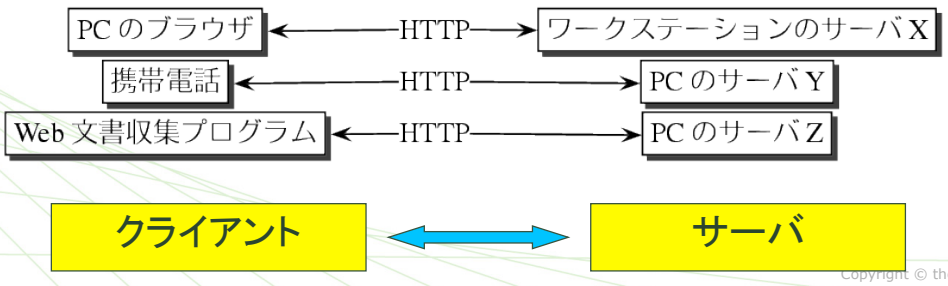
通信プロトコル

- 決められた種類の情報を伝える場合には、あらかじめ伝え方に約束事がある
  - 例)手紙: 宛名を書く場所、「気付」、差出人を書く場所、日付を書く場所、等々
    → 間違えると届かなかったり、誤解されたり
- インターネットの通信にも、種類によって約束事がある
  - 例) 電子メール: 宛先のメールアドレス、差出人のメールアドレス、日付を書く場所、等々
  - 例) WWW: URL, データの大きさ, データの種類, 更新された日時, 言語、等々

→ プロトコルと呼ぶ

プロトコル (protocol)

- 通信の意図を理解するための決めごと
  - 電話「もしもし」, トランシーバ「どうぞ」
- コンピュータ同士の通信: 人間の場合より厳密
  - WWW (HTTP, HyperText Transfer Control P.)
  - 電子メール (SMTP, Simple Mail Transfer P.)
- プロトコルを正しく使えば機器によらず通信可能

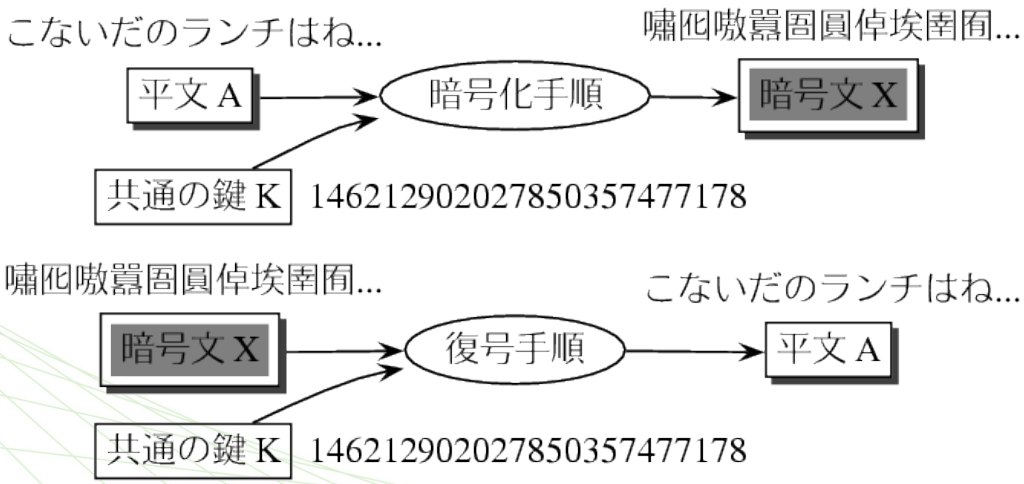


暗号

- 平文: 元のデータ。第三者に読まれたくないもの
  - 「明日のランチはね…」
- 暗号文: 変換後のデータ。盗聴されても平文を(簡単には)取り出せない。
  - 「嘯囮嗷囂圀圓倬埃圉囿…」
- 暗号化: 平文から暗号文を作成すること
- 復号: 暗号文から平文を取り出すこと
- 鍵: 暗号化や復号の際に用いられるデータ

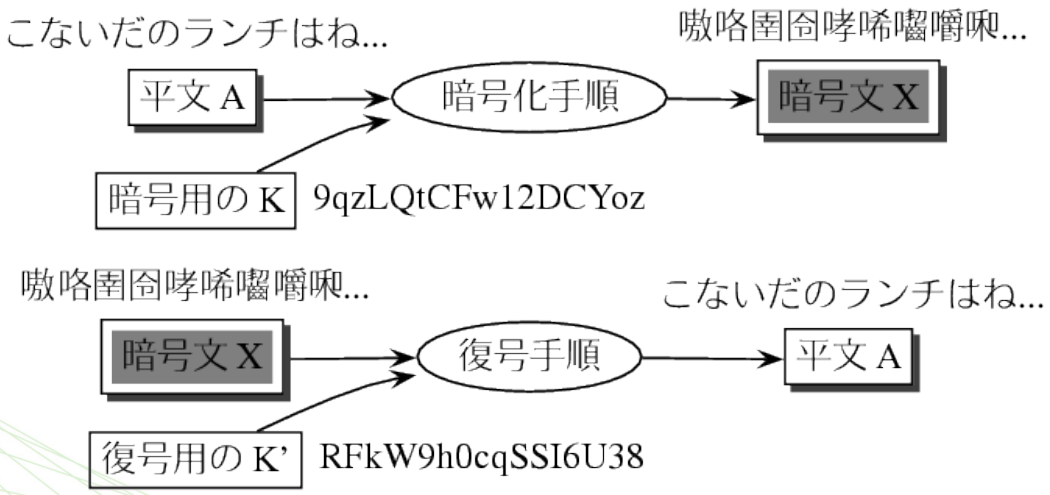
共通鍵暗号

- 一つの鍵で暗号化と復号化の両方を行うモデル
  - 鍵を秘密に保つ必要がある



公開鍵暗号

- 暗号化の鍵を公開し，復号化の鍵を秘密にする方法



3.4 インターネット

- ネットワークの集合体と通信 (3.4.1)
- 階層プロトコル (3.4.2)
- IPアドレスとポート番号 (3.4.3)

3.4 インターネット

- **インターネットを使って2つのコンピュータが通信をしているとき、実際にはどこで何が起きているのか?**

  注意:

  - 2つのコンピュータは直接つながっていない
  - 各コンピュータは、同時に色々な相手と通信している
  - 同時に色々な種類の通信もしている (例: メールとWWW)
  - 世界中のコンピュータが通信できる

インターネットの通信

- **インターネットとは広義には「ネットワークどうしをつないだネットワーク」のこと**
- **どうやって通信しているか?**
  - インターネットで通信をする場合は、情報は色々なネットワークを渡り歩いてゆく
  - 個々のネットワークの中では、コンピュータどうしが直接通信できる
  - ネットワークどうしをつなぐ機器を「ルータ」とよぶ