On Bayesian utility maximization

Erwann Rogard

2022-04-06

Abstract

The original notes are dated 2007-04-24. They were supposed to ingrate with those on the stochastic mesh[1]. They have as their starting point a paper on parametric formulation of 'Sequential Decision Analysis in Clinical Trials'[1]

1 utility

Throughout we will assume the following utility

$$u_i(\theta, a) = \begin{cases} -c_{i,T}(\theta - c_P)^+ - c_i, & a = T \\ -c_{i,P}(c_T - \theta)^+ - c_i, & a = P \\ -c_i - \infty \times 1\{i = n\}, & a = C \end{cases} \tag{1}$$

where the $c_{i,a}$'s denote penalties, $[c_P, c_T]$ denotes and indifference zone, and c_j is the cost of sampling data at period i.

Recall that $i = 0, ..., i_*$ indexes the sampling period. The sample sizes are determined in advance; let $n_1, ..., n_{i_*}$ denote the cumulative sample sizes for each period.

2 Normal mean, known variances

Parameter and data are distributed as

$$p(\theta) = N(\theta|\mu, \sigma_0^2) \tag{2}$$

$$p(\mathbf{y}_{n_i}) = \Pi_{j=1}^{i} N(\bar{y}_j|\theta, \sigma_j^2/n_j) \tag{3}$$

[1] https://github.com/erwannr/statistics/blob/main/stochmesh/stochmesh.pdf

For simplicity $n_i = 1$. The posterior distribution is $p(\theta|\mathbf{y}_i) = p(\theta|\mu(\mathbf{y}_i)) = N(\mu(\mathbf{y}_i), \boldsymbol{\sigma}_i^2)$ with

$$\mu(\mathbf{y}_i) = \boldsymbol{\sigma}_i^2(\boldsymbol{\sigma}_{i-1}^{-2}\mu(\mathbf{y}_{i-1}) + \sigma_i^{-2}y_i) \tag{4}$$

$$\boldsymbol{\sigma}_i^2 = \begin{cases} (\boldsymbol{\sigma}_i^{-2} + \sigma_i^{-2})^{-1}, & 1 \leq n \\ \sigma_0^2, & i = 0 \end{cases} \tag{5}$$

Under this model, $\mathrm{E}[u_i(\theta, a)|\mathbf{y}_i]$ for $a \in \mathcal{A}^{\mathrm{s}}$, has an explicit form in terms of $\phi(.)$ and $\Phi(.)$, the density and cumulative probability, respectively, of the standard normal distribution. We use the Beasly–Springer–Moro algorithm for a fast computation of $\Phi(.)$ For *Algorithm* 2 we also need $p(\mu(\mathbf{y}_{i+1})|\mu(\mathbf{y}_i))$ which has an explicit form in terms of $\phi(.)$.

The boundaries, parameterized as follows

$$T_i = \{\mathbf{y}_i : \mu(\mathbf{y}_i) \geq \mu_i\} \cap C_i^{\mathrm{c}} \tag{6}$$

$$P_i = \{\mathbf{y}_i : \mu(\mathbf{y}_i) < \mu_i\} \cap C_i^{\mathrm{c}} \tag{7}$$

$$C_i = \{\mathbf{y}_i : \gamma_i^- < \mu(\mathbf{y}_i) < \gamma_i^+\} \tag{8}$$

where $\mu_i = \frac{c_{i,T}c_P - c_{i,P}c_T}{c_{i,T}+c_{i,P}}$, so that $\gamma = (\tilde{\gamma}_1, ..., \tilde{\gamma}_{n-1})$ with $\tilde{\gamma}_i = (\gamma_i^-, \gamma_i^+)$. This parametrization lead is optimal.

3 Normal mean, Gamma precision

We assume (3) for the data and replace (2) by

$$p(\theta|\sigma^2) = N(\theta|\mu, \sigma^2/\kappa) \tag{9}$$

$$p(\rho) = \chi^2(\rho|\nu_0, \sigma_0^2) \tag{10}$$

$$\propto \rho^{\frac{\nu_0}{2}-1}\exp(-\nu_0\sigma_0^2\rho/2) \tag{11}$$

where $\rho \triangleq 1/\sigma^2$. The posterior distribution is

$$p(\theta, \rho|\mathbf{y}_n) = p(\theta, \rho|\bar{y}_n, s_n) \tag{12}$$

$$\propto N(\theta|\mu(\bar{y}_n), (\kappa_i\rho)^{-1})\chi^2(\rho|\nu_n, \sigma_i^2(\bar{y}_n, s_n)) \tag{13}$$

where

$$\mu(\bar{y}_n) = \frac{\kappa\mu + n\bar{y}_n}{\kappa_n} \tag{14}$$

$$\kappa_n = \kappa + n \tag{15}$$

$$\nu_n = \nu_0 + n \tag{16}$$

$$\nu_n\sigma_n^2(\bar{y}_n, s_n^2) = \frac{n\kappa}{\kappa_n}(\mu - \bar{y})^2 + (n-1)s_n^2 + \nu_0\sigma_0^2 \tag{17}$$

Equivalently,

$$p(\theta|\rho, \mathbf{y}_n) = N(\theta|\tilde{\mu}_n, (\rho\kappa_n)^{-1}) \tag{18}$$

$$p(\rho|\mathbf{y}_n) = \chi^2(\rho|\nu_n, \tilde{\sigma}_n^2) \tag{19}$$

where $\tilde{\mu}_n \triangleq \mu(\bar{y}_n)$ and $\tilde{\sigma}_n^2 \triangleq \sigma_n^2(\bar{y}_n, s_n^2)$. To compute the expected utility, we will need

$$p(\theta|\bar{y}_n, s_n^2) = t_{\nu_n}(\theta|\tilde{\mu}_n, \tilde{\sigma}_n^2/\kappa_n) \tag{20}$$

These are standard results from [2, Section 3.3]. In order to compute the weights in *Algorithm* 2 we have to determine the predictive distribution:

$$\begin{aligned}
&p(\bar{y}_{n_2}, s_{n_2}^2|\bar{y}_{n_1}, s_{n_1}^2) \qquad (21)\\
&= \mathrm{E}[N(\bar{y}_{n_2}|\theta, (n_2\rho)^{-1})\chi^2(s_{n_2}^2|n_2-1, \rho)|\bar{y}_{n_1}, s_{n_1}^2]\\
&\propto (s_{n_2}^2)^{(n_2-1)/2-1} \int\int \Gamma\left(\rho|\nu_{n_1+n_2+1}/2, \beta(\bar{y}_{n_2}, s_{n_2}^2)\right) d\theta d\rho\\
&\propto (s_{n_2}^2)^{(n_2-1)/2-1} \int \beta(\bar{y}_{n_2}, s_{n_2}^2)^{-\nu_{n_1+n_2+1}/2} d\theta\\
&\propto (s_{n_2}^2)^{(n_2-1)/2-1} (\tau^2(\bar{y}_{n_2}, s_{n_2}^2))^{-\nu_{n_1+n_2+1}/2}\\
&\quad \times \int t_{\nu_{n_1+n_2-1}}(\theta|\kappa_{n_1+n_2}^{-1}(n_2\bar{y}_{n_2} + \kappa_{n_1}\mu_{n_1}), \tau^2(\bar{y}_{n_2}, s_{n_2}^2)/\nu_{n_1+n_2})d\theta\\
&\propto (s_{n_2}^2)^{(n_2-1)/2-1} (\tau^2(\bar{y}_{n_2}, s_{n_2}^2))^{-(\nu_{n_1+n_2-1}+1)/2}
\end{aligned}$$

where

$$\begin{aligned}
\beta(\bar{y}_{n_2}, s_{n_2}^2) &\propto (\theta - \kappa_{n_1+n_2}^{-1}(n_2\bar{y}_{n_2} + \kappa_{n_1}\tilde{\mu}_{n_1}))^2 + \tau^2(\bar{y}_{n_2}, s_{n_2}^2)\\
\tau^2(\bar{y}_{n_2}, s_{n_2}^2) &= (\kappa_{n_1}n_2\kappa_{n_1+n_2}^{-2}(\bar{y}_{n_2} - \tilde{\mu}_{n_1})^2 + \kappa_{n_1+n_2}^{-1}(\nu_{n_1}\tilde{\sigma}_{n_1}^2 + (n_2-1)s_{n_2}^2))
\end{aligned}$$

It follows that

$$p(\bar{y}_{n_2}|s_{n_2}^2, \bar{y}_{n_1}, s_{n_1}^2) = t_{\nu_{n_1+n_2-1}}(\bar{y}_{n_2}|\tilde{\mu}_{n_1}, \kappa_{n_1+n_2}(n_2\kappa_{n_1}\nu_{n_1+n_2-1})^{-1}((n_2-1)s_{n_2}^2 + \nu_{n_1}\tilde{\sigma}_{n_1}^2)) \tag{22}$$

$$p(s_{n_2}^2|\bar{y}_{n_1}, s_{n_1}^2) = C \times (s_{n_2}^2)^{(n_2-1)/2-1}(s_{n_2}^2 + \nu_{n_1}\tilde{\sigma}_{n_1}^2/(n_2-1))^{-\nu_{n_1+n_2-1}/2} \tag{23}$$

$$C = ((\nu_{n_1}\tilde{\sigma}_{n_1}^2)/(n_2-1))^{\frac{1}{2}(\nu_{n_1})}\Gamma(\nu_{n_1+n_2-1}/2)/(\Gamma((n_2-1)/2)\Gamma(\nu_{n_1}/2)) \tag{24}$$

To generate the paths, for either of *Algorithm* 2 and the parametric stopping rule approaches, we can draw in two steps: $(\rho, \theta) \sim p(\rho, \theta)$, and $\mathbf{y}_i \sim p(\mathbf{y}_i|\rho, \theta)$, and map these to the $(\bar{y}_{n_i}, s_{n_i}^2)$'s.

The expected utility $u(\bar{y}_{n_1}, s^2_{n_1}, a)$ for each of $a \in \{T, P\}$ involves

$$\mathrm{E}[(\theta - c)^+ | s^2_n, \bar{y}_n] = \int_0^\infty \theta \times t_{\nu_n}(\theta | \tilde{\mu}_n - c, \tilde{\sigma}^2_n / \kappa_n) d\theta \tag{25}$$

for some c. This involves $t_{\nu_n}(.)$ and $F_{t_{\nu_n}}(.)$ which can be efficiently numerically computed. Unlike (6–7), we don't know how to express the boundaries of T_i and P_i in terms of $(\bar{y}_n, s^2_n)$. It is therefore necessary to compute $u(\bar{y}_n, s^2_n, a)$ for each of $a \in \{T, P\}$ and pick the highest value.

We postulate $\gamma = \{\alpha^-_1, \beta^-_1, \alpha^+_1, \beta^+_1, \ldots \alpha^-_{i_*-1}, \beta^-_{i_*-1}, \alpha^+_{i_*-1}, \beta^+_{i_*-1}\}$, and

$$\Gamma = \{\gamma : \alpha^-_i \leq \alpha^+_i, \beta^-_i \leq 0 \leq \beta^+_i, i = 1, \ldots, i_* - 1\} \tag{26}$$

$$C_i = \{\mathbf{y}_i : \alpha^-_i + \beta^-_i s^2_i \leq \bar{y}_i \leq \alpha^+_i + \beta^+_i s^2_i\} \tag{27}$$

Bibliography

[1] Bradley P. Carlin, Joseph B. Kadane, and Alan E. Gelfand. "Approaches for Optimal Sequential Decision Analysis in Clinical Trials". In: *Biometrics* 54 (1998), pp. 964–975.

[2] Andrew Gelman et al. *Bayesian Data Analysis*. 2nd ed. Chapman & Hall/CRC, 2004. ISBN: 0-412-03991-5.