2D Linear Model Selection, $f = 1/W$, $N = 100$

Kristyn Pantoja

8/23/2019

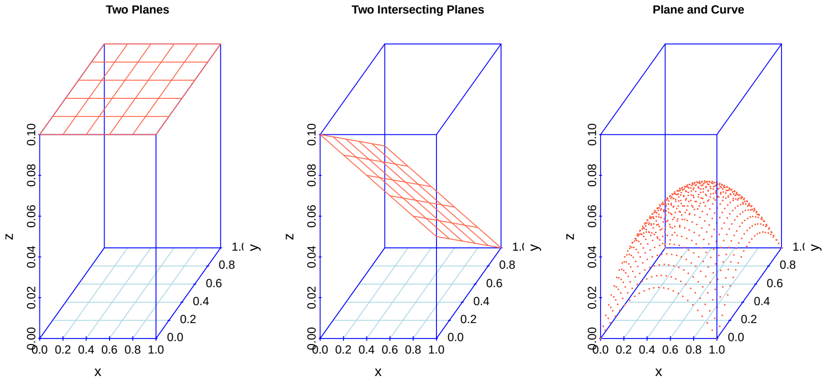
Example 1

Example 2

Example 3

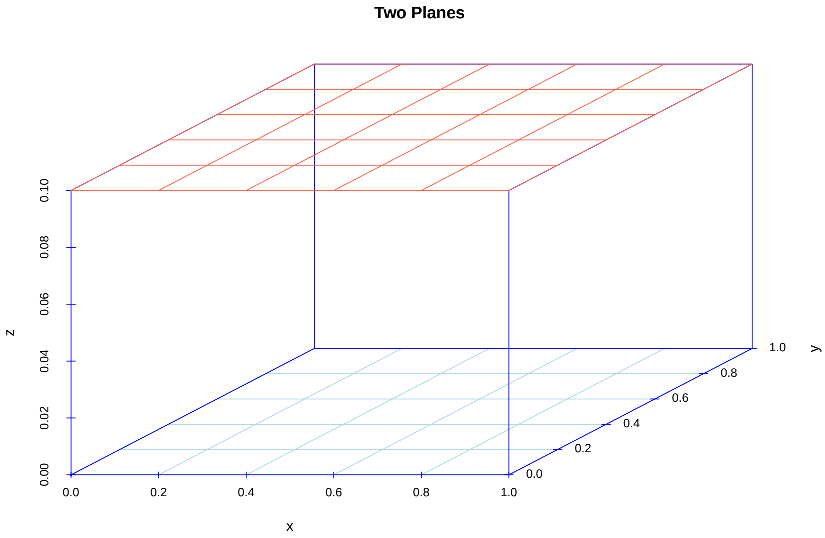
Experiments

Here I consider three examples in which two different models are compared:



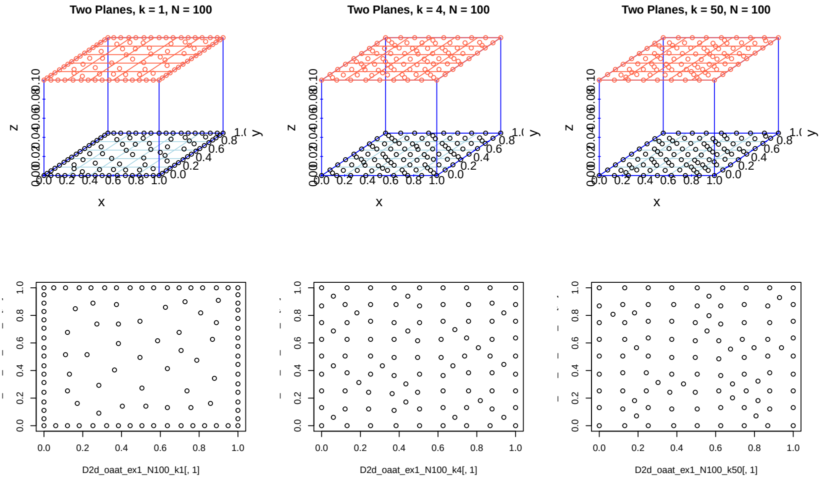
Example 1

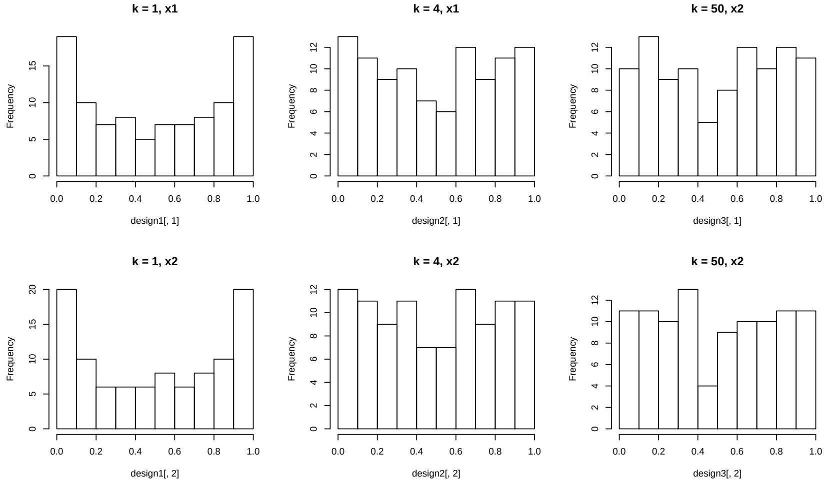
Example 1



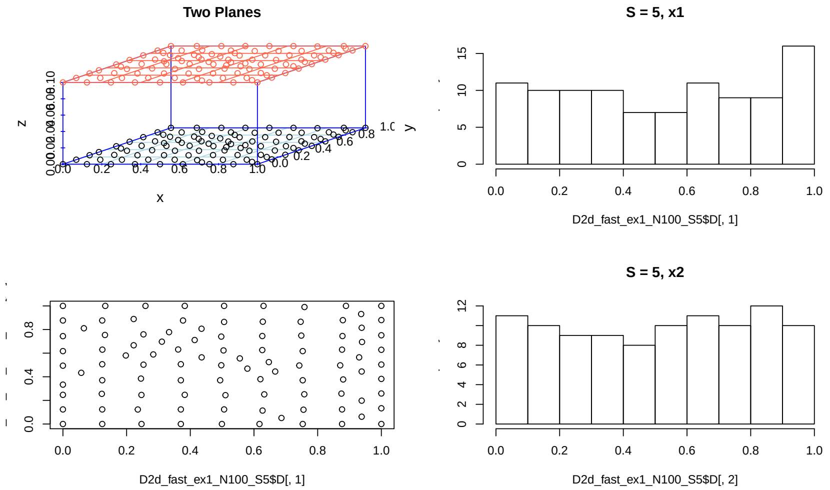
Two Planes
0.00
0.02
0.04
0.06
0.08
0.10
z
0.0
0.2
0.4
0.6
0.8
1.0
x
0.0
0.2
0.4
0.6
0.8
1.0
y

One-at-a-Time Algorithm, $k = 1, 4, 50$ and $N = 100$

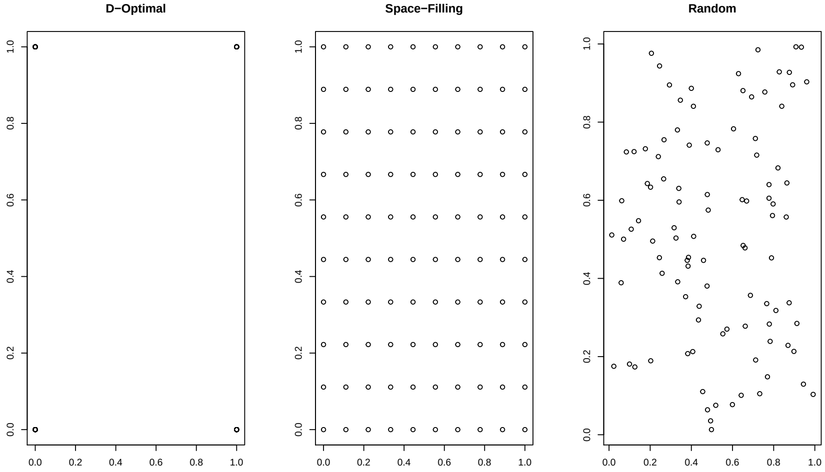


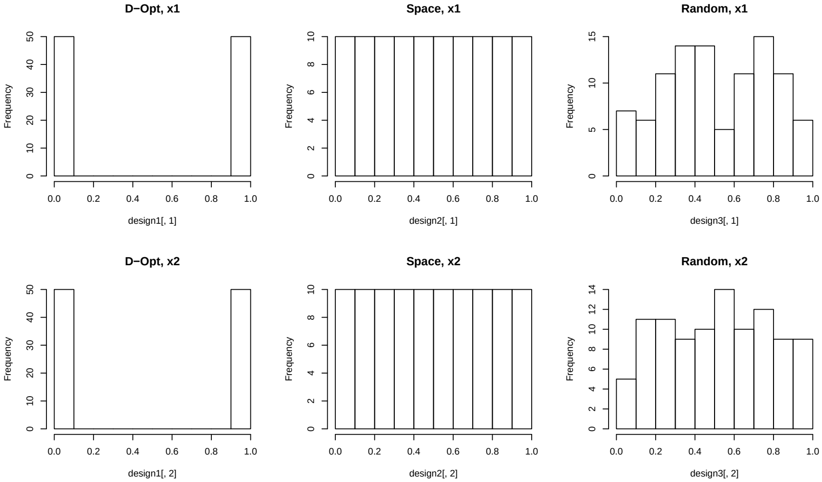

k = 1, x1
Frequency
design1[, 1]
k = 4, x1
Frequency
design2[, 1]
k = 50, x2
Frequency
design3[, 1]
k = 1, x2
Frequency
design1[, 2]
k = 4, x2
Frequency
design2[, 2]
k = 50, x2
Frequency
design3[, 2]

Fast Algorithm, S = 5, N = 100



Other Designs, N = 100



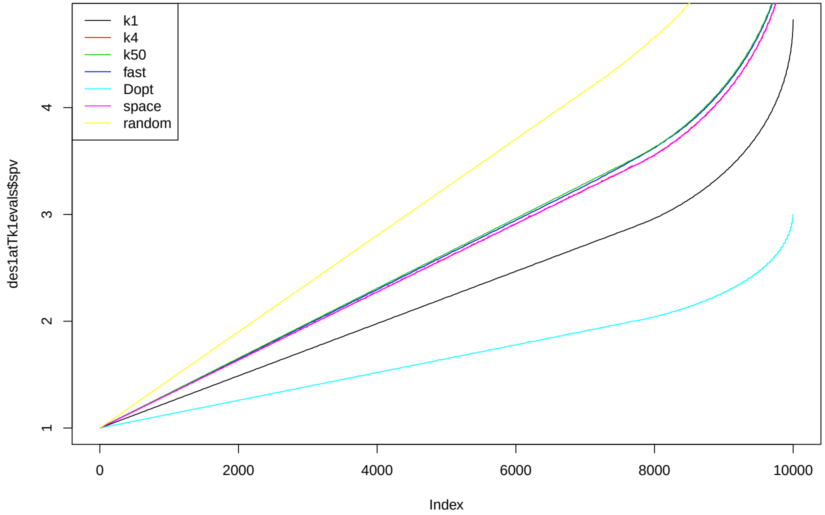

D−Opt, x1
Space, x1
Random, x1
Frequency
design1[, 1]
design2[, 1]
design3[, 1]
D−Opt, x2
Space, x2
Random, x2
design1[, 2]
design2[, 2]
design3[, 2]

Example 1 Evaluations, N = 100

	1atT,k=1	1atT,k=4	1atT,k=50	Fast,S=5	D-Opt	Space	Random
TPEx10e-3	1103.5	1182	1198.8	1196.8	Inf	1173	1531.5
Fast Crit	1980	1166.7	1166.7	1290.1	Inf	900	10266
1atT Crit (k=4)	4908	3834.9	3852.8	3886.5	Inf	3590.8	12548
E[P(H0|Y,D)|H0,D]	0.662	0.662	0.67	0.669	0.679	0.666	0.658
E[P(H1|Y,D)|H0,D]	0.338	0.338	0.33	0.331	0.321	0.334	0.342
E[BF01 | H0,D]	5.34	5.14	5.44	4.96	5.44	5.03	4.52
E[P(H0|Y,D)|H1,D]	0.376	0.385	0.388	0.386	0.387	0.384	0.384
E[P(H1|Y,D)|H1,D]	0.624	0.615	0.612	0.614	0.613	0.616	0.616
E[BF01|H1,D]	1.09	1.07	1.18	1.13	1.19	1.2	1.12
V[B0|Y,X]	0.000791	0.000885	0.000902	0.000928	0.000588	0.000884	0.00106
V[B1|Y,X]	0.0013	0.00152	0.00156	0.0015	0.000803	0.00153	0.00189
V[B2|Y,X]	0.00128	0.00154	0.00156	0.00159	0.000803	0.00153	0.00186

Scaled Prediction Variance (SPV)

$$SPV = NV[\hat{y}(x_0)]/\sigma^2 = Nx_0'(X'X)^{-1}x_0$$

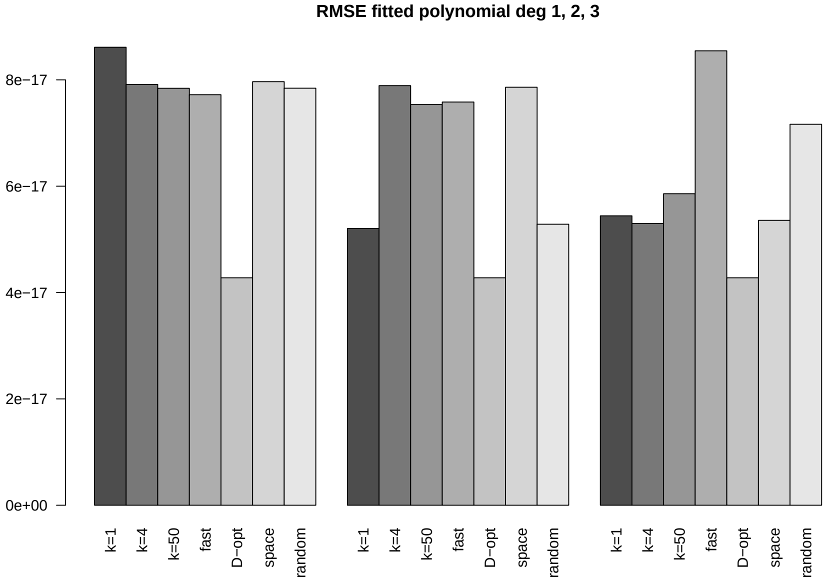


Empirical RMSE

(idea from "Bridging the Gap Between Space-Filling and Optimal Designs Design for Computer Experiments" dissertation by Kathryn Kennedy)

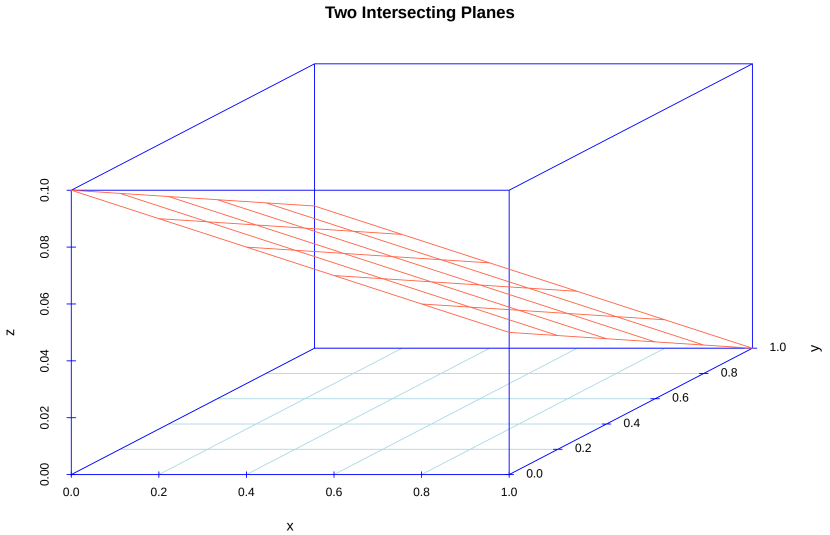
1. Create a hypothetical response variable using a test function (for my examples, I chooose the function given by H_1). (Should there be noise? Or is it supposed to be deterministic, like a computer experiment?)
2. "Analyze" the designs by fitting a polynomial (with no interaction terms). (A GP may also be used.)
3. Predict response variables for 10,000 test points (I chose them on a grid, but in her paper, she randomly chooses them from a uniform distribution) using the model created in step 2.
4. Calculate the residual error as the difference from the values determined by the test function.

Empirical RMSE

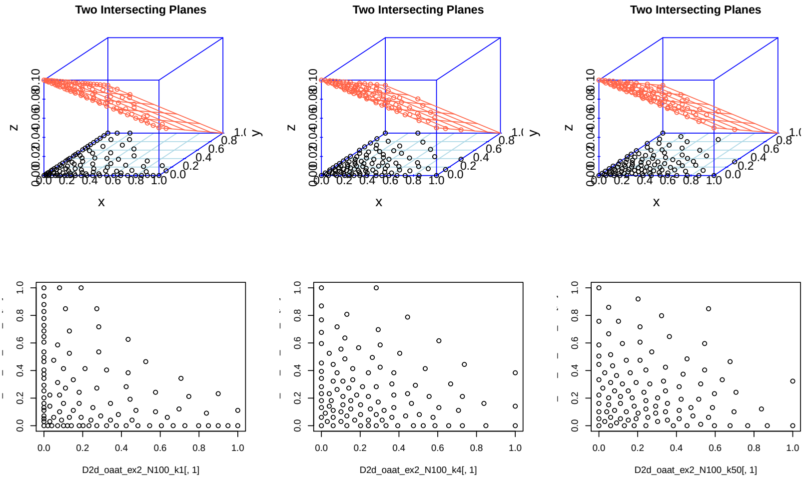


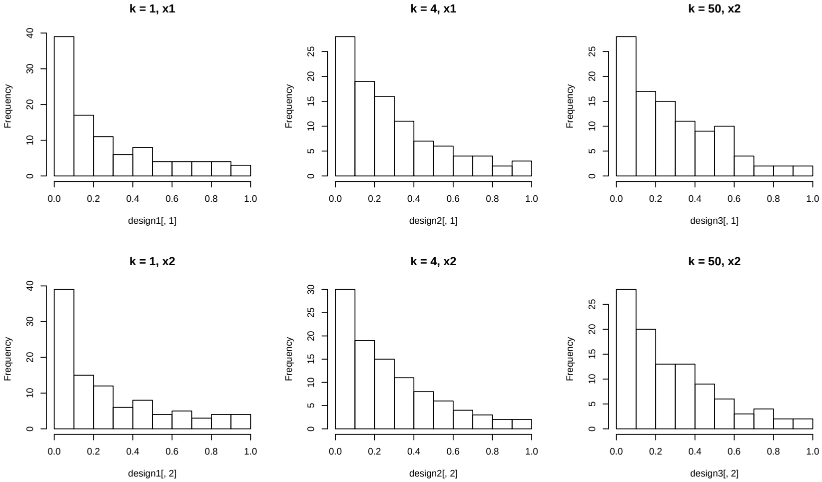
Example 2

Example 2

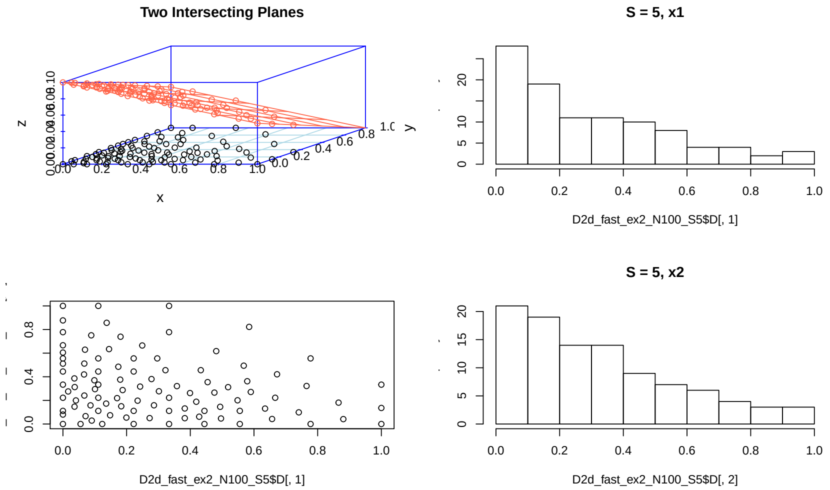

Two Intersecting Planes
0.00
0.02
0.04
0.06
0.08
0.10
z
0.0
0.2
0.4
0.6
0.8
1.0
x
0.0
0.2
0.4
0.6
0.8
1.0
y

One-at-a-Time Algorithm, $k = 1, 4, 50$ and $N = 100$




k = 1, x1
k = 4, x1
k = 50, x2
k = 1, x2
k = 4, x2
k = 50, x2
Frequency
design1[, 1]
design2[, 1]
design3[, 1]
design1[, 2]
design2[, 2]
design3[, 2]

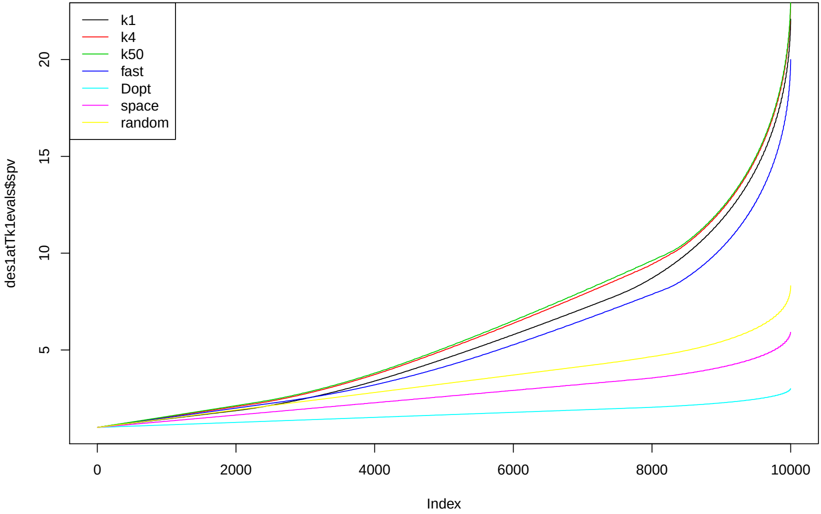
Fast Algorithm, S = 5, N = 100



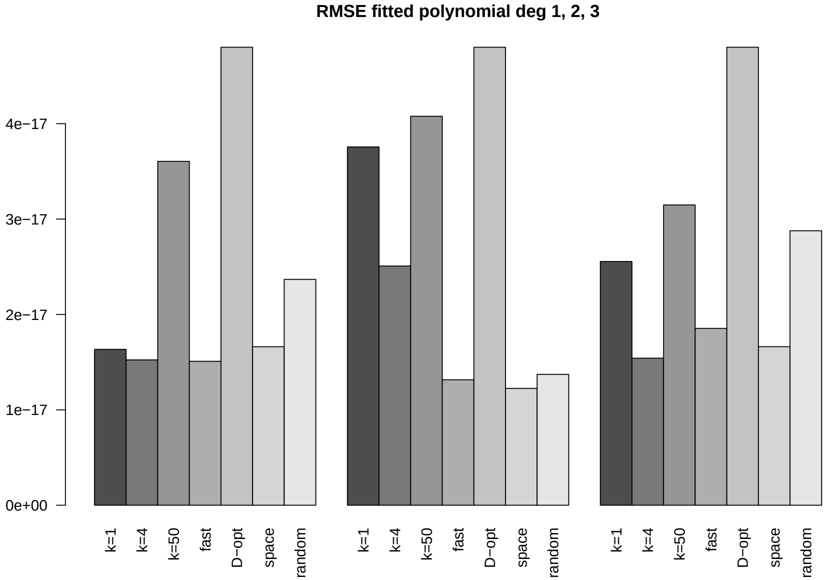
Example 2 Evaluations, N = 100

	1atT,k=1	1atT,k=4	1atT,k=50	Fast,S=5	D-Opt	Space	Random
TPEx10e-3	3093.5	3312.5	3389.4	3487	Inf	Inf	16896
Fast Crit	7539.2	3521	3291.2	4308.3	Inf	Inf	2072500
1atT Crit (k=4)	14212	10568	10636	11523	Inf	Inf	2074600
E[P(H0|Y,D)|H0,D]	0.671	0.675	0.652	0.689	0.715	0.679	0.655
E[P(H1|Y,D)|H0,D]	0.329	0.325	0.348	0.311	0.285	0.321	0.345
E[BF01 | H0,D]	5.24	5.3	5.6	5.35	8.94	5.74	4.4
E[P(H0|Y,D)|H1,D]	0.386	0.388	0.374	0.407	0.376	0.393	0.395
E[P(H1|Y,D)|H1,D]	0.614	0.612	0.626	0.593	0.624	0.607	0.605
E[BF01|H1,D]	1.24	1.5	1.39	1.29	1.72	1.49	1.42
V[B0|Y,X]	0.000525	0.000556	0.000568	0.000634	0.000588	0.000884	0.00106
V[B1|Y,X]	0.00197	0.00214	0.00224	0.00205	0.000803	0.00153	0.00189
V[B2|Y,X]	0.00192	0.00223	0.00223	0.00209	0.000803	0.00153	0.00186

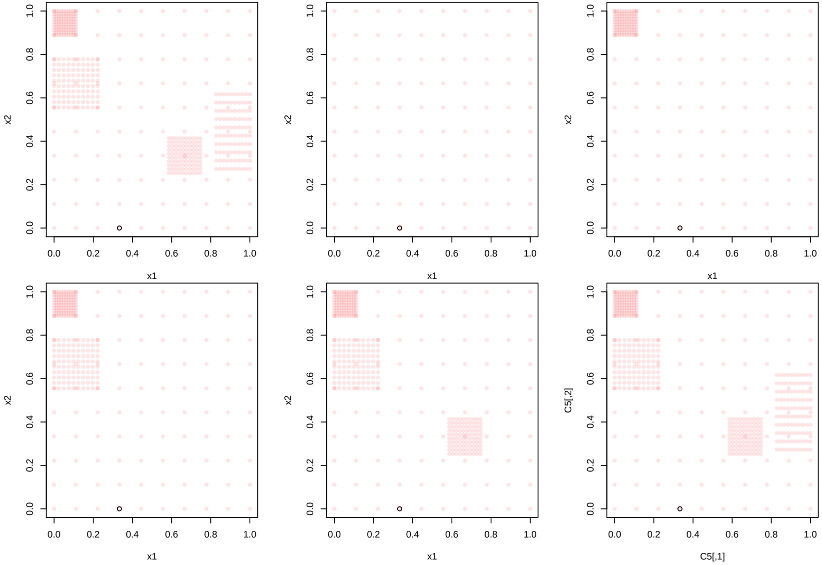
Scaled Prediction Variance (SPV)



Empirical RMSE

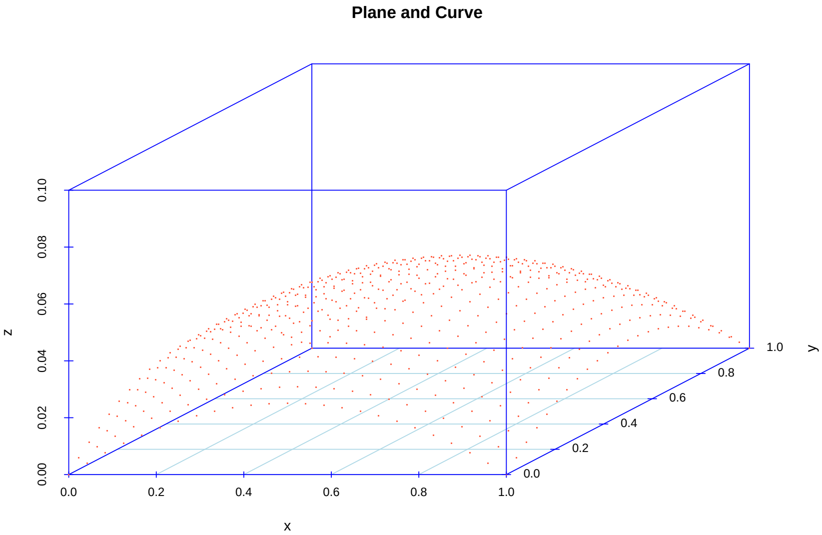


Candidates for Design Point indexed at 10



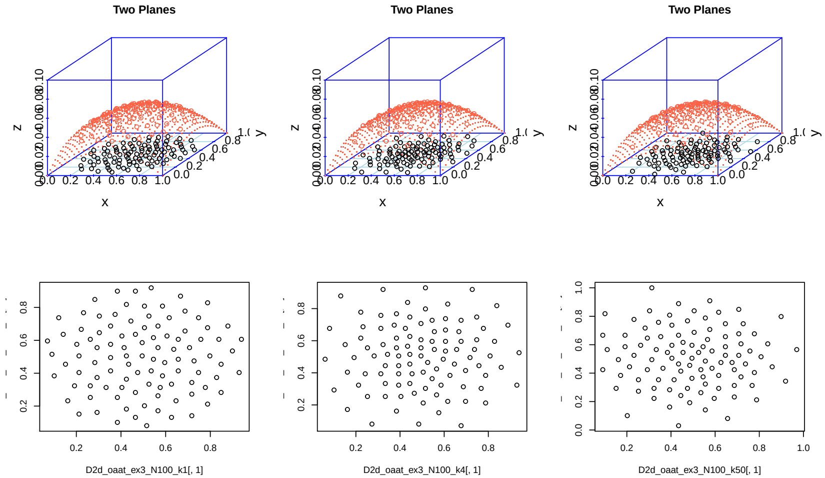
Example 3

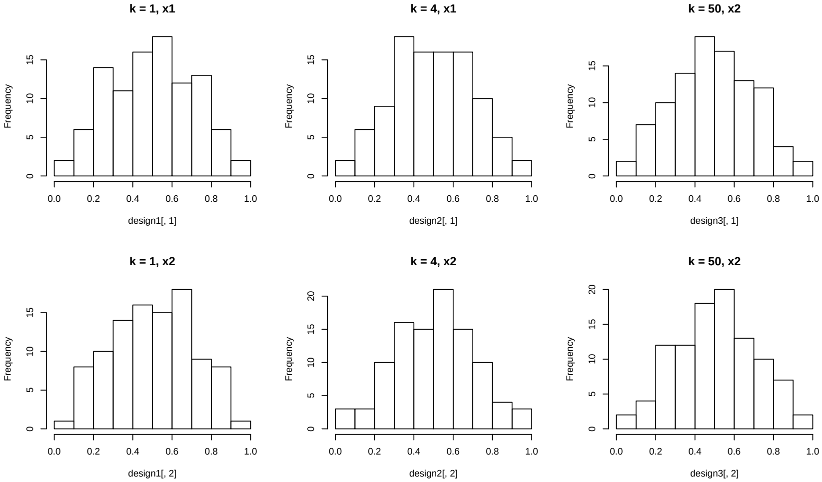
Example 3



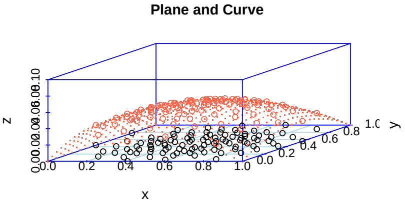
Plane and Curve
0.10
0.08
0.06
0.04
0.02
0.00
z
0.0
0.2
0.4
0.6
0.8
1.0
x
0.0
0.2
0.4
0.6
0.8
1.0
y

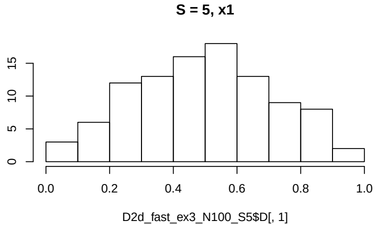
One-at-a-Time Algorithm, k = 1, 4, 50 and N = 100

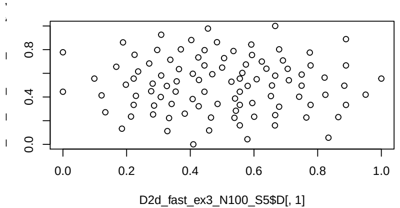


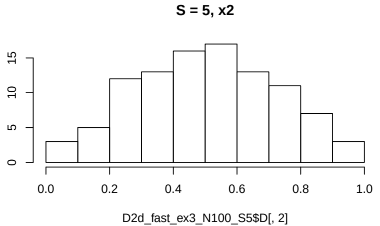

k = 1, x1
k = 4, x1
k = 50, x2
Frequency
design1[, 1]
design2[, 1]
design3[, 1]
k = 1, x2
k = 4, x2
k = 50, x2
design1[, 2]
design2[, 2]
design3[, 2]

Fast Algorithm, S = 5, N = 100









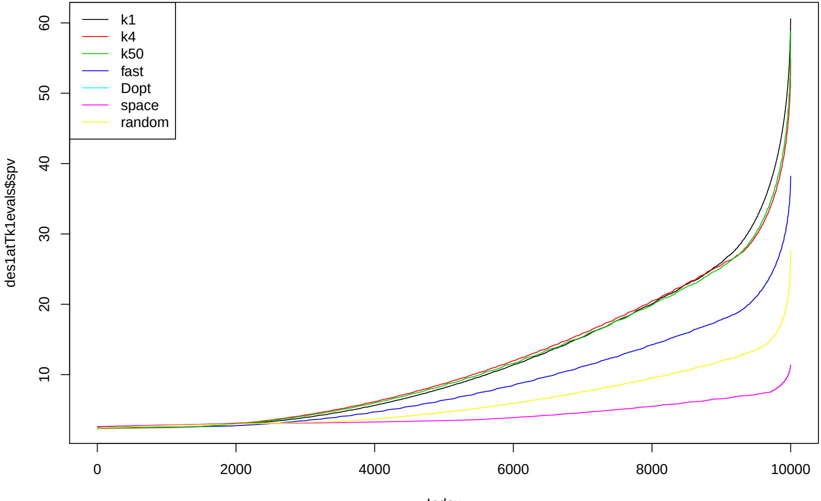
Example 3 Evaluations, N = 100

- D-optimal design tends to be worst i.t.o. $E[P(H_\ell|Y,D)|H_\ell,D], \ell = \{0,1\}$

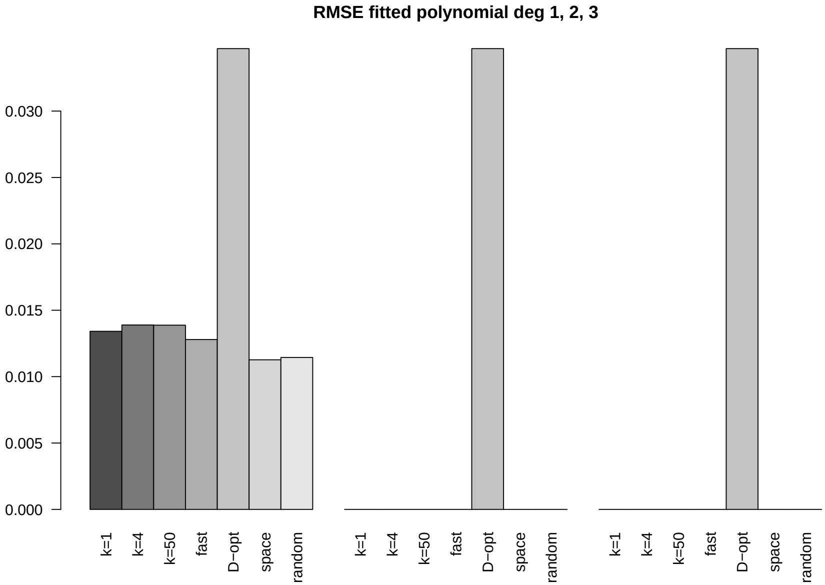
	1atT,k=1	1atT,k=4	1atT,k=50	Fast,S=5	D-Opt	Space	Random
TPEx10e-3	10376	10543	10536	10651	Inf	Inf	15060
Fast Crit	12398	9650.2	9311.2	11024	Inf	Inf	602100
1atT Crit (k=4)	32066	30192	30506	31698	Inf	Inf	602990
E[P(H0|Y,D)|H0,D]	0.568	0.568	0.572	0.581	0.529	0.595	0.581
E[P(H1|Y,D)|H0,D]	0.432	0.432	0.428	0.419	0.471	0.405	0.419
E[BF01 | H0,D]	1.74	1.76	1.81	1.86	1.23	2.1	2.12
E[P(H0|Y,D)|H1,D]	0.421	0.42	0.413	0.428	0.457	0.388	0.409
E[P(H1|Y,D)|H1,D]	0.579	0.58	0.587	0.572	0.543	0.612	0.591
E[BF01|H1,D]	0.922	0.849	0.884	0.967	1	0.951	1.04
V[B0|Y,X]	0.0012	0.00124	0.00125	0.00117	0.000618	0.000885	0.00106
V[B1|Y,X]	0.00313	0.00317	0.00317	0.0031	0.00272	0.00277	0.00295
V[B2|Y,X]	0.00326	0.00333	0.00334	0.00318	0.00272	0.00273	0.00297
V[B3|Y,X]	0.00315	0.00319	0.00318	0.00308	0.00272	0.00277	0.00294
V[B4|Y,X]	0.00332	0.00333	0.00327	0.00316	0.00272	0.00273	0.0029

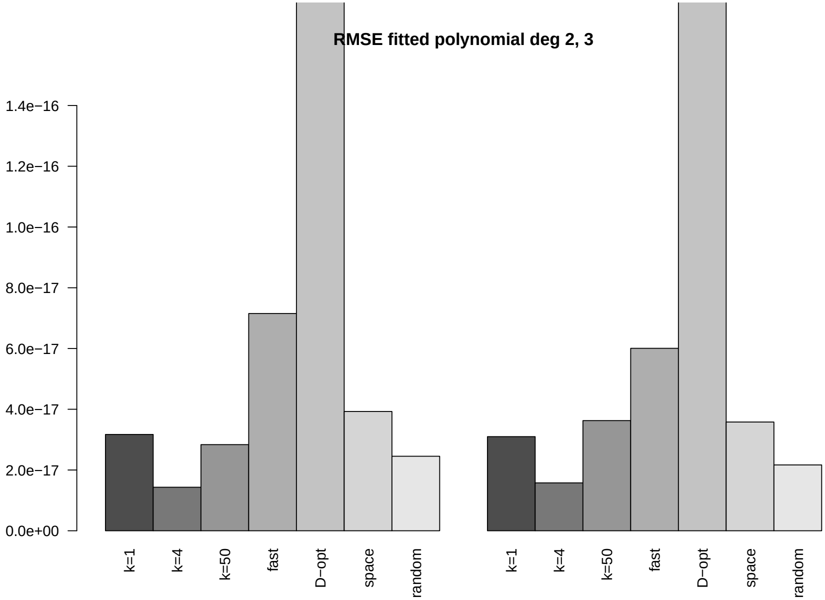
Scaled Prediction Variance (SPV)

Note: For the D-optimal design, repeating columns in the second order model causes $X'X$ to be singular, hence SPV $=$ `NA`. This is an example of when replication in the D-optimal design can cause isssues.



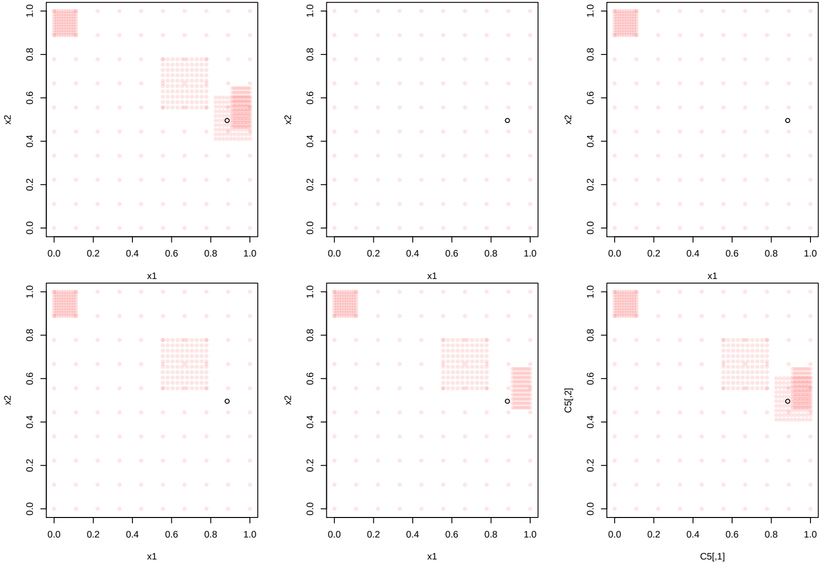
Empirical RMSE






```
## [1] "RMSE for D-opt deg2 and deg3 resp: 0.0347043064717696 ,  0.0347043064717696"
```

Candidates for Design Point indexed at 10

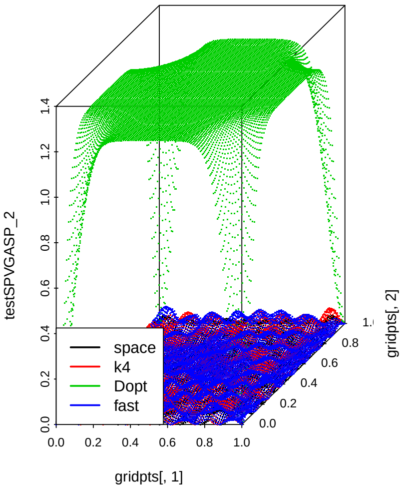


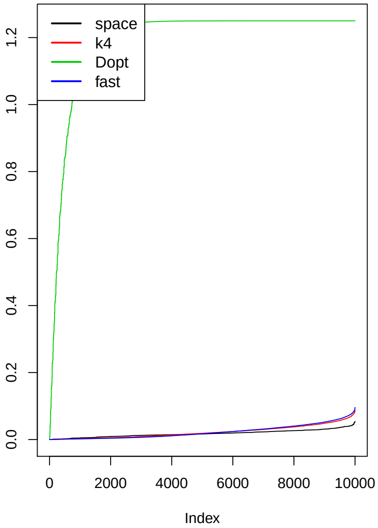
SPV for GP

"For the GP model, the relative prediction variance is dependent on the design points (and hence implicitly, the sample size and number of factors) and the unknown thetas" - Kathryn Kennedy in "Bridging the Gap Between Space-Filling and Optimal Designs Design for Computer Experiments" (dissertation)

in which space-filling does better than D-optimal design

Example 1 MEDs





Example 2 MEDs