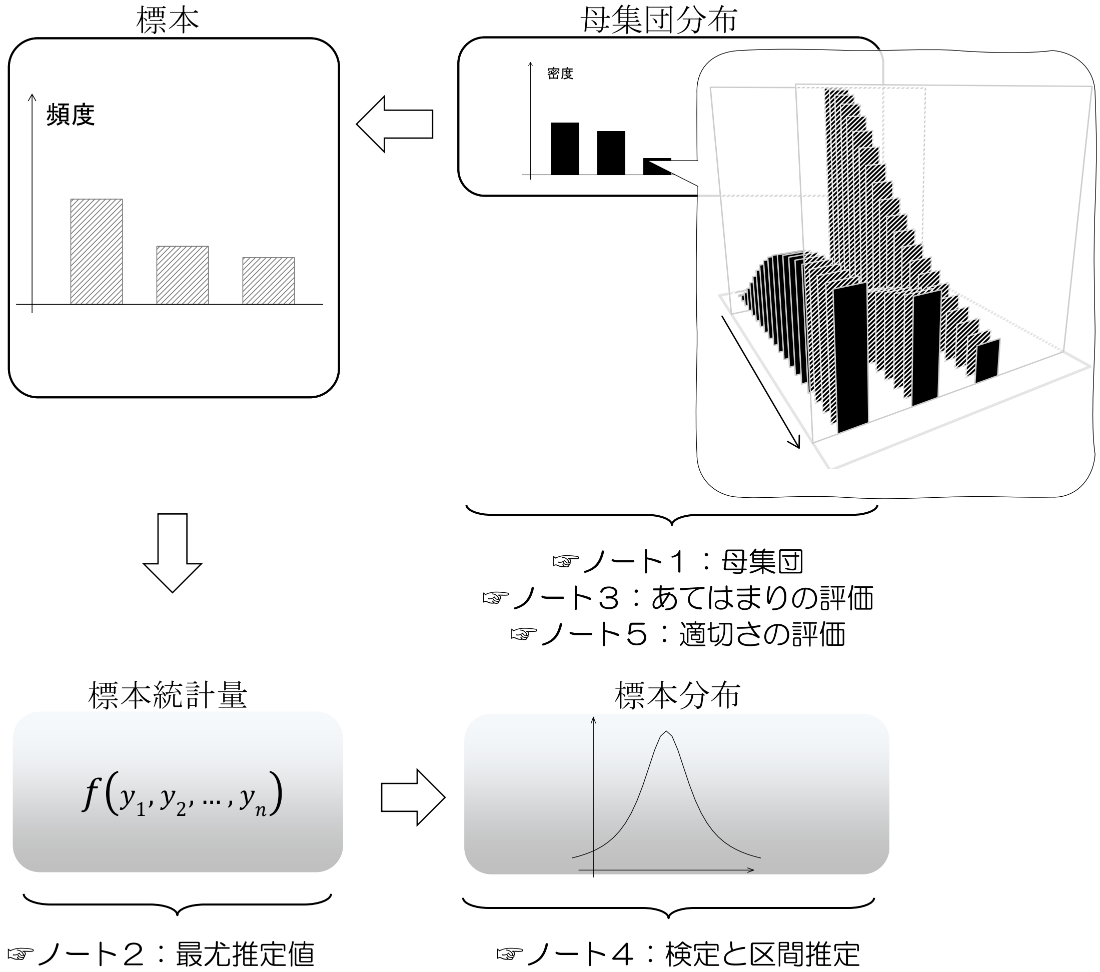
第2講　一般化線形モデル１：ロジスティック回帰

学習の目標

- □ 独立変数を用いて離散的な値を取る従属変数をモデル化する際に用いられるモデルにロジスティック回帰モデルがあることが分かる。
- □ 独立変数の線形結合はリンク関数を介して、ベルヌーイ分布（二項分布）／カテゴリカル分布（多項分布）の確率パラメータをモデル化していることが分かり、数式で表せる。
- □ リンク関数の種類は一通りではなく、無数の候補が存在することが分かる。とりわけ、よく使われるリンク関数にロジット／プロビット・リンクなどがあることが分かる。
- □ 独立変数を操作してπの値を変化させるという行為は、単体の上の位置を変化させることだということが分かる。
- □ 潜在変数を用いてロジスティック回帰の背後の構造を考えることができる。
- □ ロジスティック回帰の切片と係数の推定には、最尤推定法が用いられることが分かる。
- □ パラメータを変数として見なした同時密度関数の式が尤度関数、その対数を取ったものが対数尤度関数であることが分かり、最尤推定法の目的はこの対数尤度関数を最大にするパラメータの組を求めることであることが分かる。
- □ 対数尤度関数の導関数をスコア関数、そのさらに導関数をヘシアンと呼ぶことが分かる。
- □ 最尤推定値を求めることとは、スコア関数の値が０になる値を求めることであるということが分かる。
- □ ニュートン・ラフソン法の仕組みが分かり、説明できる。
- □ モデルのあてはまりの悪さの指標に逸脱度が用いられる頃が分かり、それが何か説明できる。
- □ 尤度に基づく検定・区間推定に、尤度比検定、ワルド検定、スコア検定があることが分かり、説明できる。
- □ モデルの評価には、感度、特異度、精度、NPV、F1 スコア、ROC 曲線、AUC などの指標が使われることが分かる。
- □ ピアソン残差や逸脱度残差から特異な挙動を示す事例を見つけることができる。
- □ モデル比較を行う際には、重回帰同様、情報量基準やクロスバリデーションが用いられることが分かる。

見取り図



データの形式

ID	説明変数	観測値（応答変数）
1	0.1	1
2	1.3	2
⋮	⋮	⋮
$\boldsymbol{n}$	3.7	0

（1）目的（リサーチクエスチョン）

あるカテゴリーに属するものの頻度（従属変数）は、独立変数たちの値によってどのように変わるのか。これに、モデルを立てて、統計的推論を行いたい。

（2）考え方

第一講では、ベルヌーイ（二項）分布やカテゴリカル（多項）分布から標本データが取られて、（相対）頻度表が作られるという話を扱った。そこでは、母集団から標本抽出を経て、（相対）頻度表／経験分布という標本／統計量が形成されるという関係を見た。

しかし、そこには、すべての観測値がいつも同じ母集団確率分布に従うという前提があり、これは、現実のサンプリングのプロセスと大きく乖離する場合がある。例えば、「TOEIC500 点以上の人の人数」は、「学年」が上がれば上がるほど増えることが予想される。中学生 1 年生と高校 3 年生が両方とも同じ分布に独立に従っているという仮定は、かなり強引なものであるはずだ。

そこで、独立変数の値が変われば、それに呼応して母集団分布の形状も変わるようなプロセスを、第 1 講の標本抽出のプロセスに加えてみることにしよう。

これは、重回帰モデルで考えたことと同じである。従属変数が独立変数の値によって変化するようにモデルを作ることで、そのモデルの推定結果をもとに、どの独立変数が従属変数の値に影響を与えているのか議論をすることができる。このような、質的な従属変数（名義尺度の従属変数）に対して、重回帰同様に複数の独立変数（と回帰係数の積）の和をつなげたモデルのことをロジスティック回帰と呼ぶ。

なお、ロジスティック回帰は、従属変数のカテゴリーがいくつあるのかで、細かい分類がされることがある。従属変数が、ベルヌーイ（二項）分布に従う場合には「二項ロジスティック回帰」、カテゴリカル（多項）分布に従う場合には「多項ロジスティック回帰」と呼ばれる。だが、その数学的な議論の構造は同じなので、ここでは両方を同じ講義の中で扱っていく。

このように、ロジスティック回帰は、重回帰分析を名義尺度従属変数へ拡張したものなので、モデルを立てた後の基本的な考え方は、これまで習ってきたものと同じである。点推定をして、パラメータを推測する。信頼区間を出す。検定でパラメータが 0 と見なせるのかどうかを吟味する。モデル比較を通して、ベストなモデルを選択する。モデルがどのくらいデータにフィットしているのかを検討する。残差などを検討し、異常な値がないか確認する…などである。

ただし、従属変数に正規分布を採用しないことで、様々なマイナーチェンジが存在する。ここでは、そういったロジスティック回帰ならではの考え方を一つ一つ丁寧に見ていくことにする。

📖 ノート０　ロジステック回帰の目的と基本的な考え方

〇目的

回帰分析と同じやりかたで、
従属変数が名義尺度のデータも扱いたい！

イメージ

	点推定値	標準誤差	p 値	
切片	1.20	0.11	0.00	***
回帰係数１	- 2.11	0.21	0.00	***
回帰係数２	0.80	0.33	0.02	*

社会言語学における利用例

① モチベーション
言語の産出を、社会的な要因から説明したい！

② 例
singing [siŋiŋ] vs. *singin'*[siŋin]のうち後者を発音する確率を、(a) 話者の発話階層、(b) 話者のジェンダー等から予測する。

音韻論における利用例

① モチベーション
言語の産出を、音韻論的な要因から説明したい！

② 例
[ŋ] vs. [n]の自由変異の選択を、(a) 直前の音韻環境、(b) 後続の音韻環境等から予測する。

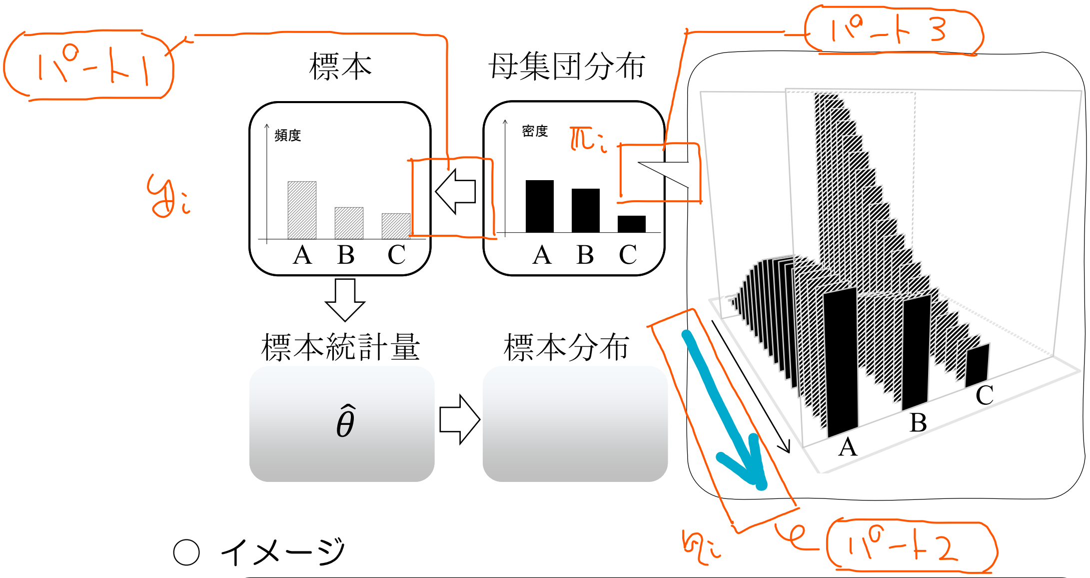
構文文法における利用例

① モチベーション
構文の選択を、音／意味／社会要因等から説明したい！

② 例
give the apple to him 型の言い方と give him the apple の選択を (a) 情報構造、(b) 直接目的語名詞句の長さ等から予測する。

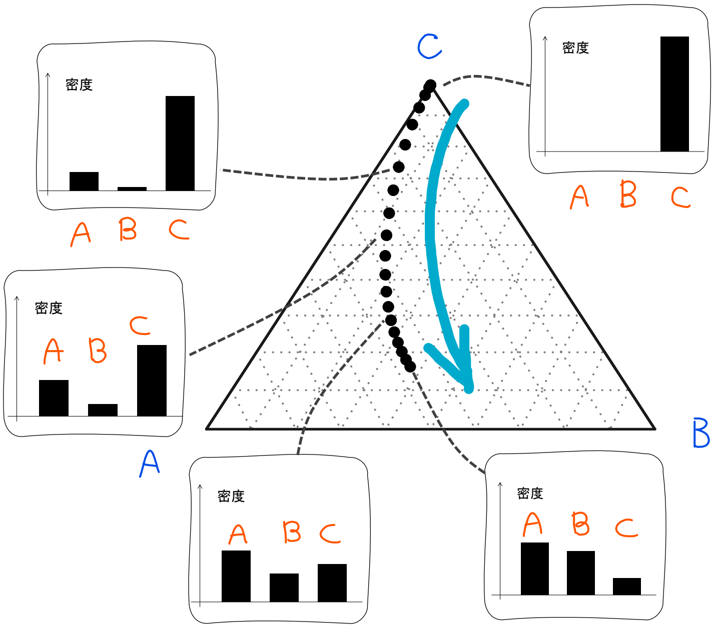
○ ポイント

独立変数をもとに母集団の確率を　構造化　する！



○ イメージ

母集団分布が独立変数を動かすと変わる
　＝単体の中の　位置　が移動するということ



📖 ノート１　母集団分布：統計モデル

○目的

「ノート０」のアイデアを数式で表現する！

（１）パート１：母集団分布からの標本抽出

① ベルヌーイ分布の場合

$$y_i \sim Bern(\pi_i)$$

② カテゴリカル分布の場合

$$y_i \sim Categorical(\boldsymbol{\pi}_i)$$

（２）パート２：独立変数たちの線形結合を作る

予測したい角度の数だけ必要、線形結合を作る必要がある。これをいったんη_iとおく。

① ベルヌーイ分布の場合

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_n x_{ni}$$

AとBのカテゴリーのうち，Bをとりやすくなる度合

② カテゴリカル分布の場合

Aに比べてどのくらいBを取りやすくなったか

$$\begin{cases} \eta_i^{(2)} = \beta_0^{(2)} + \beta_1^{(2)} x_{1i} + \cdots + \beta_n^{(2)} x_{ni} \\ \eta_i^{(3)} = \beta_0^{(3)} + \beta_1^{(3)} x_{1i} + \cdots + \beta_n^{(3)} x_{ni} \end{cases}$$

Aに比べてどのくらいCを取りやすくなったか

（３）パート３：リンク関数 Link Function

$(-\infty, \infty)$の値を取るη_iから、$(0,1)$の値を取るπ_iへ、(a) 一対一で、(b)滑らかに、変換する関数のこと。

$$\pi_i = F(\eta_i)$$

$$G(\pi_i) = \eta_i$$

Ⓟ 添え字 i

これは，1番目，2番目，…，n番目のデータで，値が変わってくるものに対して付ける．

①第1講ノート1のモデル

$y_i \sim Bern(\pi)$

添え字なし．(＝全てのデータで共通)

②第2講ノート2のモデル

$y_i \sim Bern(\pi_i)$

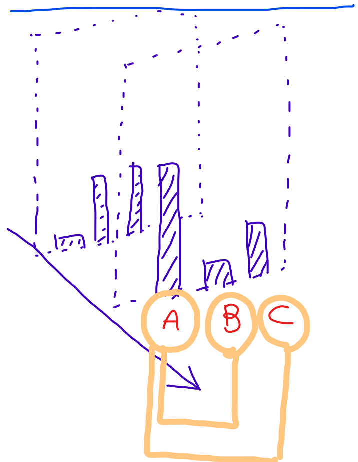
それぞれのデータごとに異なる値を取る．

Ⓟ 線形結合

$\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$

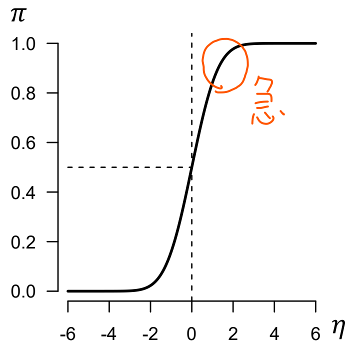
という形のこと．

Ⓟ 参照カテゴリー

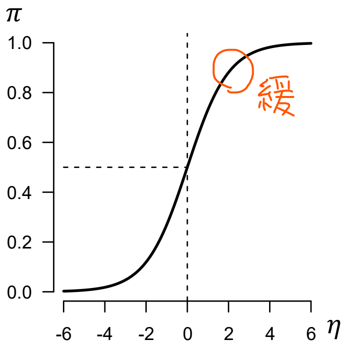


よく使われているリンク関数

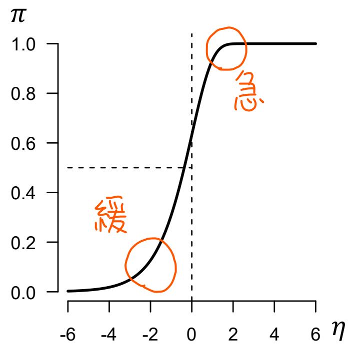
① プロビット・リンク関数



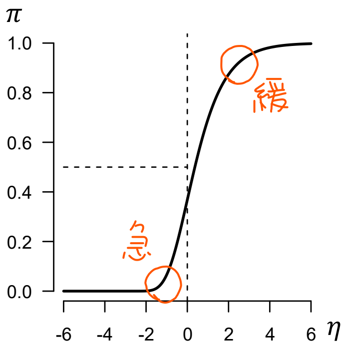
② ロジット・リンク関数



③ Complementary log-log 関数



④ log-log 関数



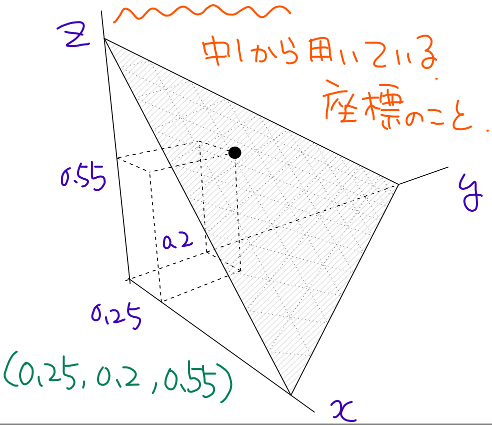
※ ロジット・リンク関数

$$\mathrm{logit}\ (\pi_i) = \log\ (\tan\theta_i) = \log\left(\frac{\pi_{1i}}{\pi_{2i}}\right) = \log\left(\frac{\pi_{1i}}{1-\pi_{1i}}\right)$$

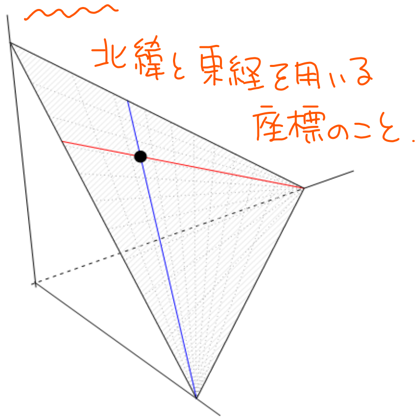
増補1　ロジスティック回帰の幾何的な理解

単体の中の 位置 を 角度 で同定する

（1）デカルト座標



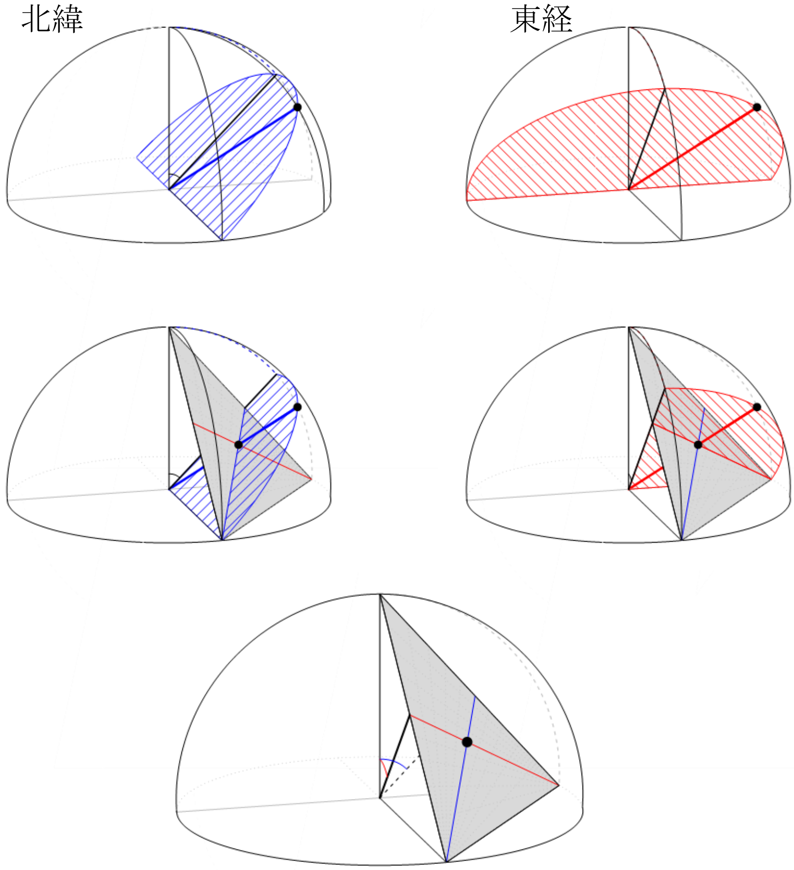
（2）極座標



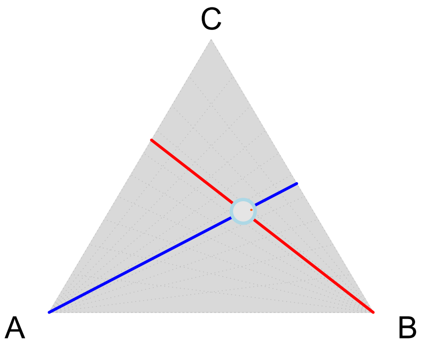
北緯と東経で座標を指定するということ

北緯

東経

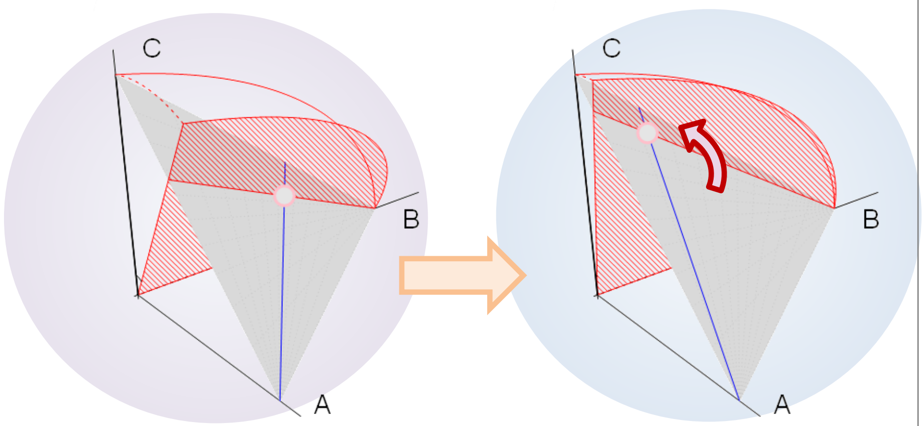


○ 単体の上での移動

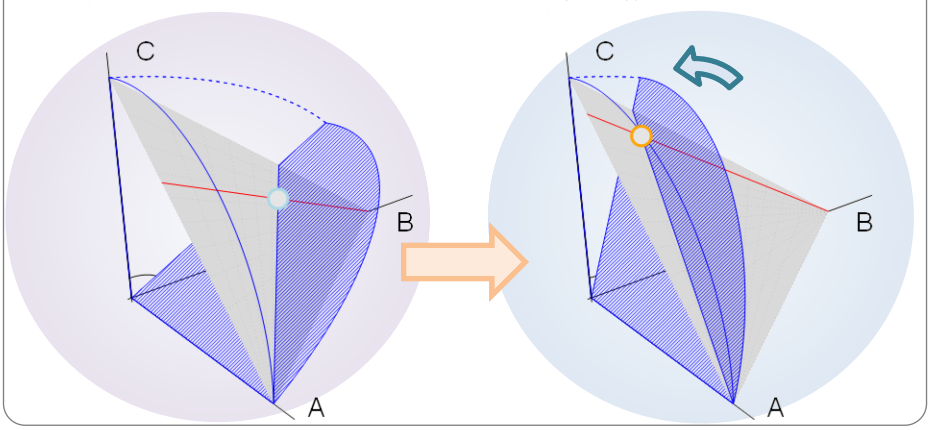


北緯と東経が変化するということ

A 軸方向への角変化（東経）



B 軸方向への角変化（北緯）



📖 増補２ ロジスティック回帰の潜在変数を用いた理解

離散的な観測変数の背後に
連続的な潜在変数が存在すると想定する！

（１）潜在変数と観測変数

① 観測変数 Observed Variable
これは、研究者が、実験や観測を通じて実際にその値を観測することのできる変数のこと。

② 潜在変数 Latent Variable
観測変数に影響を及ぼしていると、理論／モデル上想定されるものの、研究者には観測できない変数のこと。

（２）閾値モデル Threshold Model

連続的な潜在変数の値がある閾値（threshold）を超えるか超えないかで離散的な値に変換される！

① 潜在変数 y^*

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

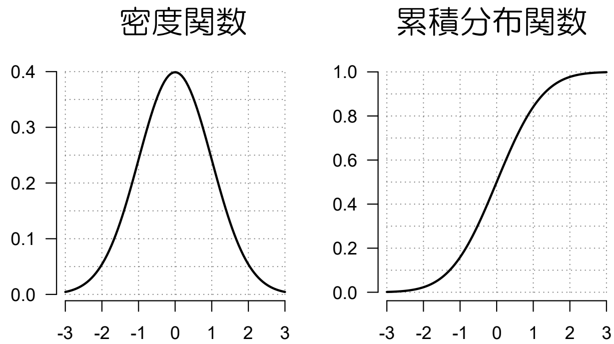
② 閾値 τ による離散化

$$y_i = I(y_i^* > \tau) \begin{cases} y_i = 0 & \text{if } y_i^* \leq \tau \\ y_i = 1 & \text{if } y_i^* > \tau \end{cases}$$

$$\begin{aligned} p(y_i = 1) &= p(y_i^* > \tau) \\ &= p(\beta_0 + \beta_1 x_{1i} + \varepsilon_i > \tau) \\ &= p(\varepsilon_i > \tau - \beta_0 - \beta_1 x_{1i}) \\ &= 1 - p(\varepsilon_i \leq \tau - \beta_0 - \beta_1 x_{1i}) \\ &= 1 - F(\tau - \beta_0 - \beta_1 x_{1i}) \end{aligned}$$

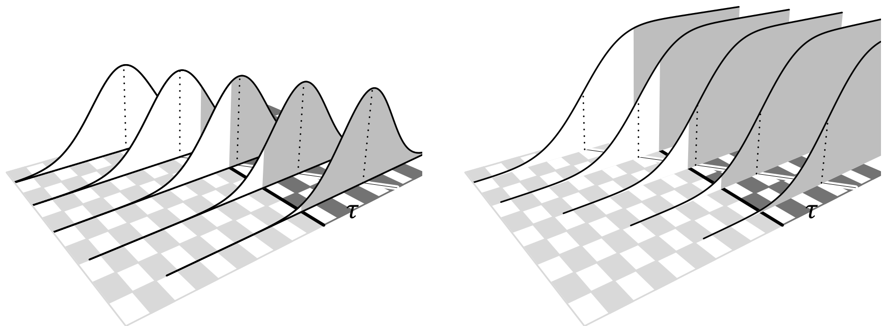
累積分布関数 Cumulative Distribution Function

例：標準正規分布

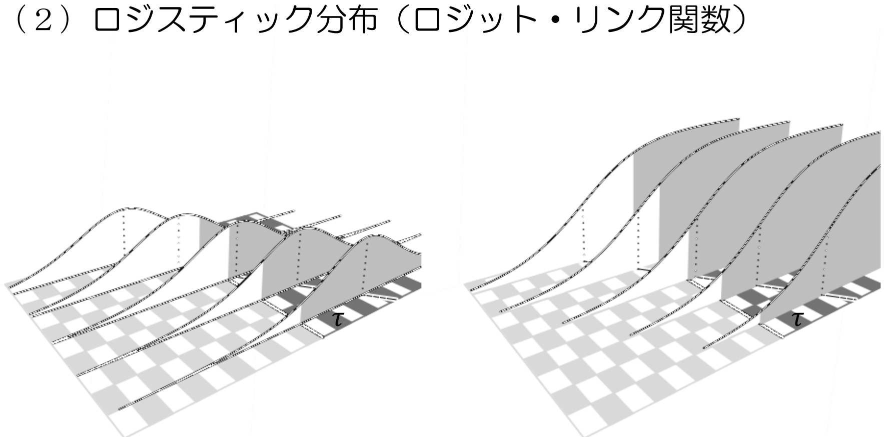


＜連続量⇒離散化＞の直感的な意味合い

（1）正規分布（プロビット・リンク関数）



（2）ロジスティック分布（ロジット・リンク関数）



※ log-log リンク関数は、ガンベル分布（the Gumbel Distribution；別名 第一種極値分布 the Type I extreme-value distribution）の累積分布関数のCDFを利用。

例題　**江戸から大正期の日本語尊敬語構文**

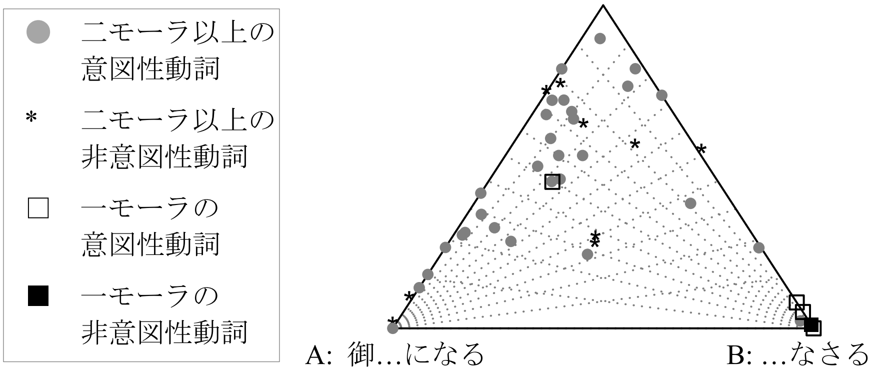
現代日本語には「御…になる」「…なさる」「御…なさる」という尊敬語が競合をしている。だが、過去の日本語でこの三つがどのように選択される関係にあったのかは明らかになっていない。

江戸から大正期のコーパスから ID1 番から 6 番の動詞たちがそれぞれの形式で何回使われたのかを表したのが次の表 1 である。このデータをもとに、多項ロジスティック回帰を用いて、選択傾向を検討してみよう。

ID	y_{i1}（御…になる）	y_{i2}（…なさる）	y_{i3}（御…なさる）	x_{1i} 1 モーラ	x_{2i} 意図性
1	56	0	0	0	0
2	18	1	9	0	1
3	13	3	14	0	1
4	49	22	53	0	1
5	57	14	88	0	1
6	31	12	31	1	1
⋮	⋮	⋮	⋮	⋮	⋮
合計	961	924	1599		

ここでは、動詞の性質によって選択傾向が変わるのではないかという仮説があらかじめ研究者によってたてられていたとしよう。とりわけ、(a)その動詞の連用形が 1 モーラなのか、(b)意図性を持つ動詞であるかどうか、という二つの要因が選択傾向に影響を与えるのではないかと、推察されていたとし、そのために、各動詞（各行）には、その動詞にまつわる二つの性質がダミー変数として、コーディングされている。

○ 散布図

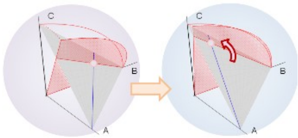


○ Rのコード

```
library(VGAM)
fit = vglm(cbind(y1, y2, y3) ~ x1 + x2, family = multinomial, data=bccwj)
summary(fit)
```

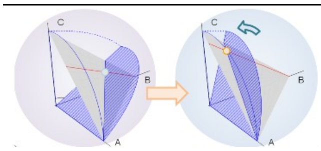
○ 推定結果

◆ A（御…になる）／ C（御…なさる）



	推定値	標準誤差	z 値	p 値	
切片	-0.38	0.08	-4.98	6.48e-07	***
x_1（1 モーラ）	0.10	0.24	0.42	0.67	
x_2（意図性）	-0.18	0.09	-1.95	0.05	.

◆ B（…なさる）／ C（御…なさる）



	推定値	標準誤差	z 値	p 値	
切片	-1.35	0.10	-13.02	2e-16	***
x_1（1 モーラ）	3.32	0.16	20.48	2e-16	***
x_2（意図性）	0.29	0.12	2.47	0.01	*

○ 解釈

プラス／マイナスの符号が
どちらを選択しやすいかという選好を表している

① 切片

2 モーラ以上、非意図性動詞では「C 御…なさる」指向

② 切片＋係数

B と C では、1 モーラで、意図性の動詞のとき-1.35 + 3.32 + 0.29 = 2.26 ポイントの B「…なさる」指向性が見られる

📖 ノート２　統計量：最尤推定法

○ 基本的な考え方

> **最尤推定法** Maximum Likelihood Estimation
>
> 母集団にモデルを仮定し、データ（標本）が得られる確率が最大　となる統計量をパラメータの推定値と考える

ケース１：切片モデル
（母集団のモデルに独立変数が存在しない）

ケース２：提案モデル
（母集団のモデルに独立変数がある）

小標本特性と大標本特性

① 小標本特性

× サンプルサイズが小さくても推定量が持っている特性

○ 繰り返し標本を抽出したとき推定量が持っている性質

（例１）不偏性

（例２）有効性

② 大標本特性

サンプルサイズが大きいとき推定量が持っている特性

（例１）漸近的不偏性
サンプルサイズが大きいとき、バイアスが限りなく０に近づくという性質。

（例２）漸近的有効性
サンプルサイズが大きいとき、不偏な推定量の中で分散が最小となるという性質。

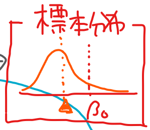
（例３）一致性
サンプルサイズが大きいとき、真の値との間に差が生じる確率が小さくなるという性質。

Ⓟ 推定量（おさらい）

標本 $y_1, \dots, y_n$ ← 母集団 $\beta_0, \beta_1, \dots$

↓

統計量 → 標本分布



b_0^*　$b_0^☆$　$b_0^?$　候補は無数

（ステップ１）推定量の候補群

基準　こういう特徴は良い特徴！

（ステップ２）推定量の選抜

最尤推定量の性質

（性質１）一致性

（性質２）不偏性は持たない！

（性質３）漸近的有効性

（性質４）不変性

不変性 Invariance

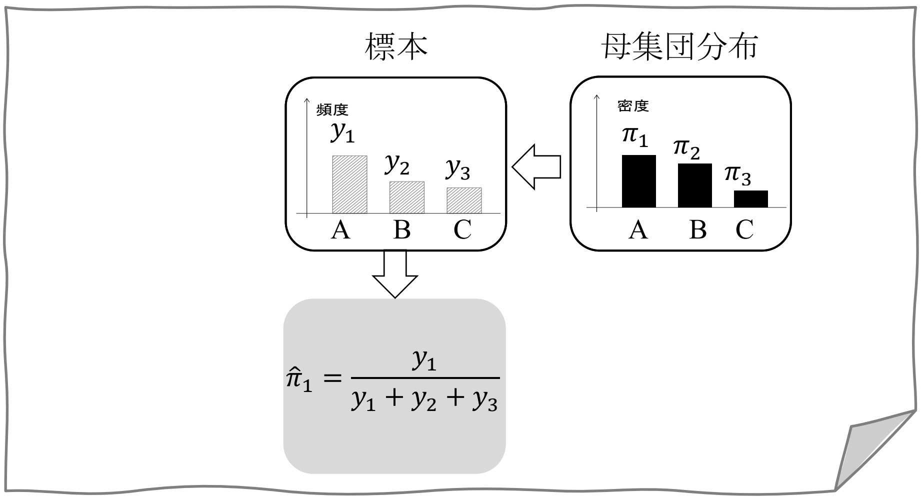
これは、変数変換に影響を受けないという性質。

すなわち、

θ の最尤推定量が$\hat{\theta}$ であるとき、

$g(\theta)$の最尤指定量が$g(\hat{\theta})$で計算できる。

■ ケース１：切片モデル（母集団のモデルに独立変数なし）



※ 独立変数がないので、この場合、推定対象となる母集団のパラメータは　πたち　である。

例１：ベルヌーイ／二項分布に従うデータ

あるコーパスで否定形丁寧語を調べると、次の結果が得られた。

「ません」30 回,「ないです」20 回

これらのデータは、全て独立で同一の二項分布に従って生成されていると想定し、「ないです」が日本語という母集団全体で生産される割合をπとする。このとき、πの最尤推定値は以下の通りである。

$$\hat{\pi} = 0.4$$

例２：カテゴリカル／多項分布に従うデータ

あるコーパスで尊敬語を調べると、次の結果が得られた。

「お…になる」50 回,「お…なさる」20 回,「…なさる」30 回

これらのデータは、全て独立で同一の多項分布に従って生成されていると想定し、「お…になる」が日本語という母集団全体で生産される割合をπ_1とする。同じように、「お…なさる」と「…なさる」が生産される母集団の確率をπ_2とπ_3とする。このとき、π_1からπ_3の最尤推定値は以下の通りである。

$$\hat{\pi}_1 = 0.5,\ \hat{\pi}_2 = 0.2,\ \hat{\pi}_3 = 0.3$$

手順１　確率を数式で表す

（１）同時確率密度関数

$$p(y|\,\mathrm{n},\pi) = {}_nC_y\,\pi^y(1-\pi)^{\mathrm{n-y}}$$

（２）対数同時確率密度関数

$$\log p(y|\,\mathrm{n},\pi) = \log\left[{}_nC_y\,\pi^y(1-\pi)^{n-y}\right]$$

ポイント：密度関数は「全知全能の視点」

密度関数は、πがどのような値であるのか知っている場合に利用できるものです。もちろん、その形が理論的に${}_nC_y\,\pi^y(1-\pi)^{\mathrm{n-y}}$になるということはわかります。しかし、人間である研究者にはπの真の値はわかりません（nの値は観測する側ももちろん把握できます）。わからないので、πに具体的な値を代入できない以上、数式は未完成なんです。この意味で、母集団に想定される密度関数とは、「全知全能の存在だったとしたらπの値が分かるから、そのような存在にとっては利用することのできる関数」でしかありません。

　ここでは、我々もそういう全知全能の視点に立つことができ π が実は 0.1 だと知っていたらという仮定に立ってみましょう。その時、$n = 50$だった場合、密度関数は、次のようになります。

$$p(y|\,50, 0.1) = {}_{50}C_y\,0.1^y(1-0.1)^{50-y}$$

すると、y に好きな値を代入すると確率を返してくれるようになります。

$y = 10$のとき　$p(10|\,50, 0.1) = {}_{50}C_{10}\,0.1^{10}(1-0.1)^{50-10}$
$= 0.01518333 \ldots$

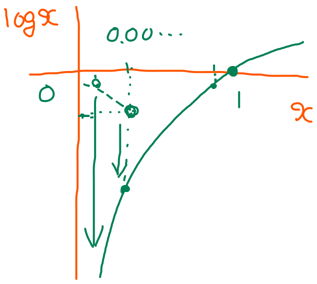
$y = 20$のとき　$p(20|\,50, 0.1) = {}_{50}C_{20}\,0.1^{20}(1-0.1)^{50-20}$
$= 0.00000001997862 \ldots$

ポイント：対数密度関数は「小さすぎる値への対策」

$y = 10$のとき　$\log\ p(10|\,50, 0.1) = \log {}_{50}C_{10}\,0.1^{10}(1-0.1)^{50-10}$
$= -4.18 \ldots$

$y = 20$のとき　$\log\ p(20|\,50, 0.1) = \log {}_{50}C_{20}\,0.1^{20}(1-0.1)^{50-20}$
$= -17.73 \ldots$

Ⓟ 対数のグラフ



手順２　尤度を数式で表す

（１）尤度関数 Likelihood Function

同時確率密度関数について、πが変数であるかのように読み替えたもの。

$$L(y|\, n,\pi) = p(y|\, n,\pi)$$
$$= {}_nC_y\, \pi^y (1-\pi)^{n-y}$$

（２）対数尤度関数 Log Likelihood Function

$$\ell(y|\, n,\pi) = \log L(y|\, n,\pi)$$
$$= \log p(y|\, n,\pi)$$
$$= \log\left[{}_nC_y\, \pi^y (1-\pi)^{n-y}\right]$$

ポイント：尤度関数は「人間の視点」

「研究者（人間）」が手にすることができるのは、n と y の値だけです。そこで、この二つを固定して、分かっていないπの方を動かしてみる（＝変数として取り扱う）ことを考えます。このように、密度関数の式をパラメータ関数として「人間目線」で捉えなおしたものことを尤度関数と言います。例えば、例１においては（50 回試行を繰り返して 20 回「ないです」が出た場合）尤度関数は、次のようになります。

$$p(20|\, 50,\pi) = {}_{50}C_{20}\, \pi^{20}(1-\pi)^{50-20}$$

これは、πの値を与えると、今回のデータが得られる確率を返してくれます。

$\pi = 0.1$のとき　$p(20|\, 50, 0.1) = {}_{50}C_{10}\, 0.1^{10}(1-0.1)^{50-10}$
$= 0.00000001997862 \ldots$

$\pi = 0.2$のとき　$p(20|\, 50, 0.2) = {}_{50}C_{20}\, 0.2^{20}(1-0.2)^{50-20}$
$= 0.0006117722 \ldots$

ポイント：対数尤度関数は「小さすぎる値への対策」

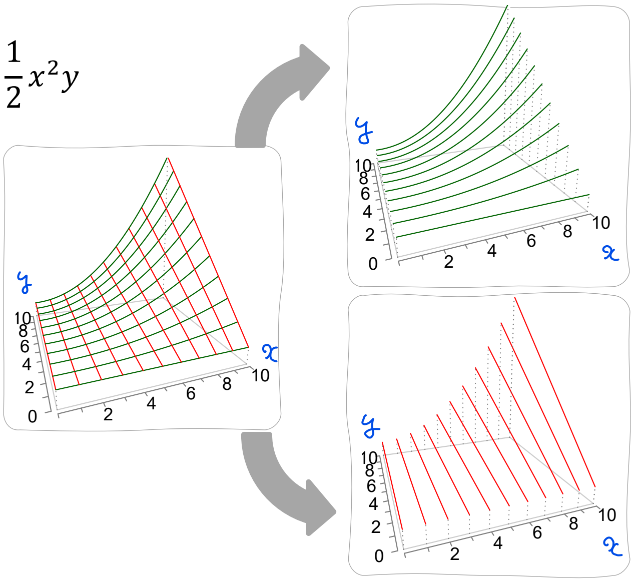
$\pi = 0.1$のとき　$\log\ p(20|\, 50, 0.1) = \log {}_{50}C_{10}\, 0.1^{10}(1-0.1)^{50-10}$
$= -17.729 \ldots$

$y = 20$のとき　$\log\ p(20|\, 50, 0.2) = \log {}_{50}C_{20}\, 0.2^{20}(1-0.2)^{50-20}$
$= -7.399 \ldots$

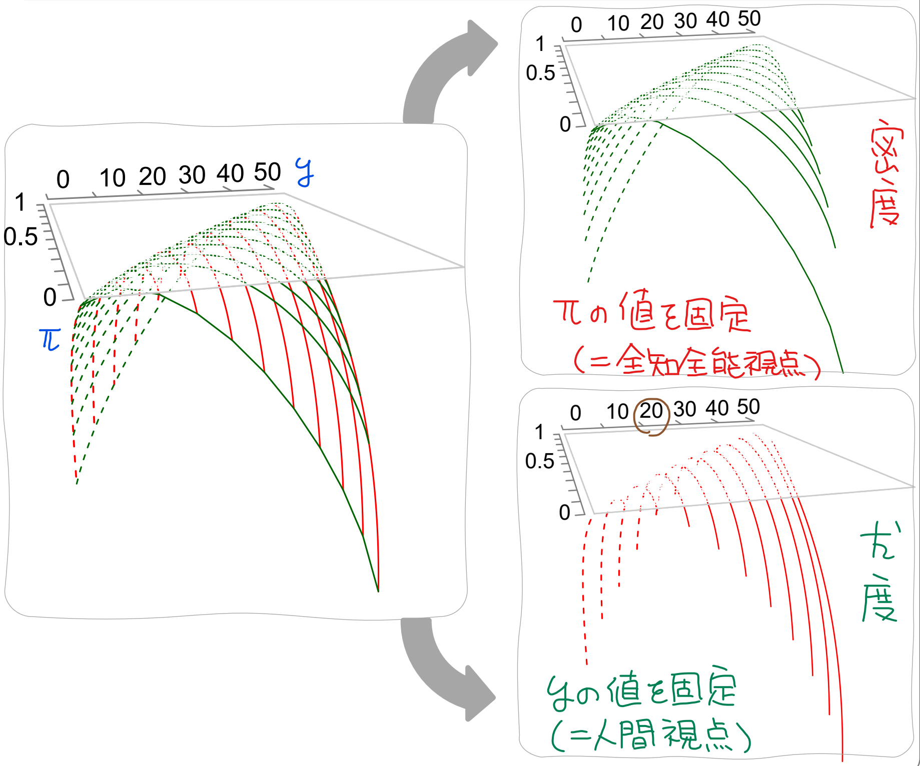
変数と見なすものを変えるということ

二つの変数を引数にとる次のような関数 f がある。この関数は x についてみると二次関数。y については一次関数。

$$f(x, y) = \frac{1}{2}x^2 y$$



（対数）密度関数と（対数）尤度関数の関係



手順３　尤度を最大にするパラメータの値を求める

（１）最尤推定値 Maximum Likelihood Estimate

これは、尤度関数を最大にするパラメータの値のこと。他の推定値と区別する場合は、添え字に MLE を付ける。

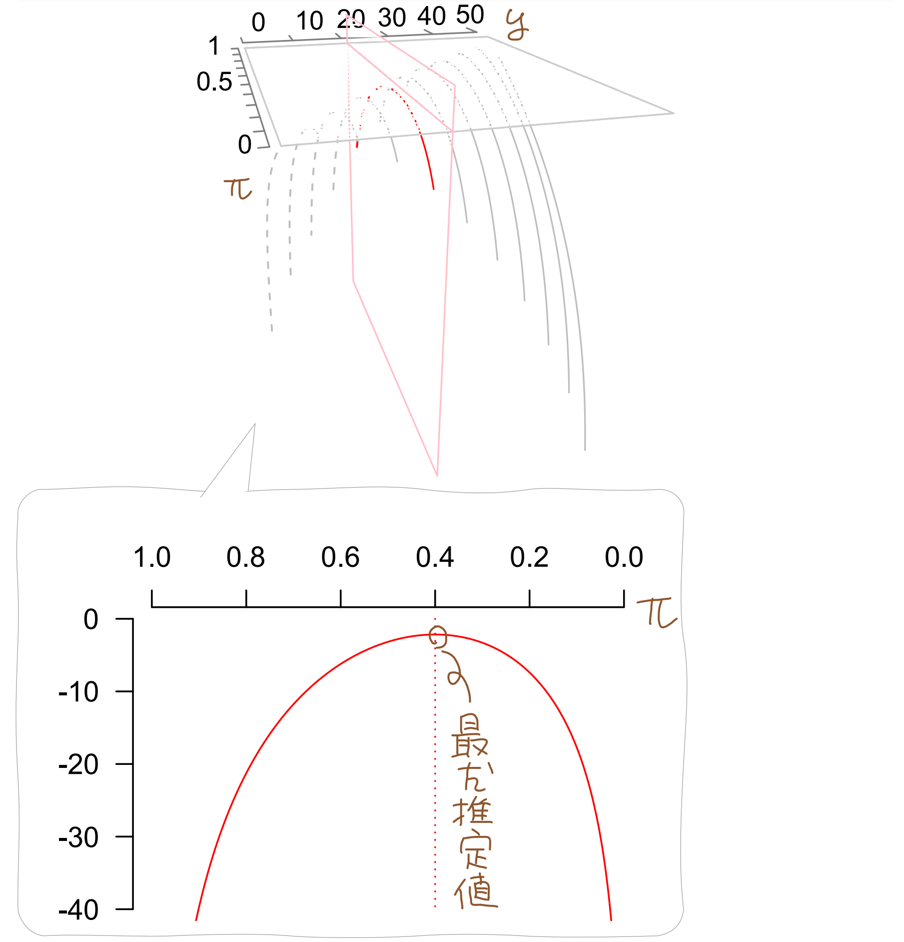
$$\hat{\pi}_{MLE} = \arg\max_{\pi} \ell(y|\, n, \pi)$$

（２）スコア関数 Score Function

これは、対数尤度関数の導関数。これがゼロになる値を探す。

$$u(\pi) = \frac{\partial}{\partial \pi} \ell(y|\, n, \pi)$$

（対数）尤度を最大にするということ

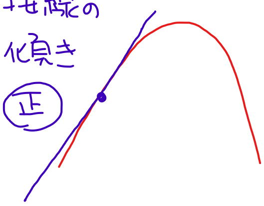


Ⓟ 接線を考える!

(ケース1)

接線の傾き

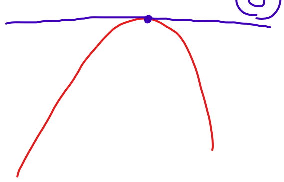
正



(ケース2)

接線の傾き

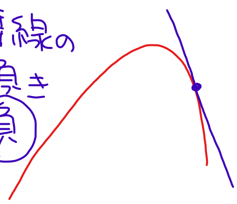
0



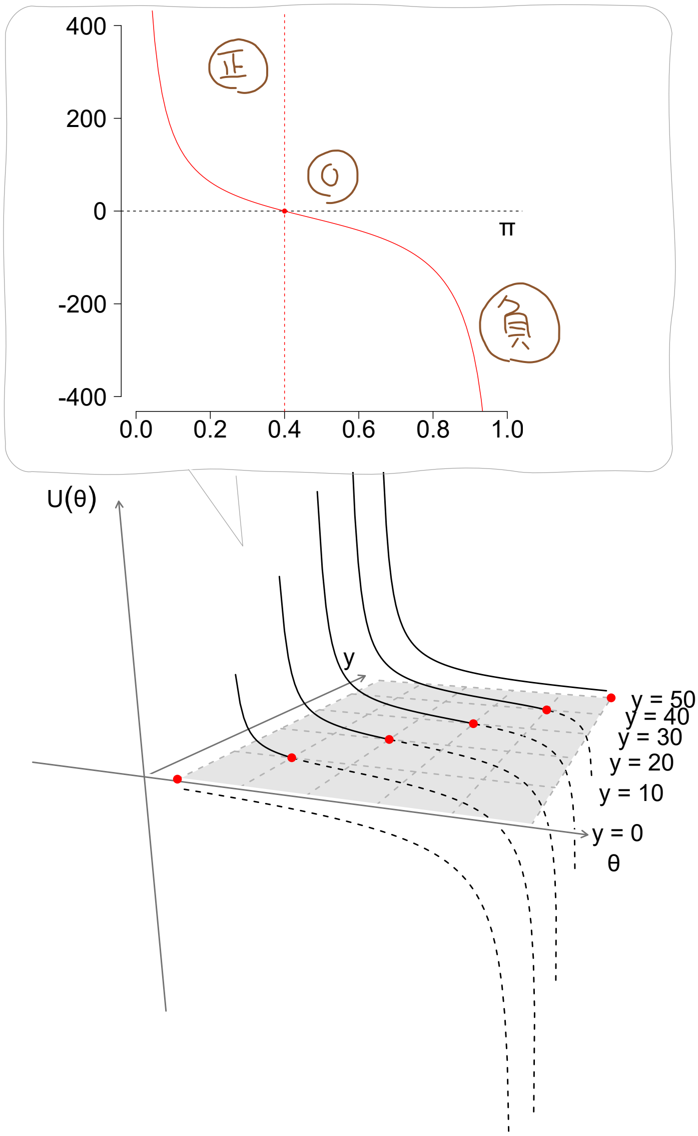
(ケース3)

接線の傾き

負



スコア関数のグラフ



スコア関数の性質

① 期待値

$$E[u(\pi)] = 0$$

② 分散

$$Var[u(\pi)] = E[u(\pi)^2]$$

※スコア関数の分散をフィッシャー情報量$\mathfrak{I}_n(\pi)$と呼ぶ。

数値計算による点推定：ニュートン・ラフソン法

① ヘシアン Hessian

これはスコア関数の導関数。

$$h(\pi) = u'(\pi)$$

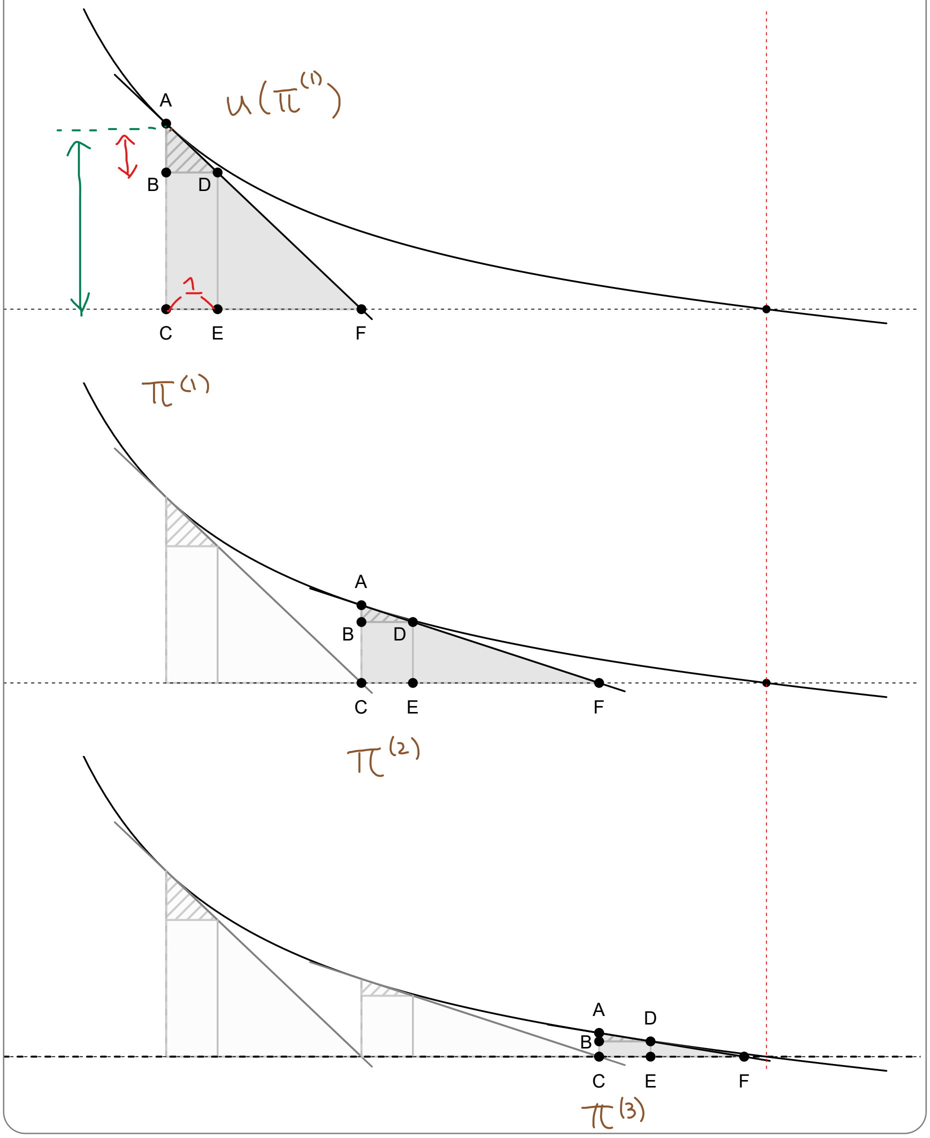
② ニュートン・ラフソン法

次の漸化式で x 軸と交わるところを探すアルゴリズム。

$$\pi^{(m)} = \pi^{(m-1)} - \frac{u(\pi)^{(m-1)}}{u'(\pi)^{(m-1)}}$$

← ACの長さ

← ABの長さ



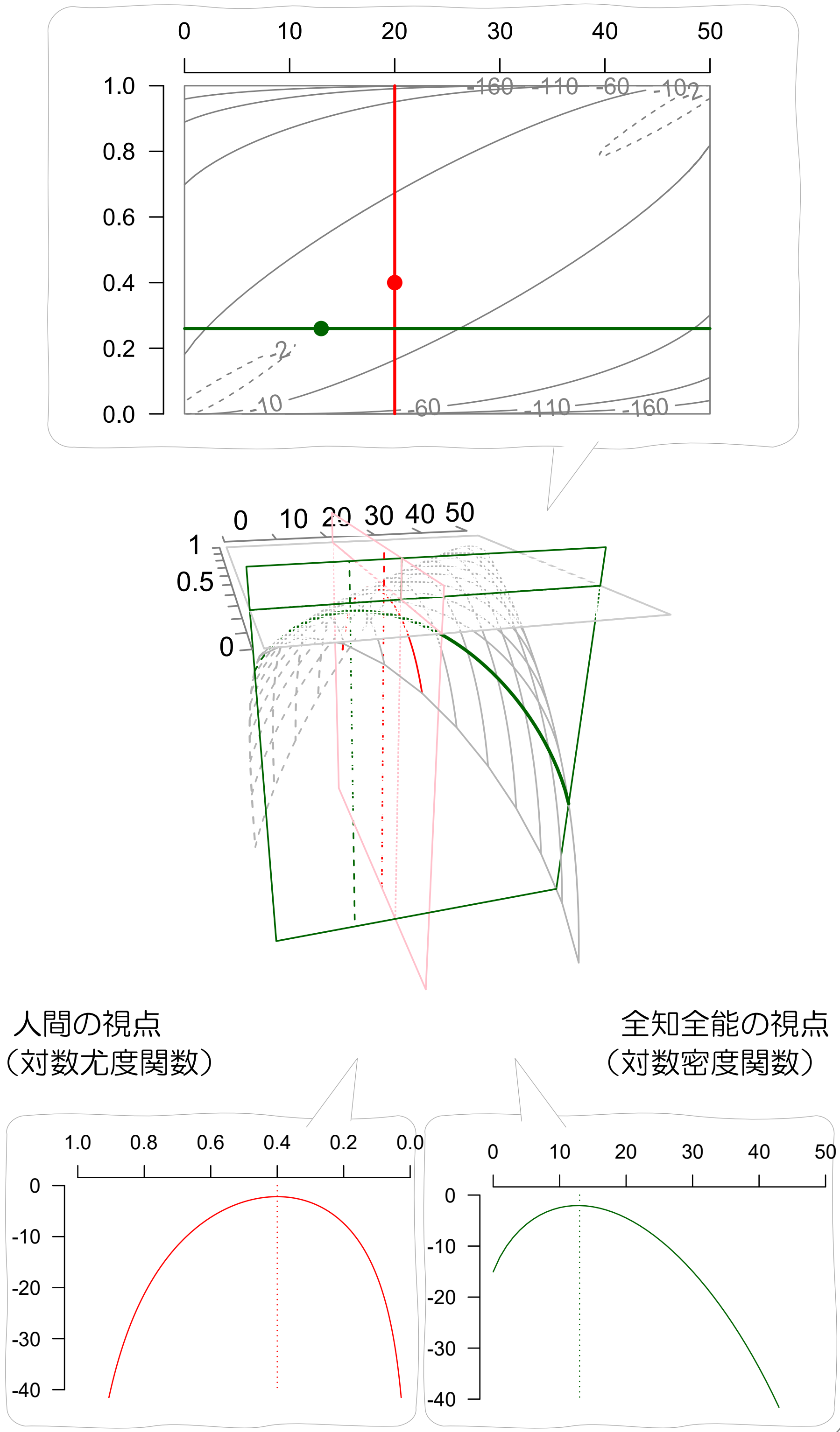
Ⓟ 三角形の相似

△ABDと△ACFと相似.

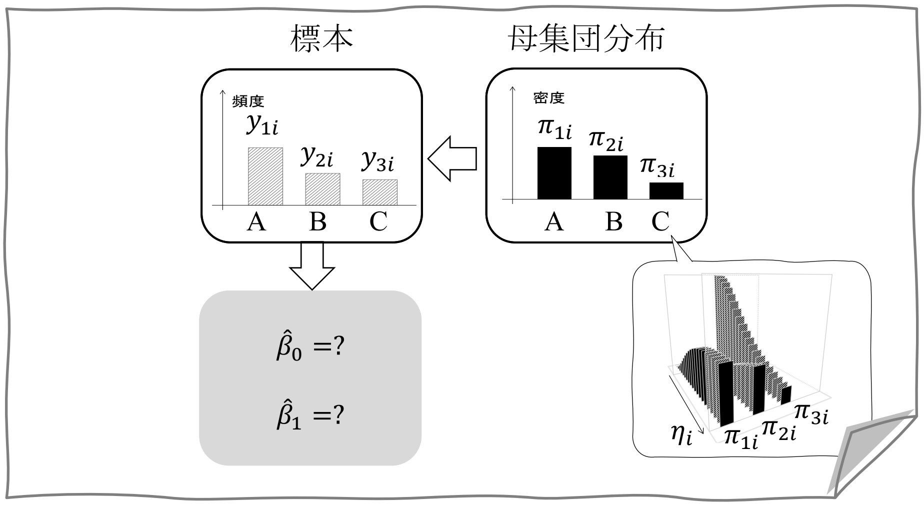
$AB : AC = BD : CF$

$$CF = \frac{AC \times BD}{AB} = \frac{AC}{AB}$$

最尤推定値と真の値の関係



■ ケース２：母集団のモデルに独立変数がある場合



手順１　確率を数式で表す

（１）同時確率密度関数

$$
\begin{aligned}
&p(\boldsymbol{y}|\,\beta_0,\beta_1,n_i)\\
&\qquad = p(y_1|\,\beta_0,\beta_1,n_i)\times\ p(y_2|\,\beta_0,\beta_1,n_i)\times\cdots\\
&\qquad\qquad\qquad\qquad\qquad\times p(y_n|\,\beta_0,\beta_1,n_i)\\
&\qquad = \prod_{i=1}^{N}\binom{n_i}{y_i}F(\beta_0+\beta_1x_i)^{y_i}\left(1-F(\beta_0+\beta_1x_i)\right)^{n_i-y_i}
\end{aligned}
$$

（２）対数同時確率密度関数

$$
\begin{aligned}
&\log p(\boldsymbol{y}|\,\beta_0,\beta_1,n_i)\\
&\qquad = \log\ [p(y_1|\,\beta_0,\beta_1,n_i)\times\ p(y_2|\,\beta_0,\beta_1,n_i)\times\cdots\\
&\qquad\qquad\qquad\qquad\qquad\times p(y_n|\,\beta_0,\beta_1,n_i)]\\
&\qquad = \log p(y_1|\,\beta_0,\beta_1,n_i)+\log\ p(y_2|\,\beta_0,\beta_1,n_i)+\cdots\\
&\qquad\qquad\qquad\qquad\qquad+\log p(y_n|\,\beta_0,\beta_1,n_i)\\
&\qquad = \sum_{i=1}^{n}\log\left[\binom{n_i}{y_i}F(\beta_0+\beta_1x_i)^{y_i}\left(1-F(\beta_0+\beta_1x_i)\right)^{n_i-y_i}\right]
\end{aligned}
$$

手順２　尤度を数式で表す

（１）尤度関数 Likelihood Function

$$L(\boldsymbol{y}|\,\beta_0,\beta_1,n_i)$$
$$=\prod_{i=1}^{N}\binom{n_i}{y_i}F(\beta_0+\beta_1x_i)^{y_i}\left(1-F(\beta_0+\beta_1x_i)\right)^{n_i-y_i}$$

（２）対数尤度関数 Log Likelihood Function

$$\ell(y|\,\beta_0,\beta_1,n_i)\quad=\log L(\boldsymbol{y}|\,\beta_0,\beta_1,n_i)$$

$$=\log\left[\prod_{i=1}^{N}\binom{n_i}{y_i}F(\beta_0+\beta_1x_i)^{y_i}\left(1-F(\beta_0+\beta_1x_i)\right)^{n_i-y_i}\right]$$
$$=\sum_{i=1}^{n}\log\left[\binom{n_i}{y_i}F(\beta_0+\beta_1x_i)^{y_i}\left(1-F(\beta_0+\beta_1x_i)\right)^{n_i-y_i}\right]$$

手順３　尤度を最大にするパラメータの値を求める

$$\left(\hat{\beta}_{0_{MLE}},\hat{\beta}_{1_{MLE}}\right)=\arg\max_{\beta_0,\beta_1}\ell(y|\,\beta_0,\beta_1,n_i)$$

カテゴリカル／多項分布の場合の同時確率密度関数

$$y_i\sim Multinom(y_i|\,\beta_0,\beta_1,n_i)$$

$$p(y_i|\,\beta_0,\beta_1,n_i)\quad=\quad\frac{n_i!}{y_1!\,y_2!\cdots y_V!}\pi_{1i}^{y_{1i}}\pi_{2i}^{y_{2i}}\cdots\pi_{Vi}^{y_{Vi}}$$

標本 ＝ N 個

$$p(\boldsymbol{y}|\,\beta_0,\beta_1,n_i)\quad=\prod_{i=1}^{N}\frac{n_i!}{y_1!\,y_2!\cdots y_V!}\pi_{1i}^{y_{1i}}\pi_{2i}^{y_{2i}}\cdots\pi_{Vi}^{y_{Vi}}$$