練 習 問 題 の 解 答 & 解 説 Ver.1

・各節の練習問題の解説のほか、発展応用問題、Q&A などを順次掲載する。
・著者サイト：http:www.qmss.jp/qmss

最終更新日'20/12/21

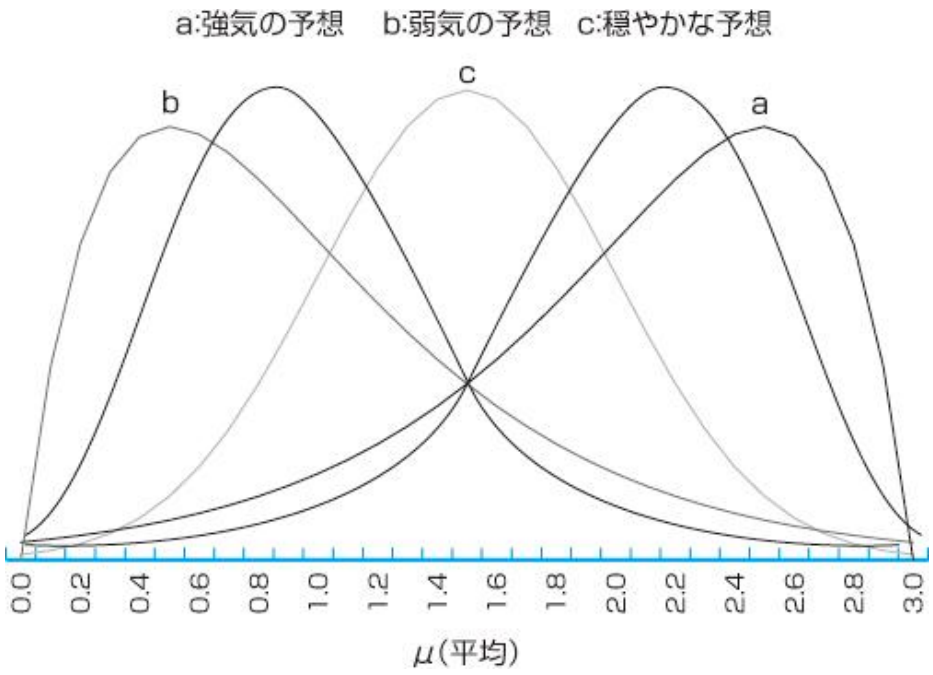
［1－1］スパム・フィルタリングはこの考え方から始まる。

① 100 × 0.80＝80、100 万 × 0.65＝65万で、快楽的サイトが約8000倍の可能性がある。
② 0.08 × 0.35＋0.15 × 0.28＋ … ＋0.14 × 0.12＝0.2268
以上のデータは問題用の仮の数字である。

［1－2］ベイズ的考え方が中心になっている分析で、ベイズ統計学のウォーミングアップ。

感度　＝a/(a＋c) = 50/131 = 0.382 (38.2%)
特異度＝1－b/(b＋d)＝d(b＋d)＝200/221＝0.905 (90.5%)

［1－3］下図、主観を表記した結果であるから、本来手書きでよい。



［1－4］本書ではじめてのベイズの定理の練習である。少数まで直したい。

a）$\frac{0.7\times(3/4)}{0.7\times(3/4)+0.3\times(1/3)}=\frac{21}{25}=0.84$，1－0.84＝0.16

b）$\frac{0.7\times(1/4)}{0.7\times(1/4)+0.3\times(2/3)}=\frac{7}{15}=0.47$，1－0.47＝0.53

［2－2］①の区別をする所が、ベイズの定理。

① A_i という原因から結果 B が生じる確率が $P(B|A_i)$，
　結果 B が生じたときその原因が A_i である確率が $P(A_i|B)$
② $\frac{0.5\times0.10}{0.5\times0.10+0.5\times0.15}=\frac{0.10}{0.25}=0.4$（愛している），$\frac{0.5\times0.15}{0.5\times0.10+0.5\times0.15}=\frac{0.15}{0.25}=0.6$（愛していない）

もちろん1－0.4＝0.6でも可。いずれも、ハンカチをくれる場合。

［2－3］「0%」でも降水がありうることの理解が重要。

161 回の「降水確率 0%」の予報に対し降水があったのは 3 回で、3/161＝0.018（1.8%）

［2－4］独立でない場合としては、最もわかりやすい例。

P(4 以上)＝3/6＝1/2，P(偶数) = 3/6 = 1/2（いずれも 6 通り中 3 通り）だが、P(4 以上かつ偶数) = 2/6 = 1/3(4,6 のみ)で、$1/3 \neq (1/2)\times(1/2)$

［2－5］SD は σ（シグマ）とも表すので、いわゆる「3 シグマ限界」の考え方。

3SD の範囲は40－3 × 2.5 = 32.5から40＋3 × 2.5＝47.5までで、図からも事実上すべてのμの範囲であり、実際99.7%確かである。

［2－6］ストーリーの重要な約束事であるが、計算のための重要な情報でもある。

アランが釈放される場合は処刑されるのはバーナードあるいはチャールズであり、この 2 人のどちらとも等しく答え得るから、バーナードが答えられる確率は 2 通り中 1 通りで1/2、また、バーナードが釈放される場合に、バーナードの処刑が答えられることはありえない。

［3－1］ポアソン分布の基本を理解する第一歩。

① POISSON において、事象の数＝0,1,2,3,4,5,6,…、期待値＝0.7、関数形式＝FALSE とすれば確率を得て、それを折れ線にすれば本文のグラフとなる。なお、数形式 FALSE, TRUE は文字通りの意味はなく、0,1 くらいの意味である。

② 次の和を作る：

$0.497 \times 0 + 0.348 \times 1 + 0.122 \times 2 + 0.028 \times 3 + 0.005 \times 4 + 0.001 \times 5 +$(以下略)$= 0.69945$

エクセルには A 列に 0, 1, 2, …と入れ、B 列には、まず SSON において平均＝0.7 と定め、関数形式を‘FALSE’として、順次 B1, B2, …に 0.497, 0.348, …と出力する（①）。計算は C 列で A 列×B 列を計算して、あとは総和Σをとればよい（②）。なお、上式で 6 までとすると、和は0.699937。

［3－2］①は省略されることが多いので、今一度確認の復習。②は前問の続き。

① $P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_k)P(B|A_k)$

② $e^{-0.5}\frac{(0.5)^4}{4!}$, $e^{-1}\frac{1}{4!}$, $e^{-2}\frac{2^4}{4!}$, $e^{-10}\frac{10^4}{4!}$, $e^{-100}\frac{100^4}{4!}$,

［3－3］［3－1］に引き続き二項分布のマスターの第一歩。前問②の続きが②。

① BINOMDIST(訂正)において、成功数＝0, 1 、…,4、試行回数＝4、成功率＝2/3、関数形式＝FALSE とすればよい。

校	0	1	2	3	4
確率	*0.012*	*0.099*	*0.296*	*0.395*	*0.198*

② $6(0.1)^2(0.9)^2$, $6(0.2)^2(0.8)^2$, $6(0.3)^2(0.7)^2$, $6(0.4)^2(0.6)^2$, $6(0.5)^2(0.5)^2$, $6(0.6)^2(0.4)^2$, $6(0.7)^2(0.3)^2$, $6(0.8)^2(0.2)^2$, $6(0.9)^2(0.1)^2$

x が定まっていて、p が変るのがベイズ統計学の考え方の特色である。

［3－4］「標準正規分布の累積分布関数」という長い名前のΦの使い勝手の良さを学ぶ。

① 『ベイズ統計解答.xlsx』のシート 1 の①を参照。

② 『ベイズ統計解答.xlsx』のシート 1 の②を参照。

③ NORMSDIST を用いると$\Phi(2)=0.97725$, $\Phi(-2)=0.02275$ で差をとって 0.9545。なお、実は$\Phi(-2)=1-\Phi(2)$なので、$2\Phi(2)-1$ でもよい。

④ NORMSINV(0.1)などから、$\Phi^{-1}(0.1)=-1.28$, $\Phi^{-1}(0.5)=0$, $\Phi^{-1}(0.9)=1.28$。なお、エクセルは計算方式の都合上、0 を極めて小さい数として出力することがある。

［3－5］正規分布のケースは多少こみ入るので、これで理解を整理する。

(1)＝11,000, (2)＝85,000, (3)＝15,000, (4)＝70,000, (5)＝75,740, (6)＝6,804

［3－6］①は待ち時間の良い理解練習。②は密度関数を視覚的にとらえる練習。

① 2.993083
② μが大きくなるにつれ左側部分で上部へ引き上げられ、また右裾が短くなる。

［4－1］例外的に非常に簡単にできるケース。［5－3］も参照。

密度関数はθに比例する直線となり、0.5 以上の面積（台形）と 0.5 以下の面積（三角形）の比は、相似の計算から 3：1。よって$P(\theta > 0.5) = 3/4 = 0.75$。*p*.114 参照。

［4－2］ガンマ分布のベイズ更新は、まぎらわしいので、理解練習。

$\Gamma(\mu, \alpha)$のベイズ更新から、事後分布では$\mu' = 3 + 7 = 10,\ \alpha' = 5 + 5 = 10$となって、$\Gamma(10,10)$が事後分布」。期待値、分散は事前分布ではそれぞれ$5/3$、$5/9$、事後分布では同じく$1$、$1/10$

［4－3］ベイズ統計学で、多変量解析の練習。

①『http://www.qmss.jp/databank/advanced/kaisetsu/2-3a.htm』を参照。制作中。

②（訂正参照）　おおよその見当をつけて判別関数を選び計算すると

A について：$F_{12} = 7.6974$,　$F_{13} = 81.9953$　∴ ヴァージニカ
B について：$F_{21} = 8.9857$,　$F_{22} = 40.0722$　∴ ヴェルシカラー
C について：$F_{31} = 89.2897$,　$F_{32} = 45.7310$　∴ セトーサ
D について：$F_{31} = 25.2555$,　$F_{32} = 0.4469$　∴ セトーサ

［4－4］
『ベイズ統計解答.xlsx』のシート 2 を参照。

［4－5］オッズ、オッズ比は医学・医薬の統計学では強調されてよいトピック。

2+98=100, 98,000+2,000=100,000 とし、問題文で「オッズ比」を「オッズ」とすると(また *p*.131 にも訂正あり)。

事前オッズ$= 100/100{,}000 = 1/100$、事後オッズ$= 98/2{,}000$より、
ベイズ因子は$\mathrm{BF} = (98/2{,}000)/(1/100) = 98{,}000/2{,}000 = 49$

［4－6］経営学、組織論に関心ある読者には絶好。

A,B,C の結果に対する「適」の事後確率はそれぞれ$21/29$、$2/3$、$3/8$。
以下、本文と同様の計算から、最適な決定の方式は A→否、B→保、C→否（d_6）となる。

［5－1］通信理論にも通じる重要トピック。

$(\mu_0 + \mu_1)/2 = (0 + 2)/2 = 1$、
$-(\sigma^2/n)\log(w_1/w_0)$
$= -(1/10)\log\{(1/3)/(2/3)\} = -(1/10)\log(1/2) = 0.0693$ で、
境界は 1.0693、
一方、データから$\bar{x} = 0.95 < 1.0693$　ゆえに$\mu_0 = 0$ と決定。

［5－2］ふつうの統計学の理解をもう一度キチンと。

よく知られた公式より、$130 \pm 1.96 \cdot 5/\sqrt{2} = 122.4, 135.4$（ベイズ信頼区間と同一)。

［5－3］代表値の理解のほか、ベータ分布の理解を深める。①は［4－1］の類題。

p.114 の図を参考とする。

① Be(2,1)に現れる三角形の面積を縦線で二分する。面積比$=\left(\text{相似比}\right)^2$だから、
$\sqrt{1/2}=1/\sqrt{2}=\sqrt{2}/2=0.7071$これがメディアンとなる。

② $\mathrm{BETAINV}(0.5,3,2,0,1)$、$\mathrm{BETAINV}(0.5,4,1,0,1)$、$\mathrm{BETAINV}(0.5,5,2,0,1)$とすると
0.164、0.686、0.736となる。
ちなみに$\mathrm{BETAINV}(0.5,5,2,1\ 0,1)=0.707$ と出る$\left(0,1\right.$ は省略してもよい$\left.\right)$。

［5－4］
『ベイズ統計解答.xlsx』のシート 3 を参照。

［5－5］
（訂正参照）。$\mathrm{se}_1=16.49,\ \mathrm{se}_2=0.34$

［6－1］「ソルバー」はエクセルの知られざる隠れた武器で、最尤推定には絶好。他でも使える。

『ベイズ統計解答.xlsx』のシート 4 を参照。

ソルバーのアドイン（ダウンロード）はエクセルのヘルプ参照。
B1 にλ、B10 に L(λ)を入れ、B1 には正の条件をつけること。
ここでは最終結果を示した。解λ＝1.143 は 8/7 に一致している。

［6－2］ベイジアン・ネットの理解も、これ位からスタート。

独立を仮定すると、例えば

$$(+b,+c)\text{の確率}=\begin{cases}0.80\times0.20=0.16 & \left(+\alpha\ \text{のとき}\right)\\ 0.20\times0.05=0.01 & \left(\neg\alpha\ \text{のとき}\right)\end{cases}$$

［6－3］事後確率が‘あっけなく’出る所がポイント。

$\lambda_n \fallingdotseq (1-\beta)/\sigma = 0.9/0.05 = 18$ となっているから $\theta_{1,} - \theta_2$の事後確率は、
それぞれ $w'(34) = \frac{1}{1+18} = 0.053$、1−0.053=0.9947 となる。

［6－4］「カルマン・フィルタ」がリアル・タイムの方法であることの確認だが、「イノベーション」の理解も関わる（全体としては、ベイジアンの考え方が効いている）。

$\hat{\theta}_1 = \hat{\theta}_0 + (y_1 - \hat{\theta}_0)/1 = (1/2)y_1 + (1/2)\hat{\theta}_0$、
$\hat{\theta}_2 = \hat{\theta}_1 + (y_2 - \hat{\theta}_1)/2 = (1/2)y_2 + (1/2)\hat{\theta}_1$、
$\quad = (1/2)y_2 + (1/4)y_{1+}(1/4)\hat{\theta}_0$
などから
$\hat{\theta}_t = (1/2)y_t + (1/2)^2 y_{t-1} + \cdots + (1/2)^t y_1 + (1/2)^t \hat{\theta}_0$

［6－5］WinBUGS の試行は別の機会とし、まず乱数シミュレーションの手始め。

① 『ベイズ統計解答.xlsx』のシート 5 を参照。
NORMSINV(RAND(　))を 100 回くり返すと、−0.46441、0.96211、…
などと得られる。

②エクセルの指定方式では、パラメータは(2, 1/2)とする。なお、 GAMMAINV(RAND(　), 2, 1/2)
を 100 回くり返すと、−0.82447、0.97599、…
などと得られる。
いずれも乱数の出力なので、実行毎に結果は異なる。
読者は、ヒストグラムでそれぞれの分布を確認しよう。