数理脳科学
2017 年 6 月 8 日

課題 3: 確率的生成モデルを用いたパターン認識

中間発表（30 秒/人）：6 月 15, 22 日（木），提出締切 7 月 13 日（木）

簡単なパターン認識の実験をしてみよう．

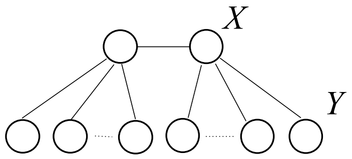

図 1: 確率変数の依存性を描いたグラフ

問題設定： 以下では，信号を $\boldsymbol{x}$，観測データを $\boldsymbol{y}$ とする．$\boldsymbol{x}$ は 2 次元ベクトルで，各要素は $0, 1$ の 2 値をとる．ただし，$\boldsymbol{x}$ は直接観測できず，データ $\boldsymbol{y}$ を通してのみ推論することができる．たとえば，$\boldsymbol{x} = (0, 1)$ という信号にノイズが加わり 7 次元の信号

$$\boldsymbol{y} \quad = \quad (-1.02, -0.83, 0.50, 0.20, 2.82, 0.58, 0.07) \tag{1}$$

が観測された場合を考えよう．$\boldsymbol{y} = (y_1^1, y_2^1, y_3^1, y_1^2, y_2^2, y_3^2, y_4^2)$ と書くと，$\boldsymbol{y}$ は $\boldsymbol{x} = (x_1, x_2)$ をもとに，確率密度関数

$$p(y_j^i|x_i) \quad = \quad \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_j^i - x_i)^2}{2\sigma^2}\right\}, i = 1, 2, j = 1, 2, 3 \tag{2}$$

にしたがい生成されたものと考える（以下では，指定がなければ $\sigma = 1.0$ とする）．0 もしくは 1 の値をとる信号 x_1, x_2 に対し，x_1 を 3 回，x_2 を 4 回，それぞれ観測をおこなって得られたデータが $\boldsymbol{y}$ であると思えばよい．問題は，$\boldsymbol{y}$ を観測し，もとの $\boldsymbol{x}$ の値を推定するのではなく，$\underline{\boldsymbol{x} = (1, 1)\text{ であるか否か}}$ を判断することにある．$\boldsymbol{x}$ が

$$\left\{\begin{array}{lcl} \Pr(X_1 = 0, X_2 = 0) & = & 0.6 \\ \Pr(X_1 = 0, X_2 = 1) & = & 0.1 \\ \Pr(X_1 = 1, X_2 = 0) & = & 0.1 \\ \Pr(X_1 = 1, X_2 = 1) & = & 0.2 \end{array}\right. \tag{3}$$

という確率分布にしたがうことは既知であるとする．以下では，$p_{11} = 0.2$ のように，$\Pr(X_1 = 0, X_2 = 1)$ を p_{01} と表記する場合がある．また，$\Pr(X_1 = x_1, X_2 = x_2)$ を $p(\boldsymbol{x})$ と書いたり，$\Pr(X_1 = 0, X_2 = 1)$ を $p(\boldsymbol{x}_{01})$ などと書く場合がある．このようなデータを生成する能力をもつモデルのことを確率的生成モデルとよび，あらかじめ与えられている $p(\boldsymbol{x})$ を事前分布，$p(\boldsymbol{y}|\boldsymbol{x})$ をデータモデルとよぶ．図 1 は確率変数の依存関係を描いたグラフである．後で見る

ように，このグラフを頭の中にイメージしておけば，事後確率分布 $p(\boldsymbol{x}|\boldsymbol{y})$，周辺分布 $p(\boldsymbol{y})$ など様々な計算や操作が容易になる．

先の例では $\boldsymbol{y}$ は 7 次元ベクトルであった．これを少しだけ一般化し，$\boldsymbol{y} = (y_1^1, y_2^1, \cdots y_{n_1}^1, y_1^2, \cdots, y_{n_2}^2)$ としておこう．つまり，x_1 については n_1 個，x_2 については n_2 個の観測データが得られる場合を考える（先の例は $n_1 = 3, n_2 = 4$）．

データ $\boldsymbol{y}$ を観測し,「$\boldsymbol{x}$ は $(1,1)$ でしょうか？」という問に Yes, No で答える機械を設計しよう．もちろん正答率の高い機械を設計したい．この課題では，自分で多数の例題を作成し，$\boldsymbol{x} = (1,1)$ かどうか推論することで，識別機械の性能を ROC カーブを描き評価する．

ROC 曲線： 識別アルゴリズムは，ベイズの公式をもとに計算した事後確率の値を利用する．この具体的な手順は後で理解するとして，はじめに，識別精度の良さを評価する際に使う ROC 曲線について，その描き方と解釈の仕方を説明しておく．

具体的に考えよう．まずは，$(\boldsymbol{x}^1, \boldsymbol{y}^1), (\boldsymbol{x}^2, \boldsymbol{y}^2), \cdots, (\boldsymbol{x}^{1000}, \boldsymbol{y}^{1000})$ と 1,000 個の例題を作成する．目的は，$\boldsymbol{y}^\alpha$ を観測し，信号源 $\boldsymbol{x}^\alpha$ が $(1,1)$ であるかどうか，Yes ($z^\alpha = 1$), No ($z^\alpha = 0$) で正確に当てる機械を設計することである（$\alpha = 1, \cdots, 1000$）．データ $\boldsymbol{y}$ を入力すると $\boldsymbol{x} = (1,1)$ であると確信する度合い $\mathcal{S} = \mathcal{S}(\boldsymbol{y})$ を返すコンピュータの関数を作ればよい．この機械は，$\mathcal{S} > \theta$，つまり，しきい値 θ よりも $\mathcal{S}$ の値が大きい場合，Yes と答える．当然であるが，機械は間違えることがある．ここで，間違え方には 2 通りあることを確認しておこう．一つは，本当は $\boldsymbol{x} = (1,1)$ であったのに No と答えてしまう場合．もう一つは，$\boldsymbol{x} \neq (1,1)$ であったのに Yes と答えてしまう場合である．これを false positive, FP とよぶ（間違えて Yes と言ってしまった，という意味）．前者は，false negative といってもよいが，通常，同じ意味をもつ correct detect という指標を用いる（正しく検出できた，という意味）．correct detect とは，$\boldsymbol{x} = (1,1)$ から生成されたデータ $\boldsymbol{y}$ を観測したときに Yes と答える場合である．1,000 個の例題があれば，false positive 率（FPR）と correct detect 率（CDR）を計算し，FPR を横軸に，CDR を縦軸にとると図に 1 つの点がプロットできる．$\mathcal{S}$ が $0 \leqq \mathcal{S} \leqq 1$ とすると，しきい値 θ を $0 \leqq \theta \leqq 1$ の間で変化させると，曲線が描ける．これを ROC 曲線とよぶ．FPR と CDR の計算の仕方を以下に整理しておこう．

$$\mathrm{FPR} = \frac{\boldsymbol{x}^\alpha \neq (1,1) \text{ の例題 } \boldsymbol{y}^\alpha \text{に対し, Yes と判定した回数}}{1,000 \text{ 個の例題のうち } \boldsymbol{x}^\alpha \neq (1,1) \text{ である個数}} \tag{4}$$

$$\mathrm{CDR} = \frac{\boldsymbol{x}^\alpha = (1,1) \text{ の例題 } \boldsymbol{y}^\alpha \text{に対し, Yes と判定した回数}}{1,000 \text{ 個の例題のうち } \boldsymbol{x}^\alpha = (1,1) \text{ である個数}} \tag{5}$$

実は，これ以上性能のよい識別機械は作れないという理想の機械が存在する．それは事後

確率 $p(\boldsymbol{x}_{11}|\boldsymbol{y})$ がある値 θ 以上のとき Yes と答える機械である．まずは，その機械の設計方法を示そう．

課題（手計算を通し用語の概念を確認する）

1. 周辺確率 $\Pr(X_1=1), \Pr(X_1=0)$ を求めよ．

$$\Pr(X_1=1) \;=\; \sum_{\tilde{x}_2=0}^{1} \Pr(X_1=1, X_2=\tilde{x}_2) = p_{10}+p_{11} = 0.3 \tag{6}$$

$$\Pr(X_1=0) \;=\; \sum_{\tilde{x}_2=0}^{1} \Pr(X_1=0, X_2=\tilde{x}_2) = p_{00}+p_{01} = 0.7 \tag{7}$$

2. 周辺確率 $\Pr(X_2=1), \Pr(X_2=0)$ を求めよ．

$$\Pr(X_2=1) \;=\; \sum_{\tilde{x}_1=0}^{1} \Pr(X_1=\tilde{x}_1, X_2=1) = p_{01}+p_{11} = 0.3 \tag{8}$$

$$\Pr(X_2=0) \;=\; \sum_{\tilde{x}_1=0}^{1} \Pr(X_1=\tilde{x}_1, X_2=0) = p_{00}+p_{10} = 0.7 \tag{9}$$

※ $\Pr(X_1=1, X_2=1) = 0.2 \neq \Pr(X_1=1)\Pr(X_2=1) = 0.09$ より，X_1, X_2 は独立ではない（どちらか一方の値が分かればもう一方を当てやすい）．

3. 事後確率 $p(\boldsymbol{x}|\boldsymbol{y})$ を，事前分布 $p(\boldsymbol{x})$，データモデル $p(\boldsymbol{y}|\boldsymbol{x})$ を使い表現せよ．

$$p(\boldsymbol{x}|\boldsymbol{y}) \;=\; \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})}{\displaystyle\sum_{\tilde{\boldsymbol{x}}} p(\tilde{\boldsymbol{x}},\boldsymbol{y})} = \frac{p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})}{\displaystyle\sum_{\tilde{\boldsymbol{x}}} p(\tilde{\boldsymbol{x}})p(\boldsymbol{y}|\tilde{\boldsymbol{x}})} \tag{10}$$

※ $p(\boldsymbol{x})$ と $p(\boldsymbol{y}|\boldsymbol{x})$ および観測データ $\boldsymbol{y}$ は与えられているので使える．これ以外に必要な情報は，これらを使って引き出す必要がある．

4. 事後確率 $p(\boldsymbol{x}|\boldsymbol{y})$ を，$p(x_1,x_2)$，$p(y_j^i|x_i)$ を用い，表現せよ．

$$p(\boldsymbol{x}|\boldsymbol{y}) \;=\; \frac{p(x_1,x_2)\displaystyle\prod_{i=1}^{2}\prod_{j=1}^{n_i} p(y_j^i|x_i)}{\displaystyle\sum_{\tilde{x}_1=0}^{1}\sum_{\tilde{x}_2=0}^{1}\left[p(\tilde{x}_1,\tilde{x}_2)\prod_{i=1}^{2}\prod_{j=1}^{n_i} p(y_j^i|\tilde{x}_i)\right]} \tag{11}$$

5. 事後確率 $p(\boldsymbol{x}|\boldsymbol{y})$ の値は 0 に近い小さな値になり，コンピュータで計算すると場合によってはアンダーフローをおこす．これを回避するには，$p(\boldsymbol{x}|\boldsymbol{y})$ の分子分母を反転し

た値を計算すればよい．$\dfrac{1}{p(\boldsymbol{x}|\boldsymbol{y})} = h_{\mathrm{G}}(\boldsymbol{x}|\boldsymbol{y})$ とおこう．$h_{\mathrm{G}}(\boldsymbol{x}|\boldsymbol{y})$ を，$p(x_1, x_2)$，$p(y^i_j|x_i)$ を用い表現せよ（実は，G は God を意味する）．

$$h_{\mathrm{G}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{\displaystyle\sum_{\tilde{x}_1=0}^{1}\sum_{\tilde{x}_2=0}^{1}\left[p(\tilde{x}_1,\tilde{x}_2)\prod_{i=1}^{2}\prod_{j=1}^{n_i}p(y^i_j|\tilde{x}_i)\right]}{\displaystyle p(x_1,x_2)\prod_{i=1}^{2}\prod_{j=1}^{n_i}p(y^i_j|x_i)} \tag{12}$$

6. この課題では，$\boldsymbol{x}$ が $\boldsymbol{x}_{11} = (1,1)$ のときの事後確率にだけに関心がある．$h_{\mathrm{G}}(\boldsymbol{x}_{11}|\boldsymbol{y})$ の分子・分母を $\displaystyle\prod_{i=1}^{2}\prod_{j=1}^{n_i}p(y^i_j|1)$ で割ってみよ．ここで，$p(y^i_j|1) = \Pr(Y^i_j = y^i_j|X_i = 1)$ である．

$$\begin{aligned}
h_{\mathrm{G}}(\boldsymbol{x}_{11}|\boldsymbol{y}) &= \frac{\displaystyle\sum_{\tilde{x}_1=0}^{1}\sum_{\tilde{x}_2=0}^{1}\left[p(\tilde{x}_1,\tilde{x}_2)\prod_{i=1}^{2}\prod_{j=1}^{n_i}\frac{p(y^i_j|\tilde{x}_i)}{p(y^i_j|1)}\right]}{\displaystyle p(\boldsymbol{x}_{11})\prod_{i=1}^{2}\prod_{j=1}^{n_i}\frac{p(y^i_j|1)}{p(y^i_j|1)}} &(13)\\
&= \frac{\displaystyle\sum_{\tilde{x}_1=0}^{1}\sum_{\tilde{x}_2=0}^{1}\left[p(\tilde{x}_1,\tilde{x}_2)\prod_{i=1}^{2}\prod_{j=1}^{n_i}\frac{p(y^i_j|\tilde{x}_i)}{p(y^i_j|1)}\right]}{p(\boldsymbol{x}_{11})} &(14)\\
&= \frac{\displaystyle\sum_{\tilde{x}_1=0}^{1}\sum_{\tilde{x}_2=0}^{1}\left[p(\tilde{x}_1,\tilde{x}_2)\prod_{i=1}^{2}\prod_{j=1}^{n_i}\exp\left\{-\frac{(y^i_j-\tilde{x}_i)^2}{2\sigma^2}+\frac{(y^i_j-1)^2}{2\sigma^2}\right\}\right]}{p(\boldsymbol{x}_{11})} &(15)\\
&= \frac{\displaystyle\sum_{\tilde{x}_1=0}^{1}\sum_{\tilde{x}_2=0}^{1}\left[p(\tilde{x}_1,\tilde{x}_2)\prod_{i=1}^{2}\prod_{j=1}^{n_i}\exp\left\{\frac{1-2y^i_j+2y^i_j\tilde{x}_i-\tilde{x}_i^2}{2\sigma^2}\right\}\right]}{p(\boldsymbol{x}_{11})} &(16)
\end{aligned}$$

と単純になる．分子の $\prod\prod$ 以降の項は $n_1 + n_2$ 個の要素のかけ算であるので，アンダーフローがおこる可能性がある．そこで，対数をとり，積を和の形で計算し，その結果を exp の肩にのせて

$$h_{\mathrm{G}}(\boldsymbol{x}_{11}|\boldsymbol{y}) = \frac{\displaystyle\sum_{\tilde{x}_1=0}^{1}\sum_{\tilde{x}_2=0}^{1}\left[p(\tilde{x}_1,\tilde{x}_2)\exp\left\{\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(\frac{1-2y^i_j+2y^i_j\tilde{x}_i-\tilde{x}_i^2}{2\sigma^2}\right)\right\}\right]}{p(\boldsymbol{x}_{11})} \tag{17}$$

と計算すればよい．この値の逆数をとれば事後確率

$$\mathcal{S}_{\mathrm{G}}(\boldsymbol{y}) = p(\boldsymbol{x}_{11}|\boldsymbol{y}) = \frac{1}{h_{\mathrm{G}}(\boldsymbol{x}_{11}|\boldsymbol{y})} \tag{18}$$

が計算できる．

7. ここまでは各要素が $0,1$ の 2 値をとる $\boldsymbol{x}$ が 2 次元の場合を考えた．いま，$\boldsymbol{x}$ が 100 次元ベクトル $\boldsymbol{x}=(x_1,x_2,\cdots,x_{100})$，であり，各 x_i が 10 値をとる場合を考えてみよう．$h_{\mathrm{G}}(\boldsymbol{x}_{11\cdots1}|\boldsymbol{y})$ の分子は $\displaystyle\sum_{\tilde{x}_1=0}^{9}\sum_{\tilde{x}_2=0}^{9}\cdots\sum_{\tilde{x_{100}}=0}^{9}\cdots$ となり，10^{100} 項を足し算する必要があり，これが現実的には計算できないことは明白である．したがって，$\boldsymbol{x}$ の次元数がある程度大きい場合に，$\mathcal{S}_{\mathrm{G}}(\boldsymbol{y})$ の代わりに使える統計量を探すことが課題となる．いろいろな統計量が考えられるだろう．例えば，$\boldsymbol{x}$ が $\boldsymbol{x}_{11\cdots1}=(1,1,\cdots,1)$ であると確信する度合いを $\boldsymbol{x}_{00\cdots0}=(0,0,\cdots,0)$ とだけ比較することは容易にできる．したがって，以下のような統計量を利用する方法が考えられる．

$$h_{\mathrm{T}}(\boldsymbol{x}_{11}|\boldsymbol{y}) \quad = \quad 1+\frac{p(\boldsymbol{x}_{00})}{p(\boldsymbol{x}_{11})}\exp\left[\sum_{i=1}^{2}\sum_{j=1}^{n_i}\left\{\frac{1-2y_j^i}{2\sigma^2}\right\}\right] \tag{19}$$

もちろん，$\boldsymbol{x}=(1,1)$ であるか否かの判定にはこれの逆数

$$\mathcal{S}_{\mathrm{T}}(\boldsymbol{y}) \quad = \quad \frac{1}{h_{\mathrm{T}}(\boldsymbol{x}_{11}|\boldsymbol{y})} \tag{20}$$

を用いる（T は Template を意味する）．これを第 2 の方法とよぼう．この統計量にはどんな意味があるのだろうか．これは，$2^{n_1+n_2}$ 通りの信号 $\boldsymbol{x}$ が存在するが，世の中には，このうち $\boldsymbol{x}_{00\cdots0}$ と $\boldsymbol{x}_{11\cdots1}$ の 2 つしか出現しないと仮定することに対応する．いわゆるテンプレートマッチングは，これに対応する．

8. 事後確率 $p(\boldsymbol{x}_{11}|\boldsymbol{y})$ の式は，以下のようにも変形できる．

$$p(\boldsymbol{x}_{11}|\boldsymbol{y}) \quad = \quad \Pr(X_1=1|\boldsymbol{y})\Pr(X_2=1|X_1=1,\boldsymbol{y}) \tag{21}$$

$$\quad = \quad \Pr(X_1=1|\boldsymbol{y}^1,\boldsymbol{y}^2)\Pr(X_2=1|X_1=1,\boldsymbol{y}^2) \tag{22}$$

ここで $\boldsymbol{y}=(\boldsymbol{y}^1\boldsymbol{y}^2)$ である．第 1 項目 $\Pr(X_1=1|\boldsymbol{y}^1,\boldsymbol{y}^2)$ を計算するのは大変である．そこで，正確でないことを承知で

$$p(\boldsymbol{x}_{11}|\boldsymbol{y}) \quad \approx \quad \Pr(X_1=1|\boldsymbol{y}^1)\Pr(X_2=1|X_1=1,\boldsymbol{y}^2) \tag{23}$$

と，第 1 項目を計算しやすい $\Pr(X_1=1|\boldsymbol{y}^1)$ と思いこんで計算を進めるのが第 3 の方法である．式（21）の右辺第 2 項 $\Pr(X_2=1|X_1=1,\boldsymbol{y})$ が式（22）では $\Pr(X_2=1|X_1=1,\boldsymbol{y}^2)$ となっているのは，$X_1=1$ という情報が既に与えられた状況では，X_2 の値の推論に $\boldsymbol{y}^1$ がもたらす情報はないからである．これは確率変数間の依存性を示

す図 1 を見ればわかりやすい．式（23）の右辺第 1 項は

$$\Pr(X_1=1|\boldsymbol{y}^1) = \frac{\Pr(X_1=1,\boldsymbol{y}^1)}{p(\boldsymbol{y}^1)} \tag{24}$$

$$= \frac{\Pr(X_1=1)p(\boldsymbol{y}^1|X_1=1)}{p(\boldsymbol{y}^1)} \tag{25}$$

$$= \frac{\Pr(X_1=1)p(\boldsymbol{y}^1|X_1=1)}{\displaystyle\sum_{i=0}^{1}\Pr(X_1=i)p(\boldsymbol{y}^1|X_1=i)} \tag{26}$$

であり，先の計算と同様にすすめると

$$\frac{1}{\Pr(X_1=1|\boldsymbol{y}^1)} = 1+\frac{\Pr(X_1=0)p(\boldsymbol{y}^1|X_1=0)}{\Pr(X_1=1)p(\boldsymbol{y}^1|X_1=1)} \tag{27}$$

$$= 1+\frac{\Pr(X_1=0)}{\Pr(X_1=1)}\exp\left[\sum_{j=1}^{n_1}\left\{\frac{1-2y_j^1}{2\sigma^2}\right\}\right] \tag{28}$$

となる．この逆数が，しきい値 θ より大きい場合にだけ，第 2 項目の計算を進めればよい．第 2 項目は

$$\Pr(X_2=1|X_1=1,\boldsymbol{y}^2) = \frac{p(\boldsymbol{x}_{11},\boldsymbol{y}^2)}{p(X_1=1,\boldsymbol{y}^2)} \tag{29}$$

$$= \frac{\Pr(X_1=1,X_2=1)p(\boldsymbol{y}^2|\boldsymbol{x}_{11})}{p(X_1=1,\boldsymbol{y}^2)} \tag{30}$$

$$= \frac{\Pr(X_1=1,X_2=1)p(\boldsymbol{y}^2|X_2=1)}{\displaystyle\sum_{i=0}^{1}\Pr(X_1=1,X_2=i)p(\boldsymbol{y}^2|X_1=1,X_2=i)} \tag{31}$$

$$= \frac{\Pr(X_1=1,X_2=1)p(\boldsymbol{y}^2|X_2=1)}{\displaystyle\sum_{i=0}^{1}\Pr(X_1=1,X_2=i)p(\boldsymbol{y}^2|X_2=i)} \tag{32}$$

となり，$p(\boldsymbol{y}^2|X_1=1,X_2=i)=p(\boldsymbol{y}^2|X_2=i)$ より，先の計算と同様にすすめると

$$\frac{1}{\Pr(X_2=1|X_1=1,\boldsymbol{y}^2)} = \sum_{\tilde{x}_2=0}^{1}\frac{\Pr(X_1=1,X_2=\tilde{x}_2)}{\Pr(X_1=1,X_2=1)}\frac{p(\boldsymbol{y}^2|X_2=\tilde{x}_2)}{p(\boldsymbol{y}^2|X_2=1)} \tag{33}$$

$$= 1+\frac{\Pr(X_1=1,X_2=0)}{\Pr(X_1=1,X_2=1)}\frac{p(\boldsymbol{y}^2|X_2=0)}{p(\boldsymbol{y}^2|X_2=1)} \tag{34}$$

$$= 1+\frac{p_{10}}{p_{11}}\exp\left[\sum_{j=1}^{n_2}\left\{\frac{1-2y_j^2}{2\sigma^2}\right\}\right] \tag{35}$$

となる．ここで文脈から明らかであるが，y_j^2 は 2 乗した値ではない．この逆数と，先に計算した値をかけ算し，

$$\mathcal{S}_{\mathrm{P}_{1\to 2}}(\boldsymbol{y}) = \Pr(X_1=1|\boldsymbol{y}^1)\Pr(X_2=1|X_1=1,\boldsymbol{y}^2) \;>\; \theta \tag{36}$$

であれば，$\boldsymbol{x} = (1,1)$ と判定する（P は Parts の意味）．これが第 3 の方法.

9. 第 3 の手法では，$\boldsymbol{y}^1$ を観測し $X_1 = 1$ かどうかを判定し，次に $\boldsymbol{y}^2$ をもとに $X_2 = 1$ かどうかを判定した．調べる順番を逆にすると（$\mathcal{S}_{\mathrm{P}_{2\to 1}}$），同じ結果にはならない．どちらを先に観測して判定するか．これは

$$k \quad = \quad \operatorname*{argmax}_i \Pr(X_i = 1|\boldsymbol{y}^i) \tag{37}$$

を計算し，大きい順に計算する方法が考えられる．これを $\mathcal{S}_{\mathrm{S}}$ としよう（S は Saccade の頭文字）．常に $1 \to 2$ の順で計算する方法と，$\mathcal{S}_{\mathrm{S}}$ を用いて判断する方法の，どちらが性能がよいかは，ROC カーブを描くと明らかになる.

課題（コンピュータシミュレーション）

パラメータの値などが指定されていない場合，適当な数字を当てはめて課題を進めてよい．課題に曖昧な点があると思う場合は，その部分については適当に解釈してよい．ただし，使用したパラメータの値や，曖昧な点をどう解釈したかをレポートに記述すること.

基本課題

1. 事前分布 $p(\boldsymbol{x})$ を式（3），データモデル $p(\boldsymbol{y}|\boldsymbol{x})$ を式（2），$n_1 = n_2 = 3, \sigma = 1.0$ とする．6 次元ベクトル $\boldsymbol{y}$ をうけとり $z' \in \{0, 1\}$ を返す関数を作成せよ．ここで，まず，$\boldsymbol{y}$ をもとに，ある統計量 $\mathcal{S}$ の値を計算し，$\mathcal{S}$ がしきい値 θ より大きいとき $z' = 1$ と判定する．$\mathcal{S}$ としては $\mathcal{S}_{\mathrm{G}}$，$\mathcal{S}_{\mathrm{T}}$，$\mathcal{S}_{\mathrm{P}_{1\to 2}}$ を用いよ.

2. 100,000 個の例題を生成し，1. で作成した関数を用い，FPR と CDR を計算せよ．ここで，しきい値 θ を，例えば，$0 \leqq \theta \leqq 1$ の範囲で 0.01 きざみで変え，実験を繰り返すことで ROC 曲線を描くことができる．$\mathcal{S}$ として $\mathcal{S}_{\mathrm{G}}$，$\mathcal{S}_{\mathrm{T}}$，$\mathcal{S}_{\mathrm{P}_{1\to 2}}$ を用い，3 本の ROC 曲線を描き，結果を比較し考察せよ.

3. $n_i = 5, 10, 20$ $(i = 1, 2)$ として，2. の実験をおこない，n_i が大きくなるにつれて，どのような変化が見られるか，得られた結果を考察せよ.

   （以下は，自由課題．ここまでの課題は必ず達成すべし．できない場合，何が障害になっているか明確にせよ.）

4. $\mathcal{S}_{\mathrm{S}}$ を用いて，4 本目の ROC 曲線を描き，結果を考察せよ.

5. $n_1 \neq n_2$ の場合についても実験してみよ（例： $n_1 = 5, n_2 = 10$）.

6. ノイズの標準偏差 σ の値を変え，実験してみよ．

7. $\boldsymbol{y}^1$ と $\boldsymbol{y}^2$ で σ の値を変えて（それぞれの式で $\sigma = \sigma_i$ などとすればよい），実験してみよ．

発展課題（個別に説明するので相談しに来てください）

8. ここまでは $\boldsymbol{x}$ は 2 次元の信号を扱った．$\boldsymbol{x}$ が 3 次元の場合について実験を試みよ．ここで，$\displaystyle\sum_{\tilde{\boldsymbol{x}}} p(\tilde{\boldsymbol{x}}) = 1$ となる事前分布 $p(\boldsymbol{x})$ をあらかじめ与えておく必要がある．

9. $\boldsymbol{x}$ が 10 次元の場合について実験してみよ．事前分布 $p(\boldsymbol{x})$ には，単純なものを選ぶ．

10. パラメータ（例えば σ）の値を変化させることで，問題の難易度を調節できる．さまざまな難易度のモデルに対し，事後確率分布の構造を詳細に調べてみよ．

11. $\cdots\cdots$ （課題を自分で作成し，考察してみよ）

レポートの最後には，感想，質問などを記述して下さい．理解しにくい点があった場合は，このプリント中の，どこの部分が分かりにくかったか，具体的に指摘してもらえれば大変助かります（来年度向けに改善するため）．

プログラミングメモ

アルゴリズム設計の手順：段階的詳細化

文字や図で説明 $\Longrightarrow$ 擬似コード $\Longrightarrow$ 特定のプログラミング言語で書かれたプログラム

まず擬似コードのレベルでアルゴリズムを設計することが重要.

アルゴリズム

Algorithm 1 問題を生成 $\rightarrow$ モデルを使い判定 $\rightarrow$ ROC カーブを描く

初期設定（パラメータ値の設定など）
for $\theta = 0.0$ to 1.0 **do**
  $n_{\mathrm{FP}} = 0, n_{\mathrm{CD}} = 0$
  $\alpha = 0$
  for $\alpha = 1$ to $100,000$ **do**
    $\boldsymbol{x}^\alpha \sim p(\boldsymbol{x})$ を生成．$\boldsymbol{x}^\alpha = (1,1)$ であれば $n_{\boldsymbol{x}_{11}} + +$.
    生成された $\boldsymbol{x}^\alpha$ をもとに $\boldsymbol{y}^\alpha \sim p(\boldsymbol{y}|\boldsymbol{x}^\alpha)$ を生成.
    統計量 $\mathcal{S}$（事後確率 $p(\boldsymbol{x}_{11}|\boldsymbol{y}^\alpha)$ など）を計算.
    $z' = 0, 1$ を判定.
    $z' = 1$ のとき，False Positive か Correct Detect かを判定．$n_{\mathrm{FP}} + +$ or $n_{\mathrm{CD}} + +$.
    $\alpha \leftarrow \alpha + 1$
  end for
  FPR と CDR を計算
end for

データ構造

$n_\alpha = 100,000$ 個の例題をもとに ROC カーブを描く場合，$\boldsymbol{x}^\alpha, \boldsymbol{y}^\alpha, \alpha = 1, \cdots, n_\alpha$ をすべて記憶していなくても ROC カーブは描ける．たとえば以下のような配列を用いて問題を表現する.

1. `x[i]`　$\boldsymbol{x} = (x_0, x_1), i = 0, 1$．後で $\boldsymbol{x} = (x_0, x_1, x_2, x_3, x_4, x_5), x_i \in \{0, 1\}$ などと拡張.
2. `y[k][j]`　$\boldsymbol{y} = (y_0^0, y_1^0, y_2^0, y_0^1, y_1^1, y_2^1)$，$k = 0, 1, j = 0, 1, 2$.
3. `p[i][j]`　$p(\boldsymbol{x}_{ij}), i, j = 0, 1 \Longrightarrow$ この表現方法では $\boldsymbol{x}$ が 10 次元になったとき，`p[x1][x2]`$\cdots$`[x10]` のように 10 次元配列を用意する必要がある（×） $\Longrightarrow$ 2 進数 $\leftrightarrow$ 10 進数変換を用意.

関数

モデルのパラメータを一括で渡せるように構造体を作っておくと便利.

```
generate_data (Pgm *model, int *x, double *y)
double compute_s (Pgm *model, double *y)
```