2 Introduction to Bayesian Thinking
入門：ベイス思考

2.1
prior, update, posteriorなどいくつかキーワードが出てきます。

2.2
ウケ狙いのheavy sleepersの記述は、ほっときましょう。

○ベイス思考の3つのポイント

1)パラメータは確率分布
■推定したいモノを「パラメータ」と言うことにします。そうすると、パラメータの確率分布を考えることになります。
■パラメータは直接観測できません。観測できるモノを「データ」と呼ぶことにします。つまり、「データ」を観測して、パラメータを推定する訳ですね。

2)データによる推定の更新
■パラメータの確率分布は、データが得られる前なら「事前分布」、データが得られた後なら「事後分布」と呼ばれます。新たなデータが得られれば、その事後分布を事前分布としてさらに新しい事後分布を推定することが可能です。このように、新しいデータによって推定を次々と更新することができるんです。
■また、このような推定の入れ子的過程が階層的モデリングに適していると言えます。つまり、事前分布の元となる「超事前分布」を導入することができます。

3)事後分布は尤度関数と事前分布の積に比例する
■尤度とは、観測されたデータが実現する確率です。ここで、尤度をパラメータの関数と考えることもできますね。つまり、「尤度関数」とは、あるパラメータの値が正しいとした時に、観測されたデータが実現する確率となります。
■さて、事前分布、尤度関数、事後分布の3つがパラメータに対する確率変数になりました。「ベイズの定理」から、これら3つの確率変数の関係が導かれ、尤度関数と事前分布から事後分布を推定することができます。もし、事前分布が何の情報も持っていない、つまり一様分布なら、この推定は「最尤推定」と同じことになりますね。なぜなら、事後分布は尤度関数そのものになるので。

ここらへんの背景は、

久保拓也さんの「生態学のデータ解析：統計学授業2007」
http://hosho.ees.hokudai.ac.jp/~kubo/ce/EesLecture.html

に詳しく解説されているので、ぜひご覧ください。ありがたや。

○ベイズの定理からのベイズ推定式の導出

ややこしいので手書きにしました。

ところで、ベイズの定理ってあまりにも当たり前のことを言っているので、こんなの役に立つの？と思われるかもしれませんが、まれな事象を推測する時には、常識的な推測がしばしば間違うことを教えてくれます。

例えば、まれな病気の検査です。病気の確率は1%としましょうP(B) = 0.01。ですから、99%あなたは病気じゃありませんP(notB) = 0.99。検査は、病気なら95%の確率で陽性P(A|B) = 0.95、病気でなければ95%の確率で陰性P(A|notB) = 0.05を出します。まあまあ信頼性が高い検査ですね（ま、ゆーい水準程度には）。ここで、検査で陽性が出た時に、ホントにあなたは病気でしょうか？この確率はP(B|A)なので、

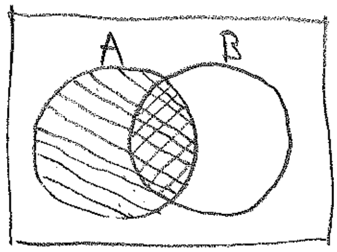
P(B|A) = (0.95 * 0.01)/(0.95 * 0.01 + 0.05 * 0.99) = 0.1610169

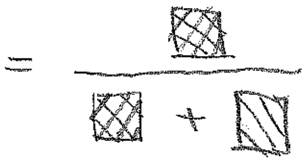
つまり、84%あなたは病気じゃありません。検査の信頼性が高いからといって、陽性が出たら病気だと決めつけない方がいいですよ。このことを偽陽性と言います。事前分布である、病気の希少性を考慮しようと言うことですね。でも、検査によって事前確率1%から事後確率16%になったんですから、検査は役に立ってます。ま、検査はこの程度のもんだ、と言うことをベイズの定理は教えてくれる訳です。

と言うことで、

ベイズの定理

$$P(B|A)=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\bar{B})P(\bar{B})}$$





ベン図ではこうなる。

$$=\frac{P(A|B)P(B)}{P(A)}$$

結局、分母は P(A)ってこと。

こう書きなおす

$$\underbrace{P(B|A)}_{\text{事後分布}}=\frac{\overbrace{P(A|B)}^{\text{尤度関数}}}{\underbrace{P(A)}_{\text{周辺尤度}}}\underbrace{P(B)}_{\text{事前分布}}$$

ここで

P(B)とは、Bが起こる確率
P(B|A)とは、Aが観測された時にBが起こる確率
なので、それぞれ事前分布と事後分布というわけ。

P(A|B)は、Bが起こるとした時にAが観測される確率、
つまり、パラメータが真の時にデータが得られる確率なので
尤度関数というわけ。

P(A)は、周辺尤度と呼ばれ、事前分布のもとで平均的に
Aが観測される確率、あるいはP(B|A)を確率変数にする
(積分して1にする)ための補正項だと思ってください。
というわけわかんない確率(ただしパラメータには依存しない量)
なので周辺尤度は省略して

$$P(B|A)\underbrace{\propto}_{\text{比例記号}} P(A|B)P(B)$$

と書かれる。

$$p(\pi|x)=\frac{p(x|\pi)}{\int p(x|\pi)p(\pi)d\pi}p(\pi)$$

ホントはこう書ける

g(p|data) ∝ g(p)L(p)

になってるわけです。2.2のウケ狙いの文章削って、ここらへんをちゃんと説明してほしいですよね。それとも、ベイズの定理は当然知っているという前提なのでしょうか。

えっと、ここで尤度関数

L(p) = p^s(1 - p)^f, 0 < p < 1

も出てきます。これはベルヌーイ試行の確率関数ですね。オモテが出る確率がp、ウラが出る確率が1 - pであるコイントスで、オモテ（成功）がs回、ウラ（失敗）がf回出たと言うことです。後で、ベータ分布が出てくるので、ベルヌーイ試行とベータ分布の関係も説明しましょうか。

と思ったけどめんどくさいので、

「入門ベイズ統計学：ファイナンス・ライブラリー10」中妻照雄（2007；朝倉書店）3,600円

から引用して済ませます。この本、金融・経済の分野ですが、ベイズ統計が丁寧に説明されていて（積分がいっぱい出てくるけど）わかりやすいです。

がわかります．ただし立て続けに納期に遅れているわけではないので，納期の遅延の頻度により事後分布の形状が1寄りになったり0寄りになったりしていくことになります．

$\begin{cases} x_i = 1 & \text{遅れた} \\ x_i = 0 & \text{守った} \end{cases}$

i

3.1.3　成功確率の事後分布の導出—データをまとめて使った場合

上では納期がくるたびに納期に遅れたかどうかを観測し，納期に遅れる確率 π の事後分布を新しく入手したデータで更新する方法を説明しました．しかし，過去の納期に関する情報をまとめて使って π の事後分布を求めることもできます．

ここで，表3.1に示されている過去5回の納期に関するデータ

$$(x_1, x_2, x_3, x_4, x_5) = (1, 0, 1, 0, 0) \tag{3.30}$$

が手元にあるとします．このデータを D と表記しましょう．確率変数 X_i $(i = 1, 2, 3, 4, 5)$ は互いに独立なので，(3.30)式のデータが実現する確率は，

$$\begin{aligned} p(D|\pi) &= p(x_1|\pi) \times p(x_2|\pi) \times p(x_3|\pi) \times p(x_4|\pi) \times p(x_5|\pi) \\ &= \prod_{i=1}^{5} \pi^{x_i}(1-\pi)^{1-x_i} = \pi^{\sum_{i=1}^{5} x_i}(1-\pi)^{5-\sum_{i=1}^{5} x_i} \end{aligned} \tag{3.31}$$

となります．D はすでに(3.30)式で与えられており，$\sum_{i=1}^{5} x_i = 2$ です．したがって，(3.31)式は

$$p(D|\pi) = \pi^2(1-\pi)^3 \tag{3.32}$$

と書き直されます．

そもそも(3.32)式は5つの確率変数 $(X_1, X_2, X_3, X_4, X_5)$ が(3.30)式で与えられる特定の値の組合わせをとる同時確率ですが，確率変数の実現値がすでに観測されているので(3.32)式を π の関数とみなすこともできます．このとき(3.32)式は**尤度 (likelihood)** と呼ばれます．尤度 $p(D|\pi)$ は仮に π が真の値であるとしたときにデータ D が実現する確率です．したがって，ある π に対して尤度 $p(D|\pi)$ が高いときはデータ D は観測されやすい数値の組合わせとなり，尤度 $p(D|\pi)$ が低いときはデータ D は観測されにくい数値の組合わせとな

ります．データ D はすでに観測されていますから，きわめて観測されにくい数値の組合わせであるとは考えにくいでしょう．よって，尤度が極端に低い π はデータと照らしあわせると真の値である可能性がきわめて低いと推測されるのです．なお，先のデータが逐次入手される場合で用いた

$$p(x_i|\pi) = \pi^{x_i}(1-\pi)^{1-x_i}$$

は1つの観測値 x_i が与えられたときの尤度と解釈されます．

データ D が実現する確率を使ったベイズの定理は

$$p(\pi|D) = \frac{p(D|\pi)p(\pi)}{\int_0^1 p(D|\pi)p(\pi)d\pi} \tag{3.33}$$

です．π の事前分布には同じく (3.11) 式を使いましょう．尤度 (3.32) と事前分布 (3.11) を (3.33) 式に当てはめると，

$$p(\pi|D) \propto p(D|\pi)p(\pi) \propto \pi^2(1-\pi)^3$$

となります．これは (3.22) 式と同じですから，π の事後分布は (3.29) 式となります．つまり，5回の納期に関するデータを一度に使って π の事後分布を求めても1回ごとに π の事後分布を更新していっても，全く同じデータ (3.30) を使って事後分布を評価していることに変わりはないので，同じ事後分布が得られることになります．

次に，過去 n 回の納期における遅延の状況がデータ $D=(x_1,\ldots,x_n)$ として与えられている場合の π の事後分布を導出しましょう．確率変数 X_i $(i=1,\ldots,n)$ が互いに独立に同じベルヌーイ分布 (3.3) に従うと仮定すると，データ D が実現する確率は

$$p(D|\pi) = \prod_{i=1}^{n} p(x_i|\pi) = \prod_{i=1}^{n} \pi^{x_i}(1-\pi)^{1-x_i} = \pi^{\sum_{i=1}^{n} x_i}(1-\pi)^{n-\sum_{i=1}^{n} x_i}$$

となります．ここで $y_n = \sum_{i=1}^{n} x_i$ と定義すると，$p(D|\pi)$ は

$$p(D|\pi) = \pi^{y_n}(1-\pi)^{n-y_n} \tag{3.34}$$

となります．これが一般の場合における π の尤度です．π の事前分布には一様

分布 (3.11) を使いましょう．事前分布 (3.11) と尤度 (3.34) に対してベイズの定理 (3.33) を適用すると，π の事後分布は

$$\boxed{p(\pi|D) = \frac{\pi^{y_n}(1-\pi)^{n-y_n}}{B(y_n+1, n-y_n+1)} \mathbf{1}_{(0,1)}(\pi)} \tag{3.35}$$

となります．(3.35) 式の $B(\cdot)$ はベータ関数です．(3.35) 式は**ベータ分布 (beta distribution)** と呼ばれる確率分布の確率密度関数です．ちなみに一般のベータ分布の確率密度関数は

$$p(x|\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \mathbf{1}_{(0,1)}(x) \tag{3.36}$$

です．代表的なベータ分布の確率密度関数 (3.36) のグラフが図 3.3 に示されています．ベータ分布は $\alpha > 1$ かつ $\beta > 1$ のときにモードを 1 つ持ちます．$\alpha > 1$ かつ $\beta \leq 1$ の場合には右上がり，$\alpha \leq 1$ かつ $\beta > 1$ の場合には右下がりのグラフになります．$\alpha < 1$ かつ $\beta < 1$ のときはグラフは U 字型です．本書ではベータ分布 (3.36) を $\mathcal{B}e(\alpha, \beta)$ と表記し，ある確率変数 X がベータ分布 $\mathcal{B}e(\alpha, \beta)$ に従うことを $X \sim \mathcal{B}e(\alpha, \beta)$ と表記します．この表記に従うと，(3.35) 式のベー

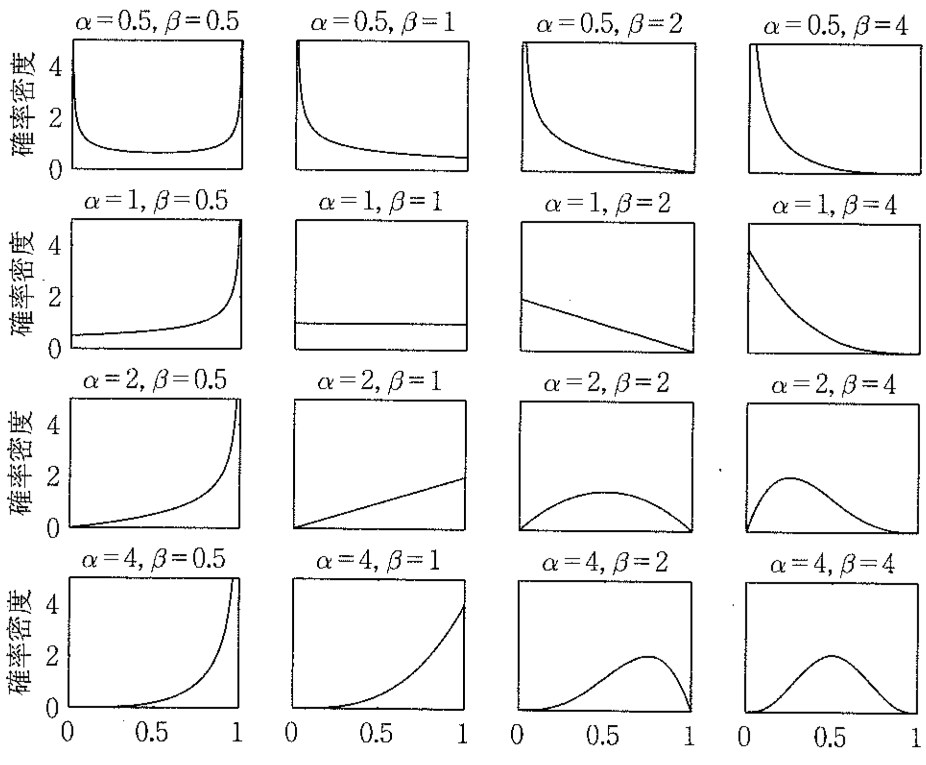


図 3.3 ベータ分布の確率密度関数

タ分布は $\mathcal{B}e(y_n+1, n-y_n+1)$ となり，$\mathcal{B}e(y_n+1, n-y_n+1)$ が π の事後分布であることは

$$\pi|D \sim \mathcal{B}e(y_n+1, n-y_n+1) \tag{3.37}$$

と表記されます．

図 3.1 では納期に遅れた回数が増すに従って事後分布が 1 の近辺に集まっていく様子を確認しました．同様のことが一般の π の事後分布 (3.35) についてもいえます．すべての納期の回数 n と納期に遅れた回数 y_n の比率を $y_n/n = 0.8$，つまり $y_n = 0.8n$ に固定しておき，n を 5，10，100，1000 と増やしていったときに π の事後分布がどのように変化していくかをみてみましょう．y_n と n の比率を 0.8 に固定しているので，(3.35) 式は

$$p(\pi|D) = \frac{\pi^{0.8n}(1-\pi)^{0.2n}}{B(0.8n+1, 0.2n+1)}\mathbf{1}_{(0,1)}(\pi) \tag{3.38}$$

と書き直されます．(3.38) 式のグラフを $n = 5, 10, 100, 1000$ の場合について描画したものが図 3.4 です．図 3.4 に示されている 4 つの事後分布のモードはす

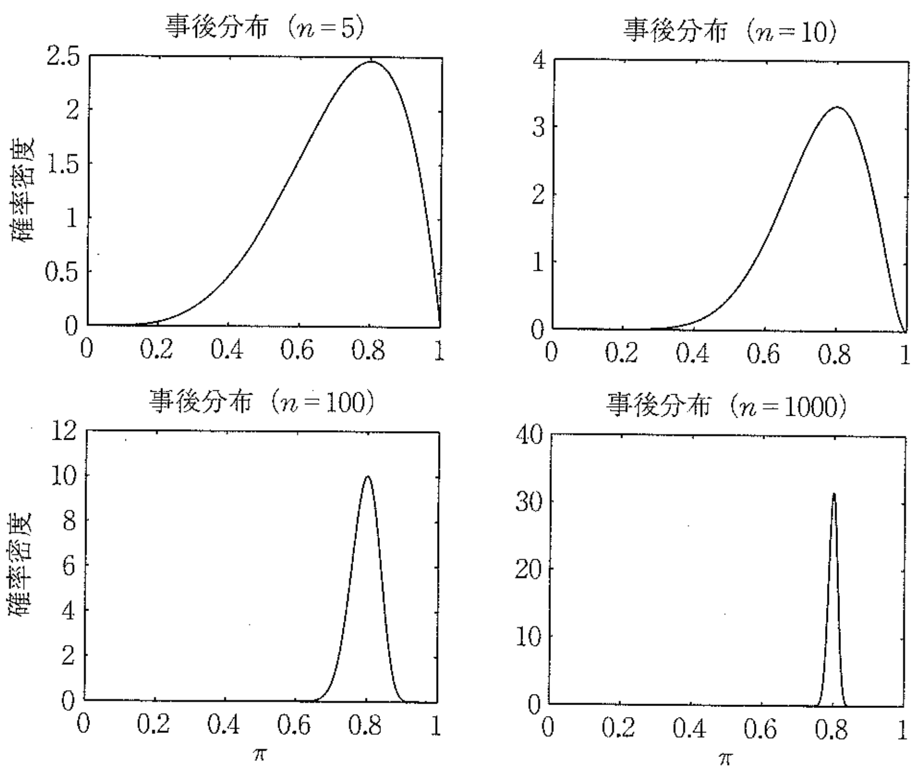


図 3.4　データの数と事後分布の関係

尤度関数 → (事前分布が一様分布ならば) 事後分布の
確率密度関数