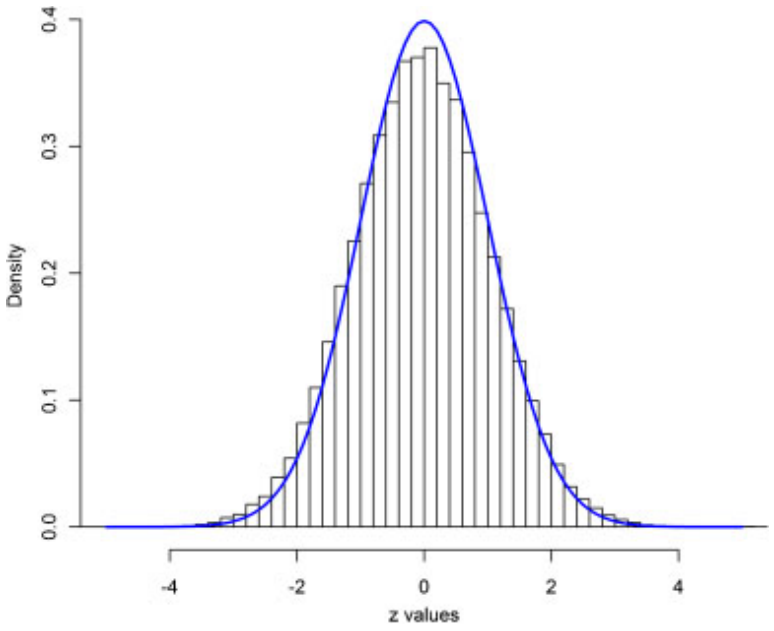
Estimating the phenotypic variance explained by GWAS SNPs using summary statistics

So, Li and Sham (2011)

Methodology

- So et al. (2011) Genet Epidemiol. 35(6):447-56
- Advantages
  - Rely on summary statistics
    - z-scores from association analysis
- Assess empirically how much the observed distribution of z-statistics differs from expected under the null hypothesis



Methodology

- The observed z statistics have underlying distribution,

$$z|\delta \sim N(\delta, 1)$$

  δ = 0 for null variants

  δ ≠ 0 for truly associated variants

- **Step 1**
  - Estimate f(z), the density function of z, from observed z's

    e.g. using kernel density estimation

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

- **Step 2**
  - Estimate f '(z) from f(z)

    e.g. using smooth.spline function in R

Methodology

- **Step 3**
  - Estimate δ from observed z by adding correction term

$$E\{\delta|z\} = z + \frac{f'(z)}{f(z)}$$

Efron (2009)

- **Step 4**
  - Convert δ to variance explained (Vg)

$$\mathrm{Vg} = \frac{[E(\delta|z)]^2}{n - 2 + [E(\delta|z)]^2}$$

- **Step 5**
  - Summate Vg to obtain overall variance explained by all SNPs

Practical on total-Vg

Variance explained by SNPs in GWAS using summary statistics

Overview

- LD pruning to remove SNPs in high LD
- Calculate the phenotypic variance explained by SNPs for the 3 quantitative traits
- Compare the estimates with GCTA

R scripts

- https://sites.google.com/site/honcheongso/software/total-vg
  - Binary trait
  - **Quantitative trait**

- So H.C., Li MX and Sham P.C. (2011)
  - Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. Genetic Epidemiology.

Practical

- Open terminal and make a new directory
  - mkdir ~/total-Vg
- Copy the data folder to your home total-Vg directory
  - cp /faculty/clara/total-Vg/* ~/total-Vg/
- Go to your total-Vg directory
  - cd ~/total-Vg

Association analysis – linear regression

```
plink --bfile example \
      --pheno example.pheno \
      --mpheno 1 \
      --linear \
      --out test.allSNPs.trait1
```

test.allSNPs.trait1.assoc.linear

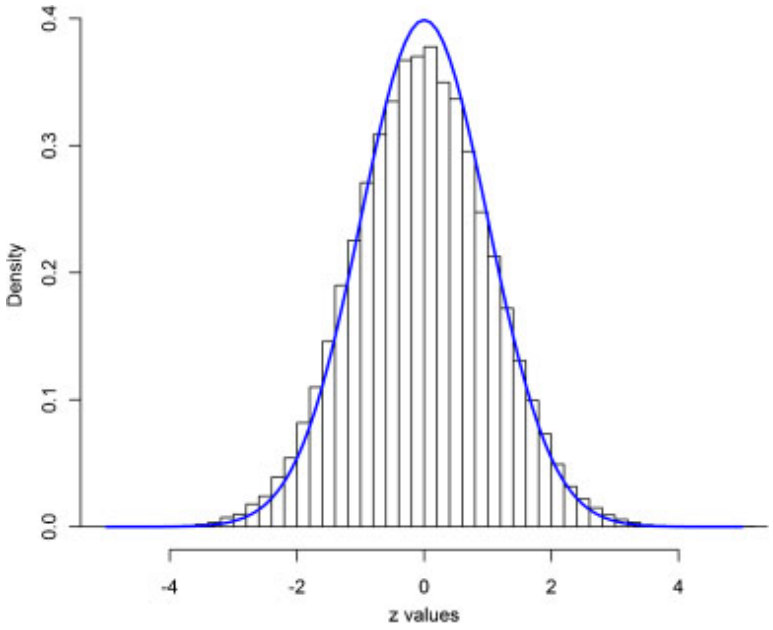
CHR	SNP	BP	A1	TEST	NMISS	BETA	STAT	P
15	rs1896801	82034088	2	ADD	2000	0.06995	0.2875	0.7737
15	15-82054371	82054371	2	ADD	2000	-0.3436	-1.038	0.2993
15	rs17158780	82057668	2	ADD	2000	-0.1449	-0.5175	0.6048
15	15-82067637	82067637	2	ADD	2000	-0.2329	-0.513	0.608
15	15-82071784	82071784	2	ADD	2000	-0.3721	-0.906	0.3651

LD Pruning

- Remove SNPs with $r^2 > 0.25$ using sliding window of 100 SNPs

```
plink --bfile example \
      --indep-pairwise 100 25 0.25 \
      --out test
```

```
R script remove-pruned-snp.R  \
 test.allSNPs.trait1.assoc.linear \
 test.prune.in \
 test.pruned.trait1.assoc.linear
```



total-Vg

- Rscript total-vg.R test.allSNPs.trait1.assoc.linear 2000
- Rscript total-vg.R test.pruned.trait1.assoc.linear 2000

V(g)/V(p)	GCTA	total-Vg (pruned)	total-Vg (unpruned)
Trait 1	0.054 (0.021)	0.063	0.207

total-Vg

- Repeat for the other 2 traits

V(g)/V(p)	GCTA	total-Vg (pruned)
Trait 1	0.054 (0.021)	
Trait 2	0.358 (0.033)	
Trait 3	0.582 (0.027)	

total-Vg

- Repeat for the other 2 traits

V(g)/V(p)	GCTA	total-Vg (pruned)
Trait 1	0.054 (0.021)	0.063
Trait 2	0.358 (0.033)	0.227
Trait 3	0.582 (0.027)	0.605

Trait 1: h2=.60, CV MAF .0005-.002

Trait 2: h2=.60, CV MAF .01-.05

Trait 3: h2=.60, CV MAF .10-.50