Validation Report for **adoptr** package

Kevin Kunzmann & Maximilian Pilz

2019-03-31

Contents

Chapter 1

Introduction

1.1 Concept

The goal of adoptrValidation is to provide a comprehensive suit of test for the adoptr package. The package is not directly inteded to be used but to automatically deploy a weekly validation report via github pages to https://kkmann.github.io/adoptrValidation/. The report is implemented as a set of vignettes which are compiled into a static web page using pkgdown. For details on the class of supported designs, see https://github.com/kkmann/adoptr.

1.2 Local validation

...

1.3 Brief Introdcution to Two-Stage Designs

In `adoptrValidation` a suitable set of cases is tested in order to validate the performance of the package `adoptr`. This package allows to compute optimal designs (adaptive two-stage, group-sequential two-stage and one-stage) for normally distributed data. For a treatment group T and a control group C where the observations $X_i^T \sim \mathcal{N}(\mu_T, \sigma^2)$, $X_i^C \sim \mathcal{N}(\mu_C, \sigma^2)$ the following hypotheses are tested:

$$\mathcal{H}_0 : \delta := \mu_T - \mu_C \leq 0 \text{ v.s. } \mathcal{H}_\infty : \delta > 0.$$

The power of a test procedure is computed on an alternative effect size $\delta_1 > 0$ where a prior distribution $\delta_1 \sim \pi(\vartheta, \tau^2)$ is imaginable.

The trial evaluation happens as follows. After n_1 patients (per group) finished the trial an interim analysis is conducted. The interim test statistic Z_1 for a standard z-test is computed and the trial is stopped early for futility, if $Z_1 < c_f$. If $Z_1 > c_e$ the null hypothesis is rejected and the trial is stopped early for efficacy. Otherwise, i.e. if $c_f \leq Z_1 \leq c_e$, the trial enters in the second stage. Due to the adaptivness of the trial design, the stage-two sample size is a function of Z_1, i.e. $n_2(Z_1)$. Also the final rejection boundary c_2 depends on Z_1. At the final analysis the stage-two test statistic Z_2 is computed and the null hypothesis is rejected if $Z_2 > c_2(Z_1)$.

A design D is a five-tuple consisting of the first-stage sample size n_1, early stopping boundaries c_f (futility) and c_e (efficacy) and stage-two functions $n_2(\cdot)$ (sample size) and $c_2(\cdot)$ (rejection boundary). All these elements can be computed optimally in `adoptr`. The incorporation of continuous priors is possible as well as including conditional and unconditional constraints.

Given a design D and a objective function f the default setting in [adoptr] is the following.

min	$f(D)$
such that	Type One Error Rate $\leq \alpha$
and	Power $\geq 1 - \beta$

Often in clinical practice one is not willing to enter in a second stage when the conditional power (i.e., the probability to reject at the final analysis given the first-stage results) is too low or too high because in these cases the stage-two result is likely predictable. Therefore, introducing conditional power constraints of the form

$$1 - \beta_2 \leq \text{Conditional Power}(z_1, D) \leq 1 - \beta_3$$

may be desirable and are supported by `adoptr`.

In `adoptrValidation` different scenarios are investigated. Each scenario is determined by the assumed effect size δ_1 and its prior distribution π. In each scenario, different tests are performed. All tests are indicated by a bullet point.

1.4 Validation strategy

adoptrValidation essentially extends the test suit of **adoptr** to cover more different scenarios. In order to generate a proper validation report the test Variants are not managed using a unit testing framework like testthat but are directly included in a set of vignettes (one per sceanrio). These vignettes are automatically built and published (here) once per week using pkgdown to keep the validation report up to date with the latest CRAN release [TODO: we currently use our master!]. The overall failure/pass status of the latest build can be checked using the Travis-CI badge. In the following, all Scenarios and their respective sub-Variants are outlined. **Scenarios** are defined by the joint distribution of the test statistic and the location parameter, while **Variants** are given by the respective optimization problem (objective, constraints).

1.4.1 Technical Setup

Initially, the both packages are loaded and the seed for simulation is set. Additionally, the options for optimization are modified by increasing the maximum number of evaluations to ensure convergence.

```
library(adoptr)
library(tidyverse)

# load custom functions in folder subfolder '/R'
for (nm in list.files("R", pattern = "\\.[RrSsQq]$"))
   source(file.path("R", nm))

# define seed value
seed <- 42

# define custom tolerance and iteration limit for nloptr
opts = list(
    algorithm = "NLOPT_LN_COBYLA",
    xtol_rel  = 1e-5,
```

```
    maxeval    = 50000
)
```

1.4.2 Scenario I

This is the default scenario.

- **Data distribution:** Two-armed trial with normally distributed test statistic
- **Prior:** $\delta \sim \delta_{0.4}$
- **Null hypothesis:** $\mathcal{H}_0 : \delta \leq 0$

1.4.2.1 Variant I.1: Minimizing Expected Sample Size under the Alternative

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1)\,|\,\delta = 0.4\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.4\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.0\big] \leq 0.025$
  3. Three variants: two-stage, group-sequential, one-stage.
- **Formal tests:**
  1. All three **adoptr** variants (two-stage, group-sequential, one-stage) comply with constraints. Internally validated by testing vs. simulated values of the power curve at respective points.
  2. ESS of optimal two-stage design is lower than ESS of optimal group-sequential one and that is in turn lower than the one of the optimal one-stage design.
  3. ESS of optimal group-sequential design is lower than ESS of externally computed group-sequential design using the rpact package.
  4. Are the ESS values obtained from simulation the same as the ones obtained by using numerical integration via `adoptr::evaluate`?
  5. Is $n()$ of the optimal two-stage design monotonously decreasing on continuation area?

1.4.2.2 Variant I.2: Minimizing Expected Sample Size under the Null Hypothesis

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1)\,|\,\delta = 0.0\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.4\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.0\big] \leq 0.025$
- **Formal tests:**
  1. Validate constraint compliance by testing vs. simulated values of the power curve at respective points.
  2. $n()$ of optimal design is monotonously increasing on continuation area. TODO
  3. ESS of optimal two-stage design is lower than ESS of externally computed group-sequential design using the rpact package.
  4. Are the ESS values obtained from simulation the same as the ones obtained by using numerical integration via `adoptr::evaluate`?

1.4.2.3 Variant I.3: Condtional Power Constraint

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1)\,|\,\delta = 0.4\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.4\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.0\big] \leq 0.025$

3. $CP := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.4, X_1 = x_1\big] \geq 0.7$ for all $x_1 \in (c_1^f, c_1^e)$

- **Formal tests:**
  1. Check $Power$ and $TOER$ constraints with simulation. Check CP constraint on three different values of x_1 in (c_1^f, c_1^e)
  2. Are the CP values at the three test-pivots obtained from simulation the same as the ones obtained by using numerical integration via `adoptr::evaluate`?
  3. Is ESS of optimal two-stage design with CP constraint higher than ESS of optimal two-stage design without this constraint?

1.4.3 Scenario II

Similar in scope to Scenario I, but with a continuous Gaussian prior on δ.

- **Data distribution:** Two-armed trial with normally distributed test statistic
- **Prior:** $\delta \sim \mathcal{N}(0.4, .3)$
- **Null hypothesis:** $\mathcal{H}_0 : \delta \leq 0$

1.4.3.1 Variant II.1: Minimizing Expected Sample Size

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1)\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta > 0.0\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.0\big] \leq 0.025$
  3. Three variants: two-stage, group-sequential, one-stage.
- **Formal tests:**
  1. All designs comply with type one error rate constraints (tested via simulation).
  2. ESS of optimal two-stage design is lower than ESS of optimal group-sequential one and that is in turn lower than the one of the optimal one-stage design.

1.4.3.2 Variant II.2: Minimizing Expected Sample Size under the Null hypothesis

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1) \,|\, \delta \leq 0\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta > 0.0\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.0\big] \leq 0.025$
- **Formal tests:**
  1. Does the design comply with $TOER$ constraint (via simulation)?
  2. Check CP constraint on three different values of x_1 in (c_1^f, c_1^e)
  3. TODO: Is the sample size function monotonously increasing?
  4. Is ESS lower than expected sample size under the null hypothesis for the optimal two stage design from Variant II-1?

1.4.3.3 Variant II.3: Condtional Power Constraint

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1)\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta > 0.0\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.0\big] \leq 0.025$
  3. $CP := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta > 0.0, X_1 = x_1\big] \geq 0.7$ for all $x_1 \in (c_1^f, c_1^e)$
- **Formal tests:**

1. Check $TOER$ constraint with simulation. Check CP constraint on three different values of x_1 in (c_1^f, c_1^e)
2. Is ESS of optimal two-stage design with CP constraint higher than ESS of optimal two-stage design without the constraint?

1.4.4 Scenario III:

- **Data distribution:** Two-armed trial with normally distributed test statistic
- **Prior:** sequence of uniform distributions $\delta \sim \text{Unif}(0.4 - \Delta_i, 0.4 + \Delta_i)$ around 0.4 with $\Delta_i = (3 - i)/10$ for $i = 0 \ldots 3$. I.e., for $\Delta_3 = 0$ reduces to a point prior on $\delta = 0.4$.
- **Null hypothesis:** $\mathcal{H}_0 : \delta \leq 0$

1.4.4.1 Variant III.1: Convergence under Prior Concentration

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1)\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta > 0.0\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.0\big] \leq 0.025$
- **Formal tests:**
  1. Simulated type one error rate is compared to $TOER$ constraint for each design.
  2. Number of iterations are checked agaist default maximum to ensure proper convergence.
  3. TODO: ESS decreases with prior variance.

Additionally, the designs are compared graphically. Inspect the plot to see convergence pattern.

1.4.5 Scenario IV: Smaller effect size, larger trials.

1.4.5.1 Variant IV.1: Minimizing Expected Sample Size under the Alternative

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1) \,|\, \delta = 0.2\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.2\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.0\big] \leq 0.025$
  3. Three variants: two-stage, group-sequential, one-stage.
- **Formal tests:**
  1. All three adoptr variants (two-stage, group-sequential, one-stage) comply with costraints. Internally validated by testing vs. simulated values of the power curve at respective points.
  2. ESS of optimal two-stage design is lower than ESS of optimal group-sequential one and that is in tunr lower than the one of the optimal one-stage design.
  3. ESS of optimal group-sequential design is lower than ESS of externally computed group-sequential design using the rpact package.
  4. Are the ESS values obtained from simulation the same as the ones obtained by using numerical integration via `adoptr::evaluate`?
  5. Is $n()$ of the optimal two-stage design monotonously decreasing on continuation area? TODO

1.4.5.2 Variant IV.2: Increasing Power

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1) \,|\, \delta = 0.2\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.2\big] \geq 0.9$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.0\big] \leq 0.025$

3. Three variants: two-stage, group-sequential, one-stage.

- **Formal tests:**
  1. Does the design respect all constraints (via simulation)?
  2. ESS of optimal two-stage design is lower than ESS of optimal group-sequential one and that is in tunr lower than the one of the optimal one-stage design.
  3. ESS of optimal group-sequential design is lower than ESS of externally computed group-sequential design using the rpact package.
  4. Are the ESS values obtained from simulation the same as the ones obtained by using numerical integration via `adoptr::evaluate`?
  5. Is $n()$ of the optimal two-stage design monotonously decreasing on continuation area? TODO

1.4.5.3 Variant IV.3: Increasing Maximal Type One Error Rate

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1)\,|\,\delta = 0.2\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.2\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.0\big] \leq 0.05$
  3. Three variants: two-stage, group-sequential, one-stage.
- **Formal tests:**
  1. Does the design respect all constraints (via simulation)?
  2. ESS of optimal two-stage design is lower than ESS of optimal group-sequential one and that is in tunr lower than the one of the optimal one-stage design.
  3. ESS of optimal group-sequential design is lower than ESS of externally computed group-sequential design using the rpact package.
  4. Are the ESS values obtained from simulation the same as the ones obtained by using numerical integration via `adoptr::evaluate`?
  5. Is $n()$ of the optimal two-stage design monotonously decreasing on continuation area? TODO

1.4.6 Scenario V: Single-arm design, medium effect size.

- **Data distribution:** One-armed trial with normally distributed test statistic
- **Prior:** $\delta \sim \delta_{0.3}$
- **Null hypothesis:** $\mathcal{H}_0 : \delta \leq 0$

1.4.6.1 Variant V.1: Sensitivity to Integration Order

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1)\,|\,\delta = 0.3\big]$
- **Constraints:**
  1. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.3\big] \geq 0.8$
  2. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.0\big] \leq 0.025$
  3. Three variants: integration order 5, 8, 11 two-stage designs [TODO: maybe more?].
- **Formal tests:**
  1. Do all designs respect all constraints (via simulation)?
  2. Do all designs converge within the respective iteration limit?
  3. Does constraint compliance get better with increased order?
  4. Does the simulated ESS get better with increased order?

1.4.6.2 Variant V.2: Utility Maximization

- **Objective:** $\lambda\, Power - ESS := \lambda\, \boldsymbol{Pr}\big[c_2(X_1) < X_2\,|\,\delta = 0.3\big] - \boldsymbol{E}\big[n(X_1)\,|\,\delta = 0.3\big]$. for $\lambda = 100$ and 200

- **Constraints:**
  1. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.0\big] \leq 0.025$
- **Formal tests:**
  1. Do both desings respect the type one error rate constraint (via simulation)?
  2. Is the power of the design with larger λ larger?

1.4.6.3 Variant V.3: n_1 penalty

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1) \,|\, \delta = 0.3\big] + \lambda\, n_1$ for $\lambda = 0.05$ and 0.2.
- **Constraints:**
  1. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.0\big] \leq 0.025$
  2. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.3\big] \geq 0.8$
- **Formal tests:**
  1. Is n_1 for the optimal design smaller than the order-5 design in V.1?

1.4.6.4 Variant V.4: n_2 penalty

- **Objective:** $ESS := \boldsymbol{E}\big[n(X_1) \,|\, \delta = 0.3\big] +$ `AverageN2`
- **Constraints:**
  1. $TOER := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.0\big] \leq 0.025$
  2. $Power := \boldsymbol{Pr}\big[c_2(X_1) < X_2 \,|\, \delta = 0.3\big] \geq 0.8$
- **Formal tests:**
  1. Is the `AverageN2` for the optimal design smaller than for the order-5 design in V.1?

Chapter 2

Scenario I: large effect, point alternative

2.1 Details

In this scenario an alternative effect size of $\delta = 0.4$ with point prior distribution is investigated. The null hypothesis is $\delta \leq 0$. Currently, `adoptr` only supports normal distributed data what is widely spread in the development of adaptive designs. We protect the one-sided type one error rate at $\alpha = 0.025$ and require the power of the design to be at least $1 - \beta = 0.8$.

2.1.1 Data distribution

Two-armed trial with normally distributed test statistic

```
datadist <- Normal(two_armed = TRUE)
```

2.1.2 Null hypothesis

The null hypothesis is $\mathcal{H}_0 : \delta \leq 0$

```
H_0 <- PointMassPrior(.0, 1)
```

2.1.3 Prior assumptions

A point mass prior with probability mass on $\delta = 0.4$ is assumed.

```
prior <- PointMassPrior(.4, 1)
```

2.2 Case I-1: Minimizing Expected Sample Size under Point Prior

2.2.1 Objective

Expected sample size under the respective prior is minimized, i.e., $\boldsymbol{E}\big[n(\mathcal{D})\big]$.

```
ess <- expected(ConditionalSampleSize(datadist, prior))
```

2.2.2 Constrains

The type one error rate is controlled at 0.025 on the boundary of the null hypothesis.

```
toer_cnstr <- expected(ConditionalPower(datadist, H_0)) <= .025
```

Power must be larger than 0.8.

```
pow_cnstr <- expected(ConditionalPower(datadist, prior)) >= .8
```

2.2.3 Initial Design

`adoptr` requires the definition of an initial design for optimization. We start with a group-sequential design from the package `rpact` that fulfills these constraints and is used later for comparison. The order of integration is set to

```
order <- 7L
```

For usage as two-stage design with variable sample size, it has to be converted to a `TwoStageDesign`.

```
init_design_gs <- rpact_design(0.4, 0.025, 0.8, TRUE, order)

init_design    <- TwoStageDesign(init_design_gs)
```

2.2.4 Optimization

The optimal design is computed in three variants: two-stage, group-sequential and one-stage. The input only differs with regard to the initial design.

```
opt_design <- function(initial_design) {
    minimize(

        ess,

        subject_to(

            toer_cnstr,
            pow_cnstr

        ),

        initial_design = initial_design,

        opts = opts

    )
}

opt1_ts <- opt_design(init_design)
opt1_gs <- opt_design(init_design_gs)
opt1_os <- opt_design(OneStageDesign(200, 2.0))
```

2.2.5 Test Cases

Check if the optimization algorithm converged in all cases.

```
iters <- sapply(list(opt1_ts, opt1_gs, opt1_os),
                function(x) x$nloptr_return$iterations)

print(iters)
```

```
## [1] 3402  985   24
```

```
testthat::expect_true(all(iters < opts$maxeval))
```

The n_2 function of the optimal two-stage design is expected to be monotonously decreasing.

```
testthat::expect_equal(
    sign(diff(opt1_ts$design@n2_pivots)),
    rep(-1, (order - 1))
)
```

Type one error rate constraint is tested for the three designs. Due to numerical issues we allow a realtive error of 2%.

```
tmp     <- sapply(list(opt1_ts, opt1_gs, opt1_os),
                  function(x) sim_pr_reject(x$design, .0, datadist))
df_toer <- data.frame(
    toer = as.numeric(tmp[1, ]),
    se   = as.numeric(tmp[2, ])
)
rm(tmp)

testthat::expect_true(all(df_toer$toer <= .025*(1.02)))

df_toer
```

```
##       toer           se
## 1 0.024951 0.0001559759
## 2 0.024978 0.0001560581
## 3 0.025116 0.0001564775
```

The power constraint can also be tested via simulation. Due to numerical issues we allow a realtive error of 2%.

```
tmp     <- sapply(list(opt1_ts, opt1_gs, opt1_os),
                  function(x) sim_pr_reject(x$design, .4, datadist))
df_pow <- data.frame(
    pow  = as.numeric(tmp[1, ]),
    se   = as.numeric(tmp[2, ])
)
rm(tmp)

testthat::expect_true(all(df_pow$pow >= .8 * (1 - 0.02)))

df_pow
```

```
##        pow           se
## 1 0.798641 0.0004010159
## 2 0.799669 0.0004002482
```

```
## 3 0.799317 0.0004005115
```

The expected sample sizes should be ordered in a specific way.

```
testthat::expect_gte(
    evaluate(ess, opt1_os$design),
    evaluate(ess, opt1_gs$design)
)

testthat::expect_gte(
    evaluate(ess, init_design_gs),
    evaluate(ess, opt1_gs$design)
)

testthat::expect_gte(
    evaluate(ess, opt1_gs$design),
    evaluate(ess, opt1_ts$design)
)
```

The expected sample size of the optimal designs is simulated and compared to the outomce of `adoptr::evaluate()`. The tolerance is set to 0.5 what is due to rounding one patient per group in the worst case.

```
ess_0 <- expected(ConditionalSampleSize(datadist, H_0))

testthat::expect_equal(
    sim_n(opt1_os$design, .0, datadist),
    evaluate(ess_0, opt1_os$design),
    tolerance = .5
)

testthat::expect_equal(
    sim_n(opt1_gs$design, .0, datadist),
    evaluate(ess_0, opt1_gs$design),
    tolerance = .5
)

testthat::expect_equal(
    sim_n(opt1_ts$design, .0, datadist),
    evaluate(ess_0, opt1_ts$design),
    tolerance = .5
)
```

Additionally, the sample sizes under the point prior are compared.

```
testthat::expect_equal(
    sim_n(opt1_os$design, .4, datadist),
    evaluate(ess, opt1_os$design),
    tolerance = .5
)

testthat::expect_equal(
    sim_n(opt1_gs$design, .4, datadist),
    evaluate(ess, opt1_gs$design),
    tolerance = .5
)
```

```
testthat::expect_equal(
    sim_n(opt1_ts$design, .4, datadist),
    evaluate(ess, opt1_ts$design),
    tolerance = .5
)
```

2.3 Case I-2: Minimizing Expected Sample Size under Null Hypothesis

2.3.1 Objective

Expected sample size under the null hypothesis prior is minimized, i.e.,

```
ess_0 <- expected(ConditionalSampleSize(datadist, H_0))
```

2.3.2 Constrains

The constraints remain the same as before.

2.3.3 Initial Design

For runtime issues the previous initial design has to be updated. It turns out that a constant c_2-starting value is much more efficient in this case. Furthermore, a more strict upper-boundary design than the default one needs to be defined because stopping for efficacy would otherwise only happen for very large values of x_1 due to optimization under the null hypothesis.

```
init_design_2 <- init_design
init_design_2@c2_pivots <- rep(2, order)


ub_design <- TwoStageDesign(
    opt1_os$design@n1,
    opt1_os$design@c1f,
    3,
    rep(300, order),
    rep(3.0, order)
)
```

2.3.4 Optimization

The optimal two-stage design is computed.

```
opt2_ts <- minimize(

        ess_0,

        subject_to(

            toer_cnstr,
```

```
            pow_cnstr

        ),

        initial_design = init_design_2,

        upper_boundary_design = ub_design,

        opts = opts

)
```

```
## Warning in minimize(ess_0, subject_to(toer_cnstr, pow_cnstr),
## initial_design = init_design_2, : initial design is infeasible!
```

2.3.5 Test Cases

Check if the optimization algorithm converged.

```
print(opt2_ts$nloptr_return$iterations)
```

```
## [1] 20640
```

```
testthat::expect_true(opt2_ts$nloptr_return$iterations < opts$maxeval)
```

The n_2 function of the optimal two-stage design is expected to be monotnously increasing.

```
testthat::expect_equal(
    sign(diff(opt2_ts$design@n2_pivots)),
    rep(1, (order - 1))
)
```

Type one error rate constraint is tested for the optimal design. Due to numerical issues we allow a realtive error of 2%.

```
tmp     <- sim_pr_reject(opt2_ts$design, .0, datadist)
df_toer2 <- data.frame(
    toer = as.numeric(tmp[1]),
    se   = as.numeric(tmp[2])
)
rm(tmp)

testthat::expect_true(all(df_toer2$toer <= .025*(1.02)))

df_toer2
```

```
##       toer           se
## 1 0.024971 0.0001560368
```

The power constraint can also be tested via simulation. Due to numerical issues we allow a realtive error of 2%.

```
tmp     <- sim_pr_reject(opt2_ts$design, .4, datadist)
df_pow2 <- data.frame(
    pow  = as.numeric(tmp[1]),
    se   = as.numeric(tmp[2])
)
```

```
rm(tmp)

testthat::expect_true(all(df_pow2$pow >= .8 * (1 - 0.02)))

df_pow2
```

```
##       pow           se
## 1 0.80175 0.0003986817
```

The expected sample size under the null should be lower than the ess under the null of the initial design derived from `rpact`.

```
testthat::expect_gte(
    evaluate(ess_0, init_design),
    evaluate(ess_0, opt2_ts$design)
)
```

The expected sample size of the optimal designs is simulated and compared to the outomce of `adoptr::evaluate()`. The tolerance is set to 0.5 what is due to rounding one patient per group in the worst case.

```
testthat::expect_equal(
    sim_n(opt2_ts$design, .0, datadist),
    evaluate(ess_0, opt2_ts$design),
    tolerance = .5
)
```

Additionally, the sample sizes under the point prior are compared.

```
testthat::expect_equal(
    sim_n(opt2_ts$design, .4, datadist),
    evaluate(ess, opt2_ts$design),
    tolerance = .5
)
```

2.4 Case I-3: Conditional Power Constraint

2.4.1 Objective

Expected sample size under the point prior is minimized and has already been defined.

2.4.2 Constrains

The constraints remain the same as before, additionally to a constraint on conditional power.

```
cp <- ConditionalPower(datadist, prior)

cp_cnstr <- cp >= .7
```

2.4.3 Initial Design

The previous initial design can still be applied.

2.4.4 Optimization

The optimal two-stage design is computed.

```
opt3_ts <- minimize(

        ess,

        subject_to(

            toer_cnstr,
            pow_cnstr,
            cp_cnstr

        ),

        initial_design = init_design,

        opts = opts

)
```

```
## Warning in minimize(ess, subject_to(toer_cnstr, pow_cnstr, cp_cnstr),
## initial_design = init_design, : initial design is infeasible!
```

2.4.5 Test Cases

Check if the optimization algorithm converged.

```
print(opt3_ts$nloptr_return$iterations)
```

```
## [1] 3316
```

```
testthat::expect_true(opt3_ts$nloptr_return$iterations < opts$maxeval)
```

Type one error rate constraint is tested for the optimal design. Due to numerical issues we allow a realtive error of 2%.

```
tmp     <- sim_pr_reject(opt3_ts$design, .0, datadist)
df_toer3 <- data.frame(
    toer = as.numeric(tmp[1]),
    se   = as.numeric(tmp[2])
)
rm(tmp)

testthat::expect_true(all(df_toer3$toer <= .025*(1.02)))

df_toer3
```

```
##      toer           se
## 1 0.02496 0.0001560033
```

The power constraint can also be tested via simulation. Due to numerical issues we allow a realtive error of 2%.

```
tmp     <- sim_pr_reject(opt3_ts$design, .4, datadist)
df_pow3 <- data.frame(
```

```
    pow  = as.numeric(tmp[1]),
    se   = as.numeric(tmp[2])
)
rm(tmp)

testthat::expect_true(all(df_pow3$pow >= .8 * (1 - 0.02)))

df_pow3
```

```
##        pow           se
## 1 0.798916 0.0004008109
```

The expected sample size under the prior should be higher than in the case without the constraint that was analyzed in I.1.

```
testthat::expect_gte(
    evaluate(ess, opt3_ts$design),
    evaluate(ess, opt1_ts$design)
)
```

The conditional power constraint needs to be tested. Select three points for this and check the constraint.

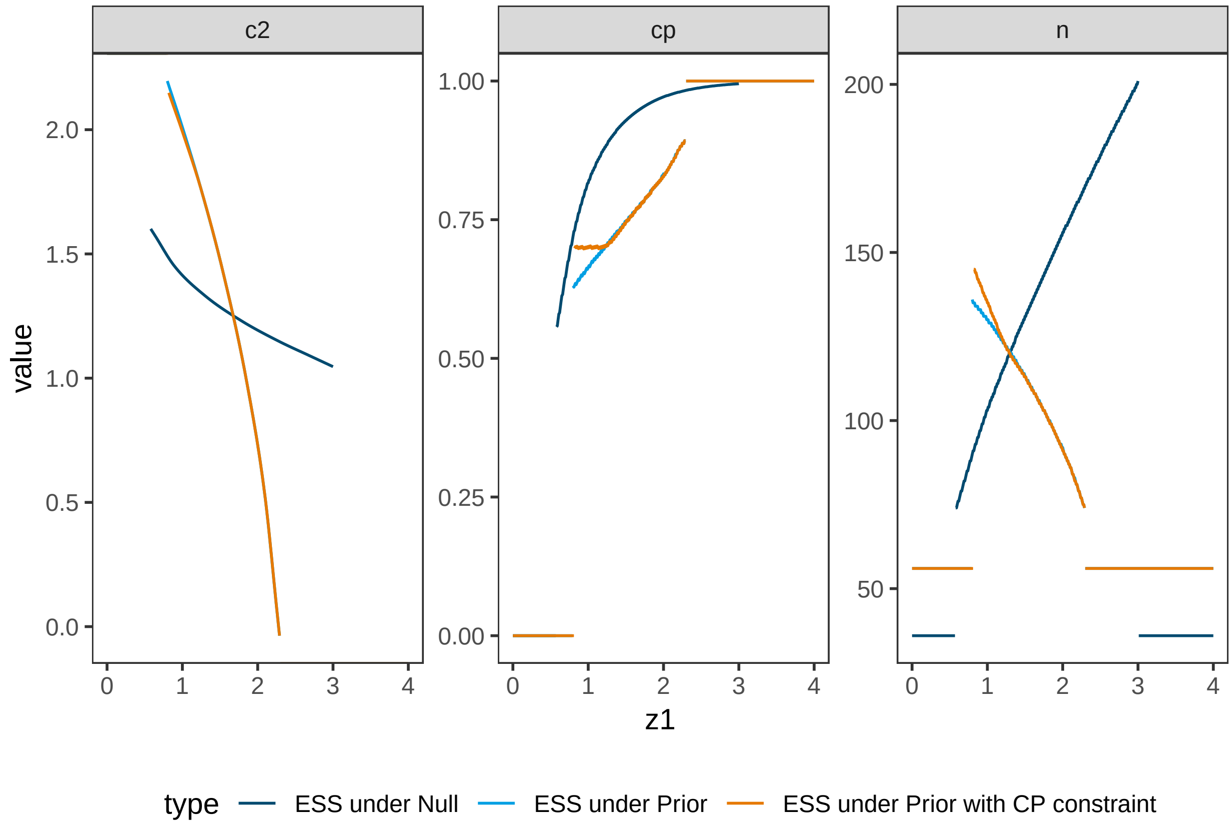
```
x <- adoptr:::scaled_integration_pivots(opt3_ts$design)[c(1, 3, 5)]

cp_val <- sapply(x, function(z) evaluate(cp, opt3_ts$design, z))

testthat::expect_true(all(cp_val >= 0.7))
```

2.5 Plot Two-Stage Designs

The optimal two-stage designs stemming from the different variants are plotted together.



c2
cp
n
value
z1
2.0
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.00
200
150
100
50
0
1
2
3
4
type
ESS under Null
ESS under Prior
ESS under Prior with CP constraint