入門ベイズ統計　by 松原　望

レジュメ作成者・朝倉利晃

* 1. 補足 〜確率の表記の違いによる意味の違いについて〜

　前回の勉強会ではMLE(maximum likelihood estimation)について、概念と簡単な具体例についてみた。ここでは、P(data), P(θ), P(data|θ), P(θ |data)とLikelihood functionの違いを明確にする。ここでdataは観測した事象、例えば生徒の点数Θ はモデルのパラメータである。例えば正規分布ならば θ = (μ,σ) となる。

　P(data), すなわちdataが起きる確率については、このままでは求めることが出来ない。我々がある事象の確率を議論するときにはその背景になんらかのモデルやパラメーターを仮定しなければいけない。よく使われる方法とはしては周辺分布(marginal distribution)を用いる方法がある。

　これは、あるモデル上で取りうる全てのパラメーターの値に対して、dataが起きる確率を評価し均一に足し合わせたものである。また、周辺分布はベイズの定理の分母に相当する。

　P(θ)に関しては、ベイズの定理の下では事前分布として取り扱われる。P(data)における議論と同様に我々はθに対して何らかのモデルやパラメーターの仮定を置かない限り、θに関して確率を議論することは出来ない。ベイズの定理の下では往々にしてP(θ)は知ることが出来ないため恣意的な選択をせざるを得ない。そこで理由不十分の原則に則り、離散分布の場合、取りうるパラメーターの確率が全て等しくなるように事前分布を与える。すなわち、であるとき任意のに関して とする。連続分布の場合は、P(θ) = 1 としたり（これはimproper prior distributionと呼ぶ。パラメーターの全区間を積分すると1にならないという意味でimproper）、分散が非常に大きい正規分布を仮定したりする。

　P(data| θ)はあるモデル、パラメータの仮定のもとでのdataが生じる確率である。一般的な確率分布が表しているのは全てこの形である。確率の基礎的な教科書では”Xが正規分布に従うとすればP(X=x)は〜”と書いたりすることがあるが、これはP(data | θ)の内、パラメーターの表記を省略している。

　僕がMLE(maximum likelihood estimation)を最初に学んだときに非常に戸惑ったのがデータが生じる確率に関する関数とLikelihood functionの違いについてだ。前者はP(X| θ)と表され、後者も同様にP(X| θ)と書かれることも多いが、ここでは明示的にL(data| θ)と書く。前者の例としては、dataが平均μ、分散σ^2の正規分布に従う変 数X( X ~ N(μ, σ^2) ) だとすると、

となり、横軸に確率変数X、縦軸にP(data| θ)のグラフが書かれる。Xに関して積分するとその値は1になる。(

また、dataが平均μ、分散σ^2の正規分布から独立に取ったn個の確率変 数Xだとすると、確率は となる。また取りうる値に関して積分すればその値は1になる。

　一方、Likelihood functionの場合は、dataが例えば1つの観測値で値が2だとする。これが分散σ^2の正規分布に従うと仮定すると、Likelihood functionは

となり、横軸にμ 、縦軸にのグラフが書かれる。今回の場合は、μに関して積分を行っても値は1になるが、likelihood functionは取りうる値について積分を行った結果が1になることは保証されていない。また、dataが複数の観測値だとすると、Likelihood function は

となる。この場合でもグラフにすれば横軸にμ、縦軸にL(data|μ,σ)が来る。

　以上の違いをまとめると、P(X| θ)は、パラメーターが既知、かつ、データの値が変化したときに確率がどう変化するかを表す。また、変数に関しての積分値は常に1である。L(data| θ)は、データが既知、かつ、データに対してパラメーターが変化したときにlikelihood(尤度、確からしさ)がどう変化するかを表す。また、変数に関しての積分値は1になるとは限らない。

　最後にP(θ |data)について考える。我々はθに関して確率を考えたいが、残念ながらdataはθに関する確率を考えるためのモデルやパラメーターに関して何の知見も与えない。そのため、何らかのモデルやパラメーターを仮定しなければいけないがここでベイズの定理を用いると、θ自身がθに関する確率のモデルやパラメーターに関して知見を与えてくれることがわかる。

　左辺を事後分布、右辺の分子の第一項目はLikelihood、第二項目は事前分布、分母は周辺分布と呼ばれる。分母が周辺分布と呼ばれる所以は(1)に示したように計算するためにはなんらかの仮定をおいた上で積分操作を行わない限り確率は求められないからだ。

　\*一言述べて置くと、ベイズの定理の式を用いるとデータがある下でパラメーターの分布を求める際にはLikelihood functionが必要となるが、これはたまたまLikelihood functionがベイズの定理と親和性があるのであって理論としては別物である。

1.7 事後分布

　ベイズの定理の確認

-離散値の場合、

原因をθ、結果をｚ(多くの場合data)とおいて事前確率を、事後確率を、likelihoodをと書けば

補足の表記に従えば

-連続値の場合、

補足の表記に従えば

分母に関しては定数となるから*MLEを行う際には*分子のみを考えればよい。正確な分布を求める際には分母の積分を行わなければならない。代表的なものの場合、分母の計算が楽な場合がある。

これらの形式でデータを取り扱う統計学を”ベイズ統計学”

これを用いた決定を”ベイズ統計”という。

1.8 事前分布

　Likelihood funcition p(z|θ)に対して“自然な共役事前分布” とは事前分布にこの分布を選べば事後分布も同じ分布族になるような分布を指す。例えば二項分布に対しては、ベータ分布を事前分布とすると事後分布もベータ分布となる。

　この章では、Likelihood functionとその自然な共役事前分布としてとして次のものが紹介されている。Likelihood function (共役事前分布) として書く。

・二項分布(ベータ分布)

・正規分布(正規分布)

・ポアソン分布(ガンマ分布)

* 1. .1 二項分布(ベータ分布)

　今、独立な確率変数で であるとする。このときdataがのときのlikelihoodは

となる。事前分布, w(θ),としてパラメーターα,βのベータ分布をとる。ベータ関数を

と書けば、

となる。したがって(1.7.2)の計算は

ここで

したがってこれはまたベータ分布となる。したがって実際に分布の更新を行う際にはパラメーターの更新だけ行えばよい。

--------例：女の子が生まれる確率------------------------------------------------------------------

ある両親から、連続して男の子が3人生まれた。次の子が女の子である確率はどれほどか。

------------------------------------------------------------------------------------------------------------------

Ans1. 通常の統計学としては男:女=1.05:1.00 の比で生まれてくる。

Ans2. ベイズ統計学としては、次のとおりになる。

事前分布: のベータ分布、

n = 3 , より

事後分布:のベータ分布

となる。

\*注：最初においた1については何の客観的な妥当性はない。

1.8.2 正規分布(正規分布)

　今、標本のそれぞれのxi が独立な確率変数で平均θ、分散σ^2の正規分布に従うとすると、標本zのlikelihood p(・|θ)は、

　expカッコ内の和について

よって

興味あるのはθに関してであるので第一因子のみが情報として寄与する。

θ の事前分布がN(μ,τ^2)に従うとする。

これらを用いれば事後分布は次のように表される。ただし、正規分布の更新の際には標本の分散は一定であると仮定している。

分子に関して計算してあげると

ここで

である。これを(1.8.9.5)の中に叩き込んであげると

となるので、結局、 (1.8.10) が言えるのである。

1.8.3 ポアソン分布 (ガンマ分布)

　今までと同様な手順でを独立な確率変数として、平均θのポアソン分布に従うとする。とすればlikelihoodは

事前分布をparameter λ,αのガンマ分布とする。

ただし、である。

これらを用いて事後分布を計算すると

したがって、 (1.8.15) が得られる。

ただし、　である。

--------例：平均火災件数-------------------------------------------------------------------------------

新人消防署長は管内の1ヶ月あたりの火災件数の平均を10土√10 (土は標準偏差)とみた。実際は、最初の四半期に23件であった。彼の予想はどう変わったか。

------------------------------------------------------------------------------------------------------------------

Ans.

まずモデルとして、事前分布にガンマ分布、それぞれの独立変数がポアソン分布に従うとする。(1.8.3の設定と同じ)。

事前分布のα,βを定めるために、

を解くと となる。

最初の四半期に23件は換言すれば、一ヶ月に起きる火災件数がポアソン分布に従うならば、ポアソン分布に従う3つの確率変数の実現値の和が23件と言える。

よって よりである。これは平均 33/4 = 5.75, 分散23/16=1.44 のガンマ分布である。