On the efficiency of two-stage adaptive designs

Björn Bornkamp (Novartis Pharma AG)

Based on:
Dette, H., Bornkamp, B. and Bretz F. (2010): On the efficiency of adaptive designs
www.statistik.tu-dortmund.de/sfb823-dp2010.html

Adaptive Designs

- Adaptive Designs:

  Phase I (CRM)

  Phase II (adaptive dose-ranging)

  Phase II/III (combination of Phase II and III results)

- Past 5-6 years increased interest in adaptive/model-based dose-ranging designs

  $\rightarrow$ FDA critical path initiative

  $\rightarrow$ ASTIN dose-finding trial (Bayesian highly adaptive)

  $\rightarrow$ PhRMA working group on adaptive dose-finding studies (José Pinheiro, Michael Krams, Vlad Dragalin & many others)

Adaptive Dose-Ranging Designs

- Allocate patients adaptively within the trial, so that information on the dose-response curve (or target dose) is maximized. Stop the trial, when "enough" information has been gathered.
- Some examples: Target function, dose-response model:
  - Müller, Berry, Grieve, Krams (2006)
    variance of the ED95, semiparametric Bayesian dose-response model
  - Dragalin et al. (2007)
    covariance matrix of parameters (D-optimality), flexible nonlinear dose-response model
  - Bornkamp et al. (2011)
    variance of threshold dose (MED), set of candidate nonlinear candidate dose-response models

Adaptive dose-finding designs: Merits

From FDA adaptive design draft guidance:

Adaptive designs may lead to studies that

- *more efficiently provide the same information (shorter duration, fewer patients)*
- *more likely to demonstrate an effect of the drug (if one exists)*
- *more informative on the treatment's effect (broader and better dose-response information, subgroup effects), which may lead to more efficient subsequent studies*

My perspective (and for adaptive dose-ranging trials):

Flexibility allows to robustify the design of the study. Refine initial assumptions within the trial, *e.g.*, on

- dose-response relationship (where to place doses to learn most, dose-range)
- expected treatment effect (stop early)

Adaptive Designs: Challenges

Logistically more challenging to plan and implement:

- simulations needed to assess operating characteristics
- drug supply/logistical questions
- recruitement rate (study duration might increase)
- more “institutional” hurdles (internal and external)

This talk:

Focus on a specific aspect regarding adaptive designs

$\hookrightarrow$ statistical estimation efficiency

Motivation

- Experience from a number of simulation studies performed with adaptive dose-finding methods:
  - moderate gains for most simulation scenarios
  - sometimes substantial gains particularly, when initial assumptions are dramatically wrong
  - but sometimes also slightly worse (!)
- Question at that time:

  Can analytical considerations confirm these findings?

  Simulations typically performed under realistic settings, *i.e.*, a number of potentially interfering parameters (logistical constraints, complex models, programming bugs etc).

  Use simplified/idealized setting to identify key factors

Considered designs

- Target: Estimate the parameters or parameter function
- Use the design that maximizes $\phi(M(\xi, \boldsymbol{\theta}))$
  $\boldsymbol{\theta}$ model parameters, $M(\xi, \boldsymbol{\theta})$ Fisher information matrix, ξ experimental design (specifying doses and allocations), ϕ differentiable "design criterion"

Compare designs:

1. ξ_F Fixed design: Take all N observations at the locally optimal design for $\boldsymbol{\theta}_0$ (parameter guess)
2. ξ_A Two-stage adaptive design: Split study into two parts, take $N_0 = p_0 N$ samples in first part, $p_0 \in (0, 1)$, calculate ML estimate $\widehat{\boldsymbol{\theta}}_1$ and allocate remaining $N - N_0$ samples to the optimal design obtained by maximizing $\phi(M(\xi, \widehat{\boldsymbol{\theta}}_1))$

Main idea for analysis

At the end calculate maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ using both parts.

Measure estimation precision in terms of the mean-squared error MSE$(\widehat{\boldsymbol{\theta}}) = E[(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \approx$ Var$(\widehat{\boldsymbol{\theta}})$, because the bias is of order $1/N^2$

Main idea for analytical considerations: Variance decomposition

$$Var(\widehat{\boldsymbol{\theta}}) = E(Var(\widehat{\boldsymbol{\theta}}|Y_1, \ldots, Y_{N_0})) + Var(E(\widehat{\boldsymbol{\theta}}|Y_1, \ldots, Y_{N_0}))$$

Derive approximations for the terms in the sum

(assume $N \to \infty$, but $p_0 = N_0/N$ constant)

Defining:

$I(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = M(\xi_{\boldsymbol{\theta}_0}, \boldsymbol{\theta})$ ($\xi_{\boldsymbol{\theta}_0}$ is the local optimal design for $\boldsymbol{\theta}_0$) and

$H(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = p_0 I(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + p_1 I(\boldsymbol{\theta}, \boldsymbol{\theta})$

One obtains for the approximate covariance matrices:

$$M(\xi_F, \boldsymbol{\theta}) \approx I(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + \frac{1}{\sqrt{N_0}} K(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + \frac{1}{N_0} L(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

$$M(\xi_A, \boldsymbol{\theta}) \approx H(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + \frac{1}{\sqrt{N_0}} \bar{K}(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + \frac{1}{N_0} \bar{L}(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$$

General observation:

- Because $H(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \geq I(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ the adaptive design is asymptotically better
- For finite sample size the terms of order $1/N_0$ and $1/\sqrt{N_0}$ are non-negligible, hence need to consider each specific model separately

For differentiable optimality criterion ϕ one obtains, when comparing the efficiency

$$\text{eff}_\phi(\xi_F, \xi_A) = \frac{\phi(M(\xi_F, \boldsymbol{\theta}))}{\phi(M(\xi_A, \boldsymbol{\theta}))} \approx \frac{\phi(I(\boldsymbol{\theta}, \boldsymbol{\theta}_0))}{\phi(p_0 I(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + p_1 I(\boldsymbol{\theta}, \boldsymbol{\theta}))} + \frac{c}{\sqrt{N_0}} + \frac{d}{N_0},$$

no information regarding the sign of the constants c and d is available in general.

For simplicity concentrate on one parameter models in what follows

$$\begin{aligned}\text{eff}(\xi_F, \xi_A) &= \frac{\text{MSE}(\hat{\theta}_F)}{\text{MSE}(\hat{\theta}_A)} \approx \frac{\text{Var}(\hat{\theta}_F)}{\text{Var}(\hat{\theta}_A)} \\ &\approx \left\{ \frac{I(\theta,\theta_0)}{H(\theta,\theta_0)} - p_1 \frac{g(\theta)(5p_0 I(\theta,\theta_0) + p_1 I(\theta,\theta))}{2N_0 H^3(\theta,\theta_0)} \right\}^{-1}\end{aligned}$$

where $g(\theta) := \nabla^2 I(\theta,\tau)\Big|_{\tau=\theta}$ is always negative

- The dominating term $\frac{I(\theta,\theta_0)}{H(\theta,\theta_0)} \leq 1$
- But for finite sample sizes, the relationship is not clear due to the second summand

Simple example

Exponential model

$$E[Y|x] = \eta(x, \theta) = e^{-\theta x}, \qquad \text{Var}\,(Y|x) = \sigma^2 > 0$$

with unknown parameter θ and initial guess θ_0

ξ_F: Fixed design

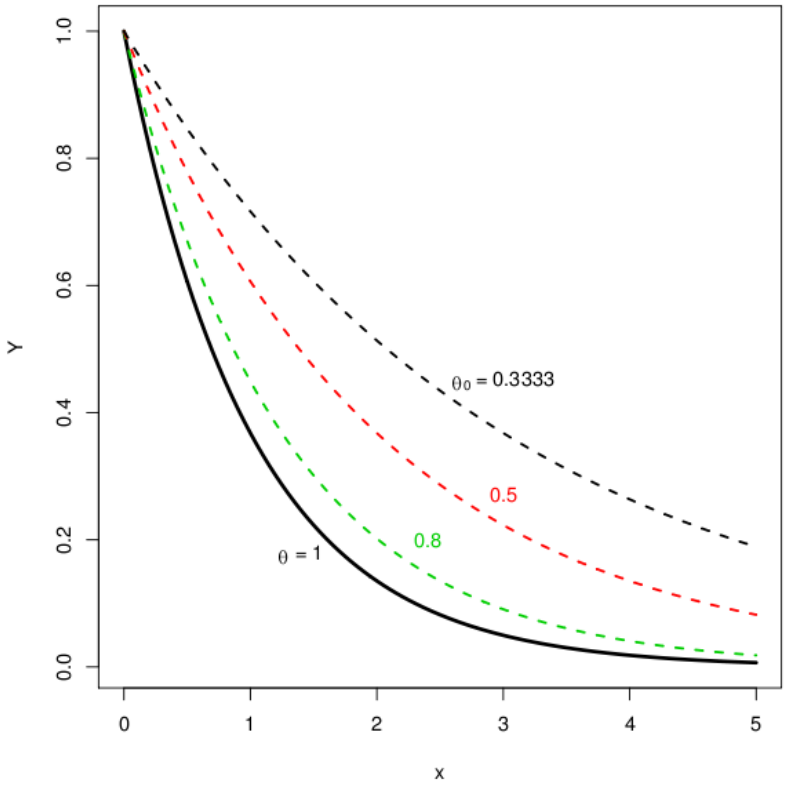
- N observations according to optimal design based on θ_0

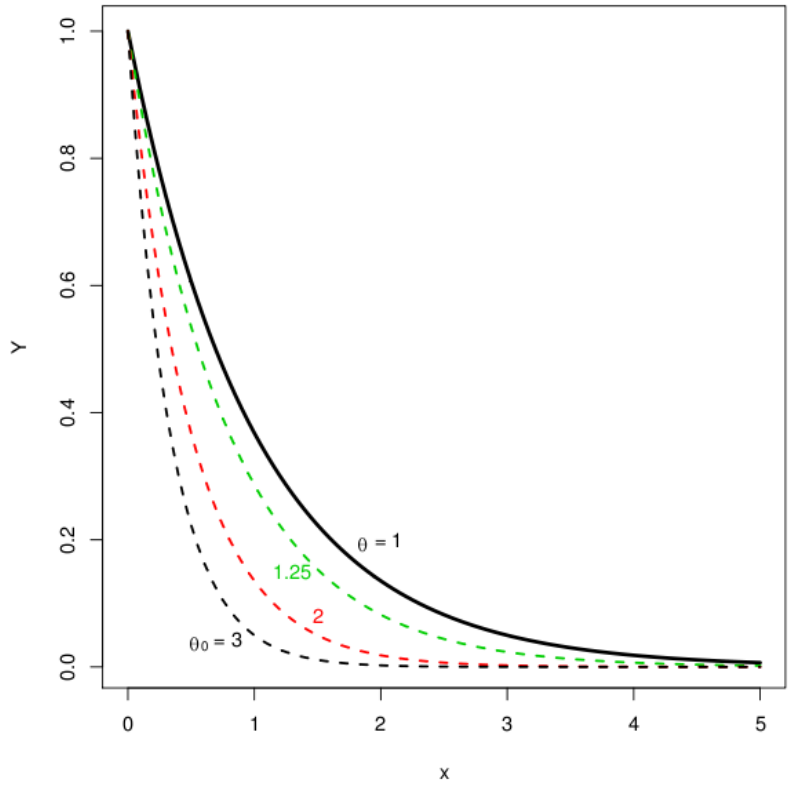
ξ_A: Two-stage adaptive design

- Stage 1: N_0 observations with design based on θ_0
- Interim: Estimate θ, resulting in $\hat{\theta}_1$
- Stage 2: $N - N_0$ observations with design based on $\hat{\theta}_1$

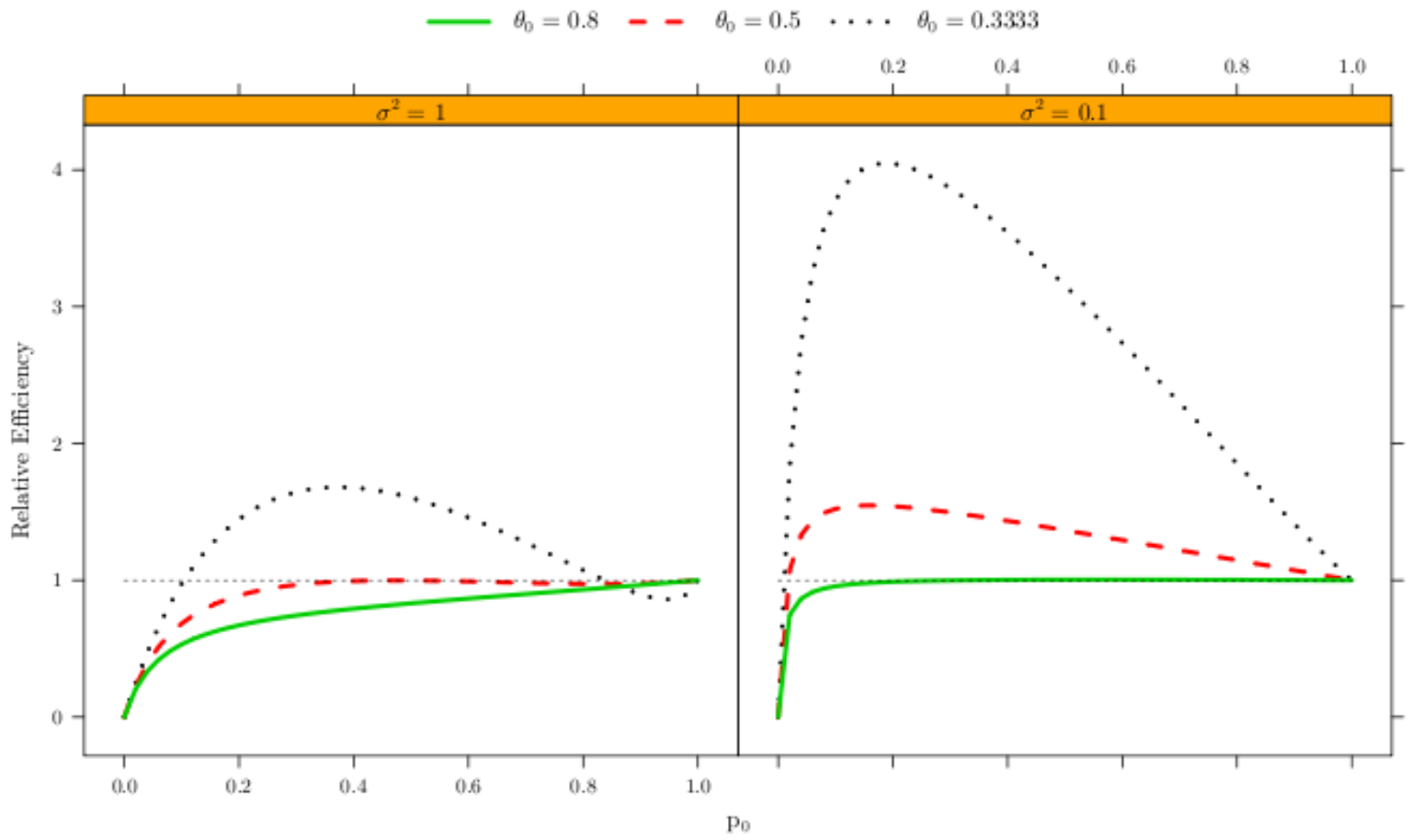
Which design is more efficient and estimates θ more precisely?

Exponential model with unknown parameter $\theta = 1$ and different initial guesses

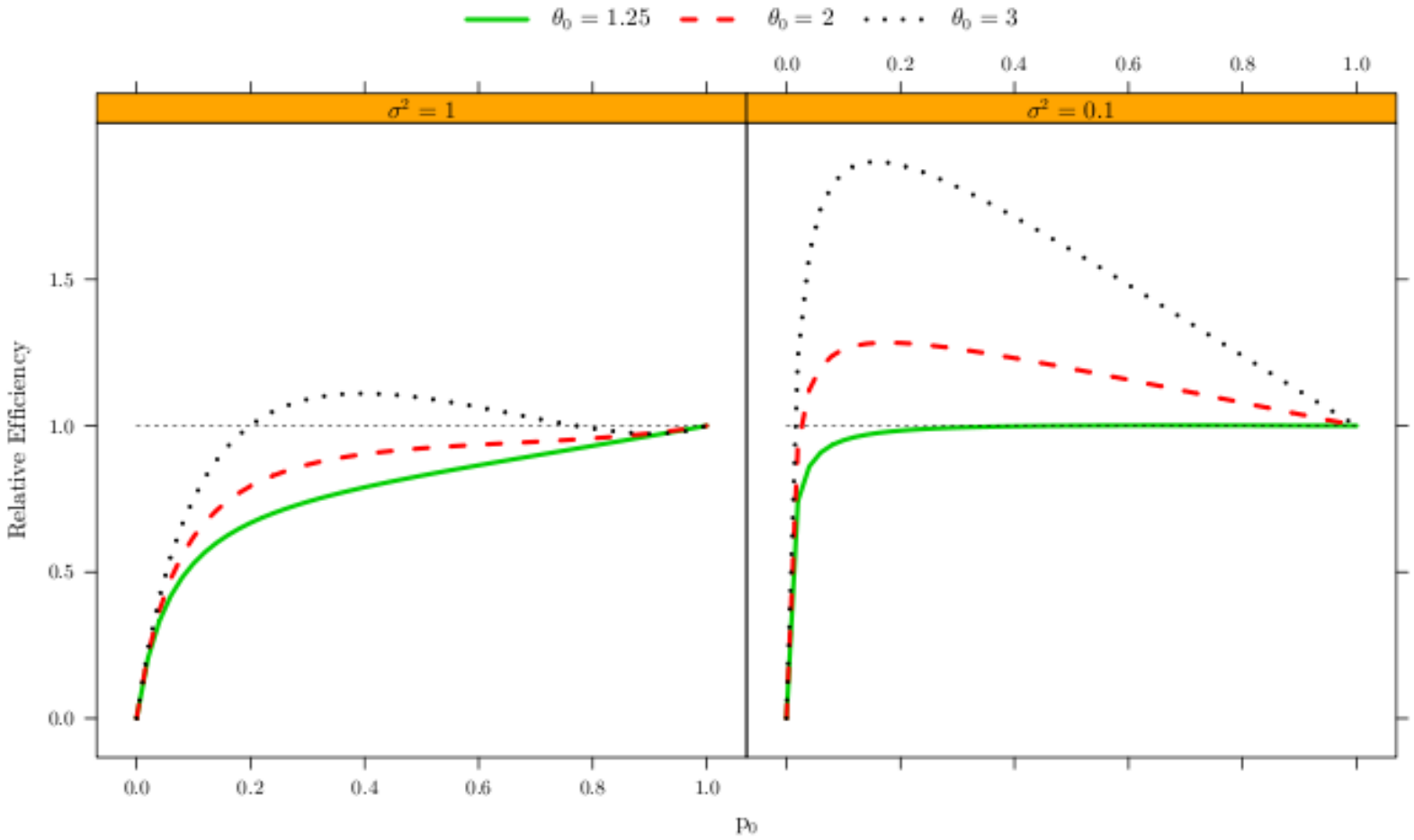




Relative efficiency of adaptive versus non-adaptive design for $N = 100$, $\theta = 1$. Efficiency > 1 indicates that the adaptive design is better.

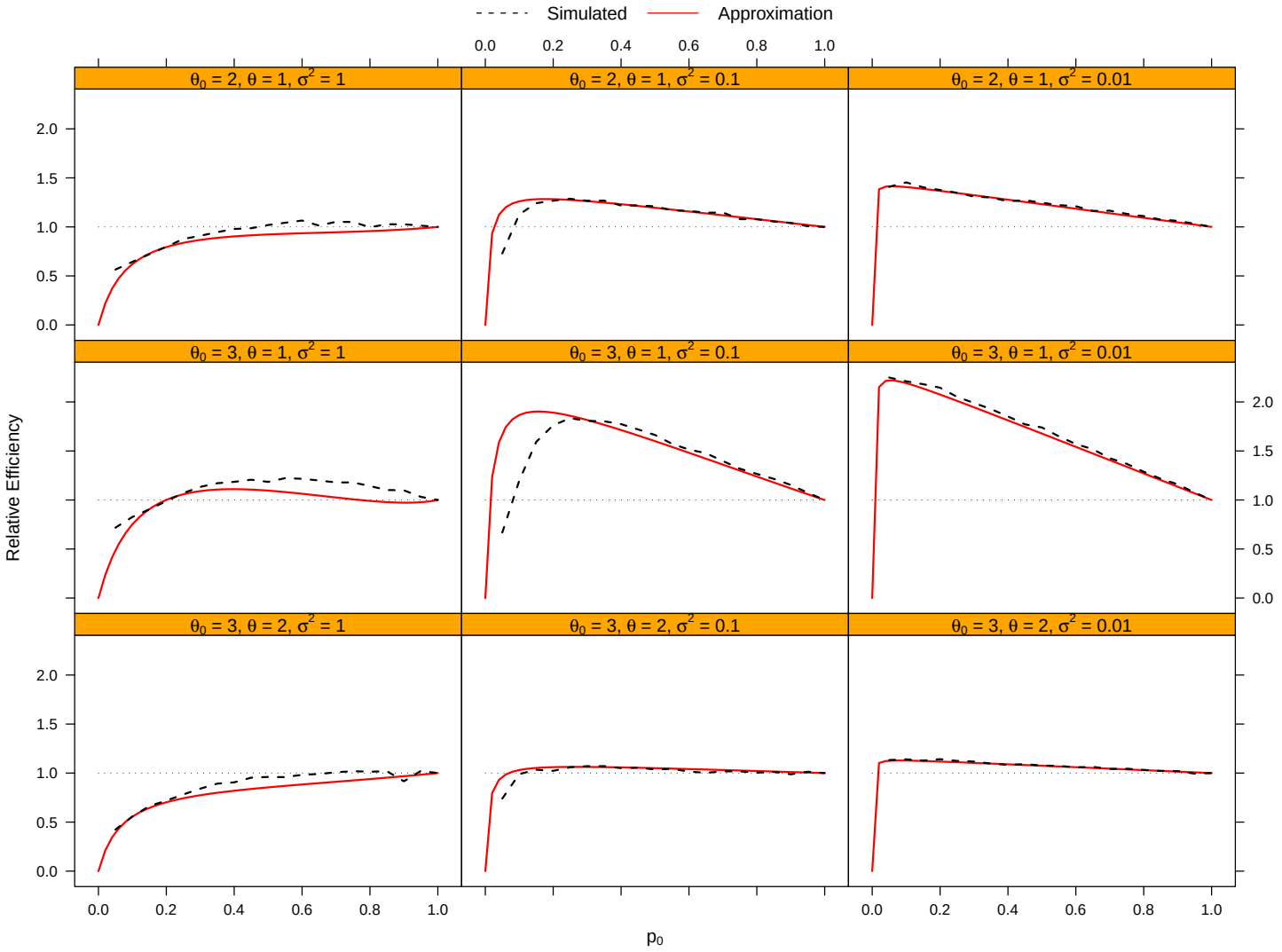


Efficiency $>$ 1 indicates that the adaptive design is better.



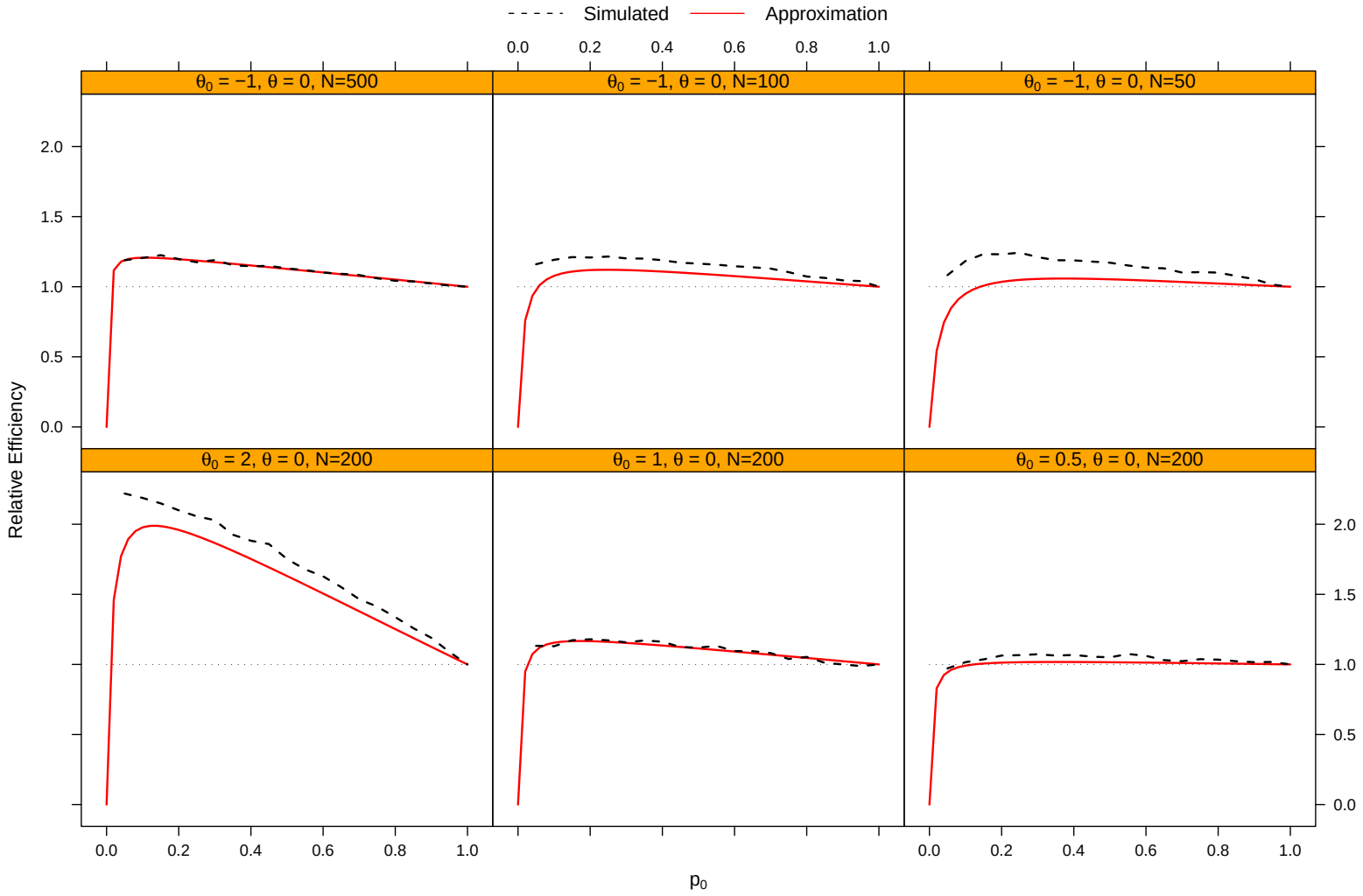
Main factors: Variability, adequacy of initial guess θ_0

Exponential Regression



Logistic regression

$$p(x,\theta) = E[Y|x] = \frac{1}{1+e^{x-\theta}}$$



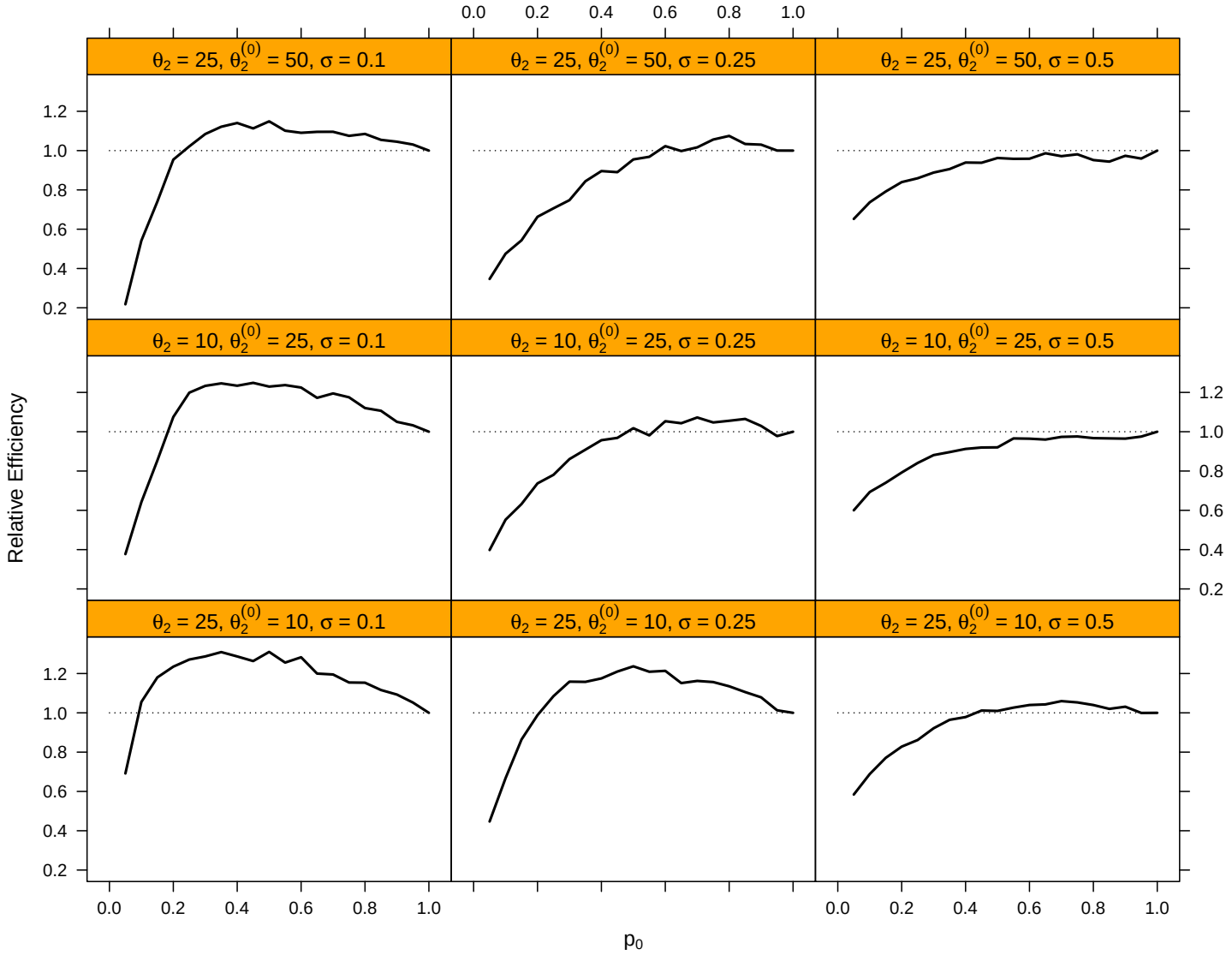
$$E[Y|d] = \eta(d, \theta) = \theta_0 + \theta_1 \frac{d}{\theta_2 + d}, \qquad \text{Var}\,(Y|d) = \sigma^2 > 0.$$

Focus on estimating the ED_{90} parameter with small variability for a trial with 100 patients.

The optimal design allocates $1/4$ of the patients on each of 0 and d_{max} the maximum dose and the remaining $1/2$ of the patients on the intermediate dose

$$\frac{d_{\max}\theta_2}{d_{\max} + 2\theta_2}.$$

Emax model


θ2 = 25, θ2(0) = 50, σ = 0.1
θ2 = 25, θ2(0) = 50, σ = 0.25
θ2 = 25, θ2(0) = 50, σ = 0.5
θ2 = 10, θ2(0) = 25, σ = 0.1
θ2 = 10, θ2(0) = 25, σ = 0.25
θ2 = 10, θ2(0) = 25, σ = 0.5
θ2 = 25, θ2(0) = 10, σ = 0.1
θ2 = 25, θ2(0) = 10, σ = 0.25
θ2 = 25, θ2(0) = 10, σ = 0.5
Relative Efficiency
p0
0.0
0.2
0.4
0.6
0.8
1.0
1.2

Summary

Despite several limitations (restriction to simple models, idealized situation, asymptotic approximations):

Analytical considerations confirm simulation results

- The adaptive design will dominate the non-adaptive design (for large enough sample size)
- But there can be situations, where adaptation makes things worse (in terms of statistical estimation efficiency)
- Results strongly depends on the specific model and situation (apparently no general results possible)

$\hookrightarrow$ Simulations need to be performed to evaluate adaptive designs

In addition: An advantage of adaptive designs is robustification (things can be different than expected), this flexibility is hard to quantify numerically

References

Mller, P., Berry, D., Grieve, A. and Krams, M. (2006) A Bayesian Decision-Theoretic Dose-Finding Trial, *Decision Analysis*, 3, 197–207

Bornkamp, B., Bretz, F., Dette, H., and Pinheiro, J. (2011) Response-adaptive dose-finding under model uncertainty, *Annals of Applied Statistics*, 5, 1611–1631

Dragalin, V., Hsuan, F. and Padmanabhan, S.K. (2007) Adaptive Designs for dose-finding studies based on the sigmoid Emax model, *Journal of Biopharmaceutical Statistics*, 17, 1051–1070