Resampling with QTL
create multiple data sets

- simulated data markers, QTL, trait phenotype recombination model; trait model even spacing: remove effect of spacing how to simulate crossovers/recombinations?

- parametric bootstrap simulate data using estimates capture real features

- analyze as if it were real create multiple data sets under same conditions calculate estimates of parameters for each run summarize estimates: how close to known? bias: compare sample mean with known value variance: how large is SD relative to mean?

statistical sampling
small sample size, effects, heritability overestimates of effects, $r^2, h^2$ Beavis (1994, 1998); Utz Melchinger (1994); Georges et al. (1995)

empirical study in maize of yield separate studies of same cross QTL located on different chromosomes is this important?
simulation study with 20 QTL all of small effect different QTL detected in each run what is goal of QTL study?

Resampling with QTL
tests and confidence intervals

- basic assumptions model may not be correct but it provides useful test statistic resample the data directly

- permutation tests sample all data without replacement under null hypothesis of no linked QTL test for presence of linked QTL shuffle trait, or residuals given cofactors Churchill Doerge work

- classical bootstrap sample directly from data with replacement under alternative hypothesis of QTL repeat multiple times data-driven confidence intervals Visscher et al. (1996)

Resampling with QTL
subdivide data to examine power & bias

- cross-validation estimate with part of data predict for rest of data training set / test set idea leave-one-out find bias, variance and power Melchinger et al. (1998); Utz et al. 2000; Zeng et al. 2000

- jackknife estimate of variance leave one datum out & compute statistic run through all data $i = 1, \ldots, n$ QTL location and/or effect mean square deviation around full data estimate Tukey (1958); Efron (1982)
Resampling with QTL cross-validation studies

Melchinger Utz Schöhn (1998)
estimation/training set of 344
validation/test set of 107
proportion of variation explained by QTL

Utz Melchinger Schöhn (2000)
5 samples, 4 environments
two independent samples
various E, G, G*E cross-validations
shows overestimate of explained variance

Zeng et al. 2000
estimate genotypic values
compare genotype & phenotype
predict genotypic value from other set
compare predicted to actual

---

Resampling with QTL
sampling from the QTL likelihood

- Monte Carlo sampling
  sample from some distribution
  use sample to study properties
  bias and variance of estimators
  estimate functions of parameters
  numerical integration (e.g. expectation)
direct or indirect samples (importance sampling)

- Markov chain Monte Carlo sampling
  may be simpler than MC to implement
distribution may require normalizing constant
distribution known only up to scale
which may be very difficult to compute
use tricks of Markov chain