

情報意味論(1)

慶應義塾大学工学部
櫻井 彰人

この講義では

- 機械学習のいくつかの代表的な手法を知る
 - 基本原理
 - 基本アルゴリズム
 - 実際に使ってみよう
 - 少しアルゴリズムに触ってみる

アルゴリズムの分類 学習の形式

- 教師付学習
- 教師なし学習
- 半教師付学習
- 強化学習

アルゴリズムの分類

- 回帰
- 事例ベース
- 正則化
- 決定木
- 統計的分類
- カーネル法
- クラスタリング
- 相関規則
- ニューラルネットワーク
- ディープラーニング
- 次元圧縮
 - トピックモデリング
- アンサンブル法
- ブースティング

回帰 regression

- Regression: 回帰と訳すが
 - 後戻り, 復帰, 後退, 退歩, 退化, 退行
 - もともとは、今の意味とは異なる、「平均への回帰」の意味で使われた
- 説明変数のある関数で、被説明変数の値を近似する。次のものに依存する
 - 関数の形
 - 誤差の形
- 学習: 訓練データで、回帰関数を作る
- 推測: 未知データを回帰関数に入れ、出力値を予測値とする

事例ベース instance-based

- 丸暗記+類推
- 学習: 事例をすべて記憶する
- 推測: 新規データに最も近い事例を取り出す
 - 「近い、遠い」の決め方にいろいろ
 - 「近い、遠い」を学習する手法もある

決定木 decision tree

- 「木」を使って、学習結果を表現する
- 分類が主であるが、回帰もできる
- 学習：ヒューリスティックな構築方法
 - 各ノードには属性1個に関する値のテスト
- 推測：未知データに決定木を適用する

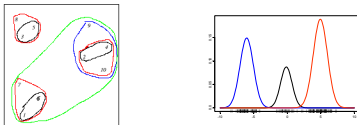


統計的分類

- 尤度最大化や事後確率最大化を図る。
 - その際、ベイズの定理を利用
- 学習：説明変数を確率変数と考え、その分布のモデルを作成する
 - モデルは、単純化する。
 - Naïve Bayes
 - 判別分析
- 推測：非説明変数の値の分布を求める。

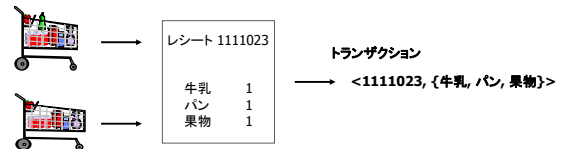
クラスタリング clustering

- 非説明変数に対する教師データはない。
 - 非説明変数はない、と言ってもよい
- 説明変数値の分布を用いて、各データをいくつかのグループ・塊り(クラスタ)に分ける
- 統計的には、隠れ変数のある統計モデルの推定問題として扱われる



相関規則 association rule

- 買い物籠1個がデータ1個
- 相関規則: If AとBを買う then Cも買う
- 発掘: 大量の買い物籠データから、信頼性と精度が高い相関規則を抽出

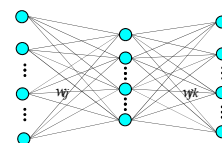


カーネル法

- 特徴量を、ある非線形関数を用いて高次元空間に写像し、そこで、線形関数を用いた分類や回帰を行う
 - 元になる手法(線形関数を用いる手法)が、カーネルトリックが有効となるような手法であるべき
 - 例: SVM
- 学習: 学習データでパラメータを推定。
 - カーネル関数は事前知識に基づいて選ぶ。ただし、情報量基準やCVを用いて選択するも可
- 推測: 未知データを入力
 - カーネルトリックを用いる故、計算量は(次元を高くしても)多くならない

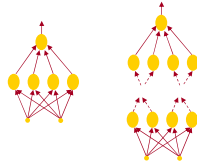
ニューラルネットワーク

- 単純な機能を持った素子(神経素子の単純なモデル)を多数結合したもの
- 学習: コスト(誤差等)が最小となるよう素子間の結合荷重を調節する
- 推測: 説明変数値を入力し、出力値を推定値とする



ディープラーニング

- 中間層数が多い(2以上)のニューラルネットワーク
2ではなかなかDLNとは認めたくない
- 基本的にはニューラルネットワーク
- 学習アルゴリズムに本質的な工夫がある



正則化 regularization

- 過学習を抑えるため、最小化すべきコストに、モデルが複雑になるほど大きくなるペナルティ項を加える
 - コスト関数 = 本来のコスト + λ ペナルティ項
 - λ の決め方に恣意性が残る

$$\min_f \sum_{i=1}^n |Y_i - f(X_i)|^2$$



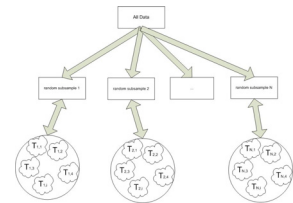
$$\min_f \sum_{i=1}^n |Y_i - f(X_i)|^2 + \lambda \|f\|^2$$

次元圧縮

- 説明変数の個数を減らす
 - 被説明変数がある場合、ない場合
 - 手法は多数あり
 - 主成分分析 (PCA)
 - 因子分析
 - 多次元尺度法 (MDS)
 - 潜在意味分析 (LSA, LSI)
 - 確率的潜在意味分析 (pLSA, pLSI)
 - Latent Dirichlet Analysis
 - 非負行列分解 (non-negative matrix factorization)
 - LASSO (least absolute shrinkage and selection operator)

アンサンブル法

- 複数の(多数の)学習器を組み合わせる
- 多数
 - ブースティング
 - バッグイング
 - AdaBoost
 - Random Forest



The top 10 algorithms in DM

- the IEEE International Conference on Data Mining (ICDM) in December 2006 で決めたもの
 - C4.5
 - k-means
 - SVM
 - A priori
 - EM
 - PageRank
 - AdaBoost
 - k-Nearest Neighbor
 - Naïve Bayes
 - CART

講義形態

- 普通の講義形態
- できるだけ、動作例を見てもらう
- シラバスから順序等多少変更あるかも
- 確率・統計の基礎はできるだけ省略
- Weka と R は道具として使うが概説のみ

評価方法

- 3回～4回のレポートに基づく

2015年度予定

1	9月28日	月	情報と意味と機械学習
2	10月5日	月	決定木と過学習
3	10月19日	月	RとWeka
4	10月26日	月	コネクショニズム
5	11月2日	月	多層神経回路網
6	11月9日	月	ベイズ学習
7	11月12日	木	モデル選択
8	11月16日	月	EMアルゴリズム
9	11月30日	月	ベイジアンネットワーク
10	12月7日	月	トピックモデル
11	12月14日	月	SVM
12	12月21日	月	Boosting
13	1月18日	月	事例ベース学習/相関規則
14	1月21日	木	Deep Learning, 強化学習

機械学習

- データから意味を抽出する作業を、従来から、機械学習とよんできた
- 機械学習 (machine learning) :
 - データ間の規則性(意味)の抽出(学習)を計算機に行なわせる
 - これは「学習」か? yes!
 - 知識獲得ともいう
 - 規則性が知識だった?
 - 適応(adaptation)でもある。
- データを集めて情報となすことにはかわりない

学習



- 少しずつ異なった意味で用いられるが
 - 外界と自分があるときに、自分を少しずつ変化させて、外界に適応する(よりよいメリットを得る)こと
 - すなわち、対象とする系の表現・表出に基づき、最適行動を計画・実行する
 - そのために、ある系の振舞い(データ)をもとに、その系を表現する(本質をとらえた一般記述)が必要

学習

- もっと一般化して考えると、学習とは
- 具体例を知り、具体例を一般化すること
 - 丸暗記という学習もある。
- 具体例(instance)を一般化する。
 - りんご1が落ちた、栗2が落ちた、、、
⇒ 物体は支えがなくなれば落ちる
 - 叩いたら痛かった: 一週間前、昨日、今日、、、
⇒ 叩くと(いつでも)痛い
 - 隣のAさんはケイタイを持っている、会社のBさんも、、、
⇒ みんなケイタイを持っている
- 特徴: 間違っているかもしれない
 - わずか(有限個の)具体例に基づくので当然。

人間: 人間は賢い

- 汎化すべきかすべきでないか、よく知っている。

鋭い一言
3歳の娘に「馬が好きならニンジンだよ。うさぎさんも好きなんだよ」と言ったら「馬を食わせようとした。すると一言。『人間だから』。(相模原市・賢くなりました:36歳)
(朝日新聞 2014年12月13日 b10面。いわせてもらおう)

機械学習

■ 「機械学習」はこの「一般化」を理論化するにあたり、結果の正しさ(という評価基準は常に必要)を、

最近の、半構造データを対象とした研究の発展に伴い、このアイデアに近い学習モデルが復活している。

- 具体例が無数になれば、正解が得られる、すなわち、
 - 具体例が無数になれば、モデルが同定できるような一般化を求めることにした。
 - 後に、この「モデル同定」でない、機械学習の特徴づけ(PAC)がなされ、機械学習のさらなる発展が起こることになる
- データ(対象とする系の動作の具体値(instance)をもとに、その系の記述を得る、その系を同定する。
- 2, 4, 6, 8, 10, 12, ..., ⇒ 偶数
 - 1, 2, 4, 8, 16, 32, ..., ⇒ 2 の冪乗

機械学習

■ まずは「学習」から離れるかもしれないが、「学習」の本質は捉えている

学習：経験(具体例)をもとにパフォーマンスを上げる

(パフォーマンスを上げるには、未経験の事例に対しても、うまく動作する必要があるのだ)
 学習：経験(具体例)をもとに未知の(類似の)事態に対応すること

そのためには、相手(外界)を知ることが必要。知るとは記述できること。
 本質：経験から(相手の)記述を帰納すること。未知事例に対して適用する。

機械学習

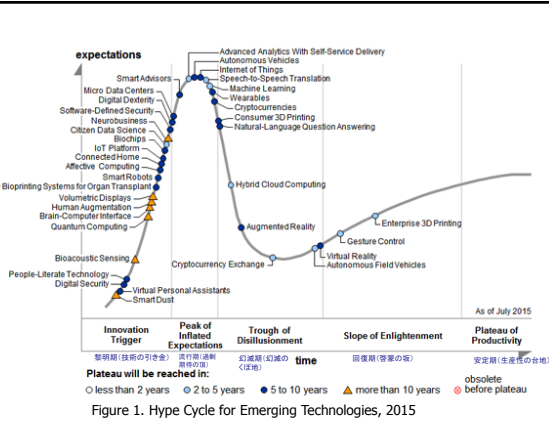
Study of algorithms that

- improve their performance P
- at some task T
- with experience E

(Tom Mitchell)

最近の「機械学習」ニュース

- Appleは人工知能と機械学習関連のエンジニアの採用を強化する予定(Techcrunch, Sep 7, 2015 by Romain Dillet)
- 機械学習によりデータ活用を強力支援、SASが2つの新製品(ニュースリリース 2015/08/25)
- Eyes, JAPAN がん超早期発見、「唾液」と「機械学習」で挑む(日経テクノロジーオンライン-2015/08/18)
- ファナック、プリファード・ネットワークスに9億円を出資-「機械学習」で差別化(日刊工業新聞 2015/08/23)
- MSやソフトバンク、「機械学習」による未来医療」に挑む(プレスリリース 2015/08/26)
- スマートニュース、機械学習及び自然言語処理における技術開発を強化~ニュース閲覧体験の向上を目指す。Preferred Infrastructure と業務提携(プレスリリース 2015/09/03)
- 自動走行車や機械学習に「過度な期待」-ガートナー先進技術ハイブサイクル(プレスリリース 2015/08/18)



学習の実例1 実世界



ロボットにペナルティキックをさせたい。もし関与するすべての物体の力学的性質が分かり、数値が測定可能かつ天候・芝の状態、キーパーの癖等がわかれば、最適なキック方法が選択できる。しかしそのようなことはない。どうするか？

自動清掃ロボットを作りたい。顧客ごとに部屋の配置を入力させるのは(入力するのは)大変だ。ロボット自身に「学習」させたい。どうしたらよいか？



ルンバは学習しない。Brooksの基本的考え

注:人工知能

- 二つの立場
 - 人間の知能そのものをもつ機械を作ろう
 - 人間が知能を使ってすることを機械にさせよう
- 後者が普通。
- 機械学習の技術も使うが、使わなくてもよい
- ロボット(知能機械)の動作に、人工知能技術は必ずしも必要ない。機械学習技術も同様
- 一方、ロボット(知能機械)でなくても、機械学習技術が必要などころはある。人工知能技術も同様



機械学習

- 機械学習 (machine learning) :
 - データ間の規則性(意味)の抽出(学習)を計算機に行なわせる
 - これは「学習」か? yes!
 - 知識獲得ともいう
 - 規則性が知識だって?
 - 適応 (adaptation) でもある。
 - 外界(自分以外の世界)の変化に自分を合わせる

規則性の記述

- 記号によるもの
 - Ex. Xには足があり、Xの上面が平らならば Xは机である。
- 統計的記述
 - Ex. K大学生の身長分布は、 $\mu=171.6\text{cm}$ 、 $\sigma=5.6\text{cm}$ の正規分布である
- 混合
 - Ex. 「過去一時間あたりの値上がり率が5%以上であれば、次の一時間あたりの値下が率が1%以上である」確率は52%である。

ところで、何故情報意味論?

- もともと、データと情報と意味を議論する講義であった(にしたかった)
 - データから意味・情報をとりだす
 - 考え方と方法
 - 取り出し方
 - 学習理論とアルゴリズム
 - 2つの方法: 記号的な方法、統計的な方法
 - 応用
 - 様々な adaptation
 - データマイニング
- その中でも「学習」に重点を置くことにした

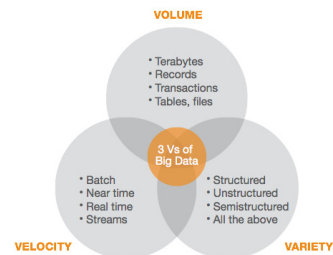
Big Data

Big Data is a **loosely defined term** used to describe data sets so **large and complex** that they become **awkward** to work with using **standard statistical software**.

Snijders, C., Matzat, U., & Reips, U.-D. (2012). 'Big Data': Big gaps of knowledge in the field of Internet. International Journal of Internet Science, 7, 1-5.

従来のデータベース管理システムなどでは記録や保管、解析が難しいような巨大なデータ群。明確な定義があるわけではなく、企業向け情報システムメーカーのマーケティング用語として多用されている。

Three V's of Big Data



Big Data Analytics Challenges Facing All Communications Service Providers
http://blog.vtira.com/bid/87945/Big-Data-Analytics-Challenges-Facing-All-Communications-Service-Providers

Big Data の取り扱い

- Big のまま扱う
 - これこそ、本道。実際、技術開発が行われている。
 - これまでのデータマイニングとは別種と考えてよい
 - データマイニングも、その当時のビッグデータを取り扱うことからスタートした
 - 解析方法を0から考えることになる
- Big data からある程度情報を抽出して、それを分析する
 - Big data の基礎的取扱い+データマイニング/機械学習
 - 多くはこちら。

Parallelization: platform choices

Platform	Communication Scheme	Data size
Peer-to-Peer	TCP/IP	Petabytes
Virtual Clusters	MapReduce / MPI	Terabytes
HPC Clusters	MPI / MapReduce	Terabytes
Multicore	Multithreading	Gigabytes
GPU	CUDA	Gigabytes
FPGA	HDL	Gigabytes

Big Data Research Progress. Chao. Jan 22, 2013.

補足: 公共データ

公共データ	サービスアイデア
事業許可情報 学校情報 工事情報 イベント情報 バリアフリー情報	<ul style="list-style-type: none"> ● 工事状況やバリアフリーなども考慮に入れて、目的地へ誘導するナビゲーションシステムの高度化 ● 買物客や観光客を案内するシステム
事業許可情報 学校、公共施設	<ul style="list-style-type: none"> ● ビッグデータ解析による出店や商品展開における高度マーケティング
事故発生情報	<ul style="list-style-type: none"> ● 事故多発の場所に近づいた際、注意を促すアプリケーション。 ● 子供が近づいた際に、近親者に連絡が行く見守りアプリケーション
気象情報	<ul style="list-style-type: none"> ● 農業の高度化 ● 流通における仕入調整等への利用
大気汚染度情報 水汚染度情報	<ul style="list-style-type: none"> ● 高付加価値な住宅情報サービス
ハローワークに登録された求人情報	<ul style="list-style-type: none"> ● 求職者のニーズに合致した求人情報を探し出す、高度なジョブマッチングサービス
製品安全・事故・リコール情報 地域で受けられる医療検診の情報 国民健康・栄養調査	<ul style="list-style-type: none"> ● 事故情報のビッグデータ解析による、事故発生の傾向分析。 ● 住民の健康を促進する情報サービス、ヘルスケアサービスの紹介等

(平成24年4月25日電子行政タスクフォース 「コンテンツ流通推進協議会事務局提出資料」)

情報とは何か？

- 英語では information
 - Inform がもとの動詞。どう使う？
- 日本語: izzgoooo 訳したか？
 - 情とは
 - 報とは

(小野厚夫, "情報という言葉を探して" (1)~(3), 情報処理(2005) を参照)

(インフォルメーションで調べてみよう)

意味とは何か

- ①記号・表現によって表される内容またはメッセージ。②物事が他との連関において持つ価値や重要性。(広辞苑)
- 動作で考えてみよう。例えば、「意味がある」行動とは？
- 次に、表現と意味との関係を考えてみよう。
 - 現実世界における「表現」は常に、冗長である。では、徹底して冗長性を排除したらどうなるか？
 - なぜ、冗長なのかも考えてみよう

情報理論における情報

- データを生み出す「データ源」の記述
 - 例1: 0は確率1/4で, 1は確率3/4でランダムに生成する
 - 例2: n番目には, n番目の素数の10進第一位を生成する

データ源の記述ができると

- データ源の記述ができると何がよいか？
 - 予測ができる
 - もしそれがノイズ源であれば、ノイズを効果的に低減することができる
 - (もっと一般的には) 制御することができる

例えば、

- 一つの音源の音を正確に採取するために、複数のマイクを使う。
 - 2つのデータ中の相関の大きな成分が当該音源の音である(ノイズには相関がない)
- 経済予測: 株価予測、売上予測
 - 潜在需要の発見とその利用(刺激して新市場創造)
- 物理現象・化学現象・社会現象の記述と予測

つまり、機械学習

- 要は、
 - 目的、方法、評価方法は様々であれ、
 - データから意味(これって、目的によって変わります)をとりだすこと
- が機械学習

データマイニングとは？

- データマイニング(データベースからの知識発見):
 - 興味深い(当たり前でない、潜在的、これまで知られていなかった、しかも、役に立つと思われる)情報あるいはパターンを大規模データベースから抽出すること
- データマイニングの別名
 - データマイニング: 命名を間違えた?
 - データベースからの知識発見(Knowledge discovery in databases, KDD)、知識抽出、データ/パターン解析、データ考古学、情報収穫、ビジネスインテリジェンス、など
- データマイニングでないのは何か？
 - (演繹)質問応答処理
 - エキスパートシステムあるいは小規模な機械学習システム/統計パッケージ

データマイニングの応用例

- データベース解析と意思決定支援システム
 - マーケット分析とマネジメント
 - ターゲット・マーケティング、CRM(customer relation management)、購入品目分析 (market basket analysis)、マーケット区分 (market segmentation)
 - 危機分析とマネジメント
 - 予測、顧客維持、保険の査定改善、品質管理、競争力分析
 - 不正検知と管理: アクセスログ解析
- 他の応用
 - テキストマイニング(電子メール、webドキュメント、ブログ)
 - Web アクセスログ解析
 - 遺伝子解析(文献解析含む)

これはデータマイニング？

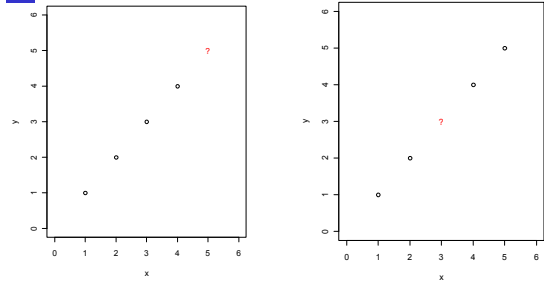
- 経済学? yes.
 - 経済的インセンティブを取り扱っている
- データマイニング? yes
 - 多量データの分析結果に基づく



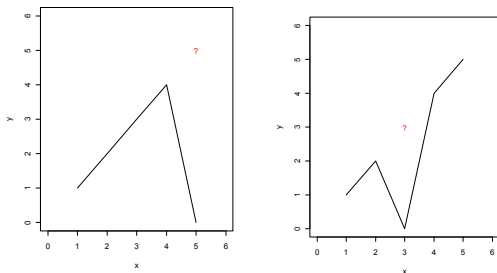
蛇足: なぜ機械学習か?

- 様々な意味で「計算能力が向上」
 - データベースマイニング: データを知識に
 - 自動カスタマイズプログラム: ニュースのフィルタ、適応的な監視カメラ
 - 行動の学習: ロボットの計画、制御の最適化、決定支援
 - プログラム困難なアプリケーション: 自動運転、音声認識
- 人間の学習や教育のよりよい理解を求めて
 - 認知科学: 知識獲得の理論 (e.g., 実践を通じて)
 - パフォーマンス向上: 推論・推測, 推薦システム
- 時は今、、、
 - 学習アルゴリズムや理論の最近の進歩は目覚ましい
 - 様々なソースから大量のオンラインデータが提供される
 - 計算機は安価・高速
 - 機械学習を用いた事業が発生・成長 (e.g., データマイニング/KDD)

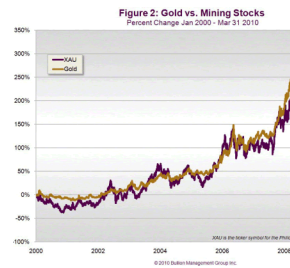
予測と推測・推定



予測と推定・推測



<http://heavenawaits.wordpress.com/god-man-and-stock-market-wave-theories/>



<http://www.safehaven.com/article/17497/why-bullion-is-outperforming-mining-stocks>

ランダムウォーク S が $2n$ 歩後に $2l$ ($-n \leq l \leq n$ とする) の地点にいる確率は

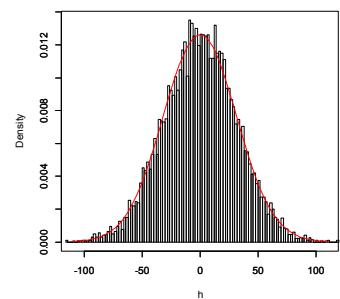
$$P(S_{2n} = 2l) = \binom{2n}{n+l} \frac{1}{2^{2n}} = \frac{(2n)!}{(n+l)!(n-l)!} \frac{1}{2^{2n}}$$

ランダムウォーク S が $2n$ 歩後に $a\sqrt{2n}$ 以上 $b\sqrt{2n}$ 以下である確率は

$$P(a\sqrt{2n} \leq S_{2n} \leq b\sqrt{2n}) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}y^2} dy$$

ただし、 $-\sqrt{2n} \leq a \leq b \leq \sqrt{2n}$

Histogram of h



```
set.seed(123)
rep <- 10000
N <- 1000
br <- 100
h <- numeric(rep)
hc <- numeric(rep); for( i in 1:rep) h[i] <- sum(rnorm(N))
hc <- hist(h, Freq=F, breaks=br)$density
ymax <- max(hc)
hist(h, Freq=F, breaks=br, ylim=c(0,ymax), xlim=c(-3.5*sqrt(N), 3.5*sqrt(N)))
par(new=T)
plot(function(x) dnorm(x,0,sqrt(N)), col=2, ylim=c(0,ymax),
      xlim=c(-3.5*sqrt(N), 3.5*sqrt(N)), xlab="", ylab="")
```

逆正弦定理

定理(逆正弦法則) ランダムウォーク S が $2n$ までの間に正の側で $2k$, 負の側で $2n-2k$ 過ごす確率 $P(n, k)$ は

$$P(n, k) = u_n u_{n-k}$$

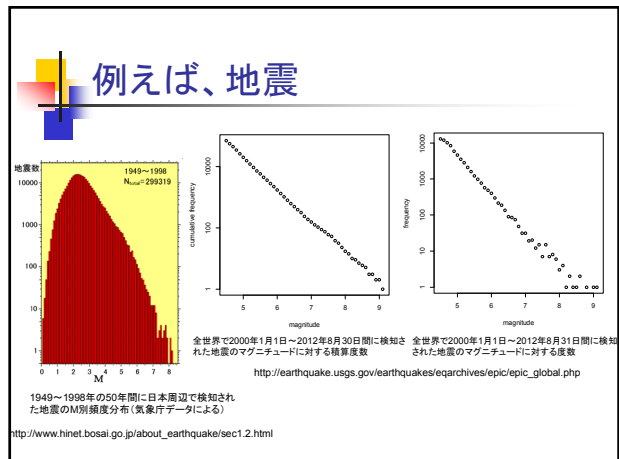
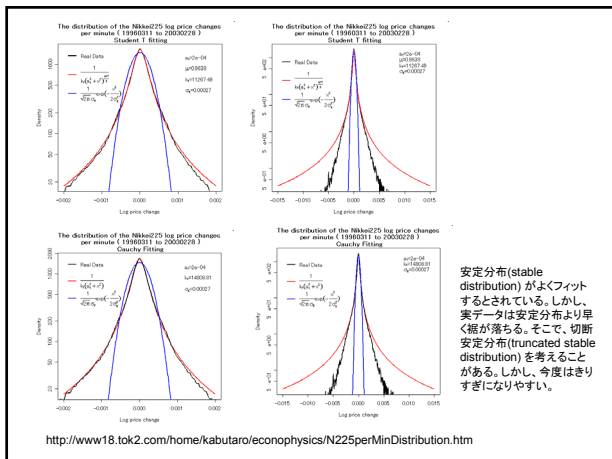
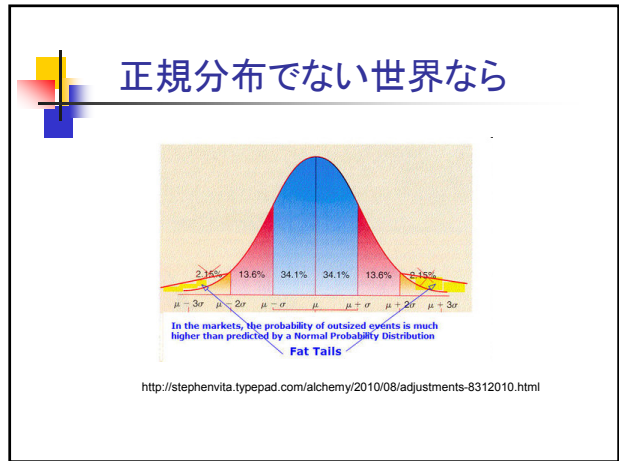
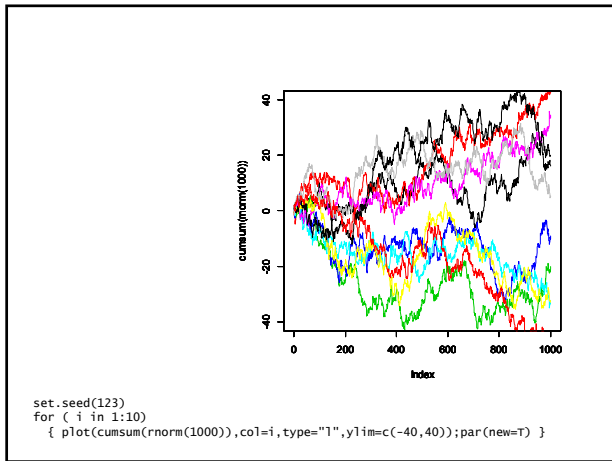
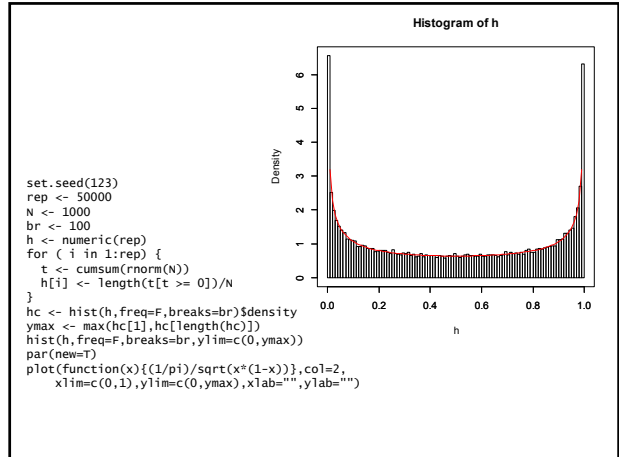
である

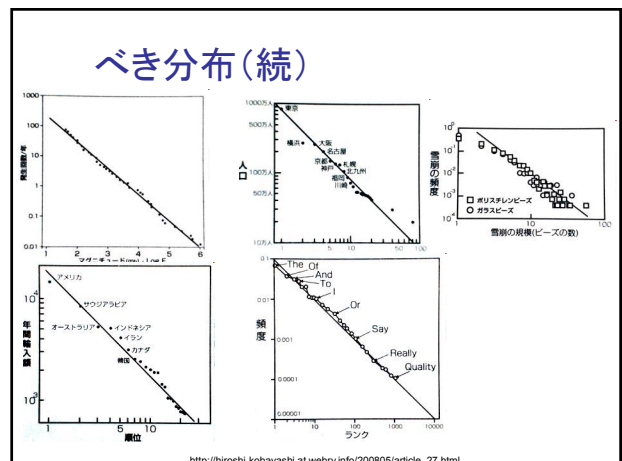
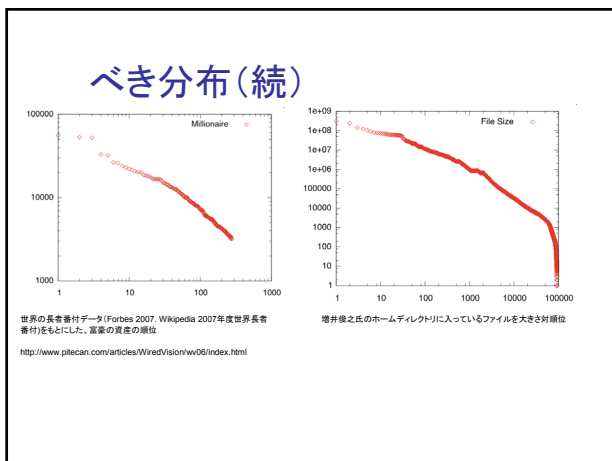
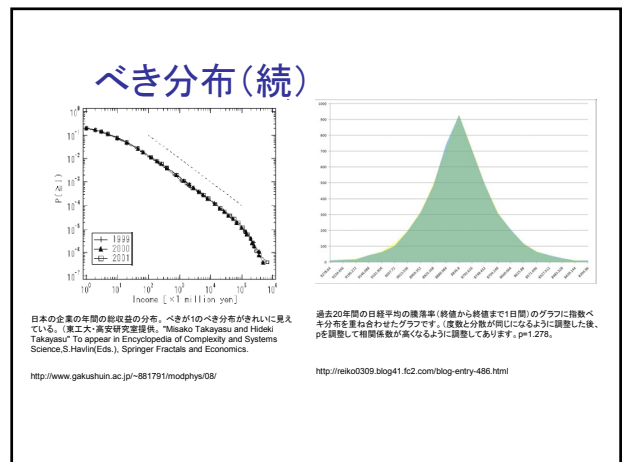
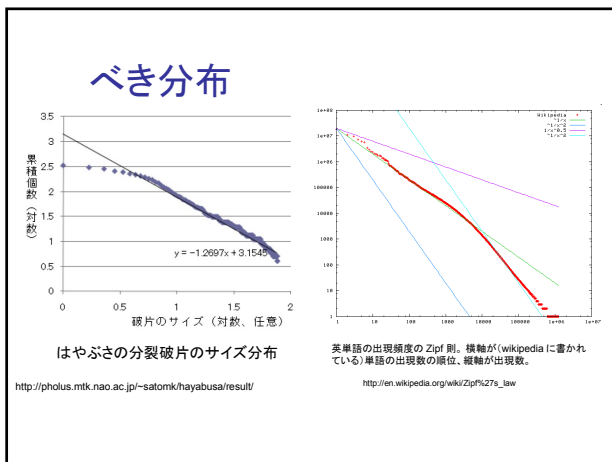
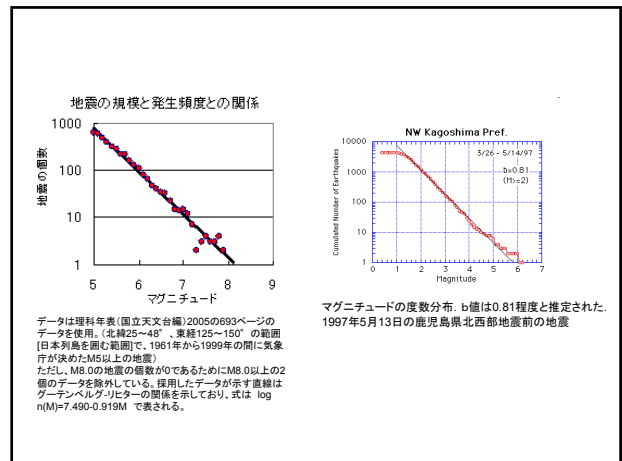
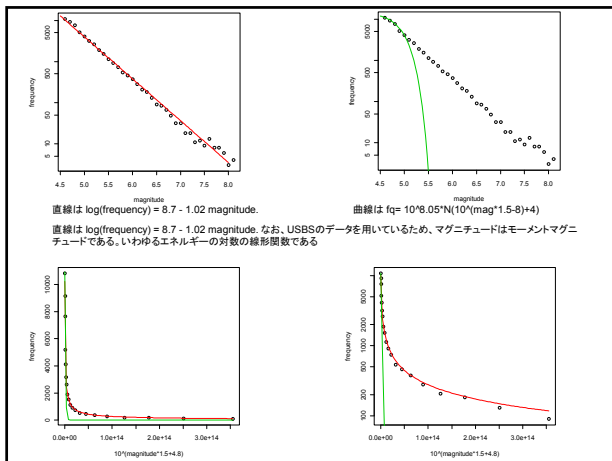
$$\text{定義} \quad u_0 = 1, u_n = \binom{2n}{n} \frac{1}{2^{2n}} = \frac{(2n)!}{n!n!2^{2n}}$$

P (ランダムウォーク S が $2n$ までの間に正の側にいる割合 $\leq \alpha$)

$$= \sum_{k=0}^n P(n, k) \approx \sum_{0 \leq \frac{k}{n} \leq \alpha} \frac{1}{\pi \sqrt{k(n-k)}} = \frac{1}{\pi} \sum_{0 \leq \frac{k}{n} \leq \alpha} \frac{1}{\sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right)}} \approx \frac{1}{\pi} \int_0^\alpha \frac{dx}{\sqrt{x(1-x)}} = \frac{2}{\pi} \arcsin \alpha^{\frac{1}{2}}$$

<http://elis.sigmath.es.osaka-u.ac.jp/~nagahata/20070816/arcsin.pdf>





現実のデータ

- 正規分布に従わないものがある
 - 冪分布に従うと、fat tail である。
 - その結果、大きく外れる予測誤り率が大きい
- しかも、現実にはデータ量が少ない
 - 絶対量が少ない場合
 - 相対量が少ない場合

世の中ビッグデータだと騒いでいるのに？

Swan and Black Swan



Swan and Black Swan

- "Black Swan" はTalebの極めて有名な著書
 - 最近は、"Black Swan" とgoogleで引くと、別のものが大量に出てきて困ります。
 - 大分とよくなりましたと言いたいが、そうでもなし
- Swanは白い鳥だと誰もが信じていた。Black Swan が発見されるまでは。
 - 「これはバブルではない、わが国経済の実力である」と誰もが信じていた。バブルが崩壊するまでは。

関連領域

- 認知科学: 言語獲得、推論の学習
- 統計学: バイアス vs. 分散, 信頼区間, 仮説検定
- ベイズの方法: ベイズの定理、欠測値の推定
- 人工知能: 記号表現、計画、知識を用いた学習
- 計算の複雑さの理論: PAC 学習、VC次元、誤差限界
- 制御理論: 最適化、動的計画、予測の学習
- 情報理論: エントロピー, MDL, 情報源符号化
- 神経科学: 人工神経回路網、脳(大脳、小脳、視床下部)
- 哲学: オッカムの剃刀、帰納的一般化
- 心理学: 練習の冪法則(Power Law of Practice) 発見的学習

機械学習環境

- Weka: Waikato大学開発
 - <http://www.cs.waikato.ac.nz/ml/weka/>
- RapidMiner:
 - <http://rapid-i.com/content/blogcategory/10/69/>
 - 旧名: Yale: yet another learning environment
 - <http://www-ai.cs.uni-dortmund.de/SOFTWARE/YALE/index.html>
- R: 統計計算用言語・パッケージ
 - <http://www.r-project.org/>
 - Rattle: <http://rattle.togaware.com/>
- Python: 機械学習用ツールがある
- 掲示板
 - <http://www.kdkeys.net/forums/>

参考書等

- パターン認識と機械学習
- Thomas Mitchell, Machine Learning, McGraw-Hill.
- Stuart Russell, Peter Norvig, エージェントアプローチ 人工知能, 共立出版
 - Artificial Intelligence: A Modern Approach (3rd edition), Prentice Hall
- <http://www.sakurai.comp.ae.keio.ac.jp/>
Google で「櫻井研究室」