Bayesian sample size determination for Type A uncertainty evaluation

Jörg Martin
joint work with Clemens Elster

Physikalisch-Technische Bundesanstalt
EMUE Workshop - Paris

January 22nd 2020

What is Sample size determination (SSD)?

Sample size determination (SSD)

Find required number of datapoints n

What is Sample size determination (SSD)?

Sample size determination (SSD)

Find required number of datapoints n to meet a certain *criterion*

What is Sample size determination (SSD)?

Experimental design

Sample size determination

What is Sample size determination (SSD)?

Bayesian experimental design

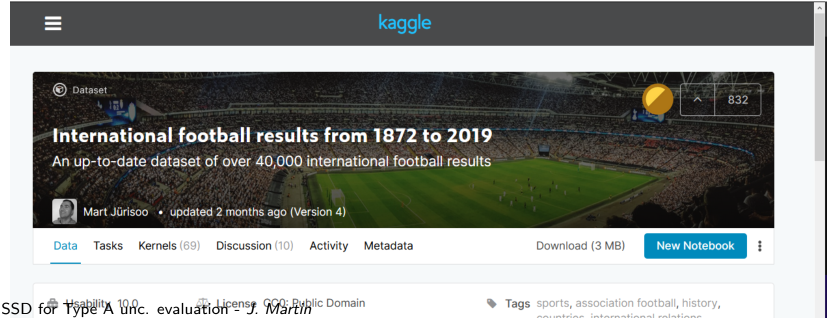
Bayesian sample size determination

Toy example: Football

Q: *Average number of goals in a football match?* $\Rightarrow \theta$

SSD: Minimal number of games to answer "Q" $\Rightarrow n$

Dataset: International football matches in the years 2014-2019

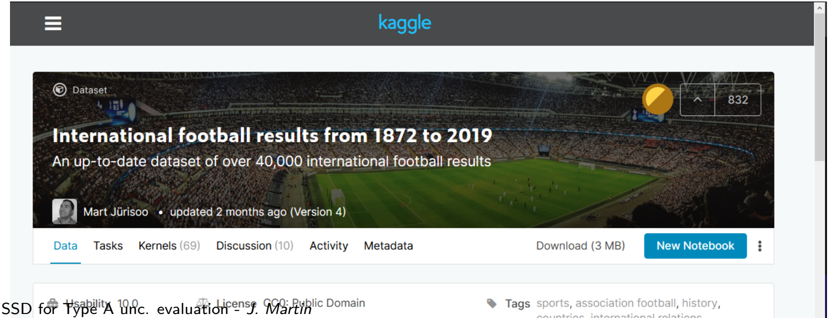


Toy example: Football

Q: *Average number of goals in a football match?* $\Rightarrow \theta$

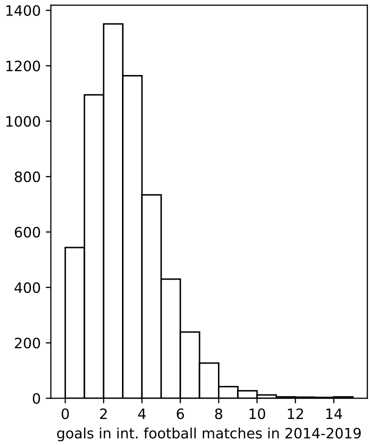
SSD: Minimal number of games to answer "Q" $\Rightarrow n$

Dataset: International football matches in the years 2014-2019



Distribution of goals between 2014-2019

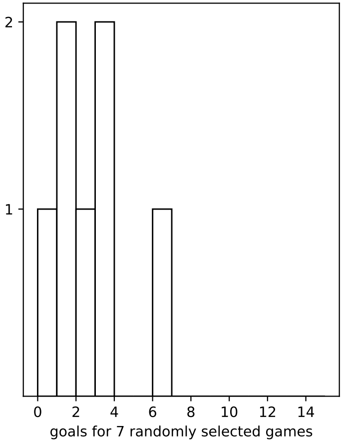
Q: Average number of goals in a football match?



Full dataset: $n = 5784$ games

Subsample of 7 games

Q: Average number of goals in a football match?



$n = 7$ games - **Is this enough to answer Q?**

What is Sample size determination (SSD)?

Sample size determination (SSD)

Find required number of datapoints n to meet a certain *criterion*

What is Sample size determination (SSD)?

Sample size determination (SSD)

Find required number of datapoints n to meet a certain *criterion*

Criterion **here**: Uncertainty below threshold

Terminology

- θ: Average number of goals in a match
- x: Number of goals in a game, e.g.

$$\text{Final score } 3:1 \qquad \Rightarrow \qquad x = 4$$

  We assume $x \sim \mathrm{Poi}(\theta)$
- For n games:

$$\mathbf{x} = (x_1, \ldots, x_n)\,,$$

  where x_i number of goals in game i

Terminology

- θ: Average number of goals in a match
- x: Number of goals in a game, e.g.

$$\text{Final score } 3:1 \qquad \Rightarrow \qquad x = 4$$

  We assume $x \sim \mathrm{Poi}(\theta)$
- For n games:

$$\mathbf{x} = (x_1, \ldots, x_n)\,,$$

  where x_i number of goals in game i

We want to find an optimal n **before** we know $\mathbf{x}$

Data analysis

Results of n games

$$\Downarrow$$

$$\mathbf{x} = (x_1, \ldots, x_n)$$

$$\Downarrow$$

Estimate for θ + Uncertainty

Example

Results of $n = 3$ games
(2:1, 0:0, 0:3)

$$\Downarrow$$

$$\mathbf{x} = (3,\ 0,\ 3)$$

$$\Downarrow$$

Estimate for θ + Uncertainty
$\theta = 2.3,\ u(\theta|\mathbf{x}) = 0.9$

Example

Results of $n = 3$ games
(2:1, 0:0, 0:3)

$$\Downarrow$$

$$\mathbf{x} = (3,\ 0,\ 3)$$

$$\Downarrow$$

Estimate for θ + Uncertainty
$\theta = 2.3,\ u(\theta|\mathbf{x}) = 0.9$

Data analysis

Example

Results of $n = 3$ games
(2:1, 0:0, 0:3)

$$\Downarrow$$

$$\mathbf{x} = (3,\ 0,\ 3)$$

$$\Downarrow$$

Estimate for θ + Uncertainty
$\theta = 2.3,\ u(\theta|\mathbf{x}) = 0.9$

SSD: Choose n s.t. $u(\theta|\mathbf{x})$ is small enough (before knowing $\mathbf{x}$)

Bayesian framework

- *Posterior distribution*

$$\underset{\text{posterior}}{\pi(\theta|\mathbf{x})} \quad \propto \quad \underset{\text{prior}}{\pi(\theta)} \quad \cdot \quad \underset{\text{likelihood}}{p(\mathbf{x}|\theta)}$$

(*recall*: $\mathbf{x} = (x_1, \ldots, x_n)$)

- **Uncertainty**: *Posterior variance*

$$u^2(\theta|\mathbf{x}) := \mathrm{Var}(\theta|\mathbf{x}) = \int (\theta - \mathbb{E}[\theta|\mathbf{x}])^2 \cdot \pi(\theta|\mathbf{x}) \, \mathrm{d}\theta$$

What is Sample size determination (SSD)?

Sample size determination (SSD)

Find required number of datapoints n to meet a certain *criterion*

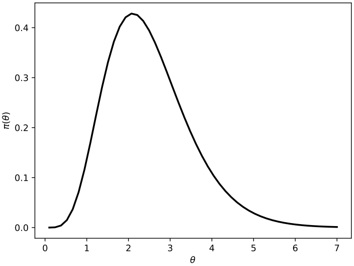
Criterion **here**: Uncertainty below threshold

Recipe for SSD

- Describe prior knowledge via *prior* $\pi(\theta)$
- Compute uncertainty $u^2(\theta|\mathbf{x})$ for **generic x** (*we don't know* **x**)

Prior knowledge

Prior knowledge about θ: "**Around** 2.50 **with unc.** 1.0"



$$\pi(\theta) = \mathrm{Gamma}(\theta|\alpha = 6.25, \beta = 2.50)$$

$$u^2(\theta|\mathbf{x}) = \frac{\alpha + \sum_{i=1}^{n} x_i}{(n+\beta)^2}$$

Recipe for SSD

Recipe for SSD

- Describe prior knowledge via *prior* $\pi(\theta)$
- Compute uncertainty $u^2(\theta|\mathbf{x})$ for **generic x** (*we don't know* $\mathbf{x}$)

Recipe for SSD

Recipe for SSD

- Describe prior knowledge via *prior* $\pi(\theta)$
- Compute uncertainty $u^2(\theta|\mathbf{x})$ for **generic x** (*we don't know* **x**)

Next task:

Choose sample size n so that uncertainty will be small enough

Uncertainty below threshold

Want something like

$$\mathrm{Var}(\theta|\mathbf{x}) = u^2(\theta|\mathbf{x}) \overset{!}{<} \varepsilon^2$$

but *before* measuring $\mathbf{x}$.

Uncertainty below threshold

Want something like

$$\mathrm{Var}(\theta|\mathbf{x}) = u^2(\theta|\mathbf{x}) \overset{!}{<} \varepsilon^2$$

but *before* measuring $\mathbf{x}$.

Idea: Average w.r.t. *prior predictive*

$$\pi_{\mathrm{pred.}}(\mathbf{x}) = \int p(\mathbf{x}|\theta)\pi(\theta)\mathrm{d}\theta$$

Average posterior variance criterion

Average posterior variance criterion (**APVC**, e.g. [Wang2002])

$$\overline{u^2} = \mathbb{E}_{\mathbf{x} \sim \pi_{\mathrm{pred.}}(\mathbf{x})} \left[u^2(\theta|\mathbf{x}) \right] \overset{!}{<} \varepsilon^2$$

- $\overline{u^2}$: Expected (squared) uncertainty
- ε^2: (subjective) threshold

Recipe for SSD

Recipe for SSD (APVC)

- Describe prior knowledge via *prior* $\pi(\theta)$
- Compute uncertainty $u^2(\theta|\mathbf{x})$ for **generic x** (*we don't know* **x**)

Recipe for SSD (APVC)

- Describe prior knowledge via *prior* $\pi(\theta)$
- Compute uncertainty $u^2(\theta|\mathbf{x})$ for **generic x** (*we don't know* **x**)
- Choose n such that $\overline{u^2} < \varepsilon^2$

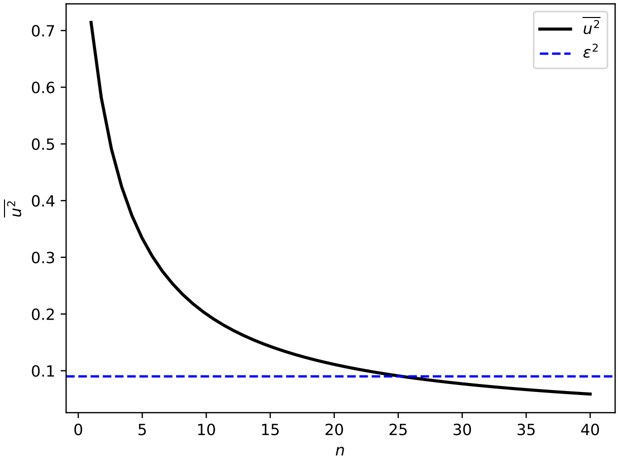
recall: $\overline{u^2} = \mathbb{E}_{\mathbf{x}\sim\pi_{\mathrm{pred.}(\mathbf{x})}}\left[u^2(\theta|\mathbf{x})\right]$

Toy example

Say we want "$u^2(\theta|\mathbf{x}) < 0.3^2$":

$$\overline{u^2} < 0.30^2$$

Toy example



$\overline{u^2}$
ε^2
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
25
30
35
40
n

Toy example

Say we want "$u^2(\theta|\mathbf{x}) < 0.3^2$":

$$\overline{u^2} < 0.30^2 \qquad \rightarrow \qquad n = 26$$

Toy example

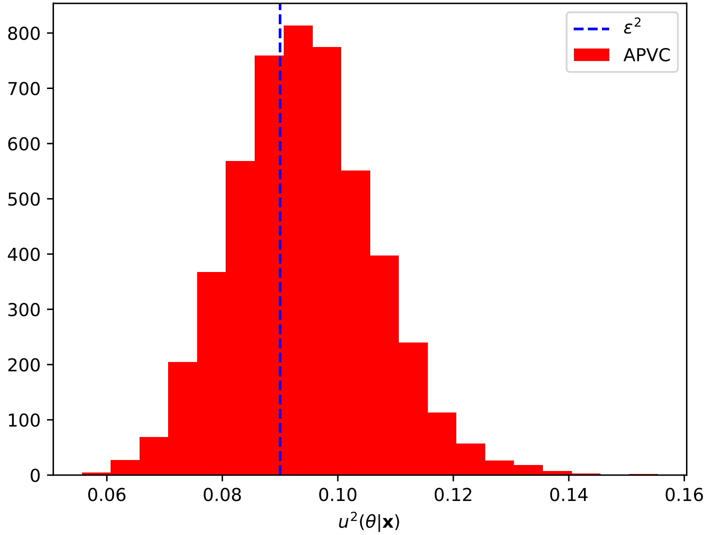
Say we want "$u^2(\theta|\mathbf{x}) < 0.3^2$":

$$\overline{u^2} < 0.30^2 \qquad \rightarrow \qquad n = 26$$

Result for 4 random choices of 26 games:

1. $\hat{\theta} = 2.36, \qquad u(\theta|\mathbf{x}) = \mathbf{0.29}$
2. $\hat{\theta} = 2.50, \qquad u(\theta|\mathbf{x}) = \mathbf{0.30}$
3. $\hat{\theta} = 2.78, \qquad u(\theta|\mathbf{x}) = \mathbf{0.31}$
4. $\hat{\theta} = 3.13, \qquad u(\theta|\mathbf{x}) = \mathbf{0.33}$

Fluctuations of u^2



Problems of the APVC

Average posterior variance criterion (APVC):

- puts $u^2(\theta|\mathbf{x})$ in right order of magnitude
- $u^2(\theta|\mathbf{x}) < \varepsilon^2$ only holds on average
- $u^2(\theta|\mathbf{x}) < \varepsilon^2$ is often violated

We propose: *Variation* of posterior variance criterion (VPVC)

VPVC

Average of posterior variance criterion (APVC)

$$\overline{u^2} < \varepsilon^2$$

Variation of posterior variance criterion (VPVC)

$$\overline{u^2} + 2 \cdot \Delta u^2 < \varepsilon^2$$

VPVC

VPVC:

$$\overline{u^2} + 2 \cdot \Delta u^2 < \varepsilon^2$$

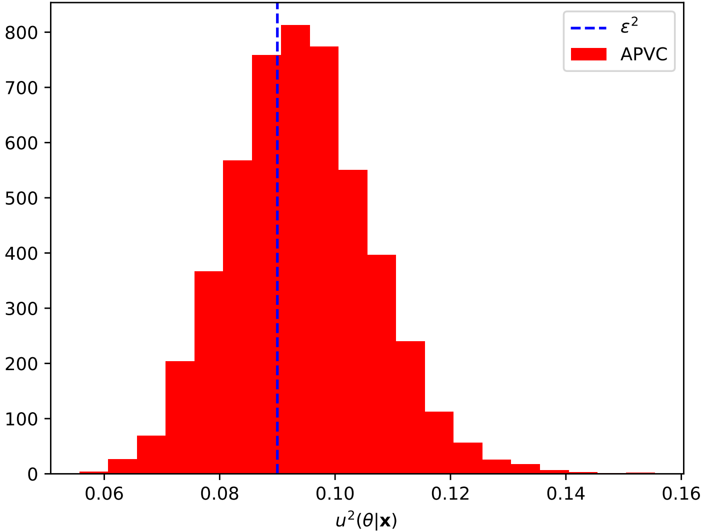
where

$$\Delta u^2 = \mathbb{E}_{\theta \sim \pi(\theta)} \left[\mathrm{Var}_{\mathbf{x} \sim p(\mathbf{x}|\theta)} \left(u^2(\theta|\mathbf{x}) \right) \right]$$

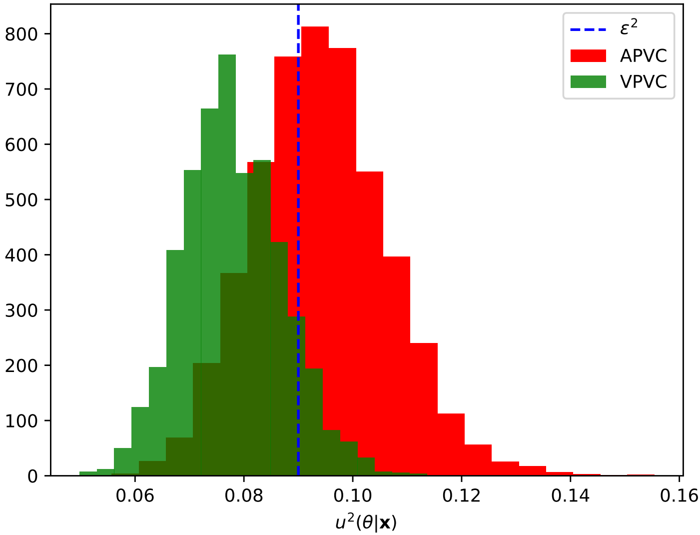
Recipe for SSD - VPVC version

Recipe for SSD (VPVC)

- Describe prior knowledge via *prior* $\pi(\theta)$
- Compute uncertainty $u^2(\theta|\mathbf{x})$ for **generic x** (*we don't know* **x**)
- Choose n such that $\overline{u^2} + 2 \cdot \Delta u^2 < \varepsilon^2$



800
700
600
500
400
300
200
100
0
0.06
0.08
0.10
0.12
0.14
0.16
$u^2(\theta|\mathbf{x})$
ε^2
APVC


800
700
600
500
400
300
200
100
0
0.06
0.08
0.10
0.12
0.14
0.16
$u^2(\theta|\mathbf{x})$
ε^2
APVC
VPVC

Sample sizes

ε	APVC	**VPVC**	freq.
0.80	2	3	6
0.70	3	5	8
0.60	5	7	10
0.50	8	11	14
0.40	14	18	21
0.30	26	32	34
0.20	60	70	72
0.10	248	267	270

Sample sizes

ε	APVC	**VPVC**	freq.
0.80	2	3	6
0.70	3	5	8
0.60	5	7	10
0.50	8	11	14
0.40	14	18	21
0.30	26	32	34
0.20	60	70	72
0.10	248	267	270

$$\text{APVC} \quad < \quad \text{VPVC} \quad < \quad \text{freq.}$$

$\uparrow$ violates ε (APVC)

$\uparrow$ no prior knowledge (freq.)

Conclusions

- SSD = planning of sample size such that a certain condition is satisfied.
  condition here: Restriction on the magnitude of uncertainty.
- Using the prior predictive $\pi_{\mathrm{pred.}}$ this can be guaranted "on average" (APVC)
- We propose a new method that limits the fluctuations (VPVC)
- Application of the VPVC yields substantially larger sample sizes than the APVC

References

- **This work**: J. Martin and C. Elster. "The variation of the posterior variance and Bayesian sample size determination." arXiv preprint arXiv:1907.12795 (2019).
- F. Wang and A. E. Gelfand. "A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models." Statistical Science 17.2 (2002): 193-208.
- T. Pham-Gia and N. Turkkan. "Sample size determination in Bayesian analysis." Journal of the Royal Statistical Society: Series D (The Statistician) 41.4 (1992): 389-397.