Posterior Distribution vs Tolerance intervals for Sampling Plan Determination in Pharmaceutical Manufacturing

Bayes2016, Leuven

20/05/2016

Martin Otava

Manufacturing, Toxicology and Applied Statistical Sciences EU
Janssen Pharmaceutica, Belgium

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

Process development

1. Engineering runs: to make the process running
2. Characterization phase: to explore its basic properties
3. Factors optimization: designed experiments, optimization
4. Validation phase: to evaluate the final process setup
5. Production phase: to be able to detect any possible issues occurring in the process

Process performance qualification (PPQ) protocol

- Protocol of final validation experiment
- Data available from previous stages (engineering runs, characterization study, DoE)
- Different settings for different data sources

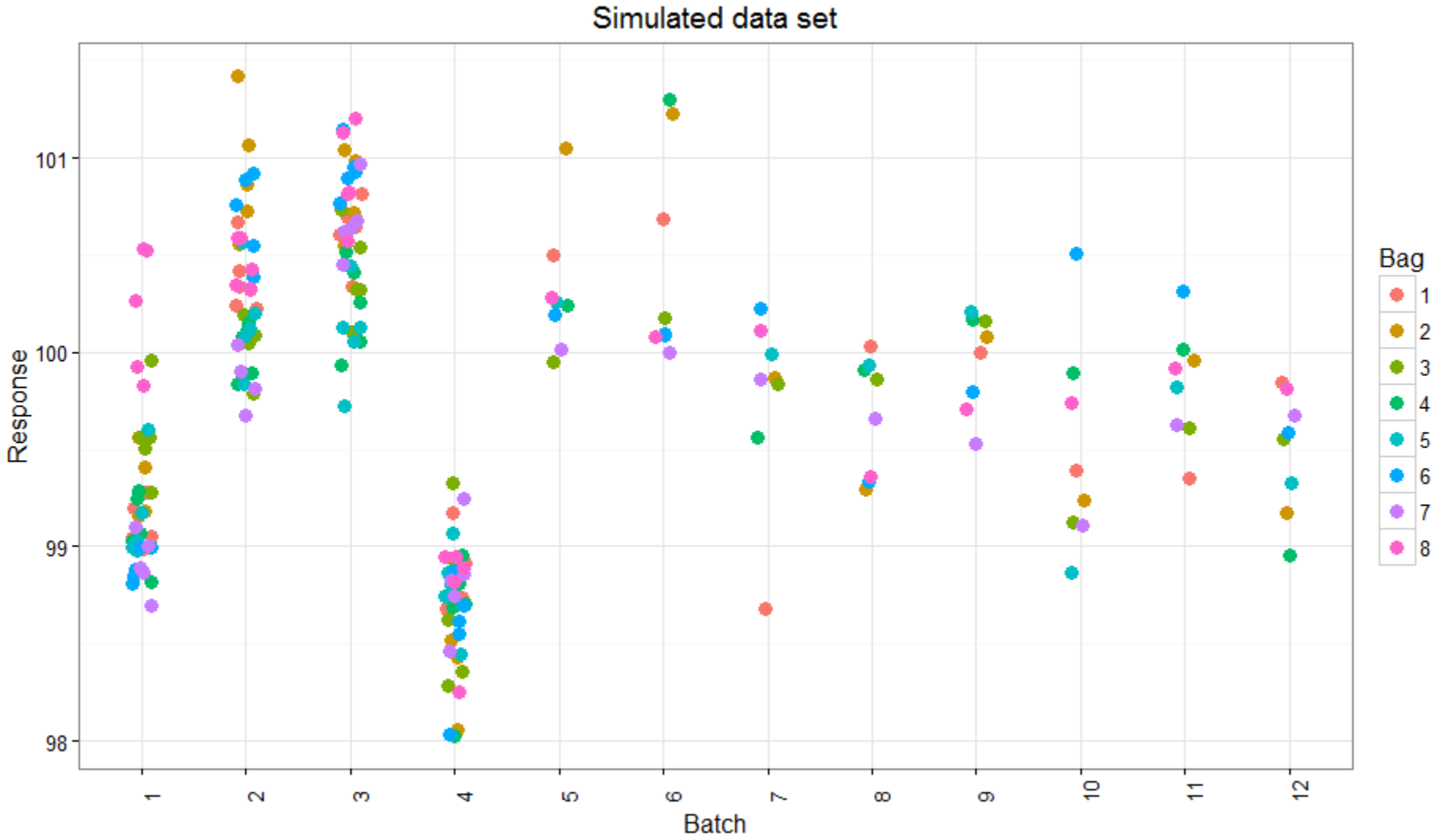
Example Data Set (simulated):

- 12 Batches from pre-PPQ studies of varying purpose
- 4 Batches: 8 Bags, 5 observations per bag
- 8 Batches: 8 Bags, only 1 observation per Bag

- Simulated data set based on structure of real data set (but all the value of parameters are completely artificial)

- Response: % of label claim
- Acceptance criteria: each individual value between 90-110
- Future experiment: 3 Batches, 10 Bags, 1 Sample

Data set


Simulated data set
Response
101
100
99
98
1
2
3
4
5
6
7
8
9
10
11
12
Batch
Bag
1
2
3
4
5
6
7
8

PPQ protocol

- Main question:
- Sampling plan for validation experiment

- Our focus:
- Asses the quality of the process
- Estimate probability of passing validation experiment (given the sampling plan)

Methodology

Frequentist framework

- $Y = N(\mu, \sigma^2)$
- Point estimate of μ
- 95% **Confidence interval** on μ: confidence statement on parameter estimate
- 99% **Prediction interval** for Y: interval containing future observation with confidence of 99%
- 99%/95% **Tolerance interval** for Y: interval containing 99% future observations with confidence of 95%

99% / 95% Tolerance interval

- One-sided: confidence statement about 99% quantile

- Two-side: more complex problem
- Typically centred around the mean
- Normal case approximately (one of many formulas):

$$\hat{\mu} \pm \sqrt{\frac{\nu\left(1+\frac{1}{N}\right) z^2{}_{(1-p)/2}}{\chi^2{}_{1-\alpha,\nu}}} \sqrt{\widehat{\sigma^2}}$$

Bayesian framework

- $Y = N(\mu, \sigma^2)$
- Posterior distribution of parameters available
- Sampling from posterior can be done (within fitting the chain)
- Samples of individual values can be obtained
- Simulation of future experiment can be done

Posterior probabilities of “success”

1. At each iteration of MCMC simulate future experiment

2. Apply acceptance criteria on the simulated experiment (simple threshold or complex decision tree)

3. Record indicator of passing/failing the criteria

4. Mean of indicator = Posterior probability of passing the test for future experiment

- Assumption: future process will behave similarly to current one
- Typically conservative solution

“Bayesian” 99% / 95% tolerance intervals

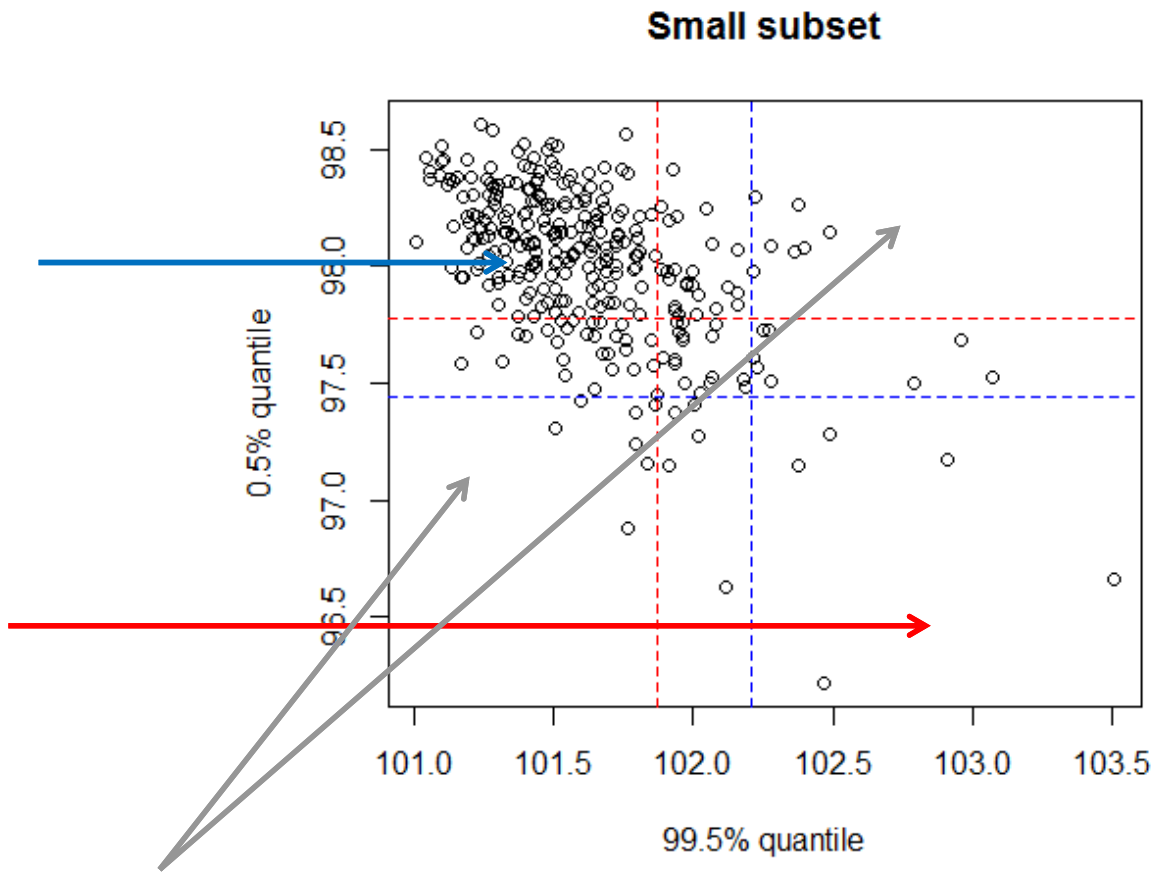
1. At each iteration, sample parameters μ, σ^2
2. Compute quantiles of respective normal distribution
3. Obtain posterior distribution of quantiles
4. Estimation **99% / 95% tolerance** interval

- Wolfinger vs Krishnamoorthy & Mathew: how to determine two-sided tolerance interval?

Comparison: W vs K&M

K&M: 95% of points in top-left rectangle, i.e. within bounds

W: 5% of points in bottom-right rectangle, i.e. outside of the bounds



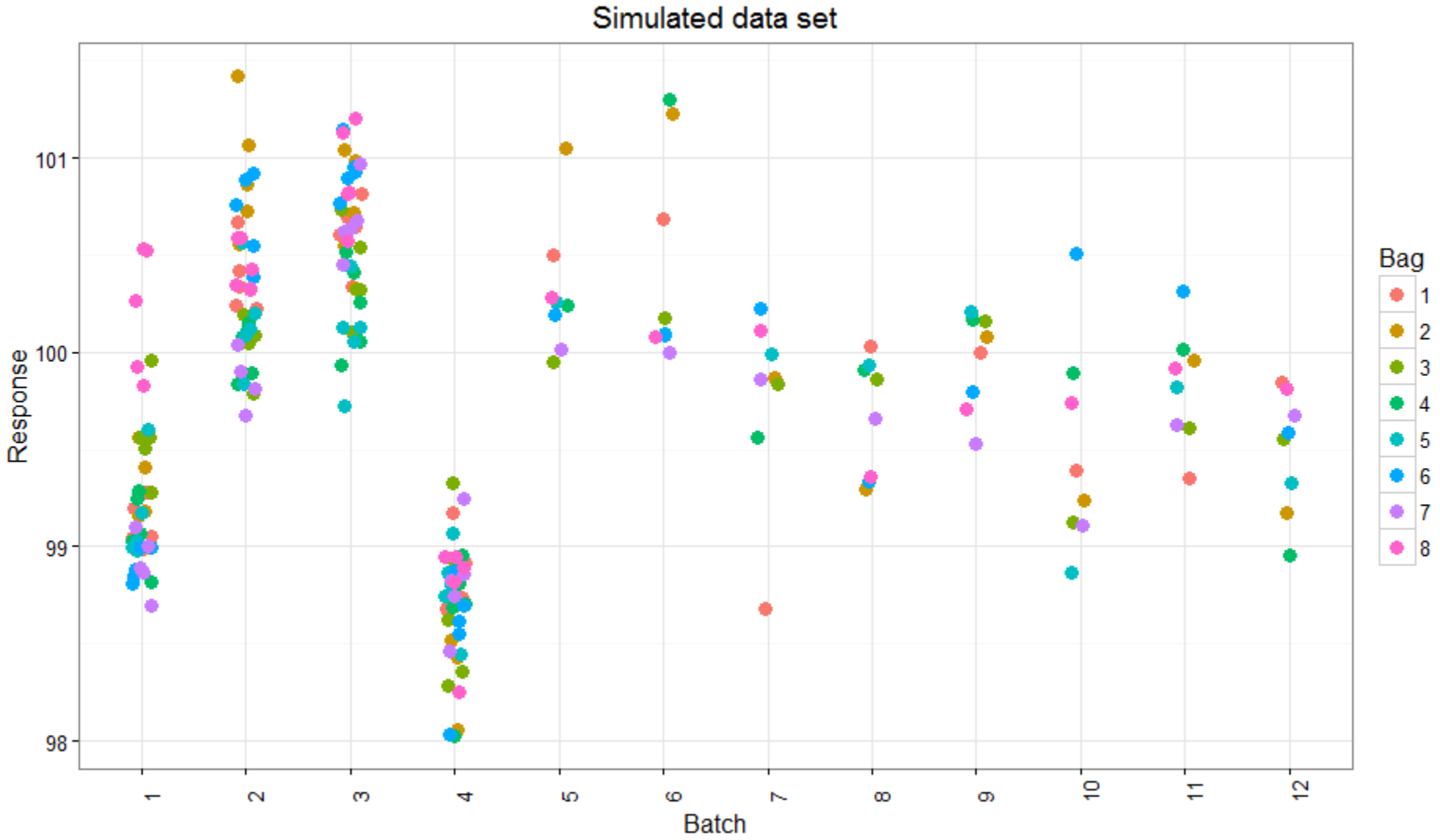
Points in these areas causes difference between methods.

Alternative Bayesian tolerance intervals

- Sampling from the posterior distribution of individual values
- 99% **content Tolerance interval**: 99% credible interval on individual values
- Related to 99% prediction interval
- Interpretation of uncertainty in terms of posterior distribution

Application

Data set


Simulated data set
Response
101
100
99
98
1
2
3
4
5
6
7
8
9
10
11
12
Batch
Bag
1
2
3
4
5
6
7
8

Model

- The following linear mixed model was fitted to the data of all batches:

$$y_{ijk} = \mu_0 + b_i + c_{ij} + \varepsilon_{ijk}$$

where

- y_{ijk} = response for the *k*-th value of *j*-th bag in the *i*-th batch
- μ_0 = the process mean
- b_i = random effect of *i*-th batch: $b_i \sim N(0,\sigma_b^2)$
- c_{ij} = random effect of *j*-th bag of *i*-th batch: $c_{ij} \sim N(0,\sigma_c^2)$
- ε_{ijk} = residual error: $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$.

Note on priors

- Typically “non-informative” type of priors
- Sometimes experiments from previous stages (DoE) are used
- Their relevance is questionable
- Their questionable relevance corresponds tothe reason why they are not included in the data set to be analysed

Output

Estimates	Mean	2.5% CI*	97.5% CI	Truth
Process mean	**99.82**	99.48	100.16	**100.00**

Source of variability	Median	2.5% CI	97.5% CI	Truth
Batch	**0.55**	0.36	0.92	**0.50**
Bag	**0.31**	0.24	0.38	**0.30**
Residual	**0.25**	0.22	0.28	**0.25**

*CI here stand for Credible Interval

Output: tolerance intervals

Tolerance intervals	Lower	Upper
99%/95% Wolfinger	97.60	102.05
99%/95% K & M	97.20	102.44
99% content Tolerance interval	97.82	101.84
True quantiles (0.5%, 99.5%)	98.37	101.63

- We do expect that TIs will be wider than true quantiles
- We do expect that 99%/95% TIs will be wider than 99% content
- We do expect to see difference between W and K&M

Output: Posterior probability

Posterior probability	
Passing acceptance criteria	1*

*Not really 1

- Typical result in practical applications in late stage of development
- Real meaning: “it has never happened in my simulated MCMC that…”
- Depends on number of chains, iteration, thinning & correlation
- In this example (worst case scenario) **>99.91%**

Conclusions

Summary

- Posterior probability and TIs connected, but approaching main question from different viewpoints

- Reporting both of them has added value

- **Bayesian approach** allows us to**:**
  - estimate both quantities directly
  - fit more complex models without much extra effort
  - use of prior information

  - BUT, be careful with:
    99%/95% TIs computation
    specification of priors
    computational time

References

- Wolfinger, R. D. (1998): Tolerance intervals for variance component models using Bayesian simulation. *Journal of Quality technology*, 30(1)

- Krishnamoorthy, K. & Mathew, T. (2009): *Statistical Tolerance Regions: Theory, Applications, and Computation.* ISBN: 978-0-470-38026-0

Thank you

motava@its.jnj.com

20/05/2016


janssen
PHARMACEUTICAL COMPANIES
OF Johnson&Johnson