正規分布モデルの共役事前分布によるベイズ統計

- 黒木玄
- 2022-09-03

目次

In [1]:

```julia
using Distributions
using StatsPlots
default(fmt=:png, size=(400, 250),
    titlefontsize=10, tickfontsize=6, guidefontsize=9,
    plot_titlefontsize=10)
using SymPy
```

In [2]:

```julia
# Override the Base.show definition of SymPy.jl:
# https://github.com/JuliaPy/SymPy.jl/blob/29c5bfd1d10ac53014fa7fef468bc8deccadc2fc/src/types.

@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::SymbolicObject)
    print(io, as_markdown("\\displaystyle " *
            sympy.latex(x, mode="plain", fold_short_frac=false)))
end
@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
    function toeqnarray(x::Vector{Sym})
        a = join(["\\displaystyle " *
                sympy.latex(x[i]) for i in 1:length(x)], "\\\\")
        """\\left[ \\begin{array}{r}$a\\end{array} \\right]"""
    end
    function toeqnarray(x::AbstractArray{Sym,2})
        sz = size(x)
        a = join([join("\\displaystyle " .* map(sympy.latex, x[i,:]), "&")
                for i in 1:sz[1]], "\\\\")
        "\\left[ \\begin{array}{" * repeat("r",sz[2]) * "}" * a * "\\end{array}\\right]"
    end
    print(io, as_markdown(toeqnarray(x)))
end
```

1 正規分布モデルの共役事前分布とその応用

1.1 逆ガンマ正規分布

平均 $\mu \in \mathbb{R}$, 分散 $v = \sigma^2 \in \mathbb{R}_{>0}$ の正規分布の確率密度函数を次のように表す:

$$p_{\mathrm{Normal}}(y|\mu, v) = \frac{1}{\sqrt{2\pi v}}\exp\left(-\frac{1}{2v}(y-\mu)^2\right) \quad (y \in \mathbb{R}).$$

分散パラメータ σ^2 を v に書き直している理由は, σ^2 を1つの変数として扱いたいからである.

パラメータ $\kappa, \theta > 0$ の逆ガンマ分布の確率密度函数を次のように書くことにする:

$$p_{\mathrm{InverseGamma}}(v|\kappa, \theta) = \frac{\theta^\kappa}{\Gamma(\kappa)} v^{-\kappa-1} \exp\left(-\frac{\theta}{v}\right) \quad (v > 0).$$

v がこの逆ガンマ分布に従う確率変数だとすると,

$$
\frac{1}{v} \sim \mathrm{Gamma}\left(\kappa, \frac{1}{\theta}\right) = \frac{1}{2\theta}\mathrm{Gamma}\left(\frac{2\kappa}{2}, 2\right) = \frac{1}{2\theta}\mathrm{Chisq}(2\kappa),
$$
$$
E[v] = \frac{\theta}{\kappa - 1}, \quad \mathrm{var}(v) = \frac{E[v]^2}{\kappa - 2}.
$$

A と B が μ, v に関する定数因子の違いを除いて等しいことを $A \propto B$ と書くことにする.

逆ガンマ正規分布の密度函数を次のように定義する:

$$
\begin{aligned}
p_{\mathrm{InverseGammaNormal}}(\mu, v|\mu_*, v_*, \kappa, \theta) &= p_{\mathrm{Normal}}(\mu|\mu_*, v_* v) p_{\mathrm{InverseGamma}}(v|\kappa, \theta) \\
&\propto v^{-(\kappa+1/2)-1} \exp\left(-\frac{1}{v}\left(\theta + \frac{1}{2v_*}(\mu - \mu_*)^2\right)\right).
\end{aligned}
$$

この逆ガンマ正規分布の密度函数に従う確率変数を μ, v と書くと,

$$
E[v] = \frac{\theta}{\kappa - 1}, \quad \mathrm{var}(v) = \frac{E[v]^2}{\kappa - 2}, \quad \mathrm{cov}(\mu, v) = 0, \quad E[\mu] = \mu_*, \quad \mathrm{var}(\mu) = v_* E[v].
$$

この逆ガンマ正規分布が正規分布の共役事前分布になっていることを次の節で確認する.

1.2 共役事前分布のBayes更新

データの数値 $y_1, \dots, y_n$ が与えられたとき, 正規分布モデルの尤度函数は

$$
\prod_{i=1}^n p_{\mathrm{Normal}}(y_i|\mu, v) \propto v^{-n/2} \exp\left(-\frac{1}{2v}\sum_{i=1}^n (y_i - \mu)^2\right)
$$

の形になる. このとき,

$$
\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2.
$$

とおくと,

$$
\sum_{i=1}^n (y_i - \mu)^2 = n(\mu - \bar{y})^2 + n\hat{\sigma}^2
$$

なので, 尤度を最大化する μ, v は $\mu = \bar{y}, v = \hat{\sigma}^2$ になることがわかる.

さらに, 次が成立することもわかる:

$$
\begin{aligned}
&\prod_{i=1}^n p_{\mathrm{Normal}}(y_i|\mu, v) \times p_{\mathrm{InverseGammaNormal}}(\mu, v|\mu_*, v_*, \kappa, \theta) \\
&\propto v^{-n/2} \exp\left(-\frac{n}{2v}\left((\mu - \bar{y})^2 + \hat{\sigma}^2\right)\right) \times v^{-(\kappa+1/2)-1} \exp\left(-\frac{1}{v}\left(\theta + \frac{1}{2v_*}(\mu - \mu_*)^2\right)\right) \\
&= v^{-(\kappa+n/2+1/2)-1} \exp\left(-\frac{1}{v}\left(\theta + \frac{n}{2}\left(\hat{\sigma}^2 + \frac{(\bar{y} - \mu_*)^2}{1 + nv_*}\right) + \frac{1 + nv_*}{2v_*}\left(\mu - \frac{\mu_* + nv_*\bar{y}}{1 + nv_*}\right)^2\right)\right).
\end{aligned}
$$

ゆえに共役事前分布から得られる事後分布のパラメータは次のようになる:

$$
\begin{aligned}
&\tilde{\kappa} = \kappa + \frac{n}{2} = \frac{n}{2}\left(1 + \frac{2\kappa}{n}\right), \\
&\tilde{\theta} = \theta + \frac{n}{2}\left(\hat{\sigma}^2 + \frac{(\bar{y} - \mu_*)^2}{1 + nv_*}\right) = \frac{n\hat{\sigma}^2}{2}\left(1 + \frac{2\theta}{n\hat{\sigma}^2} + \frac{(\bar{y} - \mu_*)^2}{(1 + nv_*)\hat{\sigma}^2}\right), \\
&\tilde{\mu}_* = \frac{\mu_* + nv_*\bar{y}}{1 + nv_*} = \bar{y}\frac{1 + \mu_*/(nv_*\bar{y})}{1 + 1/(nv_*)}, \\
&\tilde{v}_* = \frac{v_*}{1 + nv_*} = \frac{1}{n}\frac{1}{1 + 1/(nv_*)}.
\end{aligned}
$$

In [3]:
```
@vars n ȳ v̂ μ v μ0 v0 κ θ
```

Out[3]: (n, ȳ, v̂, μ, v, μ0, v0, κ, θ)

In [4]:
```
negloglik = n/2*log(v) + n/(2v)*((μ - ȳ)^2 + v̂)
```
Out[4]:
$$\frac{n\log(v)}{2} + \frac{n\left(\hat{v} + (-\bar{y} + \mu)^2\right)}{2v}$$

In [5]:
```
neglogpri = (κ + 1//2 + 1)*log(v) + 1/v*(θ + 1/(2v0)*(μ-μ0)^2)
```
Out[5]:
$$\left(\kappa + \frac{3}{2}\right)\log(v) + \frac{\theta + \frac{(\mu-\mu_0)^2}{2v_0}}{v}$$

In [6]:
```
neglogpost = (κ + n/2 + 1//2 + 1)*log(v) +  1/v*(
    θ + n/2*(v̂ + 1/(1+n*v0)*(ȳ - μ0)^2) +
    (1 + n*v0)/(2v0)*(μ - (μ0 + n*v0*ȳ)/(1 + n*v0))^2)
```
Out[6]:
$$\left(\frac{n}{2} + \kappa + \frac{3}{2}\right)\log(v) + \frac{\frac{n\left(\hat{v} + \frac{(\bar{y}-\mu_0)^2}{nv_0+1}\right)}{2} + \theta + \frac{\left(\mu - \frac{nv_0\bar{y}+\mu_0}{nv_0+1}\right)^2(nv_0+1)}{2v_0}}{v}$$

In [7]:
```
simplify(negloglik + neglogpri - neglogpost)
```
Out[7]: 0

1.3 μの周辺事前・事後分布および事前・事後予測分布

確率密度函数

$$p(\mu|\mu_*, v_*, \kappa, \theta) = \int_{\mathbb{R}_{>0}} p_{\text{InverseGammaNormal}}(\mu, v|\mu_*, v_*, \kappa, \theta)\,dv$$

で定義されるμ の周辺事前分布は次になる:

$$\mu \sim \mu_* + \sqrt{\frac{\theta}{\kappa}v_*}\ \text{TDist}(2\kappa).$$

確率密度函数

$$p_*(y_{\text{new}}|\mu_*, v_*, \kappa, \theta) = \iint_{\mathbb{R}\times\mathbb{R}_{>0}} p_{\text{Normal}}(y_{\text{new}}|\mu, v) p_{\text{InverseGammaNormal}}(\mu, v|\mu_*, v_*, \kappa, \theta)\,d\mu\,dv$$

で定義される y_{new} の事前予測分布は次になる:

$$y_{\text{new}} \sim \mu_* + \sqrt{\frac{\theta}{\kappa}(1+v_*)}\ \text{TDist}(2\kappa).$$

パラメータをBayes更新後のパラメータ

$$\begin{aligned}
\tilde{\kappa} &= \kappa + \frac{n}{2} = \frac{n}{2}\left(1 + \frac{2\kappa}{n}\right),\\
\tilde{\theta} &= \theta + \frac{n}{2}\left(\hat{\sigma}^2 + \frac{(\bar{y}-\mu_*)^2}{1+nv_*}\right) = \frac{n\hat{\sigma}^2}{2}\left(1 + \frac{2\theta}{n\hat{\sigma}^2} + \frac{(\bar{y}-\mu_*)^2}{(1+nv_*)\hat{\sigma}^2}\right),\\
\tilde{\mu}_* &= \frac{\mu_* + nv_*\bar{y}}{1+nv_*} = \bar{y}\frac{1 + \mu_*/(nv_*\bar{y})}{1 + 1/(nv_*)},\\
\tilde{v}_* &= \frac{v_*}{1+nv_*} = \frac{1}{n}\frac{1}{1+1/(nv_*)}.
\end{aligned}$$

に置き換えればこれは μ の周辺事後分布および事後予測分布になる.

その事後分布を使った区間推定の幅は

- n が大きいほど狭くなる.
- κ が大きいほど狭くなる.
- θ が大きいほど広くなる.
- $|\bar{y}-\mu_*|/\hat{\sigma}$ が大きいほど広くなる.
- $|\bar{y}-\mu_*|/\hat{\sigma}$ が大きくても, v_* がさらに大きければ狭くなる.

1.4 Jeffreys事前分布の場合

パラメータ空間が $\{(\mu, v) = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0}\}$ の 2 次元の正規分布モデルのJeffreys事前分布 $p_{\text{Jeffreys}}(\mu, v)$ は

$$p_{\text{Jeffreys}}(\mu, v) \propto v^{-3/2}$$

になることが知られている. ただし, 右辺の $(\mu, v) \in \mathbb{R} \times \mathbb{R}_{>0}$ に関する積分は ∞ になるので, この場合のJeffreys事前分布はimproperである.

逆ガンマ正規分布の密度函数

$$p_{\text{InverseGammaNormal}}(\mu, v | \mu_*, v_*, \kappa, \theta) \propto v^{-(\kappa+1/2)-1} \exp\left(-\frac{1}{v}\left(\theta + \frac{1}{2v_*}(\mu - \mu_*)^2\right)\right).$$

と比較すると, Jeffreys事前分布に対応する共役事前分布のパラメータ値は形式的に次になることがわかる:

$$\kappa \to 0, \quad \theta \to 0, \quad v_* \to \infty.$$

そのとき, Bayes更新後のパラメータの公式は次のようにシンプルになる:

$$\tilde{\kappa} = \frac{n}{2}, \quad \tilde{\theta} = \frac{n\hat{\sigma}^2}{2}, \quad \tilde{\mu}_* = \bar{y}, \quad \tilde{v}_* = \frac{1}{n}.$$

さらに, 前節の公式から, $n \to \infty$ のとき, 一般のパラメータ値に関するBayes更新の結果は, $n \to \infty$ のとき漸近的にこのJeffreys事前分布の場合に一致する.

さらに, Jeffreys事前分布の場合には

$$\frac{\tilde{\theta}}{\tilde{\kappa}} = \hat{\sigma}^2, \quad \tilde{v}_* = \frac{1}{n}, \quad 2\tilde{\kappa} = n.$$

ゆえに, μ に関する周辺事後分布は

$$\mu \sim \bar{y} + \frac{\hat{\sigma}}{\sqrt{n}}\ \mathrm{TDist}(n)$$

になり, 事後予測分布は次になる:

$$y_{\text{new}} \sim \bar{y} + \hat{\sigma}\sqrt{1 + \frac{1}{n}}\ \mathrm{TDist}(n).$$

1.5 平均と対数分散について一様な事前分布の場合

平均 μ と分数の対数 $\log v = \log \sigma^2$ に関する一様な事前分布は

$$p_{\text{flat}}(\mu, v) \propto v^{-1}$$

になる. ただし, 右辺の $(\mu, v) \in \mathbb{R} \times \mathbb{R}_{>0}$ に関する積分は ∞ になるので, この事前分布はimproperである.

逆ガンマ正規分布の密度函数

$$p_{\text{InverseGammaNormal}}(\mu, v | \mu_*, v_*, \kappa, \theta) \propto v^{-(\kappa+1/2)-1} \exp\left(-\frac{1}{v}\left(\theta + \frac{1}{2v_*}(\mu - \mu_*)^2\right)\right).$$

と比較すると, 平均と対数分散について一様な事前分布に対応する共役事前分布のパラメータ値は形式的に次になることがわかる:

$$\kappa \to -\frac{1}{2}, \quad \theta \to 0, \quad v_* \to \infty.$$

このとき, Bayes更新後のパラメータの公式は次のようになる:

$$\tilde{\kappa} = \frac{n-1}{2}, \quad \tilde{\theta} = \frac{n\hat{\sigma}^2}{2}, \quad \tilde{\mu}_* = \bar{y}, \quad \tilde{v}_* = \frac{1}{n}.$$

この場合には

$$\frac{\tilde{\theta}}{\tilde{\kappa}} = \frac{n\hat{\sigma}^2}{n-1} = s^2, \quad \tilde{v}_* = \frac{1}{n}, \quad 2\tilde{\kappa} = n - 1.$$

ここで, s^2 はデータの数値 $y_1, \ldots, y_n$ の不偏分散

$$
s^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2 = \frac{n\hat{\sigma}^2}{n-1} > \hat{\sigma}^2
$$

であり, s はその平方根である.

ゆえに, μ に関する周辺事後分布は

$$
\mu \sim \bar{y} + \frac{s}{\sqrt{n}}\ \mathrm{TDist}(n-1)
$$

になり, y_{new} に関する事後予測分布は次になる:

$$
y_{\text{new}} \sim \bar{y} + s\sqrt{1+\frac{1}{n}}\ \mathrm{TDist}(n-1).
$$

したがって, 前節の結果と比較すると, Jeffreys事前分布の事後分布と予測分布による区間推定よりもこの場合の区間推定は少し広くなる.

1.6 通常の信頼区間と予測区間との比較

通常の t 分布を使う平均の信頼区間と次の値の予測区間の構成では以下を使う:

$$
\frac{\bar{y}-\mu}{s/\sqrt{n}} \sim \mathrm{TDist}(n-1), \qquad \frac{y_{\text{new}}-\bar{y}}{s\sqrt{1+1/n}} \sim \mathrm{TDist}(n-1).
$$

ここで, s^2 はデータの数値の不偏分散であり, s はその平方根である.

したがって, 前節の結果と比較すると, 通常の信頼区間と予測区間は, 平均と対数分散に関する一様事前分布に関する事後分布と予測分布を用いた区間推定に一致する.

1.7 データの数値から事前分布を決めた場合

$a, b > 0$ であると仮定する.

データの数値から共役事前分布のパラメータを次の条件によって決めたと仮定する:

$$
E[\mu] = \mu_* = \bar{y}, \quad E[v] = \frac{\theta}{\kappa-1} = \hat{\sigma}^2, \quad \mathrm{var}(\mu) = v_* E[v] = a\hat{\sigma}^2, \quad \mathrm{var}(v) = \frac{E[v]^2}{\kappa-2} = b\hat{\sigma}^4.
$$

これは次と同値である:

$$
\mu_* = \bar{y}, \quad v_* = a, \quad \kappa = 2 + \frac{1}{b}, \quad \theta = \hat{\sigma}^2\left(1+\frac{1}{b}\right).
$$

これのBayes更新の結果は以下のようになる:

$$
\begin{aligned}
\tilde{\kappa} &= 2 + \frac{1}{b} + \frac{n}{2} = \frac{n}{2}\left(1 + \frac{2(2+1/b)}{n}\right) && \to 2 + \frac{n}{2},\\
\tilde{\theta} &= \hat{\sigma}^2\left(1+\frac{1}{b}+\frac{n}{2}\right) + \frac{n}{2}\frac{(\bar{y}-\bar{y})^2}{1+na} = \frac{n\hat{\sigma}^2}{2}\left(1+\frac{2(1+1/b))}{n}\right) && \to \hat{\sigma}^2\left(1+\frac{n}{2}\right),\\
\tilde{\mu}_* &= \frac{\bar{y}+nv_*\bar{y}}{1+nv_*} = \bar{y} && \to \bar{y},\\
\tilde{v}_* &= \frac{a}{1+na} = \frac{1}{n}\frac{1}{1+1/(na)} && \to \frac{1}{n}.
\end{aligned}
$$

以上における $\to$ は $a \to \infty$, $b \to \infty$ での極限を意味する.

以上の構成のポイントは, $\mu_* = \bar{y}$ となっているおかげで, $\tilde{\mu}_*$ も $\tilde{\mu}_* = \bar{y}$ となってバイアスが消え, さらに, $\tilde{\theta}$ の中の $\dfrac{n}{2}\dfrac{(\bar{y}-\mu_*)^2}{1+na}$ の項が消えて, 区間推定の幅が無用に広くならずに済むことである.

ただし, この場合には

$$
\frac{\tilde{\theta}}{\tilde{\kappa}} = \hat{\sigma}^2\frac{1+2(1+1/b)/n}{1+2(2+1/b)/n} < \hat{\sigma}^2, \quad v_* = \frac{1}{n}\frac{1}{1+1/(na)} < \frac{1}{n}
$$

なので, 区間推定の幅はJeffreys事前分布の場合よりも少し狭くなる.

しかし, n が大きければそれらの違いは小さくなる.

In [ ]:

```
1
```