Predictively Consistent Prior Effective Sample Sizes

Beat Neuenschwander

Novartis Oncology, Basel, Switzerland

Joint work with Sebastian Weber, Heinz Schmidli (Novartis), and Anthony O'Hagan (Sheffield University)

1st Italian Bayesian Day for Clinical Research

Italian Society of Pharmaceutical Medicine and Italian Biostatistics Group

May 10, 2019, Torino, Italy

NOVARTIS

Outline

Problem statement and motivation

Methods

- Conjugate analyses
- General case
- ESS comparisons: examples
- Predictive consistency
- Computations

Applications

1. ESS for historical data prior
2. Hierarchical subgroup analyses

Summary and Outlook

Problem statement

Assume we have

- a statistical model $p(Y|\theta)$
- a prior distribution $p(\theta)$
- a one-dimensional parameter θ

We want to quantify the prior information as

- an equivalent number of observations (events for the time-to-event setting)
- the prior *effective sample size (ESS)*, an intuitive and useful metric

Motivation: two applications

1. Assume we have historical control data from previous trials.

   Knowing the *ESS* from the respective historical data prior for the control parameter helps designing an upcoming randomized trial with a smaller control group.

2. Hierarchical (borrowing) analyses for small subgroups.

   The posterior *ESS* is a useful metric to quantify the amount of information for each subgroup.

NOVARTIS

Outline

NOVARTIS

ESS for conjugate priors

The *ESS* is well-understood for conjugate Bayesian analyses, for example

- data $N(\theta, \sigma^2)$, known σ, prior $\theta \sim N(m, \sigma^2/n_0)$, $ESS = n_0$
- data $\text{Bin}(n, \theta)$, prior $\text{Beta}(a, b)$, $ESS = a + b$
- data $\text{Pois}(\theta)$, prior $\text{Gamma}(a, b)$, $ESS = b$
- data $\text{Exp}(\theta)$ ($\theta = mean$), prior $\text{Gamma}(a, b)$, $ESS = a$

There are three justifications for the respective *ESS*.

Justification 1: prior-to-posterior updating rule

The *ESS* follows directly from the updated posterior parameters.

Binomial-Beta example:

- posterior: $\theta | r \sim Beta(a + r, b + n - r)$
- updating rule: $a \to a + r$, $b \to b + n - r$
  (a,b = “prior number of successes and failures”)
- $\Rightarrow ESS = a + b$

Justification 2: posterior mean (a weighted average)

Posterior mean = weighted average of prior mean and sample mean (standard parameter estimate from data)

Binomial-Beta example:

- posterior mean: $\frac{a+r}{a+b+n} = \frac{(a+b)\frac{a}{a+b}+n\frac{r}{n}}{a+b+n}$
- the non-normalized weights of prior and data are $a+b$ and n, respectively
- $\Rightarrow ESS = a+b$

NOVARTIS

Justification 3: variance ratio

Ratio of

- the expected variance from one observation.
- and the prior variance of the mean parameter θ

$$ESS = \frac{E_\theta\{\mathrm{Var}(Y_1|\theta)\}}{\mathrm{Var}(\theta)}$$

Binomial-Beta example:

- one-unit variance:
  $E_\theta\{\mathrm{Var}(Y_1|\theta)\} = E_\theta\{\theta(1-\theta)\} = \frac{ab}{(a+b)(a+b+1)}$
- prior variance: $\mathrm{Var}(\theta) = \frac{ab}{(a+b)^2(a+b+1)}$
- $\Rightarrow ESS = a + b$

NOVARTIS

Outline

ESS: general case

We now move to the general, non-conjugate case.

- A conjugate prior does exist but another prior is used.

  For example, priors derived from historical data via a hierarchical model are not conjugate.
- A conjugate prior does not exist.

ESS conjugacy requirement:

if the prior is conjugate, the general *ESS* method (formula) should give the known *ESS*.

Methods

In the following, three methods will be discussed. They build on the variance/precision-ratio justification 3 from the conjugate case.

- variance/precision-ratio methods
- Morita-Thall-Müller method [4]
- *expected local-information-ratio (ELIR)* method (new, [7])

One-unit Fisher information and prior information

Data: the information for one observation unit is the Fisher information

$$i_F(\theta) = E_{Y_1|\theta}\{i_F(Y_1,\theta)\} = -E_{Y_1|\theta}\{\frac{d^2 \log p(Y_1|\theta)}{d\theta^2}\}$$

Analogously, the prior information is defined as

$$i(p(\theta)) = -\frac{d^2 \log p(\theta)}{d\theta^2}$$

Note: these are functions of the parameter, not single numbers: so we can't simply take the ratio of the two to obtain the *ESS*.

Variance/precision-ratio *ESS* (ESS_{VR}, ESS_{VR})

Following justification 3, the variance-ratio and precision-ratio *ESS* are defined as:

$$ESS_{VR} = \frac{E_\theta\{i_F^{-1}(\theta)\}}{\mathrm{Var}(\theta)}, \quad ESS_{PR} = \frac{\mathrm{Var}^{-1}(\theta)}{E_\theta\{i_F(\theta)\}} \qquad (1)$$

Morita-Thall-Müller ESS (ESS_{MTM})

- Morita, Thall and Müller [4] proposed another (more complicated) method, which turns out to be a version of the precision-ratio method.
- Idea: find sample size m such that

information of prior $p(\theta)$

=

expected (under $p(\theta)$) posterior information

(under a vague prior $p_0(\theta)$)

- Vague prior $p_0(\theta)$: $E(p_0(\theta)) = E(p(\theta)) = \tilde{\theta}$

NOVARTIS

Morita-Thall-Muüller ESS (ESS_{MTM})

Formally: ESS_{MTM} is the integer m that minimizes

$$|i(p_0(\tilde{\theta})) + E_{Y_m}\{i_F(Y_m;\tilde{\theta})\} - i(p(\tilde{\theta})| \qquad (2)$$

...the distance (evaluated at the prior mean $\tilde{\theta}$) between

- the first two terms: the expected posterior information for a sample of size m based on the vague prior $p_0(\theta)$
- third term: the information of the actual prior

ESS_{MTM}...comments

1. Evaluation at the mean! MTM point out that this is necessary to fulfill the *ESS conjugacy requirement*, i.e., correct *ESS* under conjugacy.
2. The vague prior p_0 is not unique: a minor issue because the respective prior information $i(p_0(\tilde{\theta}))$ is small anyway (usually 1 or 0).
3. Restriction to integers not really needed
4. MTM give an algorithm to obtain the *ESS*

NOVARTIS

ESS_{MTM}...simplified

- Note that $E_{Y_m}\{i_F(Y_m;\tilde{\theta})\} = m \cdot E_{Y_1}\{i_F(Y_1;\tilde{\theta})\}$. This avoids the original minimization problem and leads to

$$ESS_{MTM} = \frac{i(p(\tilde{\theta})) - i(p_0(\tilde{\theta}))}{E_{Y_1}\{i_F(Y_1;\tilde{\theta})\}} \quad (3)$$

- $i(p_0(\tilde{\theta})) = 0$ and replacing $E_{Y_1}\{i_F(Y_1;\tilde{\theta}\}$ by $i_F(\tilde{\theta})$ leads to a further simplication (Gene Penello, CDRH-FDA)

$$ESS_{MTM.P} = \frac{i(p(\tilde{\theta}))}{i_F(\tilde{\theta})} \quad (4)$$

NOVARTIS

Expected local-information-ratio (ESS_{ELIR})

We propose the following *ESS*:

- the *expected local-information-ratio ESS*

$$ESS_{ELIR} = E_{\theta}\{\frac{i(p(\theta))}{i_F(\theta)}\} \quad (5)$$

- Note the similarity to

$$ESS_{MTM.P} = \frac{i(p(\tilde{\theta}))}{i_F(\tilde{\theta})} \qquad \tilde{\theta} = \text{prior mean} \quad (6)$$

  a plug-in version of ESS_{ELIR}

- Result: ESS_{ELIR} fulfills the *ESS conjugacy requirement*

NOVARTIS

Example: ESS_{ELIR} for binomial data with Beta prior

- Binomial-Beta example: $r|\theta \sim \text{Bin}(n,\theta), \theta \sim \text{Beta}(a,b)$
- Fisher and prior information

$$i_F(\theta) = \frac{1}{\theta(1-\theta)}, \quad i(p(\theta)) = \frac{a-1}{\theta^2} + \frac{b-1}{(1-\theta)^2}$$

- $ESS_{ELIR} =$

$$E_\theta\{(a-1)\frac{1-\theta}{\theta}+(b-1)\frac{\theta}{1-\theta}\} = (a-1)\frac{b}{a-1}+(b-1)\frac{a}{b-1}$$

$$= a+b \qquad (a,b>1)$$

NOVARTIS

ESS_{ELIR} for natural parameter in exponential family

- One-parameter exponential family with natural parameter η, and conjugate prior:

$$p(y|\eta) = \exp\{y\eta - M(\eta)\}, \quad p(\eta) = \exp\{n_0 m_0 \eta - n_0 M(\eta)\}$$

- Fisher and prior information, ESS_{ELIR}

$$i_F(\theta) = d^2M(\theta)/d\theta^2, i(p(\theta)) = n_0 \cdot d^2M(\theta)/d\theta^2, ESS_{ELIR} = n_0$$

- Binomial-Beta example:

$$\eta = \log\{\theta/(1-\theta)\}, M(\eta) = \log\{1 + \exp(\eta)\},$$

$$n_0 = a + b \qquad (a, b > 0)$$

NOVARTIS

Outline

Comparison: two examples

The methods

- look very similar
- fulfill the *ESS conjugacy requirement* for the standard one-parameter exponential families
- will now be compared for two examples:
  1. normal data and Student-t prior
  2. exponential data and generalized Gamma prior

Example 1: normal data, Student-t prior

Normal data with mean θ, known variance σ^2, t(df) prior with scale s.
The heavy-tailed prior is robust against prior-data conflict (O'Hagan [8], O'Hagan and Pericchi [9])

$$i_F(\theta) = 1/\sigma^2, \quad i(p(\theta)) = \frac{1}{s^2}\frac{df+1}{df}\frac{1-\theta^2/df}{(1+\theta^2/df)^2}$$

$$ESS_{VR} = ESS_{PR} = (\sigma/s)^2\frac{df-2}{df} \qquad (df > 2)$$

$$ESS_{MTM} = (\sigma/s)^2\frac{df+1}{df} \qquad (df > 1)$$

$$ESS_{ELIR} = (\sigma/s)^2\frac{df+1}{df+3}$$

ESS for example 1

ESS for normal data with Student-t prior: $(\sigma/s)^2 = 100$

df	VR	PR	MTM	MTM.P	ELIR
2	—	—	150	150	60
3	33	33	133	133	67
4	50	50	125	125	71
5	60	60	120	120	75
10	80	80	110	110	85
50	96	96	102	102	96

Example 2: exponential data, gen-Gamma prior

Exponential data with hazard θ, generalized Gamma prior with shape, scale, and family parameter a, s, and f:

$$p(\theta) = \frac{f\theta^{a-1}\exp\{-(\theta/s)^f)\}}{s^a\Gamma(a/f)}$$

More flexible than Gamma prior, may be useful to fit a prior with expert-elicited median and interquartile range.

Special cases: Gamma ($f = 1$) and Weibull ($f = a$)

$$i_F(\theta) = 1/\theta^2, \quad i(p(\theta)) = (a-1)/\theta^2 + f(f-1)\theta^{f-2}/s^f$$

$$ESS_{ELIR} = af - 1$$

The other *ESS* can also be obtained analytically.

NOVARTIS

ESS for example 2

ESS for exponential data with gen-Gamma prior $(a, s = 1, f)$

distribution	a	f	VR	PR	MTM	MTM.P	ELIR
Gamma	9.00	1.00	10.0	6.2	9.0	8.0	8.0
Weibull	3.00	3.00	8.6	3.5	7.3	6.3	8.0
gen-Gamma	2.54	3.54	7.9	2.3	6.4	5.4	8.0
Gamma	25.00	1.00	26	22	25	24	24
Weibull	5.00	5.00	20	15	18	17	24
gen-Gamma	4.52	5.52	19	14	16	15	24
Gamma	121.00	1.00	122	118	121	120	120
Weibull	11.00	11.00	84	79	77	76	120
gen-Gamma	10.51	11.51	81	76	74	73	120

Outline

NOVARTIS

The predictive consistency criterion

So far, we have seen various *ESS* methods, which

- fulfill the *ESS conjugacy requirement*
- can differ considerably for non-conjugate priors

This is a major problem. Which method should be used?
More than the *ESS conjugacy requirement* is needed.

Predictive consistency:
for a sample of size M, the expected posterior *ESS* must be the sum of the prior *ESS* and M.

Do the methods fulfill this basic requirement?

Predictive consistency? Normal-t example

Prior ESS and expected posterior $ESS - M$

	method	prior ESS	(expected posterior ESS)–M M=10	M=100	M=1000
df=2	VR	—	36	54	60
	MTM	150	95	74	63
	ELIR	**60**	**60**	**60**	**60**
df=5	VR	60	63	72	76
	MTM	120	107	87	77
	ELIR	**75**	**75**	**75**	**75**
df=50	VR	96	96	96	97
	MTM	102	101	99	97
	ELIR	**96**	**96**	**96**	**96**

Only ESS_{ELIR} is predictively consistent.

NOVARTIS

Predictive consistency? Exponential-Weibull example

		prior ESS	(expected posterior ESS)–M		
a = 3	VR	8.6	9.6	10	10
	PR	3.5	5.6	6.0	6.2
	MTM	7.3	8.2	8.8	9.0
	ELIR	**8**	**8**	**8**	**8**
a = 7	VR	36	41	49	50
	PR	32	37	45	46
	MTM	33	37	45	49
	ELIR	**48**	**48**	**49**	**48**
a = 11	VR	84	91	111	121
	PR	79	86	107	117
	MTM	77	82	100	116
	ELIR	**120**	**120**	**121**	**121**

NOVARTIS

ESS_{ELIR}: proof of predictive consistency

Y_M : predictive data of size M, with posterior ESS

$$ESS(p(\theta|Y_M)) = E_{\theta|Y_M}\{\frac{i(p(\theta)) - d^2 \log p(Y_M|\theta)/d\theta^2}{i_F(\theta)}\}$$

Expected posterior ESS under prior predictive distribution

$$E_{Y_M}\left[E_{\theta|Y_M}\{\frac{i(p(\theta)) - d^2 \log p(Y_M|\theta)/d\theta^2}{i_F(\theta)}\}\right]$$

$$= E_{\theta}\left[E_{Y_M|\theta}\{\frac{i(p(\theta)) - d^2 \log p(Y_M|\theta)/d\theta^2}{i_F(\theta)}\}\right]$$

$$= E_{\theta}\{\frac{i(p(\theta)) + Mi_F(\theta)}{i_F(\theta)}\} = ESS(p(\theta)) + M$$

NOVARTIS

Outline

Computational aspects

- If the integral in ESS_{ELIR} is not available analytically:
  - obtain ESS_{ELIR} by simulation
  - for large S, simulate θ_s ($s = 1, \ldots, S$) from the prior and then take the mean of the $i(p(\theta_s))/i_F(\theta_s)$ ratios
- If the prior is only available as an MCMC sample
  - approximate the prior by a mixture of standard distributions
  - Diaconis and Ylvisaker [3] showed that this can be done to any degree of accuracy
  - Software: e.g., *RBesT* R-package [15], *SAS PROC FMM* [11]

NOVARTIS

Outline

NOVARTIS

1. Prior ESS for historical data prior

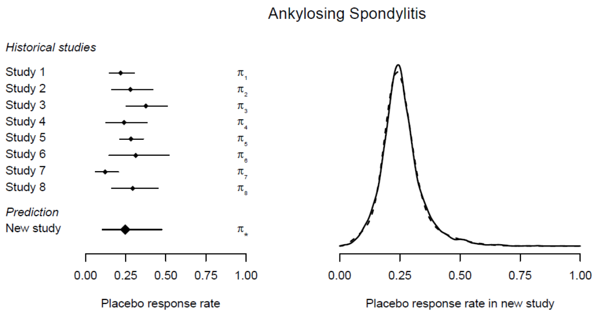
A small proof-of-concept (PoC) trial (Baeten et al. [1])

- Disease: *ankylosing spondylitis*, a chronic inflammatory disease
- Binary endpoint: response at week 6
- Randomized trial: *secukinumab (T)* vs. placebo (C)
- Standard design: would require $n = 24$ per arm
- Historical placebo data from 8 trials $\rightarrow$ prior $p(\pi_C)$
- What is the prior's ESS?
- A historical data design was used with $n_T = 24, n_C = 6$

NOVARTIS

1. Historical data and MAP prior

Median and 95%-intervals for event and MAP event rate for new trial (left panel), and MAP prior density (solid line) with two-component Beta mixture approximation (dashed line) (right panel).



NOVARTIS

1. Historical data prior: MAP prior

- Data model: $r_j|\pi_j \sim \text{Bin}(n_j, \pi_j)$
- Parameter model: on log-odds scale, $\theta_j = \log\{\pi_j/(1-\pi_j)\}$

$$\theta_1, \ldots, \theta_8, \theta\star|\mu, \tau \sim N(\mu, \tau^2)$$

- Prior distributions: $\mu \sim N(0, 10^2)$, $\tau \sim$ half-N(scale=1)
- meta-analytic-predictive prior for new study (Spiegelhalter et al. [13], N et al. [5], Schmidli et al. [12], Viele et al. [16]):

$$p(\theta_\star|r = (r_1, \ldots, r_8))$$

- Appproximations to the MCMC MAP prior:
  single moment-matching Beta (poor approximation),
  2- and 3-component Beta mixtures

1. ESS for MAP prior

Prior ESS for historical data prior

MAP prior approximation	ESS_{ELIR}
$\text{Beta}(6.3, 18.3)$	25
$0.67 \cdot \text{Beta}(16.30, 49.74) + 0.33 \cdot \text{Beta}(3.1, 8.1)$	35
$0.53 \cdot \text{Beta}(6.1, 18.1) + 0.37 \cdot \text{Beta}(30.0, 91.7) + 0.10 \cdot \text{Beta}(2.1, 4.7)$	36

ESS for the other (predictively inconsistent) ESS

- variance-ratio: 26 (same for the three priors)
- MTM: 25, 55, 79 (!)

Outline

2. Posterior ESS for hierarchical subgroup analyses

A phase II (Chugh et al. [2], Thall et al. [14])

- Single arm design: to assess the effect of *imatinib* in 10 histological subtypes of sarcoma
- Binary endpoint: clinical benefit response (CBR)
- 179 patients: subtype sampe sizes between 2 and 29
- Similar response rates expected
- Design (Thall et al. [14]) based on a hierarchical model (HM)
- How much information can be gained by the HM analysis?
- What is the posterior *ESS* for each subgroup?

2. Hierarchical models

- HM-100: the original model, same (full exchangeability) model as for application 1:

$$\theta_1, \ldots, \theta_{10} | \mu, \tau \sim N(\mu, \tau^2)$$

- Robust versions: mixtures with weights w_j for the above model and $1 - w_j$ for independent priors for each θ_j.

  Three robust models:

  HM-90, HM-75, HM-50 ($w_j = 0.9, 0.75, 0.5$)

2. Posterior ESS for HM analyses

Substantial information gains even for robust HM analyses

			HM-100	HM-90	HM-75	HM-50
Subtype	r/n	(%)		ESS		
Angiosarcoma	2/15	(13)	65	60	51	35
Ewing	0/13	(0)	56	46	36	24
Fibrosarcoma	1/12	(8)	61	55	45	30
Leiomyosarcoma	6/28	(21)	78	71	62	47
Liposarcoma	7/29	(24)	75	66	57	44
MFH	3/29	(10)	74	68	61	46
Osteosarcoma	5/26	(19)	77	72	62	48
MPNST	1/5	(20)	55	49	39	23
Rhabdomysarcoma	0/2	(0)	52	44	33	18
Synovial	2/20	(15)	71	66	58	42

Summary and Outlook

- *ESS* is an intuitive metric and is particularly useful for sample size determinations in clincial trials with prior information.
- There are various, similar precision-ratio methods available, which fulfill the minimal *ESS conjugacy requirement.*
- Somewhat surprisingly, the *ESS* can differ considerably.
- Only the newly proposed *expected local-information-ratio* ESS_{ELIR} fulfills the *predictive consistency* requirement.

Summary and Outlook

More research needed:

- What is the *ESS* if ESS_{ELIR} does not exist (undefined integral)?
- Are there other definitions of *ESS* that fulfill both requirements?
- For the multivariate case, $\theta = (\theta_1, \ldots, \theta_m)$, what is the *ESS* for each parameter and the vector θ?

References I

- Baeten D et al. Anti-interleukin-17A monoclonal antibody secukinumab in ankylosing spondylitis: a randomized, double-blind, placebo-controlled trial. *The Lancet* 2013; 382: 1705–1713.
- Chugh R et al. Phase II multicenter trial of imatinib in 10 histologic subtypes of sarcoma using a Bayesian hierarchical statistical model. *Journal of Clinical Oncology* 2009; 27(19): 3148–3153.

NOVARTIS

References II

- P. Diaconis and D. Ylvisaker.
  Quantifying prior opinion.
  Bayesian Statistics (Proceedings of the Second Valencia International Meeting) 1984; 2:133–148.
- Morita S, Thall PF and Müller P. Determining the effective sample size of a parametric prior. *Biometrics* 2008; 64: 595–602.
- Neuenschwander B, Capkun-Niggli G, Branson M, and Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. *Clinical Trials* 2010; 7: 5–18.

References III

- Neuenschwander B, Wandel S, Roychoudhury S, and Bailey S. Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical Statistics* 2015; 15: 123-134.
- Neuenschwander B, Weber S, Schmidli H, O'Hagan, A. Predictively consistent prior effective sample sizes. (submitted).
- O'Hagan, A. On outlier rejection phenomena in Bayes inference. *Journal of the Royal Statistical Society, Series B* 1979; **41,** 358–367.

References IV

- O'Hagan A and Pericchi L. Bayesian heavy-tailed models and conflict resolution: a review. *Brazilian Journal of Probability and Statistics* 2012; 26: 372–401.
- Pennello G and Thompson L. Experience with reviewing Bayesian medical device trials. *Journal of Biopharmaceutical Statistics* 2008; 18(1): 81–115.
- SAS Institute.
  SAS user guide: Statistics. The FMM procedure.
  Cary, NC: SAS Institute Inc., 2014.

References V

- Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan O, Spiegelhalter DJ and Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* 2014; 70: 1023–32.
- Spiegelhalter DJ, Abrams KR and Myles JP. *Bayesian Approaches to Clinical trials and Health-Care Evaluation*. 2004; Chichester: Wiley.
- Thall PF, Wathen J, Bekele B, Champlin R, Baker L, and Benjamin R. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Journal of Clinical Oncology* 2003; 22: 763–780.

References VI

- S. Weber.
  RBesT: R Bayesian Evidence Synthesis Tools, 2017.
  R package version 1.2-3.
- Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, Hobbs B, Ibrahim JG, Kinnersley N, Lindborg S, Micallef S, Roychoudhury S and Thompson L. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat* 2014; 13(1): 41–54.