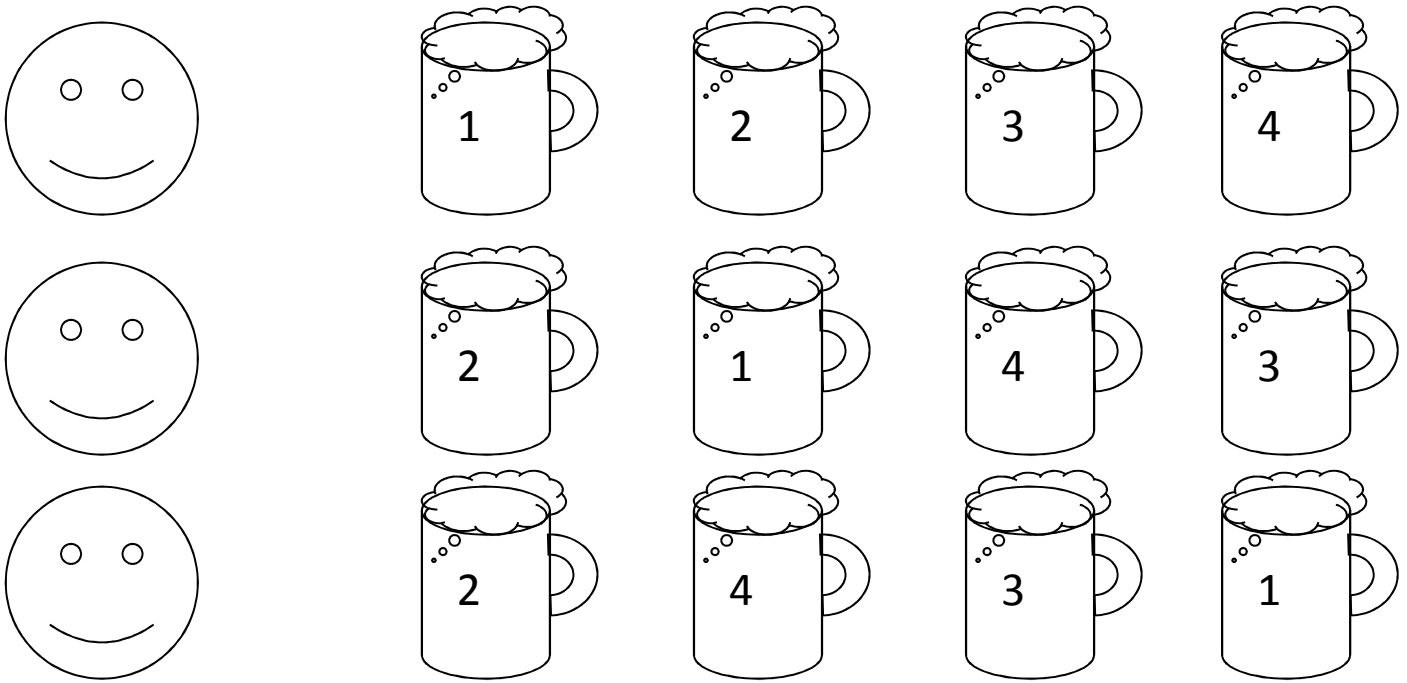
Optimal crossover designs in a model with interactions between treatments and subjects

Joachim Kunert

with

Andrea Bludowsky and John Stufken

In crossover studies, each subject gets treated with a series of treatments, one after the other.



Most papers on optimal crossover designs consider a model with carryover effects, such that

$$y_{dij} = \tau_{d(i,j)} + \rho_{d(i,j-1)} + \beta_i + \alpha_j + e_{ij},$$

where the errors $e_{11}, \dots, e_{1p}, e_{21}, \dots, e_{np}$ are i.i.d.

This model is often critizised because it implies that (neglecting carryover effects)

an experiment with $n = 2$ subjects and $p = 100$ observations per subject
provides as much information as
an experiment with $n = 100$ subjects and $p = 2$ observations per subject.

This clearly is not realistic :

For instance in preference studies,
one consumer comparing products A and B 50 times,
provides less information than 50 consumers comparing A and B just once.

Any consumer prefering A, will most likely prefer it all the time.

In consumer studies,
therefore often an interaction between product and consumer is modelled.

We consider this extended model, where the treatment effect $\tau_{d(i,j)}$ becomes

$$\tau_{d(i,j)} + \xi_{d(i,j),i}$$

Here
$\tau_{d(i,j)}$ is a fixed effect of the treatment, which is the same for all subjects and
$\xi_{d(i,j),i}$ is a random interaction between subject and treatment,
with expectation 0 and variance σ_ξ^2.

Hence, if an assessor gets the same product more than once,
the two (or more) observations of the same product by the same assessor
are correlated, with correlation γ, where

$$\gamma = \frac{\sigma_\xi^2}{\sigma_e^2 + \sigma_\xi^2}.$$

Therefore, if subject i gets teatment-sequence

1 2 3 3,

then y_{di3} and y_{di4} are correlated, with correlation γ.

If, however, subject i gets sequence

1 2 3 4

(i.e. there are no repeated treatments on this unit) then the errors are i.i.d..

Although this model often is used in practice
(at least for preference studies),
there appear to be no published papers on optimal designs for this model.

Here is an attempt to fill this gap.

We consider the model

$$y_{dij} = \tau_{d(i,j)} + \xi_{d(i,j),i} + \rho_{d(i,j-1)} + \beta_i + \alpha_j + e_{ij},$$

where the random components $e_{11}, \dots, e_{1p}, e_{21}, \dots, e_{np}, \xi_{1,1}, \dots, \xi_{1,n}, \xi_{2,1}, \dots, \xi_{t,n}$ are all independent.

Note that we assume that β_i, the effect of unit i, is a fixed effect.
(This corresponds to random unit effects with a large variance.)

If we rewrite

$$y_{dij} = \tau_{d(i,j)} + \rho_{d(i,j-1)} + \beta_i + \alpha_j + \xi_{d(i,j),i} + e_{ij},$$

we see that the fixed part of the model is the same as in the traditional model.

The difference lies in the covariance structure (which depends on the design).

The model in vector notation:

$$y = T_d\tau + F_d\rho + U\beta + P\alpha + \epsilon$$

where

$$Cov(\epsilon) = \Sigma_d = \begin{bmatrix} S_{d1} & & \\ & S_{d2} & \\ & & \ddots \end{bmatrix}$$

There are no correlations between measurements at different subjects,
the covariance matrix S_{di} at subject i depends on the sequence of treatments for subject i.

Since each observation is under one treatment and at one subject,
we have

$$T_d 1 = U1 = 1$$

and τ is not estimable.

Kiefer (1975):

In a linear model

$$y = A_d\alpha + B_d\beta + e$$

where $A_d 1 \in image(B_d)$

consider

$\hat{\alpha}$, the unbiased estimate for $\left(I - \frac{1}{t}11^T\right)\alpha$.

Information matrix:

$$C_d = A_d^T \omega^{\perp}(B_d) A_d.$$

This is the Moore-Penrose generalized inverse of $Cov(\hat{\tau})$.

Note that C_d has row- and column-sums zero.

Kiefer (1975):

If a design d^* is such that

(i) $C_{d^*} = aI_t + b1_t 1_t^T$
(ii) $trace\ C_{d^*} = \max_{d \in \Delta} trace\ C_d$

then d^* is universally optimal over Δ.

Weighted least squares estimate:

Use Q_d, such that $Q_d \Sigma_d Q_d = I$.

Defining

$$\tilde{y} = Q_d y, \tilde{T}_d = Q_d T_d, \tilde{F}_d = Q_d F_d, \tilde{P} = Q_d P, \tilde{U} = Q_d U,$$

we get the transformed model

$$\tilde{y} = \tilde{T}_d \tau + \tilde{F}_d \rho + \tilde{U} \beta + \tilde{P} \alpha + e,$$

where e is a vector of i.i.d. errors.

The information matrix is then

$$C_d = \tilde{T}_d^T \omega^{\perp}\left(\left[\tilde{F}_d, \tilde{U}, \tilde{P}\right]\right)\tilde{T}_d,$$

which is not easy to handle.

To proceed, we need Kushner's method.

We consider approximate designs,
where the design-points are the possible sequences we might give to the subjects
and the weights are the proportions of subjects receiving these sequences.

Remember that we try to find a design d^*
maximizing the trace of the information matrix.

For approximate designs, $trace\ C_d = c_{d11} - \frac{c_{d12}^2}{c_{d22}}$,

where $c_{d11} = trace\ \tilde{T}_d^T \omega^{\perp}(\tilde{U}) \tilde{T}_d$,
$c_{d12} = trace\ \tilde{T}_d^T \omega^{\perp}(\tilde{U}) \tilde{F}_d$,
$c_{d22} = trace\ \tilde{F}_d^T \omega^{\perp}(\tilde{U}) \tilde{F}_d$.

Each of the c_{dij} can be written as $c_{dij} = \sum_{\ell} \pi_{\ell} c_{ij}(\ell)$,
where
ℓ goes over all possible sequences of treatments,
$c_{ij}(\ell)$ is what we would get if all units received sequence ℓ, and
π_{ℓ} is the proportion of sequences receiving ℓ.

(To be more precise: ℓ indicates an equivalence class of sequences, all of which produce the same $c_{ij}(\ell)$.)

So the c_{dij} are a linear combination of the $c_{ij}(\ell)$, but

$$tr\ C_d = c_{d11} - \frac{c_{d12}^2}{c_{d22}}$$

is not.

Kushner (1997) observed that

$$tr\ C_d = c_{d11} - \frac{c_{d12}^2}{c_{d22}}$$

is the minimum of

$$H_d(x) = c_{d11} + 2c_{d12}\ x + c_{d22}\ x^2$$

and that

$$H_d(x) = \sum_{\ell} \pi_\ell H_\ell(x)\,,$$

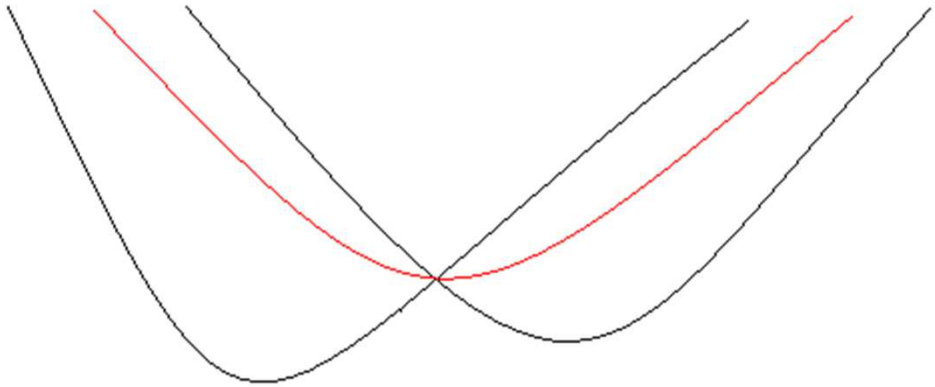
where

$$H_\ell(x) = c_{11}(\ell) + 2c_{12}(\ell)\ x + c_{22}(\ell)\ x^2$$

is the polynomial of the sequence class ℓ.

This implies that the $H_d(x)$ cannot be larger than the maximum of the $H_\ell(x)$

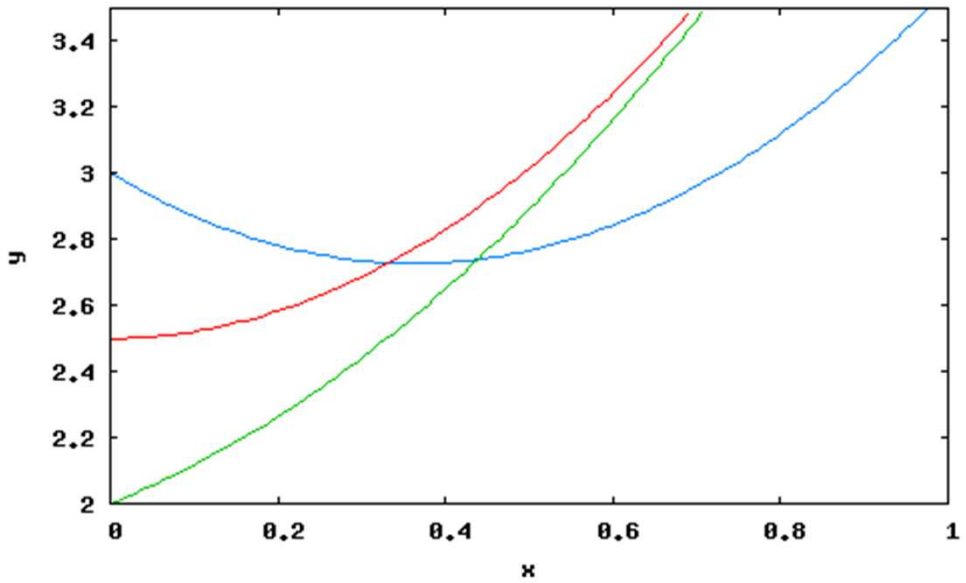
and $\min_x H_d(x) \leq \min_x \max_\ell H_\ell(x)$.



The equivalence classes of sequences for $p = 4$.

class-number k	representative sequence $z(k)$	$c_{11}(z(k))$	$c_{12}(z(k))$	$c_{22}(z(k))$
1	[1 2 3 4]	3	$-\frac{4}{3}$	$\frac{9t-3}{4t}$
2	[1 2 3 3]	$\frac{5+\gamma}{2+\gamma}$	$-\frac{\gamma}{4+2\gamma}$	$\frac{(9-2\gamma-\gamma^2)t-3+2\gamma+\gamma^2}{(4-2\gamma-2\gamma^2)t}$
3	[1 2 3 2]	$\frac{5+\gamma}{2+\gamma}$	$-\frac{2}{2+\gamma}$	$\frac{(9-2\gamma-\gamma^2)t-3+2\gamma+\gamma^2}{(4-2\gamma-2\gamma^2)t}$
4	[1 2 3 1]	$\frac{5+\gamma}{2+\gamma}$	$-\frac{4+\gamma}{4+2\gamma}$	$\frac{(9-4\gamma-2\gamma^2)t-3}{(4-2\gamma-2\gamma^2)t}$
5	[1 2 2 3]	$\frac{5+\gamma}{2+\gamma}$	$-\frac{1}{4+2\gamma}$	$\frac{(7-\gamma^2)t-3+2\gamma+\gamma^2}{(4-2\gamma-2\gamma^2)t}$
6	[1 2 2 2]	$\frac{3}{2+\gamma}$	$\frac{1}{4+2\gamma}$	$\frac{(7+\gamma)t-3+3\gamma}{(4-2\gamma-2\gamma^2)t}$
7	[1 2 2 1]	$\frac{2}{1+\gamma}$	$-\frac{1}{2+2\gamma}$	$\frac{(7+5\gamma)t-3-\gamma}{(4-4\gamma^2)t}$
8	[1 2 1 3]	$\frac{5+\gamma}{2+\gamma}$	$-\frac{5}{4+2\gamma}$	$\frac{(7-4\gamma)t-3}{(4-2\gamma-2\gamma^2)t}$
9	[1 2 1 2]	$\frac{2}{1+\gamma}$	$-\frac{3}{2+2\gamma}$	$\frac{(7-3\gamma)t-3-\gamma}{(4-4\gamma^2)t}$
10	[1 2 1 1]	$\frac{3}{2+\gamma}$	$-\frac{3}{4+2\gamma}$	$\frac{(7+\gamma)t-3-\gamma}{(4-2\gamma-2\gamma^2)t}$
11	[1 1 2 3]	$\frac{5+\gamma}{2+\gamma}$	$-\frac{1+\gamma}{4+2\gamma}$	$\frac{(7-2\gamma-2\gamma^2)t-3}{(4-2\gamma-2\gamma^2)t}$
12	[1 1 2 2]	$\frac{2}{1+\gamma}$	$\frac{1}{2+2\gamma}$	$\frac{(7+5\gamma)t-3-\gamma}{(4-4\gamma^2)t}$
13	[1 1 2 1]	$\frac{3}{2+\gamma}$	$-\frac{3}{4+2\gamma}$	$\frac{(7+\gamma)t-3-\gamma}{(4-2\gamma-2\gamma^2)t}$
14	[1 1 1 2]	$\frac{3}{2+\gamma}$	$-\frac{1}{4+2\gamma}$	$\frac{(3+\gamma)(t-1)}{(4-2\gamma-2\gamma^2)t}$
15	[1 1 1 1]	0	0	$\frac{3(t-1)}{4t(1-\gamma)}$

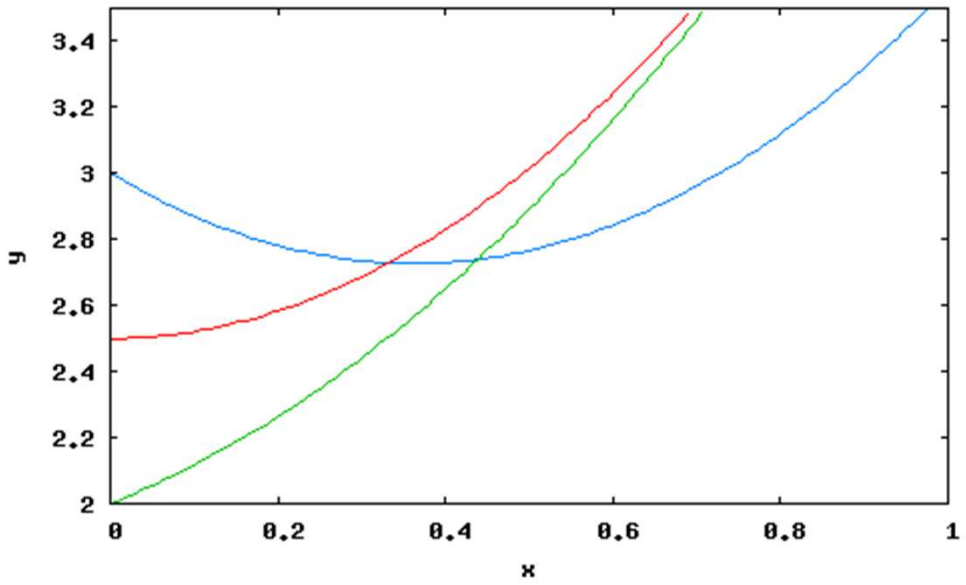
As an example, we consider the situation that $p = t = 4$.
In this situation, there are 15 equivalence classes of sequences. The three most interesting are
$\ell = 1$: $[1\ 2\ 3\ 4]$ (blue)
$\ell = 2$: $[1\ 2\ 3\ 3]$ (red)
$\ell = 12$: $[1\ 1\ 2\ 2]$ (green)



We plot their polynomials in the special instance that $\gamma = 0$.
Note that this corresponds to the situation that there is no treatment by unit interaction, i.e to the traditional model.

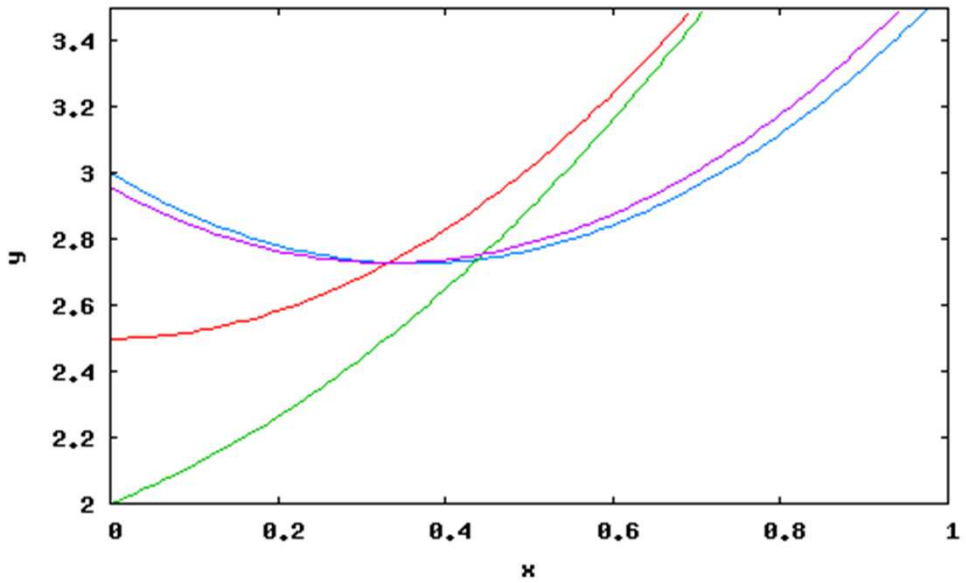
The graph explains the well-known fact that the optimal design in the traditional model uses sequences 1 and 2.

The proportion of sequences from class 2 is small (1/12).



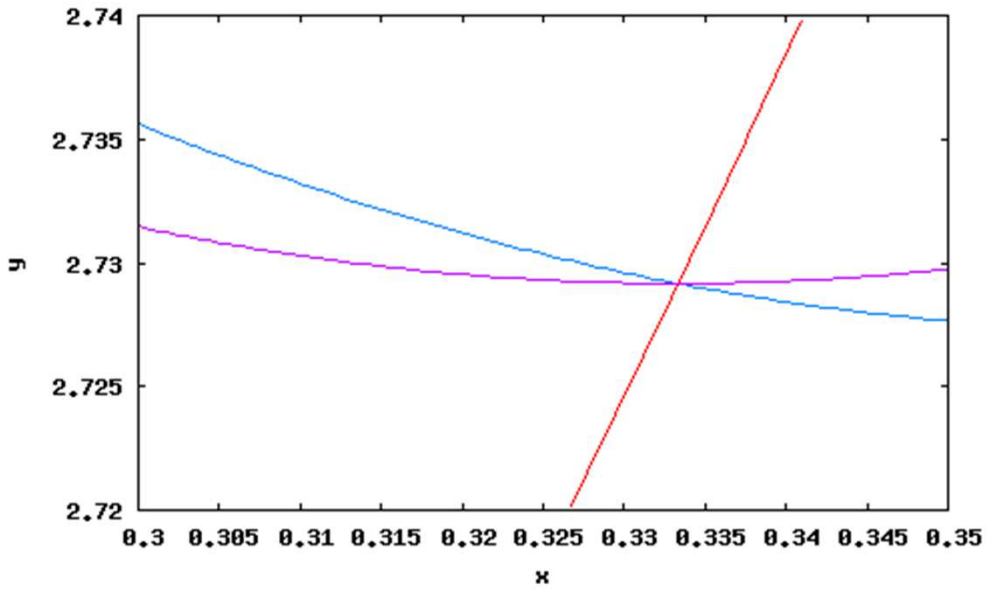
The graph explains the well-known fact that the optimal design in the traditional model uses sequences 1 and 2.

The proportion of sequences from class 2 is small (1/12).



The violet line is the linear combination.

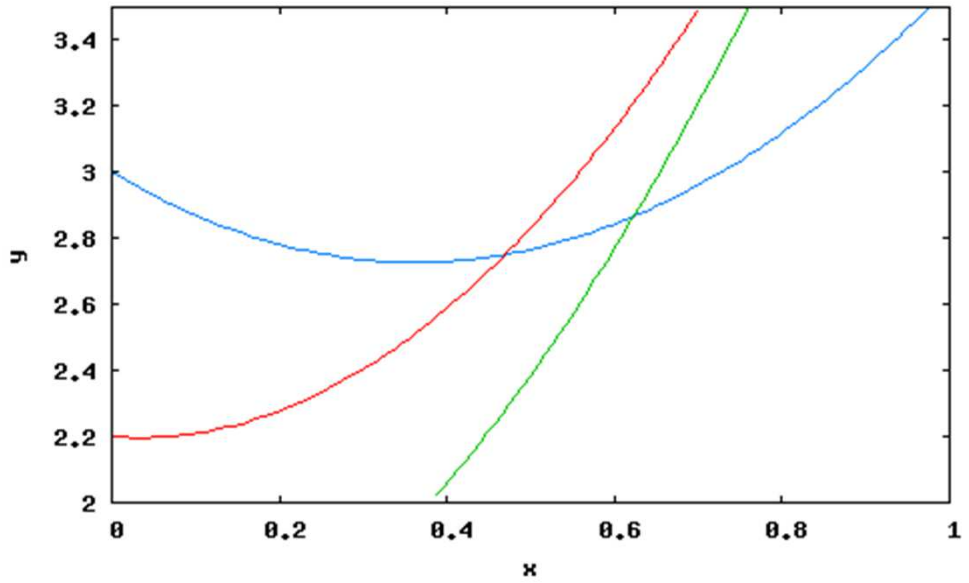
We can see it better if we zoom into the graph:



The combination of the two sequences has its minimum at the intersection point.

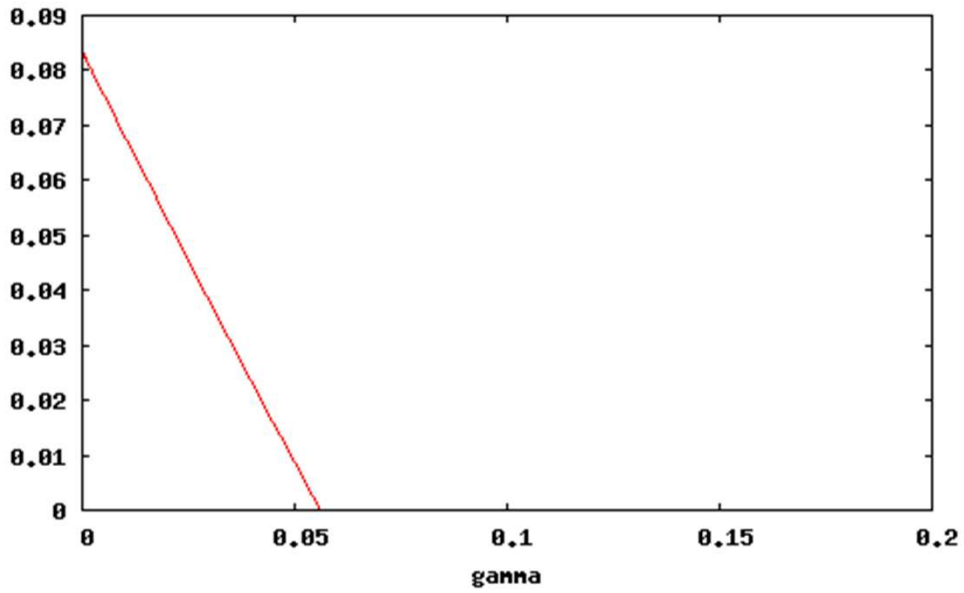
(Note that we extended the y-axis very much.)

If there are interactions, the situation changes. For instance, consider the situation where $\gamma = 0.5$.



The blue line corresponding to [1 2 3 4] remains unchanged, but the other two start at a lower point, and do not reach the blue line before it starts increasing again. Hence: the optimal design uses only [1 2 3 4].

Proportion of sequence 1 2 3 3 in the optimal design.

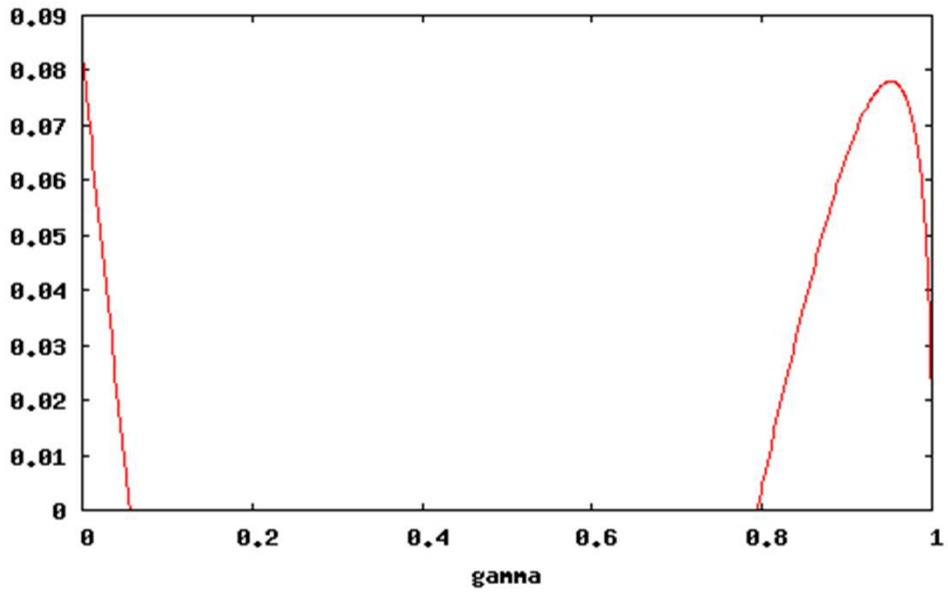


Since this proportion starts at 1/12 for $\gamma = 0$, it has to be nonzero for some small γ.

But it disappears quite fast.

This is exactly what I expected,
when I asked Andrea Bludowsky to work on this model for her thesis.

I was therefore rather surprised, when Andrea showed me the complete picture:



The sequence $1\ 2\ 3\ 3$ reappears for large γ!

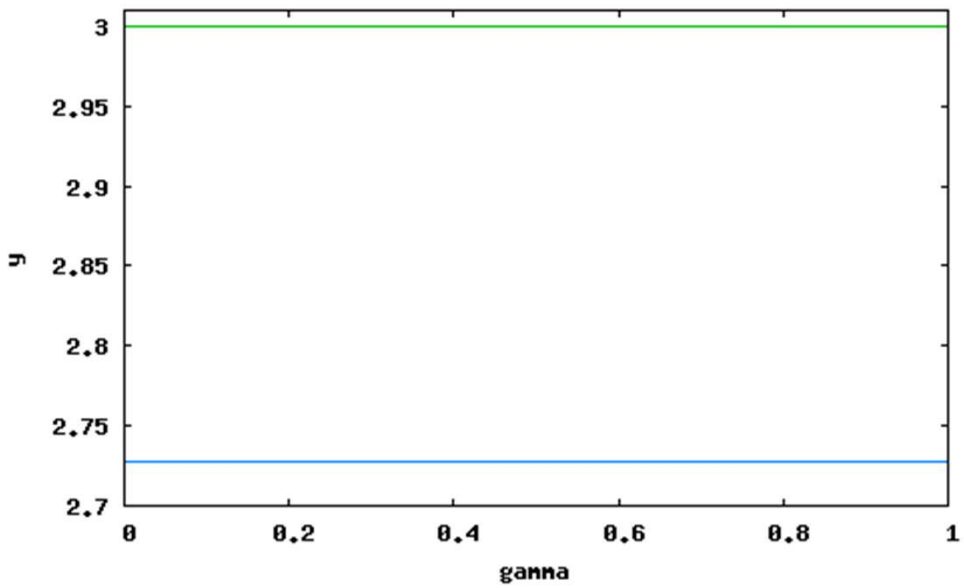
We have considered the situation that $p = 3, 4, 5$ or 6.

For all the p that we considered,
if γ is large, the optimal design uses pairs of identical treatments.

In fact, for larger p, even sequences like $1\ 2\ 3\ 3\ 4\ 4$ appear.

Is there a general stucture?

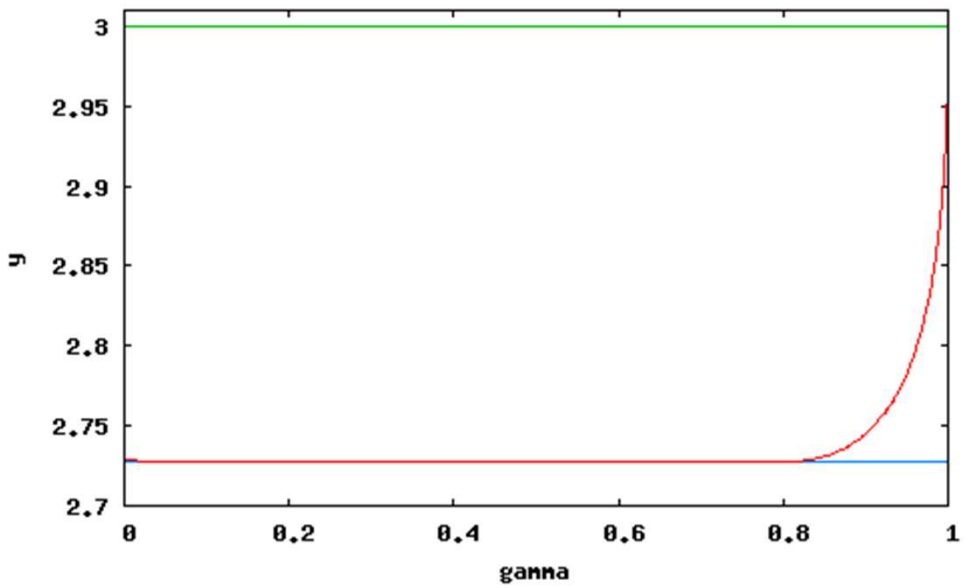
That is where we started at INI in 2011 .



Result 1:

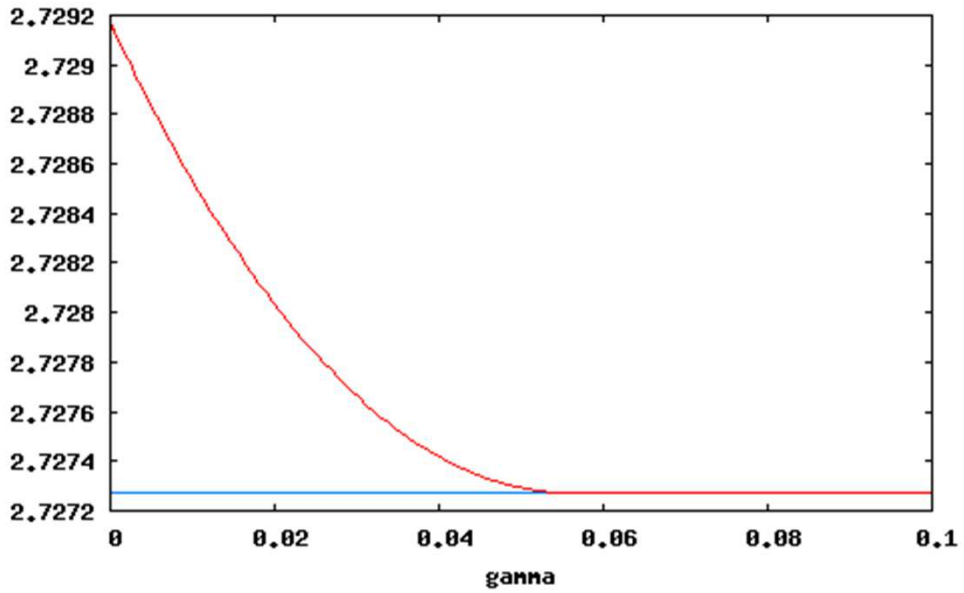
For all p, for all t and for all γ,
any design d must have $trC_d \leq p - 1$.

A neighbour balanced BIBD has $trC_d = (p-1)(1 - \frac{p-1}{p^2-p-\frac{p}{t}})$.

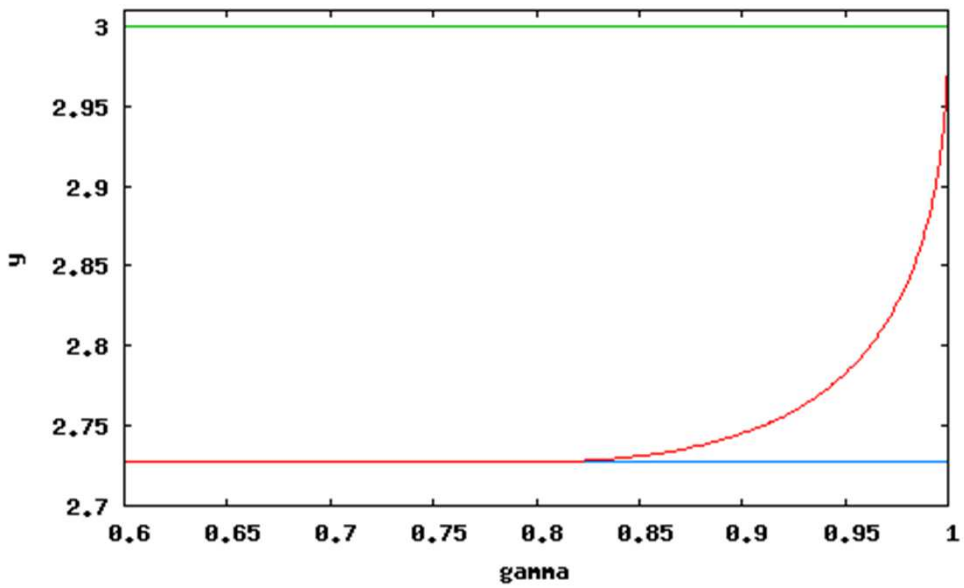


The blue line is the performance of the binary design
the green line is the bound
the red line is the performance of the optimal design

for γ between 0 and 1.



For small γ, the optimal design does not perform much better than the binary design,
the difference can only be seen if we zoom into the graph.



For $\gamma \to 1$, the optimal design performs a lot better,
it gets almost as good as the bound.

How can this be explained?

Observe that any observation has
$Var(y_{dij}) = \sigma^2$, where $\sigma^2 = \sigma_e^2 + \sigma_\xi^2$.

Now, consider a sequence i where treatment 2 appears in periods j and $j+1$.

Then $y_{dij} = \tau_2 + \rho_1 + \beta_i + e_{ij} + \xi_{2i}$
and $y_{dij+1} = \tau_2 + \rho_2 + \beta_i + e_{ij} + \xi_{2i}$.
Therefore, $y_{dij} - y_{dij+1} = \rho_1 - \rho_2$.

The correlation between y_{dij} and y_{dij+1} is γ and, therefore,
$Var(y_{dij} - y_{dij+1}) = \sigma^2(1-\gamma)$.

Therefore, in the limit as $\gamma \to 1$, a small proportion of sequences with identical pairs allows to estimate the carryover effects with variance almost zero.

The other sequences can then be used to estimate the direct effects with the same precision as if there were no carryover effects.

Result 2:

For all p, for all t. If $\gamma \to \infty$,
then for the optimal design d^* we get $trC_{d^*} \to p - 1$.

References:

Kiefer, J. (1975): Construction and optimality of generalized Youden designs. In: A Survey of Statistical Design and Linear Models (J.N. Srivastava, ed.), North Holland, Amsterdam

Kushner, H.B. (1997): Optimal Repeated Measurements Designs: The Linear Optimality Equations. Annals of Statistics 25, 2328 - 2344

Bludowsky, A. (2009): Optimal Crossover Designs with interactions between treatments and units. Thesis, TU Dortmund

Bludowsky, A., Kunert, J., Stufken,J. (2015): Optimal designs for the carryover model with random interactions between subjects and treatments. Australian and New Zealand Journal of Statistics, to appear