KYOTO UNIVERSITY

統計的モデリング基礎⑨
～さまざまな確率モデル～
（ロジスティック回帰の発展と生存期間のモデル）

鹿島久嗣
（情報学科 計算機科学コース）

DEPARTMENT OF INTELLIGENCE SCIENCE
AND TECHNOLOGY

今回の話題：
さまざまな確率モデル

- ロジスティック回帰モデルの発展：
  - 多クラスロジスティック回帰モデル
  - 順序回帰モデル
  - ランキング（一対比較）モデル
  - ニューラルネットワーク
- 生存期間のモデル
  - ハザード関数
  - 生存期間モデルの最尤推定

ロジスティック回帰モデルの発展

ロジスティック回帰モデルの発展：
従属変数の型に合わせた発展

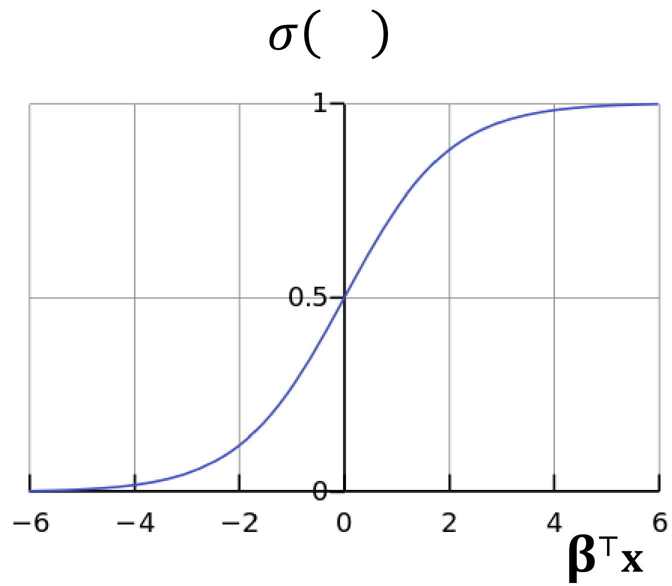
- 確率モデルはデータの生成モデル
- 分析対象のデータに合わせてモデルが変わる
  - 質的変数（ダミー変数0, 1）の場合：ロジスティック回帰モデル
    → 選択肢が複数の場合：多値ロジスティック回帰
  - 量的変数（連続値）の場合：線形回帰モデル
    → 順序尺度（例えば5段階評価）の場合：順序回帰
  - 比較：一対比較のモデル（例：2つのうちどちらがよいか？）
- 多層化による非線形モデルの実現：ニューラルネットワーク

ロジスティック回帰モデル：
ダミー変数を従属変数とするモデル

- 従属変数Yが（2値の）ダミー変数であるモデル
- ロジスティック回帰モデル： $Y = +1$となる確率

$$P(Y = 1|\mathbf{x}, \boldsymbol{\beta}) = \frac{1}{1+\exp(-\boldsymbol{\beta}^\top \mathbf{x})} = \sigma(\boldsymbol{\beta}^\top \mathbf{x})$$

  - σ：ロジスティック関数（$\sigma: \mathbb{R} \to (0,1)$）



多値ロジスティック回帰モデル：多値の従属変数を説明するモデル

- 従属変数Yが多値である場合（$Y \in \mathcal{Y} = \{1,2,\dots,k\}$）
  - ただし、$1,2,\dots,k$ の並び順に意味はないことに注意
- 多値ロジスティック回帰モデル：$Y = y$である確率
  - 各$y \in \mathcal{Y}$ごとにパラメータ$\boldsymbol{\beta}_y$をもつ

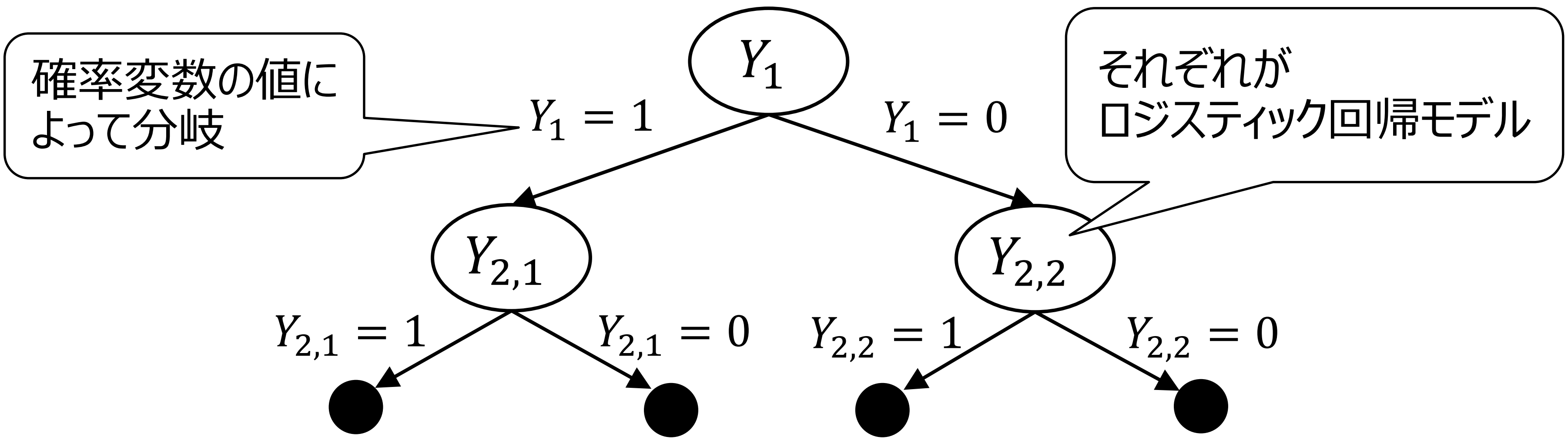
いわゆるソフトマックス関数

$$P\left(Y = y \middle| \mathbf{x}, \{\boldsymbol{\beta}_{y'}\}_{y' \in \mathcal{Y}}\right) = \frac{\exp\left(\boldsymbol{\beta}_y^\top \mathbf{x}\right)}{\sum_{y' \in \mathcal{Y}} \exp\left(\boldsymbol{\beta}_{y'}^\top \mathbf{x}\right)}$$

  - $\mathcal{Y} = \{+1, -1\}$のときは$\boldsymbol{\beta} = \boldsymbol{\beta}_{+1} - \boldsymbol{\beta}_{-1}$とすると通常のロジスティック回帰に一致

多段ロジスティック回帰モデル：
多段に連結されたロジスティック回帰モデル

- 複数の連続した従属変数：
  - 階層的な分類（大カテゴリ→中カテゴリ→小カテゴリ）
  - 段階的な意思決定プロセス（購入の有無→商品）
- ロジスティック回帰モデルを連結する



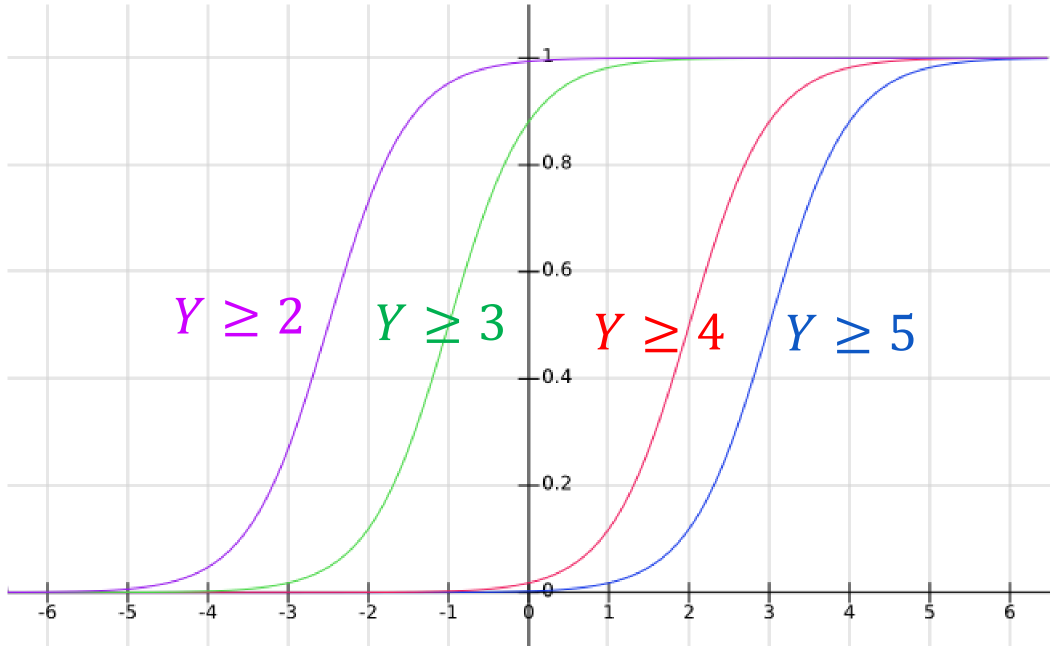
順序回帰：
順序尺度をもつ従属変数を説明するモデル

この順序に意味がある場合
（5段階評価など）

- Yが多値で順序尺度をもつ場合（$Y \in \mathcal{Y} = \{1,2,\ldots,k\}$）
- $Y \geq y$となる確率を与えるモデルを考える：

$$P(Y \geq y|\mathbf{x}, \boldsymbol{\beta}) = \frac{1}{1+\exp\left(-\boldsymbol{\beta}^\top \mathbf{x} + \alpha_y\right)} = \sigma(\boldsymbol{\beta}^\top \mathbf{x} - \alpha_y)$$

  - パラメータ$\boldsymbol{\beta}$はyによらず共通
  - 「切片」がyごとに異なる：
    $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_k$
    （平行移動）



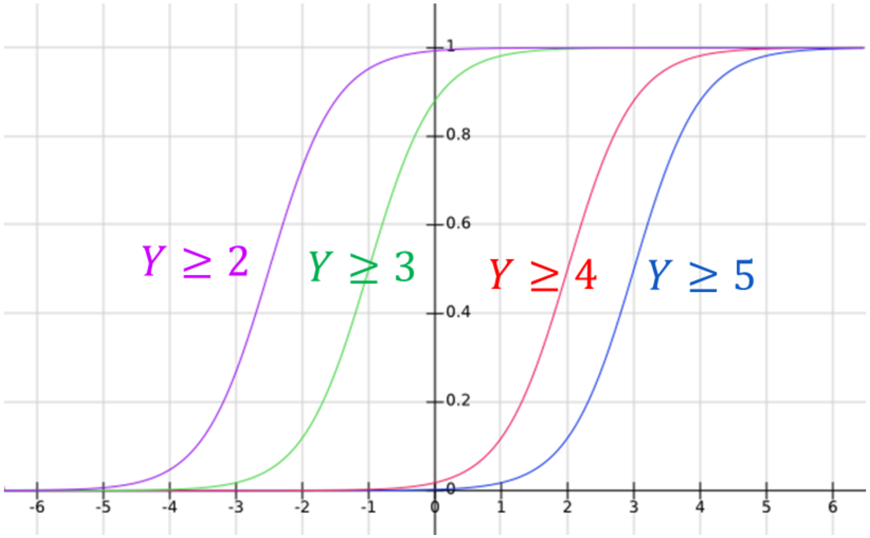
順序回帰モデル：
順序尺度をもつ従属変数を説明するモデル

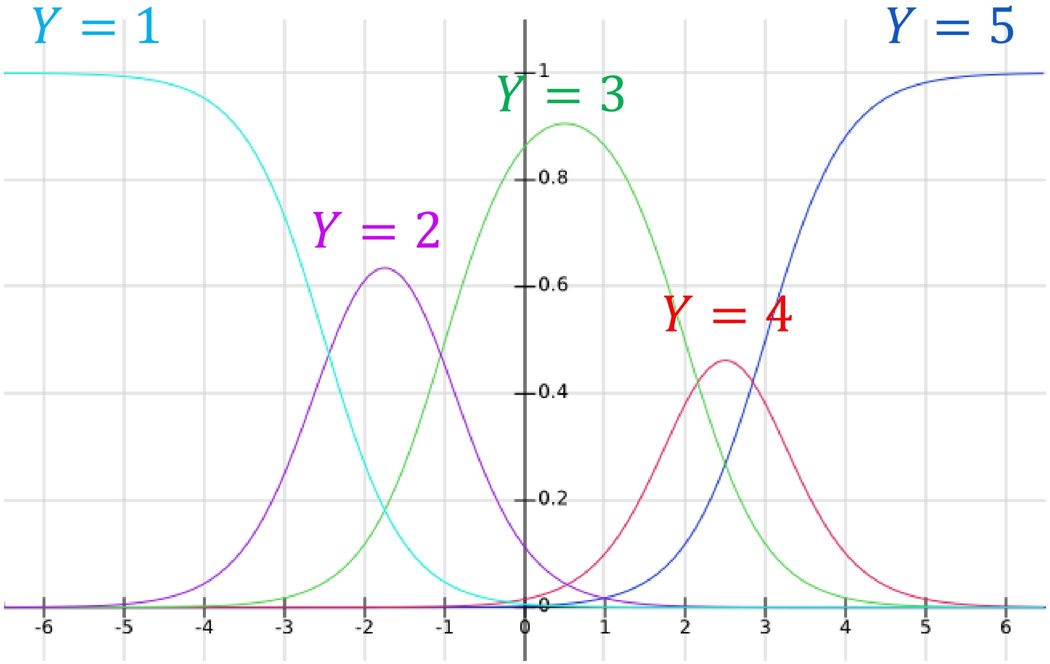
- 順序回帰モデルは$Y \geq y$となる確率がロジスティック回帰モデル：

$$P(Y \geq y | \mathbf{x}, \boldsymbol{\beta}) = \sigma(\boldsymbol{\beta}^\top \mathbf{x} - \alpha_y)$$

- $Y = y$ である確率は上記モデルの差を用いて表現できる：

$$P(Y = y | \mathbf{x}, \boldsymbol{\beta}) = P(Y \geq y | \mathbf{x}, \boldsymbol{\beta}) - P(Y \geq y - 1 | \mathbf{x}, \boldsymbol{\beta})$$



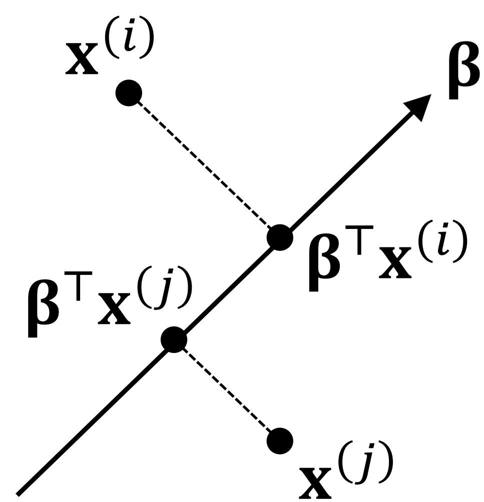


ランキングの確率モデル：一対比較のモデル

- 感性評価などは絶対評価を与えにくい
- 一対比較：「どちらがよいか」のほうが答えやすい
- データiがデータjよりも上位である（$i \succ j$）確率：

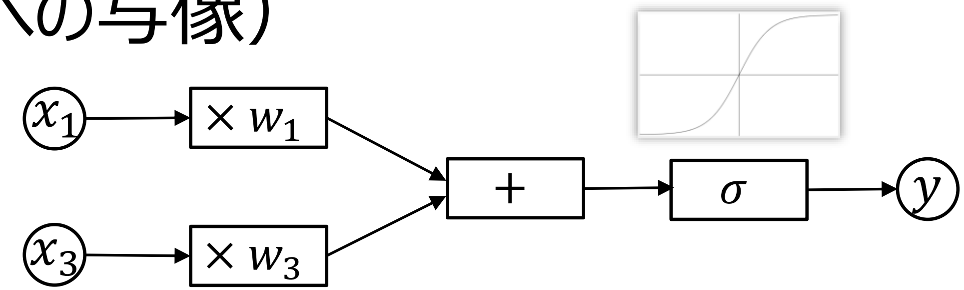
$$P\left(i \succ j \middle| \mathbf{x}^{(i)}, \mathbf{x}^{(j)}, \boldsymbol{\beta}\right) = \frac{\exp\left(\boldsymbol{\beta}^\top \mathbf{x}^{(i)}\right)}{\exp(\boldsymbol{\beta}^\top \mathbf{x}^{(i)}) + \exp(\boldsymbol{\beta}^\top \mathbf{x}^{(j)})}$$
$$= \sigma(-\boldsymbol{\beta}^\top\left(\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right))$$

  - 2つのデータの独立変数の差を、ペアに対する独立変数としたロジスティック回帰モデル

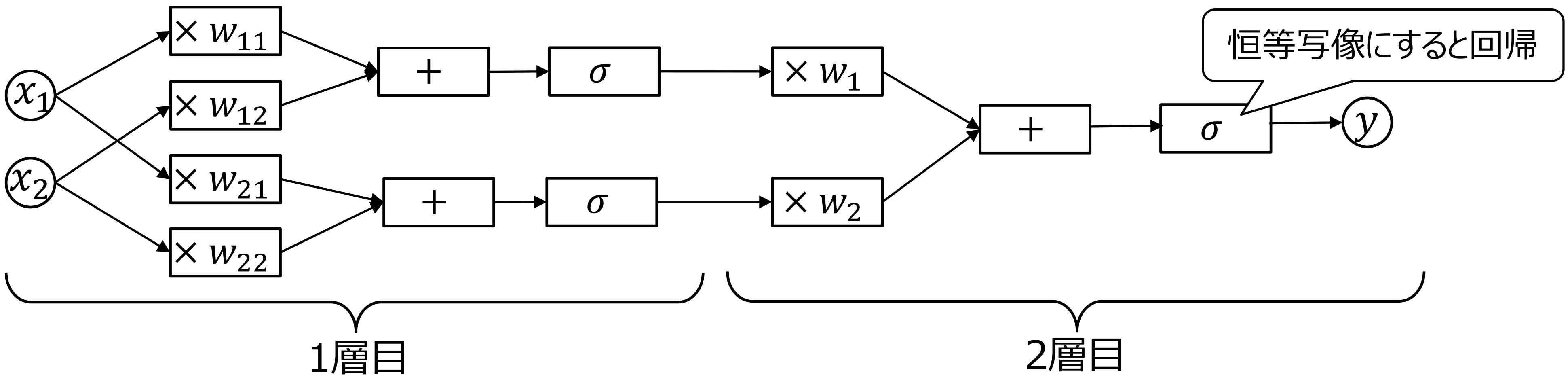


ニューラルネットワーク：
ロジスティック回帰の多層化

- ロジスティック回帰は線形回帰モデルの出力に非線形写像を適用（(0,1)区間への写像）



- ニューラルネットワークはこれを多層化し非線形性を導入したもの
  - 非線形写像は必ずしもロジスティック関数である必要はない



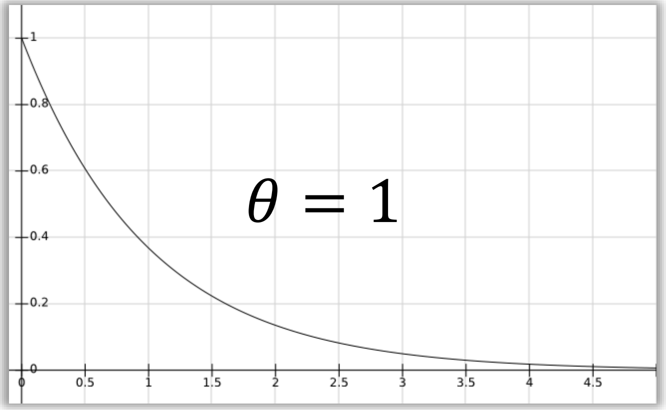
生存期間のモデル

生存期間のモデル：
期間を確率変数とするモデル

- 期間（非負の実数）を確率変数とするようなモデル：
  - 商品の寿命、患者の生存期間、…
  - 一方、ポアソン分布は回数（非零の整数）のモデル
- 生存期間の確率変数T：$\Pr(T \leq t) = F(t) = \int_0^t f(\tau)\,\mathrm{d}\tau$
  - 確率密度関数$f(t)$：時刻tまで生存していて、時刻$t+\Delta t$までの間の死亡確率が$f(t)\Delta t$（「時刻tまで生存」かつ「$t \sim t+\Delta t$で死亡」）
    - $f(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t}$
  - $\Pr(T > t) = S(t) = 1 - F(t)$：時刻$t$以降も生存する確率（少なくとも時刻$t$までは生存する）
    - 生存関数

指数分布モデル：
もっとも単純な生存期間のモデル

- $f(t)$ ：時刻tまで生存していて、時刻$t+\Delta t$までの間の死亡確率が$f(t)\Delta t$ （「時刻tまで生存」かつ「 $t \sim t+\Delta t$ で死亡」）
- 指数分布モデル：$f(t) = \theta \exp(-\theta t)$
  - $\theta > 0$ ：モデルパラメータ



- 生存期間T：

$$\Pr(T \leq t) = F(t) = \int_0^t f(\tau)\,\mathrm{d}\tau = 1 - \exp(-\theta t)$$

  - $\mathrm{E}[T] = \frac{1}{\theta}, \mathrm{Var}[T] = \frac{1}{\theta^2}$
- 独立変数によってパラメータが変わる場合：$\theta = \exp(\boldsymbol{\beta}^\top \mathbf{x})$

ハザード関数：
ある時刻の死亡リスクを表す関数

- $f(t)$ は「時刻tまで生存している」かつ「次の瞬間に死亡する」可能性を表す（ちょっと解釈しにくい）

- 瞬間瞬間の死亡リスクをみたほうがわかりやすい？
  - 「時刻tまで生存している」という条件のもとでの「次の瞬間に死亡する」可能性（条件付確率）をみる

- ハザード関数： $h(t) = \dfrac{f(t)}{S(t)} = -\dfrac{\mathrm{d}\log S(t)}{\mathrm{d}t}$

$S(t) = 1 - F(t)$ ：
生存関数（少なくとも時刻tまでは生存する確率）

- ハザード関数の時間変化：
$\dfrac{\mathrm{d}h(t)}{\mathrm{d}t} > 0$ のとき、リスクが時間とともに増加（ < 0 であれば減少）

ワイブル分布：指数分布の一般化

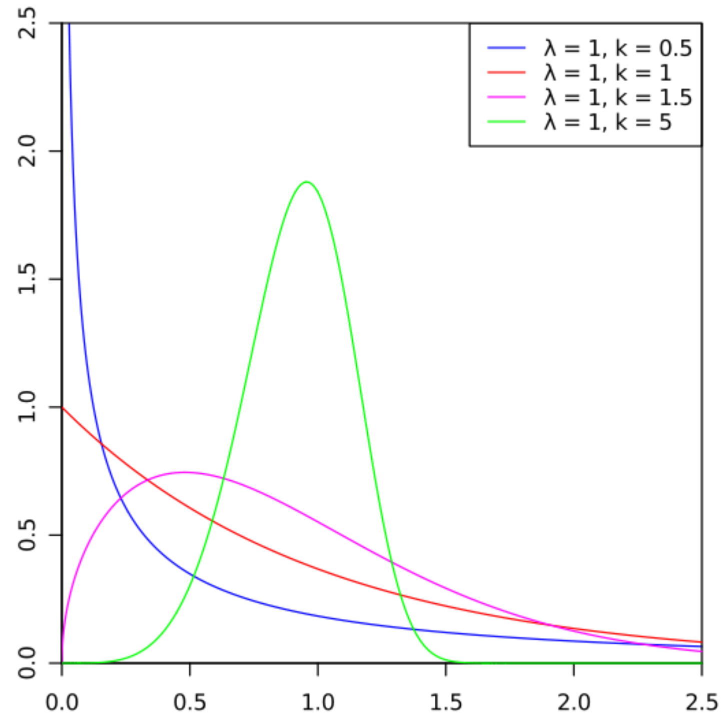
- 指数分布モデルはリスクが時間に関わらず一定
  - 指数分布のハザード関数： $h(t) = \frac{f(t)}{S(t)} = \frac{\theta \exp(-\theta t)}{\exp(-\theta t)} = \theta$ (定数)

- ワイブル分布モデル：

$$f(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} \exp\left\{-\left(\frac{t}{\lambda}\right)^{k}\right\}, k, \lambda > 0$$

  - $k = 1$のとき指数分布（$\theta = 1/\lambda$）
    $f(t) = \theta \exp(-\theta t)$

- 独立変数を取り込む場合：
  $\lambda = \exp(\boldsymbol{\beta}^\top \mathbf{x})$



https://en.wikipedia.org/wiki/Weibull_distribution#/media/File:Weibull_PDF.svg

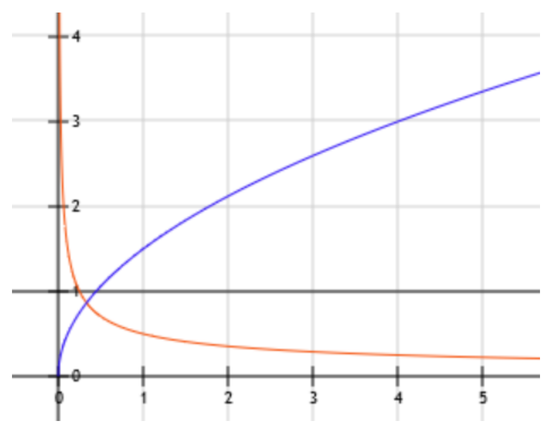
ワイブル分布：
パラメータによってハザード関数の時間的増減が決まる

- ワイブル分布の生存関数：

$$S(t) = \int_t^\infty \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} \exp\left\{-\left(\frac{t}{\lambda}\right)^k\right\} \mathrm{d}t = \exp\left\{-\left(\frac{t}{\lambda}\right)^k\right\}$$

- ハザード関数： $h(t) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1}$
  - $k = 1$ のとき $h(t) = 1/\lambda$
  - $k > 1$ のとき $\frac{\mathrm{d}h(t)}{\mathrm{d}t} > 0$
  - $k < 1$ のとき $\frac{\mathrm{d}h(t)}{\mathrm{d}t} < 0$

kによって決まる



生存時間モデルの最尤推定：
生存期間の確率密度関数$f(t)$を最尤推定

- データ $\{t^{(1)}, t^{(2)}, \dots, t^{(N)}\}$ ：
  - N個の独立な観測（生存期間がちょうど$t^{(i)}$）
- 尤度関数$L(\theta) = \prod_{i=1}^{N} f\left(t^{(i)}\right)$

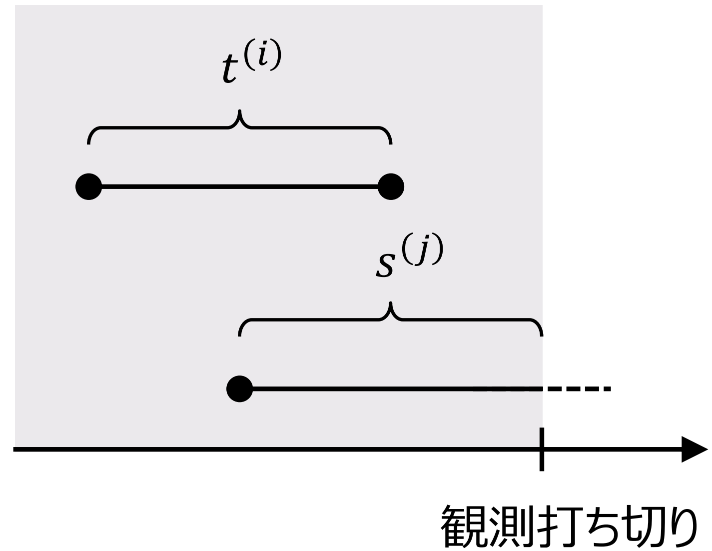
言い換えれば、$t^{(i)}$まで生きていて
次の瞬間死亡したという観測データ

  - 確率密度関数$f(t)$ ：時刻tまで生存していて、時刻$t + \Delta t$までの間の死亡確率が$f(t)\Delta t$
  - 指数分布モデルの場合：$L(\theta) = \prod_{i=1}^{N} \theta \exp\left(-\theta t^{(i)}\right) \cancel{\Delta t}$
    - 対数尤度にすると $\log L(\theta) = N \log \theta - \theta \sum_{i=1}^{N} t^{(i)}$
    - 最尤推定量は $\hat{\theta} = \dfrac{N}{\sum_{i=1}^{N} t^{(i)}}$

打ち切りがある場合の最尤推定：
打ち切りデータに対して生存関数 $S(t)$ を当てはめる

- データ $\left\{t^{(1)}, t^{(2)}, \ldots, t^{(N)}\right\} \cup \left\{s^{(1)}, s^{(2)}, \ldots, s^{(M)}\right\}$：
  - N 個の生存期間データに加えて、
    M 個の打ち切りデータ（少なくとも $s^{(i)}$ 期間生存）
- 尤度関数：

$$L(\theta) = \prod_{i=1}^{N} f\left(t^{(i)}\right) \cdot \prod_{j=1}^{M} S\left(t^{(j)}\right)$$



  - 生存関数 $S(t) = \int_t^{\infty} f(t)\mathrm{d}t$：
    少なくとも t 期間は生存している確率
  - 指数分布の場合：$S(t) = \exp(-\theta t)$
    - $\log L(\theta) = N \log \theta - \theta\left(\sum_{i=1}^{N} t^{(i)} + \sum_{j=1}^{M} s^{(j)}\right)$

今回の話題：
さまざまな確率モデル

- ロジスティック回帰モデルの発展：
  - 多クラスロジスティック回帰モデル
  - 順序回帰モデル
  - ランキング（一対比較）モデル
  - ニューラルネットワーク
- 生存期間のモデル
  - ハザード関数
  - 生存期間モデルの最尤推定