`ldDesign` 2: EXPERIMENTAL DESIGN FOR GENOME-WIDE ASSOCIATION STUDIES

RODERICK D. BALL

Contents

1. Introduction 2
2. Quantitative traits 4
2.1. Factors affecting power 4
2.2. Bayes factor, B 4
2.3. Linkage disequilibrium 4
2.4. Examples 5
3. Case-control studies 6
3.1. Factors affecting power 6
3.2. Odds ratios 7
3.3. Examples 7
4. Design of Association Studies 9
4.1. Genomic associations 9
4.2. Genome-wide association studies 9
4.3. Prior elicitation 9
4.4. Genome scans 10
4.5. Candidate gene studies 12
4.6. Dense markers and whole genome sequencing 12
4.7. Choice of population 12
4.8. Choice of trait 12
4.9. Multi-stage trials and replication 12
5. Functions useful for general Bayesian analysis 13
5.1. `SS.oneway.bf()` 13
5.2. `oneway.bf.alpha()` 14
5.3. `calc.alphaB.ABF()` 14
5.4. `calc.Balpha.ABF()` 15
References 15

1. Introduction

This vignette illustrates the use of the `ldDesign` R package for design of experiments for detecting genomic associations. Version 1 of this package implements the method for quantitative traits (Ball 2005: "Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies" Genetics 2005). There are two main functions: `ld.power()` and `ld.design()` for determining the power and sample size respectively. Version 2 of this package also implements the method for case-control studies (Ball 2011; "Experimental designs for robust detection of effects in genome-wide case-control studies" Genetics 2011). For case-control studies there are two main functions: `cc.power()` and `cc.design()` for determining the power and sample size respectively.

While `ldDesign` was written primarily for genetic applications, `ldDesign` also provides generally useful functions like `SS.oneway.bf()` for calculating Bayes factors from one-way analysis of variance models, `oneway.bf.alpha()` for calculating the Bayes factor for a given α level in one-way analysis of variance models, `calc.Balpha.ABF()` for calculating the approximate Bayes factor corresponding to a given α level and `calc.alphaB.ABF()` for calculating the α level corresponding to an approximate asymptotic Bayes factor. These functions are useful for retrospective calculations: *e.g.* determining the Bayes factor when p-values are published, or to give Bayesian measures of evidence when using frequentist methods like analysis of variance or *vice versa*. Readers interested in these functions for non-genetic analyses may wish to skip to Section. 5.

This package implements a Bayesian experimental design methodology designed to ensure robust detection of effects in genome-wide association studies (GWAS). In GWAS, hundreds of thousands of loci are being tested on thousands of individuals. To robustly detect effects we need sufficient power to detect the effects with sufficiently strong evidence to overcome the low prior odds for any given polymorphism to be associated with a detectable size effect.

Traditional approaches design experiments with power to detect effects with a given p-value or less ($P < \alpha$ for some α). However the p-value as a measure of evidence can be misleading and is not well calibrated for detecting effects in general. This is particularly problematic in genome-wide association studies where sample sizes are large and effect sizes are small. Traditional approaches using p-values have lead to many published spurious associations (Altshuler et al., 2000; Terwilliger and Weiss, 1998; Emahazion et al., 2001; Ball, 2005, 2007a,b, 2011) Even very low p-values (*e.g.* 5×10^{-7}, or 5×10^{-8}) may not be sufficient for respectable posterior odds in large studies or meta-analyses (Ball, 2011).

Traditional frequentist approaches adjust for multiple tests. However, gene mapping is, intrinsically, a model selection problem not a hypothesis testing problem. The model we wish to determine is ultimately something approximating the true genetic model—the set of loci affecting the trait and their modes of action. See *e.g.* (Ball, 2001; Sillanpää and Corander, 2002, and references therein)

Often the true model is not unequivocally determined by the data. Neglecting this leads to problems with spurious associations and selection bias, where the estimated effects are over-estimated unless the power to detect the true size of effect is high or independent data is used for detection and estimation. We recommend Bayesian 'model selection' approaches that consider alternative possible models according to their probabilities (*e.g.* Ball (2001)). A non-MCMC implementation of Bayesian model selection for QTL mapping is given in the R package `BayesQTLBIC` (Ball, 2009).

The power calculations in this package are for single marker tests, which are widely used. This will be a good approximation as long as the resolution of the study and the extent of linkage disequilibrium is such that effects are approximately independent. When multiple markers within the extent of linkage disequilibrium are affecting the trait multi-locus methods for analysis are recommended. In this case, the power calculations in `ldDesign` would be approximate and conservative.

This package uses a novel approach: unlike traditional power calculations that use p-values as a measure of evidence, we use Bayes factors, which, unlike p-values, have a direct interpretation as strength of evidence independent of sample size, experimental design, and test set-up. Designing experiments with power to obtain a sufficiently large Bayes factor, B, where B is chosen large enough to obtain respectable posterior odds enables us to design experiments with sufficient power to robustly detect associations at unknown positions in the genome.

Note that by obtaining the Bayes factor in single marker tests we are still considering multiple models: corresponding to the null and alternative hypotheses H_0, H_1 — the Bayes factor combined with prior odds determines the posterior probabilities. If posterior odds are not high an unbiased estimator would be the product of the estimate conditional on H_1 and the posterior probability for H_1. This would often result in considerable shrinkage of effect estimates.

Association studies detect *linkage disequilibrium* between an observed marker locus and an unobserved trait locus (QTL or QTN). Since linkage disequilibrium between 2 loci decays exponentially each generation at a rate proportional to the recombination rate between the loci, association mapping using population based samples offers potentially higher resolution (resolution 100s to 1000s of base

pairs) than family based QTL mapping approaches (centi-Morgans to 10s of centi-Morgans) which detect linkage disequilibrium generated within a pedigree. However achieving this resolution requires much larger sample sizes. Spurious associations can be generated by population structure or simply because the strength of evidence is insufficient to overcome the low prior odds for genomic associations in diverse populations. For further information see (Ball, 2005, 2007a,b, and references therein)

2. Quantitative traits

2.1. **Factors affecting power.** For a given marker and causal locus, factors affecting power of an experiment to detect linkage disequilibrium between the marker and causal locus include:

- sample size (n)
- allele frequency at the marker locus (p)
- allele frequency at the causal locus (q)
- linkage disequilibrium coefficient (D, or D')
- effect QTL heritability (h^2_q)
- genetic model (dominance ratio, ϕ)
- Bayes factor required (B)

corresponding to the arguments of `ld.power()`:

```
> library(ldDesign)
> args(ld.power)

function (n, p, q, D, h2, phi, Bf, missclass.rate = 0)
NULL
```

The function `ld.power()` implements the method from Ball (2005) which uses the Spiegelhalter and Smith (1982) Bayes factor—an analytical formula for one way analysis of variance models (implemented in the R function `SS.oneway.bf()`, *cf.* Section 5), in conjunction with a frequentist power calculation adapted from Luo (1998).

Note: the Spiegelhalter and Smith (1982) Bayes factor—an analytical formula for one way analysis of variance models (function `SS.oneway.bf()`), so does not incorporate prior variance. In our experience, use of `SS.oneway.bf()` is approximately equivalent to assuming prior information equivalent to a single sample point ($a = 1$ in the arguments to `cc.power()` below). In principle `ld.power()` could incorporate prior variance using the methods used by `cc.power()`. This may be incorporated in a future version of `ld.power()`.

2.2. **Bayes factor, B.** The Bayes factor for comparing two models (M_0, M_1) is the ratio

$$B = \frac{\Pr(\text{data} \mid M_1)}{\Pr(\text{data} \mid M_0)} \tag{1}$$

Prior and posterior odds are related by:

$$\text{posterior-odds} = B \times \text{prior-odds} \tag{2}$$

Hence, *e.g.* if the prior odds were 1:50000 and we want posterior odds of 20:1, we should have a Bayes factor of

$$\frac{20}{1/50,000} = 1,000,000 \tag{3}$$

2.3. **Linkage disequilibrium.** We do not necessarily observe the causal locus, but a marker-trait association is induced by *linkage disequilibrium* between the marker and a causal locus. Linkage disequilibrium (Weir, 1996) is non-independence between genetic loci or positions on the genome. We will consider bi-allelic loci, *e.g.* a marker with alleles A, a and a causal locus with alleles Q, q. The pairwise probabilities for the alleles

$$\Pr(AQ) = \Pr(A)\Pr(Q) + D = pq + D \tag{4}$$

$$\Pr(aQ) = \Pr(a)\Pr(Q) - D = (1-p)q - D \tag{5}$$

$$\Pr(Aq) = \Pr(A)\Pr(q) - D = p(1-q) - D \tag{6}$$

$$\Pr(aq) = \Pr(a)\Pr(q) + D = (1-p)(1-q) + D \tag{7}$$

where D is the linkage disequilibrium coefficient and p, q are the allele frequencies at the marker and causal locus, respectively. Linkage disequilibrium can also be specified as D' which is D expressed as a proportion of the maximum (resp. minimum) disequilibrium if D is positive (resp. negative).

It is often more convenient to specify D' (D divided by its maximum absolute value for the given allele frequencies and sign of D) because otherwise we have to work out the maximum or minimum values of D for the given allele frequencies.

Another useful quantity is r^2, related to D, p, q by:

$$r^2 = \frac{D^2}{p(1-p)q(1-q)} \tag{8}$$

Similar to r^2 in linear regression, r^2 gives approximately the proportion of variance explained by the marker in predicting the causal locus. The approximation is to first order for $r^2 \approx 1$, however in practice r^2 may be significantly less than 1. For this reason our power calculations use exact quantities in expressed in terms of D, p, q.

2.4. **Examples.**

1. Find the power to detect an effect with QTL heritability $h^2_q = 0.05$, with Bayes factor 10^6, marker and causal allele frequencies 0.3, 0.2, linkage disequilibrium coefficient $D = 0.1$, and sample size 1000. Assume an additive model.

```
> ld.power(B=1e6, h2=0.05, D=0.1, p=0.3, q=0.2, n=1000, phi=0)
        n  power
[1,] 1000 0.0134
attr(,"parms")
           p            q            D           h2
       3e-01        2e-01        1e-01        5e-02
         phi           Bf missclass.rate
       0e+00        1e+06        0e+00
```

2. Find the power to detect an effect with QTL heritability $h^2_q = 0.05$, with Bayes factor 10^6, marker and causal allele frequencies 0.3, 0.2, linkage disequilibrium coefficient $D = 0.1$, and sample size 1000. Assume a dominant model.

```
> ld.power(B=1e6, h2=0.05, D=0.1, p=0.3, q=0.2, n=1000, phi=1)
        n    power
[1,] 1000 0.008843
attr(,"parms")
           p            q            D           h2
       3e-01        2e-01        1e-01        5e-02
```

```
          phi             Bf missclass.rate
        1e+00          1e+06          0e+00
```

3. As per [2.] above find the sample size required for power 0.8 and print the power curve:

```
> ld.design(B=1e6, h2=0.05, D=0.1, p=0.3, q=0.2, phi=1, power=0.8,
+           nmin=1730, nmax=4620, ninterp=20, print.it=TRUE)
          n   power
 [1,] 1730 0.09797
 [2,] 1822 0.12026
 [3,] 1918 0.14662
 [4,] 2020 0.17760
 [5,] 2127 0.21337
 [6,] 2240 0.25425
 [7,] 2359 0.30015
 [8,] 2484 0.35092
 [9,] 2616 0.40603
[10,] 2755 0.46465
[11,] 2901 0.52576
[12,] 3055 0.58785
[13,] 3217 0.64953
[14,] 3388 0.70892
[15,] 3568 0.76467
[16,] 3757 0.81514
[17,] 3956 0.85941
[18,] 4166 0.89672
[19,] 4387 0.92700
[20,] 4620 0.95049
attr(,"parms")
             p              q              D             h2
         3e-01          2e-01          1e-01          5e-02
           phi             Bf missclass.rate
         1e+00          1e+06          0e+00
[1] 3695
```

3. Case-control studies

3.1. **Factors affecting power.** For a given marker and causal locus, factors affecting power of an experiment to detect linkage disequilibrium between the marker and causal locus include:

- Bayes factor required (B)
- odd ratio(s) (OR) (or relative risk(s) (R))
- linkage disequilibrium coefficient (D, or D')
- allele frequency at the marker locus (p)
- allele frequency at the causal locus (q)
- disease prevalence (or baseline risk)
- sample sizes (number of cases, number of controls)
- genetic model (additive, dominant, recessive or general)
- prior variance for effects (a, or σ^2_η)

corresponding to the arguments of `cc.power()`:

```
> library(ldDesign)
> args(cc.power)
```

```
function (B, OR = NULL, D, p, q, baseline.risk, Dprime = NULL,
    R = NULL, prevalence = NULL, n.cases, n.controls, model = c("additive",
        "dominant", "recessive", "general"), a = 1, sigma2.eta = NULL,
    verbose = FALSE, amalgamate.cells = FALSE, show.attributes = FALSE)
NULL
```

3.2. **Odds ratios.** The odds ratio for 2 genotypes g_1, g_2 is the ratio

$$\frac{\Pr(\text{case} \mid g_2)/\Pr(\text{control} \mid g_2)}{\Pr(\text{case} \mid g_1)/\Pr(\text{control} \mid g_1)} \tag{9}$$

For the additive, dominant and recessive models a single odds ratio or relative risk is specified. For the general model a vector of 2 odds ratios or relative risks are specified.

3.3. **Examples.**

1. Find the power to detect an effect with odds ratio 1.6, with Bayes factor 10^6, marker and causal allele frequencies 0.3, 0.2, linkage disequilibrium coefficient $D = 0.1$, baseline risk 0.1, 1000 cases and 1000 controls. Assume an additive model, and a prior with information equivalent to a single sample point.

```
> cc.power(B=1e6, OR=1.6, D=0.1, p=0.3, q=0.2, baseline.risk=0.1,
+          n.cases=1000, n.controls=1000, model="additive", a=1)
[1] 0.01322
```

2. (Illustrating vectorisation of sample sizes). As per [1.] above but find the power for sample sizes from 2000 to 12000 in steps of 2000.

```
> cc.power(B=1e6, OR=1.6, D=0.1, p=0.3, q=0.2, baseline.risk=0.1,
+          n.cases=1000*seq(2,12,by=2),
+          n.controls=1000*seq(2,12,by=2),
+          model="additive", a=1)
[1] 0.2113 0.8895 0.9973 1.0000 1.0000 1.0000
```

3. Find the sample size required to detect an effect with odds ratio 2.0 with Bayes factor 10^6, marker and causal allele frequencies 0.3, 0.2, linkage disequilibrium coefficient $D = 0.1$, baseline risk 0.1. Assume the number of controls is 50% more than the number of cases. Assume an additive model, and a prior with information equivalent to a single sample point. Print the power curve with power ranging from 0.1 to 0.99.

```
> cc.design(B=1e6, OR=2.0, D=0.1, p=0.3, q=0.2, power=0.9,
+           baseline.risk=0.1, n.cases=2000, n.controls=3000,
+           model="additive", a=1, pmin=0.1, pmax=0.99,
+           ninterp=20, print.power.curve=TRUE)
Power curve:
      n.controls n.cases  power
 [1,]        681     454 0.1000
 [2,]        723     482 0.1253
 [3,]        768     512 0.1557
 [4,]        816     544 0.1918
 [5,]        867     578 0.2339
 [6,]        921     614 0.2822
 [7,]        978     652 0.3366
 [8,]       1038     692 0.3966
 [9,]       1103     735 0.4610
[10,]       1171     781 0.5286
[11,]       1244     830 0.5974
[12,]       1322     881 0.6653
```

```
[13,]       1404     936 0.7301
[14,]       1491     994 0.7895
[15,]       1584    1056 0.8420
[16,]       1682    1121 0.8862
[17,]       1786    1191 0.9218
[18,]       1897    1265 0.9489
[19,]       2015    1344 0.9684
[20,]       2141    1427 0.9816
[21,]       2274    1516 0.9900
1719 controls and 1146 cases for power 0.9
        n n.controls    n.cases
     2865       1719       1146
```

4. As per [3.] but assume a general model with 2 independent odds ratios of 1.5, 1.5.

```
> cc.design(B=1e6, OR=c(1.5,1.5), D=0.1, p=0.3, q=0.2, power=0.9,
+           baseline.risk=0.1, n.cases=2000, n.controls=3000,
+           model="general", a=1, print.power.curve=FALSE)
8120 controls and 5414 cases for power 0.9
        n n.controls    n.cases
    13534       8120       5414
```

5. (Show attributes.) As per [1.] show attributes including non-centrality parameter (`ncp`) and the optimal weighing used (`c1.opt`).

```
> cc.power(B=1e6, OR=1.6, D=0.1, p=0.3, q=0.2, baseline.risk=0.1,
+          n.cases=1000, n.controls=1000, model="additive", a=1,
+          show.attributes=TRUE)
[1] 0.01322
attr(,"model")
[1] "additive"
attr(,"n.cases")
[1] 1000
attr(,"n.controls")
[1] 1000
attr(,"prevalence")
[1] 0.1214
attr(,"baseline.risk")
[1] 0.1
attr(,"relative.risk")
[1] 1.509
attr(,"odds-ratio")
[1] 1.6
attr(,"p")
[1] 0.3
attr(,"q")
[1] 0.2
attr(,"ncp")
[1] 12.63
attr(,"B")
[1] 1e+06
attr(,"alphac")
[1] 7.735e-09
attr(,"c1.opt")
[1] 0.6792
```

```
attr(,"ps")
            bb     Bb      BB
control 0.4986 0.4155 0.08587
case    0.4274 0.4527 0.11988
```

A brief explanation of the attributes is shown in Table 1. Other attributes are as given in the function call, except that odds ratios and relative risks are both shown and baseline risk and prevalence are both shown regardless of which were specified.

TABLE 1. Attributes returned by `cc.power()`

`ncp`	the non-centrality parameter (12.63)
`alphac`	the value of α threshold corresponding to the given Bayes factor (7.73×10^{-9})
`c1.opt`	`c1.opt` $= 0.679$ is the weight placed on the first odds ratio (`Bb` vs `bb`) and $1 -$ `c1.opt` $= 0.321$ is the weight placed on the second odds ratio (`BB` vs `Bb`) when estimating the odds ratio in the additive model
`ps`	table of expected marker allele frequencies. A 2x3 table for the additive or general models and a 2x2 table (with functionally equivalent genotypes amalgamated) for the dominant or recessive models.

4. Design of Association Studies

Here we discuss application to design of GWAS for detecting genomic associations, including elicitation of the prior parameters.

4.1. **Genomic associations.** The genome contains many (*e.g.* 3×10^9 for humans) loci. Typically about 1/1000 loci are polymorphic, *i.e.* differ between individuals in a species. For a given trait, only a small proportion or polymorphic loci will be causal loci with practically significant effects on the trait.

4.2. **Genome-wide association studies.** Genome-wide association studies (GWAS) aim to detect associations between SNP markers spaced along the genome and causal loci. If a marker is in linkage disequilibrium with a causal locus, then the marker genotypes will be associated with trait variation. The strength of association depends on the size of the causal effect and the linkage disequilibrium coefficient D.

4.3. **Prior elicitation.** Prior odds and prior variance for the effect(s) being tested are critical factors in determining the Bayes factor required. Ascertaining the prior is, inevitably, subjective. This does not mean, however, that we can 'choose' the prior arbitrarily. The prior should represent our prior knowledge before carrying out the experiment or observing the data. This is a subjective Bayesian approach as opposed to 'objective' Bayesian approaches that try to use a 'default' or non-informative prior. Using $a = 1$ in `cc.power()` is an example of a default prior. While it is useful to have a good conservative default it is not recommended in general to always use this. The process of ascertaining the prior is known as prior elicitation.

Prior elicitation (O'Hagan et al., 2006), where prior information is elicited from experts is an important but neglected area. This is likely to become more important (and interesting) as more associations are detected.

In general the prior odds depend on the number of loci expected to significantly affect the trait (ignoring effects that are undetectable), the number of markers and the extent of linkage disequilibrium. If marker spacing is comparable to the extent of LD, then we consider the prior probability per interval of this size around a marker. If there are fewer markers the number of markers is limiting, and we again consider the prior probability per interval of this size around a marker. If there are substantially more markers we may increase the minimum D that we design for.

Prior odds also depend on previous results from the same or similar loci and/or traits in the same or related populations and/or species. In practice one never observes true replicates but experiments, data or information that have varying degrees of relevance to the experiment or decision in question. What weight to give various elements of information is the subjective aspect, where experts can be useful. This is central to the application of statistics in general since we are always trying to infer or predict something in a new or similar but different situation using information from past observations.

4.4. **Genome scans.** First consider the scenario as in Table 2 with approximately one marker per interval of length equal to the extent of LD.

TABLE 2. Determination of prior odds and Bayes factor required.

Bayes factor from previous studies	500
genome length	3×10^9
expected number of loci	10
extent of LD	6kb
number of SNP markers	500000
prior odds per marker	$6 \times 10^3/3 \times 10^9 = 1/500000$
posterior odds required	20:1
combined Bayes factor required	10^6
study Bayes factor required	$10^6/500 = 2000$

Note:

- The posterior probability per marker is the probability that the marker is within the extent of LD of a causal marker.
- We are free to choose the criterion for defining the extent of LD, *e.g.* $D' > 0.5$ or $r^2 > 0.8$. However if choosing a lower threshold this will reduce power, and if choosing a higher threshold this will result in smaller sized intervals requiring more markers to be genotyped for acceptable coverage.
- Conversely if we have more markers we may use a higher value of D in the power calculations.
- The expected number of loci, should be the expected number of loci with detectable size effects. If the experiment is sufficiently powerful the expected number of effects may increase. For example, if the intended experiment is powerful enough to detect effects explaining 1% of the phenotypic variance and the trait heritability is 30%, there can be at most 30 such effects. In practice the amount of variation explained by additive effects of 1% or greater may be less, *e.g.* only 10%, due to the possible existence of many smaller effects, and non-additive variation, justifying and expected number of loci of 10.

- We also need to consider the allele frequencies for loci we wish to detect. Rare alleles may be poorly tagged by common SNPs *e.g.* from the HapMap project (`www.hapmap.org`) (e.g Yang et al. (2010)). Large recessive effects may not yet be detected Ball (2011), *e.g.* to detect an effect of odds ratio 3 with minor allele frequency 5% when tagged by a marker with allele frequency 10% at $D' = 0.5$ requires over 10^6 cases and controls:

```
> cc.design(B=1e6, OR=3, Dprime=0.5, p=0.1, q=0.05,
+           baseline.risk=0.1, n.cases=1000, n.controls=1000,
+           model="recessive", a=1, power=0.8,
+           print.power.curve=FALSE)
688379 controls and 688379 cases for power 0.8
         n n.controls    n.cases
   1376758     688379     688379
```

Recent analyses for human height (*e.g.* Gudbjartsson et al. (2008), Lango Allen et al. (2010),Weedon et al. (2008)) have of the order of 15,000 samples in combined meta-analyses. Hundreds of putative effects were identified although few were replicated across all three analyses. For example, (Weedon et al., 2008, Table 1, p577) reported 20 SNP effects with combined p-values $P < 5 \times 10^{-7}$ putatively collectively explaining 3% of the variance from a meta-analysis of human height. We re-examine the power, using a prior odds based on 100 expected loci:

```
> ld.power(B=5e5, h2=0.001, D=0.25, p=0.5, q=0.5, n=13665, phi=1)
        n    power
[1,] 13665 0.003232
attr(,"parms")
            p             q             D            h2
      5.0e-01       5.0e-01       2.5e-01       1.0e-03
          phi            Bf missclass.rate
      1.0e+00       5.0e+05       0.0e+00
```

we see the power is low to obtain respectable posterior odds, even for a well tagged locus, hence these effects are not robustly detected. At the very least, the effects need to be re-estimated in an independent population.

The Bayes factor corresponding to the threshold used is:

```
> calc.Balpha.ABF(alpha=5e-7, n=13665, a=1)
[1] 2617
```

representing strong evidence, but not strong enough for respectable posterior odds without replication at a similar strength of evidence.

The sample size required for good power is from 50,000 for a well tagged locus ($r^2 \approx 1$) or 110,000 for a less well tagged locus ($r^2 = 0.5$) or 460,000 for a poorly tagged locus ($r^2 = 0.12$):

```
> ld.design(B=1e5, h2=0.001, D=0.25, p=0.5, q=0.5, power=0.8,
+           phi=1)
[1] 52410
> # sample size for a less well tagged locus, D=Dmax, MAF=0.05
> ld.design(B=1e5, h2=0.001, D=0.045, p=0.1, q=0.05, power=0.8,
+           phi=1, nmin=50000, nmax=200000)
[1] 108472
> # sample size for a poorly tagged locus, D=Dmax/2, MAF=0.05
> ld.design(B=1e5, h2=0.001, D=0.045/2, p=0.1, q=0.05, power=0.8,
+           phi=1, nmin=50000, nmax=500000)
[1] 466401
```

Note the use of arguments `nmin` and `nmax` here. These values should be chosen to bracket the required sample size, as interpolation is used on values calculated by `ld.power()` for sample sizes from `nmin` to `nmax`. This may require some iteration. Having found a solution, the solution can be refined by specifying `nmin` and `nmax` closer to the solution.

4.5. **Candidate gene studies.** `ldDesign` applies equally well to candidate gene studies, where polymorphisms are sourced from 'candidate genes' considered likely to affect the trait. Depending on the trait and candidate gene, the polymorphism may have higher prior probability than a random genomic polymorphism.

As a default starting value we may assume similar prior odds to a genome scan.

Substantially higher prior odds than this require some justification. Just being in a gene or associated promoter alone does not necessarily increase prior odds by much since there are $\sim 50,000$ genes and a number (*e.g.* up to ~ 10 ??) of polymorphisms in each gene, and we have to allow for the possibility that a proportion of causal effects are not in the candidate gene set.

A polymorphism mapping into a QTL region may have higher prior odds depending on strength of evidence for the QTL, the posterior distribution for QTL location, and relative position of the polymorphism and the QTL (Ball, 2007b).

Bioinformatics, *e.g.* where similar sequences have been found in other species, and knowledge of gene action and pathways that affect the trait, may be useful.

4.6. **Dense markers and whole genome sequencing.** With dense markers or whole genome sequencing we may be in the luxurious position of having multiple or many markers within the extent of LD of any given marker or causal locus. This means there may be multiple markers within the extent of LD of a causal locus competing to explain the variation. A work-around is to consider the prior odds for a single representative marker within the interval.

Whole genome sequencing raises additional challenges *e.g.* errors in marker genotyping when coverage is low or limited sample sizes when coverage is high. The argument `missclass.rate` in `ld.power()` can be used where there is an estimate of genotyping error rate.

On the '+' side of the ledger having whole genome sequences means we will have the causal locus, *i.e.* some polymorphism in complete LD with a causal locus. We just have to identify which ones. This should give more power, quantifying this is a topic for future research.

4.7. **Choice of population.** Where sample size is limited experimenters may be better off choosing a less diverse population, *e.g.* an isolated Finnish or island population with a relatively small number of founders, rather than a diverse African population.

The tradeoff is lower resolution but smaller sample size and number of markers needed to obtain that resolution. This means more power to detect effects but a number of effects (particularly rare variants) may not be present in the small population. If studying a disease, the disease would obviously have to be present in the study population. A population with high prevalence may be promising.

Effects could then be validated in other populations requiring genotyping only a limited number of markers.

4.8. **Choice of trait.** A trait closer to basic biochemistry *e.g.* resulting from a known biochemical pathway is likely to be influenced by only a moderate number of genes. Prior odds would be much higher for polymorphisms in the known genes.

Human height or traits such as tree growth rate or wood density may be influenced by many small effect genes. Putative associations from GWAS to date explain only a small proportion of the variation in human height and complex diseases.

Some of these may in fact be larger effect rare alleles poorly tagged by current SNPs (*cf.* the example in subsection 4.4).

4.9. **Multi-stage trials and replication.** `ldDesign` does not yet cater explicitly for multi-stage designs. However, when detecting and validating effects in several stages we may partition the Bayes factor required among stages, *e.g.* if $B > 1000$ in each of 2 independent samples this combines, conservatively, to give $B > 10^6$, of the order required for genomic associations.

It is not sufficient just to say the results have been replicated for some α level (which is usually much less than the 5×10^{-7} used in genome scans circa 2007 (*e.g.* WTCCC2007)). It is vital that the replication sample(s) have a sufficiently large Bayes factor so that the combined Bayes factor is large enough to achieve the required posterior odds. Since the p-value 5×10^{-7} corresponds to a Bayes factor around 1000 this value replicated *twice* would be of the order of magnitude needed.

The current *de facto* standard of 5×10^{-8} is the threshold for the combined p-value over all replicates. This corresponds to a Bayes factor of around 24000 for Weedon et al. (2008)($n = 13665$) or a Bayes factor of around 9000 if $n = 100,000$:

```
> calc.Balpha.ABF(alpha=5e-8, n=13665, a=1)
[1] 24245
> calc.Balpha.ABF(alpha=5e-8, n=100000, a=1)
[1] 8971
```

5. Functions useful for general Bayesian analysis

5.1. `SS.oneway.bf()`. This function, used by `ld.power()`, calculates the Spiegelhalter and Smith (1982) Bayes factor corresponding to a one-way analysis of variance with given group sizes and F-statistic:

$$B = \frac{\left(1 + \frac{m-1}{n-m}F\right)^{n/2}}{\sqrt{\frac{m+1}{2n}\prod_{i=1}^{m} n_i}} \tag{10}$$

where m is the number of groups, n_i the sample size within the ith group, $n = \sum n_i$ is the total sample size. This can be used in general statistical applications where a Bayesian measure of evidence is desired, either for retrospective analysis where p-values have been used or where it is convenient to use existing software like R analysis of variance. We find this frequently useful as a quick and easy way to complement a frequentist analysis with a Bayesian analysis measure of evidence, that does not seem to be generally realised.

Consider the following R analysis of variance for the `Oats` dataset:

```
> library(nlme)
> data(Oats)
> summary(aov(yield ~ nitro* Variety + Error(Block), data=Oats))

Error: Block
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  5  15875    3175

Error: Within
              Df Sum Sq Mean Sq F value  Pr(>F)
nitro          1  19536   19536   81.52 7.6e-13 ***
```

```
Variety        2   1786     893    3.73    0.03 *
nitro:Variety  2    168      84    0.35    0.71
Residuals     61  14620     240
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> with(Oats, table(Variety))
Variety
Golden Rain  Marvellous     Victory
         24          24          24
```

Noting that there are 3 groups (3 varieties) of size 24, and an F-statistic of 3.73, the Spiegelhalter and Smith (1982) Bayes factor for testing and effect of `Variety` is calculated as:

```
> SS.oneway.bf(group.sizes=c(24,24,24), Fstat=3.73)
[1] 2.055
```

A Bayes factor of 2 indicates that the evidence for `Variety` is very weak indeed. Traditionally this would have been regarded as 'significant'. Early users of significance tests were fortunate that their prior odds were generally high.

Note: Although not strictly a one-way analysis of variance, the method can nevertheless be applied here if it is assumed the F-statistic represents a comparable strength of evidence to the same F-statistic obtained without the additional experimental structure. (This seems reasonable here, and underlies most frequentist analysis, but don't take our word for it! We leave it to the reader to find a proof or derive an adjustment.)

5.2. `oneway.bf.alpha()`. This function, used by `ld.power()`, calculates the Spiegelhalter and Smith (1982) Bayes factor corresponding to a given α threshold. For example: the Bayes factors corresponding to $\alpha = 0.05, 0.01, 0.001$ for testing for an effect of `Variety` in the Oats data are calculated as:

```
> oneway.bf.alpha(n=72, group.sizes=c(24,24,24), alpha=c(0.05,0.01,0.001))
[1]  1.163  6.234 68.908
```

5.3. `calc.alphaB.ABF()`. This function, used by `cc.power()` calculates the α threshold corresponding to a given Bayes factor when using asymptotic approximate Bayes factors, for given prior and sample size. For example, to calculate the α significance level corresponding to a Bayes factor $B = 1000$ for a sample size $n = 20, 100, 1000$, with prior information equivalent to a single sample point ($a = 1$):

```
> calc.alphaB.ABF(B=1000, n=20, a=1, alpha.start=1e-3)
[1] 2.582e-05
> calc.alphaB.ABF(B=1000, n=100, a=1, alpha.start=1e-3)
[1] 1.600e-05
> calc.alphaB.ABF(B=1000, n=1000, a=1, alpha.start=1e-3)
[1] 5.247e-06
```

Note the decreasing thresholds with increasing sample size. Early users of significance tests were fortunate that their prior odds were generally high and their sample sizes relatively low.

It is sometimes useful to specify an approximate starting value (`alpha.start`) as above. The starting value is used as an initial upper bound in searching for a solution. If the function returns `NA`, try a higher or lower starting value.

The function uses interpolation to find a solution. The solution can be refined by choosing a starting value slightly above the solution and using a smaller value of `reduction.factor` *e.g.*

```
> calc.alphaB.ABF(B=1000, n=1000, a=1, alpha.start=6e-6,
+                 reduction.factor=1.05)
[1] 5.247e-06
```

This makes little difference in this case confirming the accuracy of the initial solution.

5.4. `calc.Balpha.ABF()`. This function is the inverse of `calc.alphaB.ABF()` *i.e.* calculates the approximate Bayes factor corresponding to a given α level for given prior and sample parameters.

```
> calc.Balpha.ABF(alpha=0.01, n=100, a=1)
[1] 2.657
> calc.alphaB.ABF(B=2.657, n=100, a=1, alpha.start=0.02)
[1] 0.01000
```

Note: Approximate Bayes factors for genetic analysis similar to those used in `calc.Balpha.ABF()` were first derived by the author in (Ball, 2007a, p166–167) for S-TDT tests and subsequently by Wakefield (2007) for asymptotically normal test statistics. Our derivation uses the Savage–Dickey approximation (*cf.* Ball (2011)), which gives the Bayes factor for nested models in certain conditions as a ratio of prior to posterior at zero. Assuming a test statistic with sampling variance $1/n$ and a prior with mean 0 and variance $1/a$ equivalent to a sample points the approximate Bayes factor is given as:

$$B \approx \frac{\sqrt{a}}{\sqrt{n+a}} \exp\Big(\frac{n^2 Z_n^2}{2(n+a)}\Big) \tag{11}$$

References

D. Altshuler, J. N. Hirschhorn, M. Klannemark, C. M. Lindgren, M.-C. Vohl, J. Nemesh, C. R. Lane, S. F. Schaffner, S. Bolk, C. Brewer, T. Tuomis, D. Gaudet, T. J. Hudson, M. Daly, L. Groop, and E. S. 2000 Lander. The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nature Genetics*, 26:76–80, 2000.

R. D. Ball. Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian Information Criterion. *Genetics*, 159:1351–1364, 2001.

R. D. Ball. Experimental designs for reliable detection of linkage disequilibrium in unstructured random population association studies. *Genetics*, 170:859–873, 2005.

R. D. Ball. Statistical analysis and experimental design. In N. C. Oraguzie, E. H. A. Rikkerink, H. N. De Silva, and S. E. Gardiner, editors, *Association Mapping in Plants*, pages 133–196, New York, 2007a. Springer.

R. D. Ball. Quantifying evidence for candidate gene polymorphisms—Bayesian analysis combining sequence-specific and QTL co-location information. *Genetics*, 177:2399–2416, 2007b.

R. D. Ball. `BayesQTLBIC`—Bayesian multi-locus QTL analysis based on the BIC criterion, 2009. URL `http://cran.r-project.org/web/packages/BayesQTLBIC/index.html`. Accessed Nov 7, 2011.

R. D. Ball. Experimental designs for robust detection of effects in genome-wide case-control studies. *Genetics (to appear)*, 2011. URL `http://www.genetics.org/content/early/2011/09/13/genetics.111.131698.abstract`.

T. Emahazion, L. Feuk, M. Jobs, S. L. Sawyer, D. Fredman, D. St. Clair, J. A. Prince, and A. J. Brookes. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends in Genetics*, 17:407–413, 2001.

D. F. Gudbjartsson, G. B. Walters, G. Thorleifsson, H. Stefansson, B. V. Halldorsson, P. Zusmanovich, P. Sulem, S. Thorlacius, A. Gylfason, S. Steinberg, and A. Halgadottir. Many sequence variants affecting diversity of adult human height. *Nature Genetics*, 40:609–615, 2008.

HapMap. The HapMap project. URL `www.hapmap.org`.

H. Lango Allen, K. Estrada, and G. Lettre *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467:832–838, 2010.

Z. W. Luo. Linkage disequilibrium in a two-locus model. *Heredity*, 80:198–208, 1998.

A. O'Hagan, C.E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain judgements: Eliciting experts' probabilities*. Wiley, Hoboken, NJ, 2006.

M J Sillanpää and J Corander. Model choice in gene mapping: what and why. *Trends in Genetics*, 18:301–307, 2002.

D. Spiegelhalter and A.F.M. Smith. Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society, Series B*, 44(3), 1982.

J.D. Terwilliger and K.M. Weiss. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr. Opin. Biotechnol.*, 9:578–594, 1998.

The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447: 661–678, 2007.

J. Wakefield. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.*, 81:208–227, 2007.

M.N. Weedon, H. Lango, C.M. Lindgren, C. Wallace, D.M. Evans, M. Mangino, R.M. Freathy, J.R.B. Perry, S. Stevens, A.S. Hall, J.J. Samani, and B. Shields. Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics*, 40:575–583, 2008.

B. S. Weir. *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA, 1996.

J. Yang, B. Beben, D.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42:565–569, 608, 2010.