Bayesian vs. frequentist sample sizes for multi-arm studies

Philip Pallmann

November 2, 2015

In this vignette we compare the Bayesian sample sizes calculated using the package `BayesMAMS` with sample sizes calculated under the frequentist paradigm. Similar comparisons are discussed in section 3 of Whitehead et al. (2015).

We consider a scenario where $k = 2$ experimental treatments are to be compared with a common control group. The allocation ratio is $\sqrt{k}$ to 1 in favour of control. For simplicity, we choose an anticipated precision of $\nu = 1$, which translates to a variance of $\sigma^2 = 1$ and also a standard deviation of $\sigma = 1$. The precision is assumed to be known.

Criterion 1

The posterior probability of one ore more experimental treatments being better than control is at least η, or else the posterior probability of none of the treatments being better than control (by a relevant margin δ^*) is at least ζ.

Bayesian

For the Bayesian sample size calculation subject to criterion 1, we define the relevant treatment effect as $\delta^* = 0.5$ and set the probabilities $\eta = 0.95$ and $\zeta = 0.90$. Further we assume a prior precision of $q_0 = 0$ for all groups i.e., no prior information about ν.

```
library("BayesMAMS")
ssbayes(k=2, nu=1, q0=c(0, 0, 0), eta=0.95, zeta=0.90, deltastar=0.5, prec="known",
        crit="1")
```

```
##
## Control: 102
## Group A:  72
## Group B:  72
```

Frequentist

For the frequentist sample size calculation, we choose the common type I error rate of $\alpha = 0.05$ and a desired power of $1 - \beta = 0.90$.

```
k <- 2
alloc <- sqrt(k)
nu <- 1
deltastar <- 0.5
alpha <- 0.05
power <- 0.90
```

Using a Bonferroni adjustment for the multiplicity of comparisons, we get exactly the same sample sizes as with the Bayesian approach.

```
ssfreq_bon <- ((qnorm(1 - alpha/k) + qnorm(power)) / (sqrt(nu) * deltastar))^2 *
              (1 + 1/sqrt(k))
ceiling(c(sqrt(k) * ssfreq_bon, rep(ssfreq_bon, k)))
```

```
## [1] 102  72  72
```

With a Dunnett-type adjustment that accounts for correlation among tests, the required sample sizes are slightly lower.

```
library("mvtnorm")
rho <- 1 / (1 + alloc)
corr <- matrix(rho, k, k) + diag(1 - rho, k)
quan <- qmvnorm(0.95, mean=rep(0, k), corr=corr)$quantile
ssfreq_dun <- ((quan + qnorm(power)) / (sqrt(nu) * deltastar))^2 * (1 + 1/alloc)
ceiling(c(sqrt(k) * ssfreq_dun, rep(ssfreq_dun, k)))
```

```
## [1] 100  71  71
```

Criterion 2

The posterior probability of at least one (any) experimental treatment being better than control is at least η, or else the posterior probability of none of the treatments being better than control (by a relevant margin δ^*) is at least ζ.

Bayesian

Leaving all other parameters unchanged, the Bayesian sample size for criterion 2 is considerably lower than for criterion 1.

```
ssbayes(k=2, nu=1, q0=c(0, 0, 0), eta=0.95, zeta=0.90, deltastar=0.5, prec="known",
        crit="2")
```

```
##
## Control: 83
## Group A: 59
## Group B: 59
```

Frequentist

This is comparable to a frequentist sample size when multiplicity of comparisons is not adjusted for.

```
ssfreq_unadj <- ((qnorm(1 - alpha) + qnorm(power)) / (sqrt(nu) * deltastar))^2 *
                (1 + 1/sqrt(k))
ceiling(c(sqrt(k) * ssfreq_unadj, rep(ssfreq_unadj, k)))
```

```
## [1] 83 59 59
```

References

Dunnett CW (1955) A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association*, **50**(272), 1096–1121.

Whitehead J, Cleary F, Turner A (2015) Bayesian sample sizes for exploratory clinical trials comparing multiple experimental treatments with a control. *Statistics in Medicine*, **34**(12), 2048-2061.