The power of replications in difference tests

Per Bruun Brockhoff

Dept. of Mathematics and Physics

The Royal Veterinary and Agricultural University

Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark

Result

A few reps á day makes the low power go away!

Settting

- Panel size: n
- Each assessor perform k difference(triangle) test
- The N=nk binomial test is OK (Meyners & Kunert, 1999)
- The power of this test is unknown!

Outline

- Introduce the power of the usual binomial test with replications
- Present and compare different statistical models for replications
- Calculate and compare the power within these models
- Give limits for this power

The power of the binomial test

- The probability of claiming a difference when in fact it is there:
  - $P(X=x_{critical})$, where
  - X=total number of correct answers
  - and assuming some ”alternative situation”:
  - some (at least one) on the individual p_is are larger than 1/3.

Model types for the alternative

- Beta-binomial (Ennis & Bi, 1998)
- Generalized Linear Mixed Models (Brockhoff, 1997, Hunter et al., 2000)
- Binomial mixture models (Meyners & Kunert, 1999)

The Beta-binomial model:

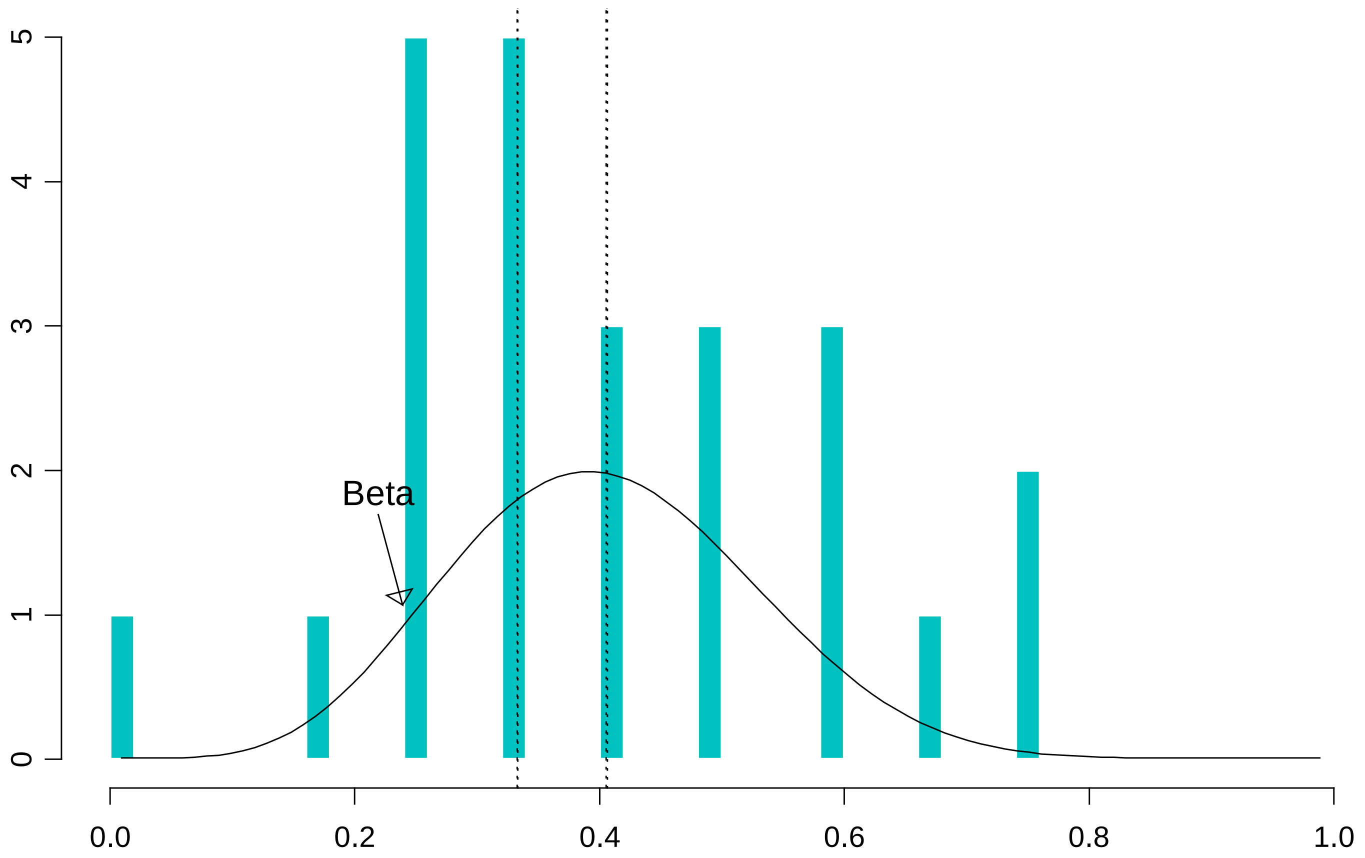
For each assessor:

$$X_i \sim \mathrm{bin}(k, p_i)$$

The individual probabilities are randomly distributed:

$$p_i \sim \mathrm{Beta}(\alpha, \beta)$$

Data: n=24, k=12



The Generalized Linear Mixed Model (GLMM)

For each assessor:

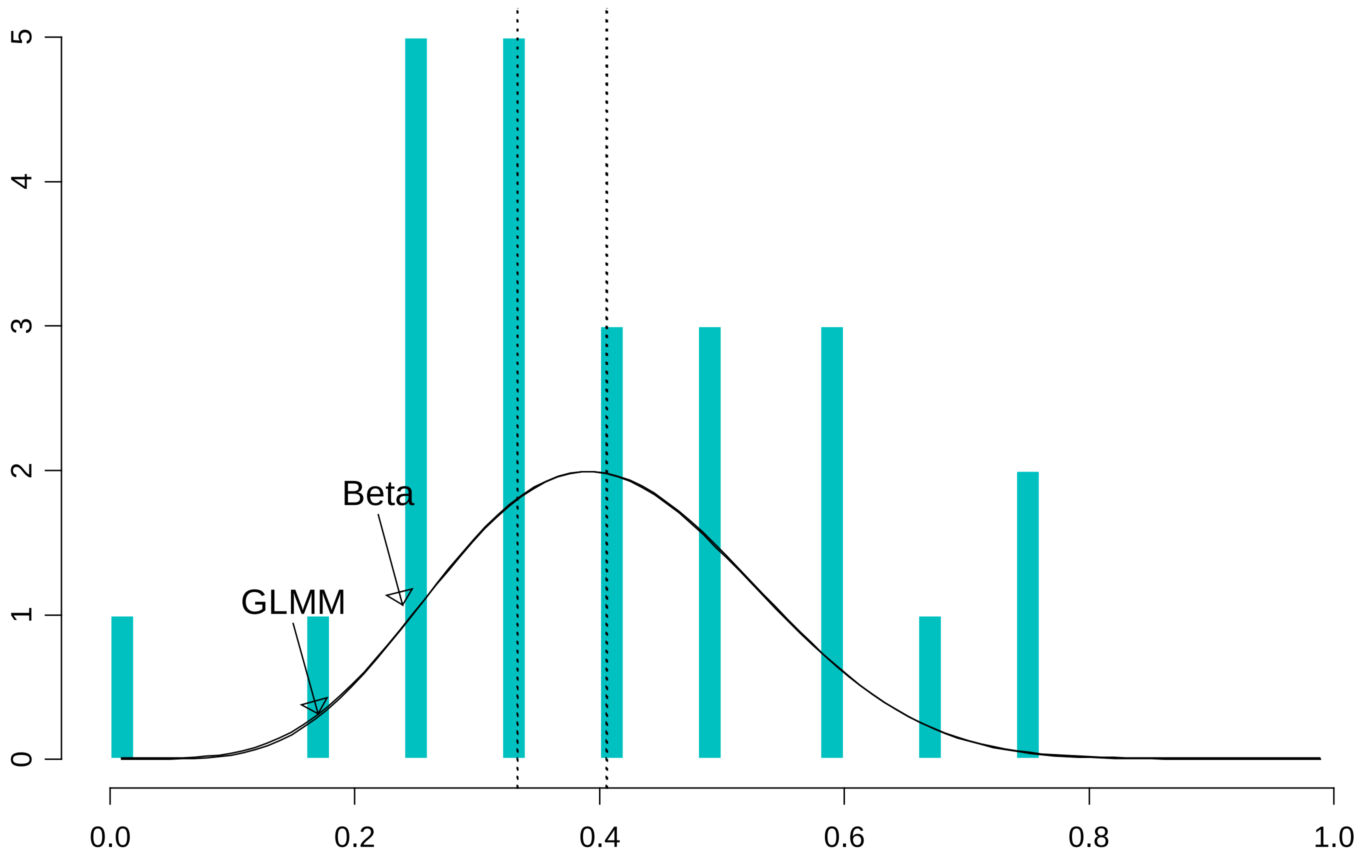
$$X_i \sim \text{bin}(k, p_i)$$

The individual probabilities are randomly distributed:

$$p_i = \Phi\left(d + Z_i\right)$$

$$Z_i \sim N(0, \sigma^2)$$

Data: n=24, k=12



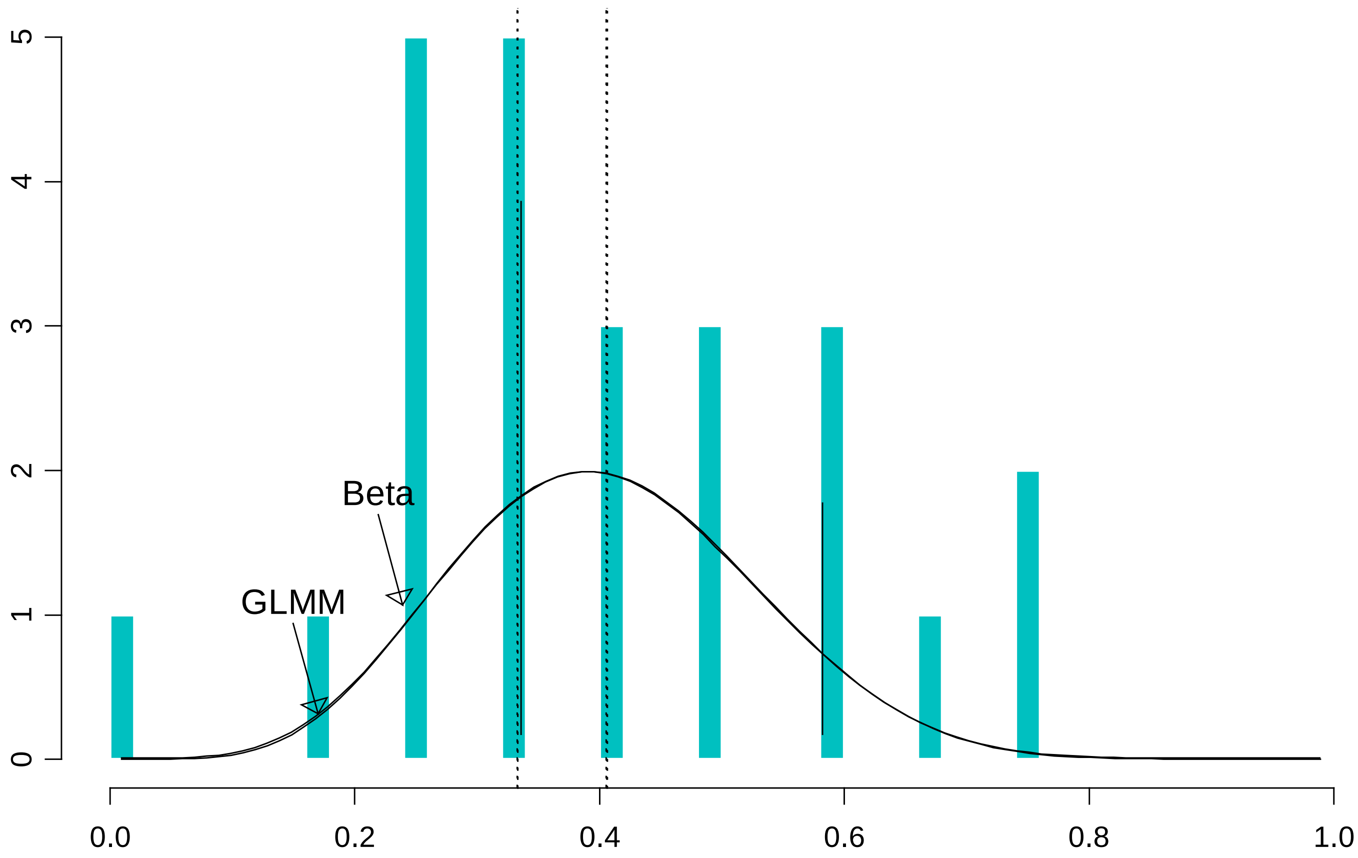
The binomial mixture model

For each assessor: $X_i \sim \mathrm{bin}(k, p_i)$

The individual probabilities are randomly distributed:

$$p_i = \begin{cases} \frac{1}{3} & \text{with probability 1-?} \\ \frac{1}{3} + (1 - \frac{1}{3})\pi & \text{with probability ?} \end{cases}$$

Data: n=24, k=12



The Corrected Beta-binomial model:

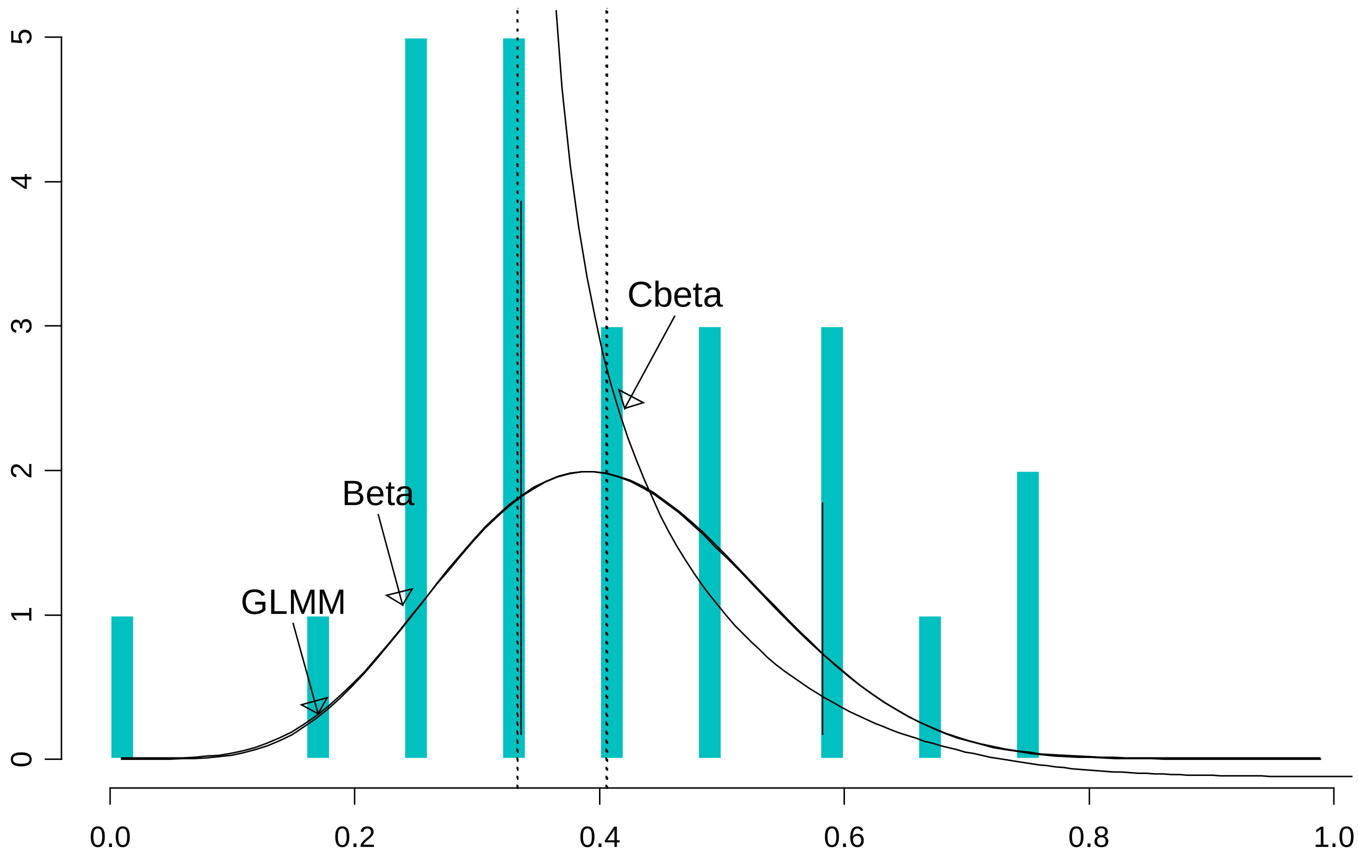
For each assessor: $X_i \sim \text{bin}(k, p_i)$

The individual probabilities are randomly distributed:

$$p_i = \frac{1}{3} + (1 - \frac{1}{3})\pi_i$$

$$\pi_i \sim \text{Beta}(\alpha, \beta)$$

Data: n=24, k=12



The Corrected Generalized Linear Mixed Model (CGLMM)

For each assessor:

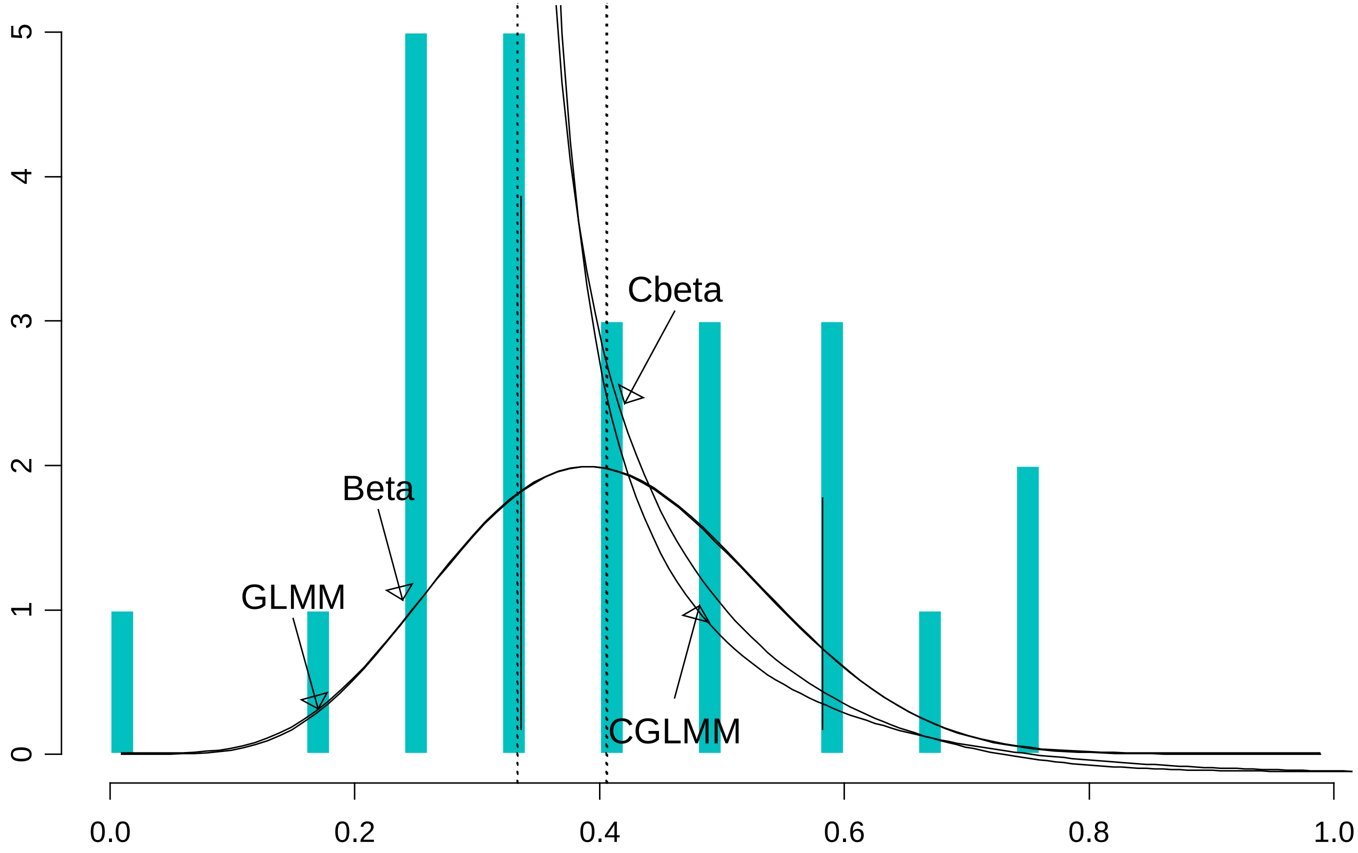
$$X_i \sim \mathrm{bin}(k, p_i)$$

The individual probabilities are randomly distributed:

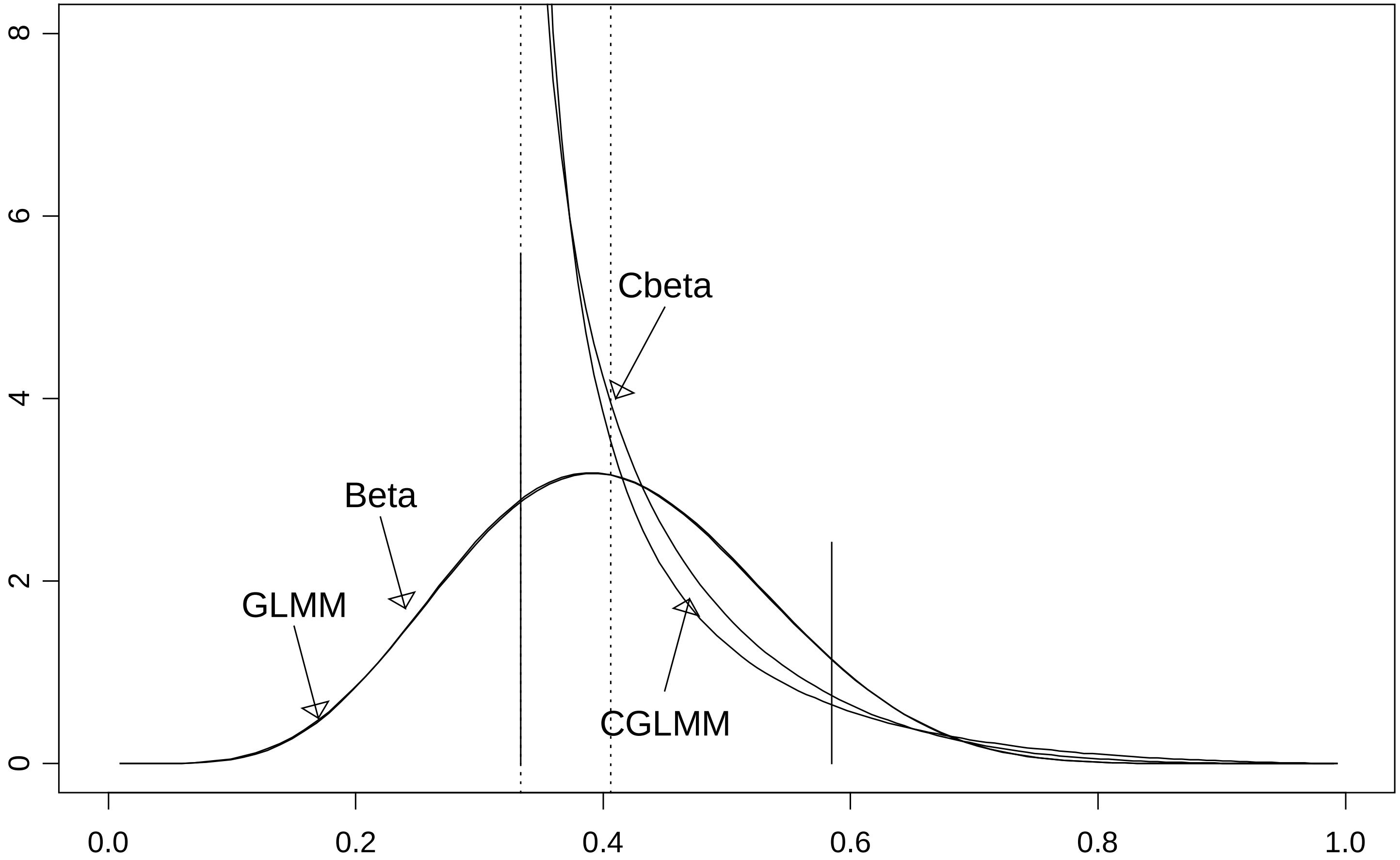
$$p_i = \frac{1}{3} + (1 - \frac{1}{3})\Phi(d + Z_i)$$

$$Z_i \sim N(0, \sigma^2)$$

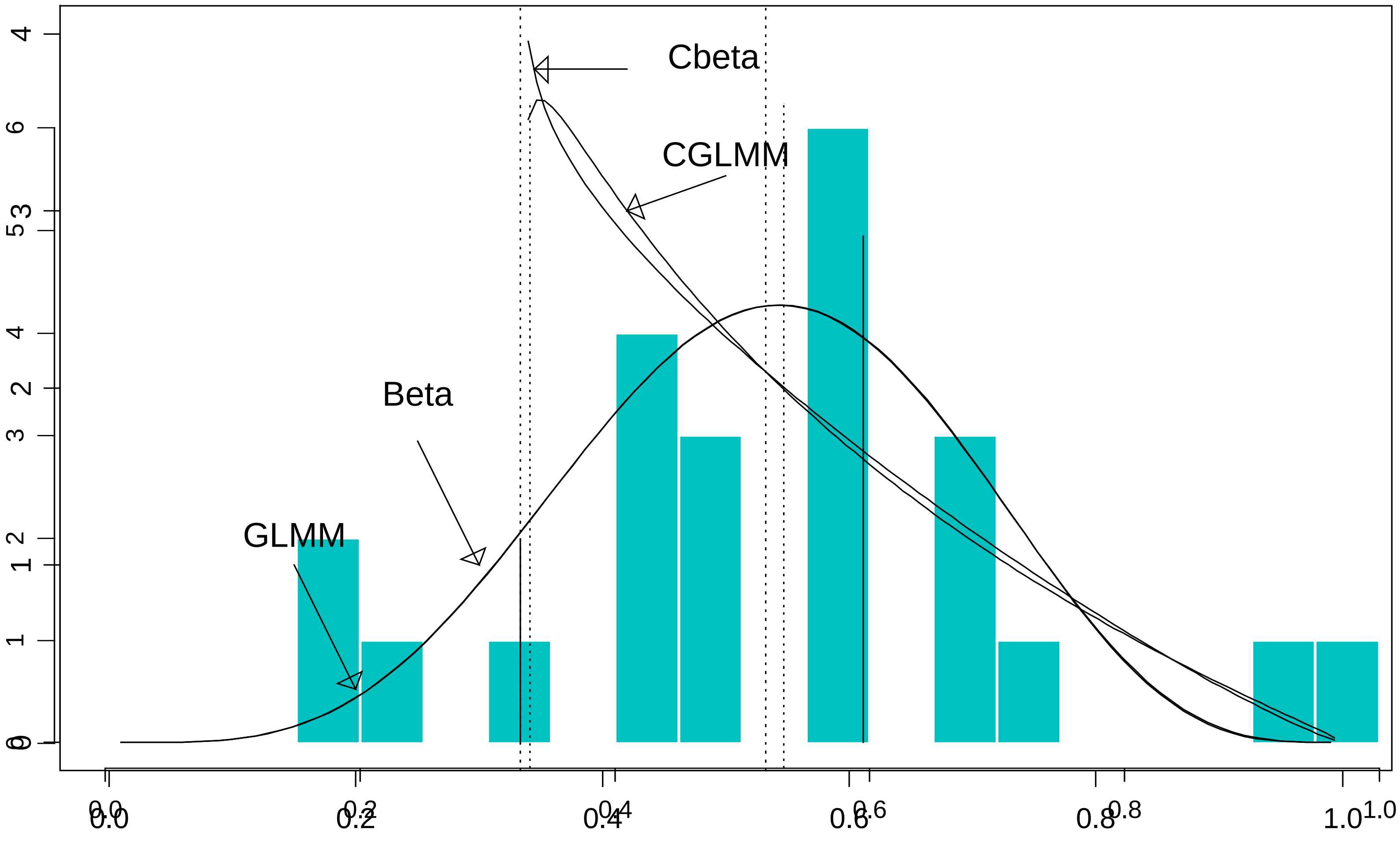
Data: n=24, k=12



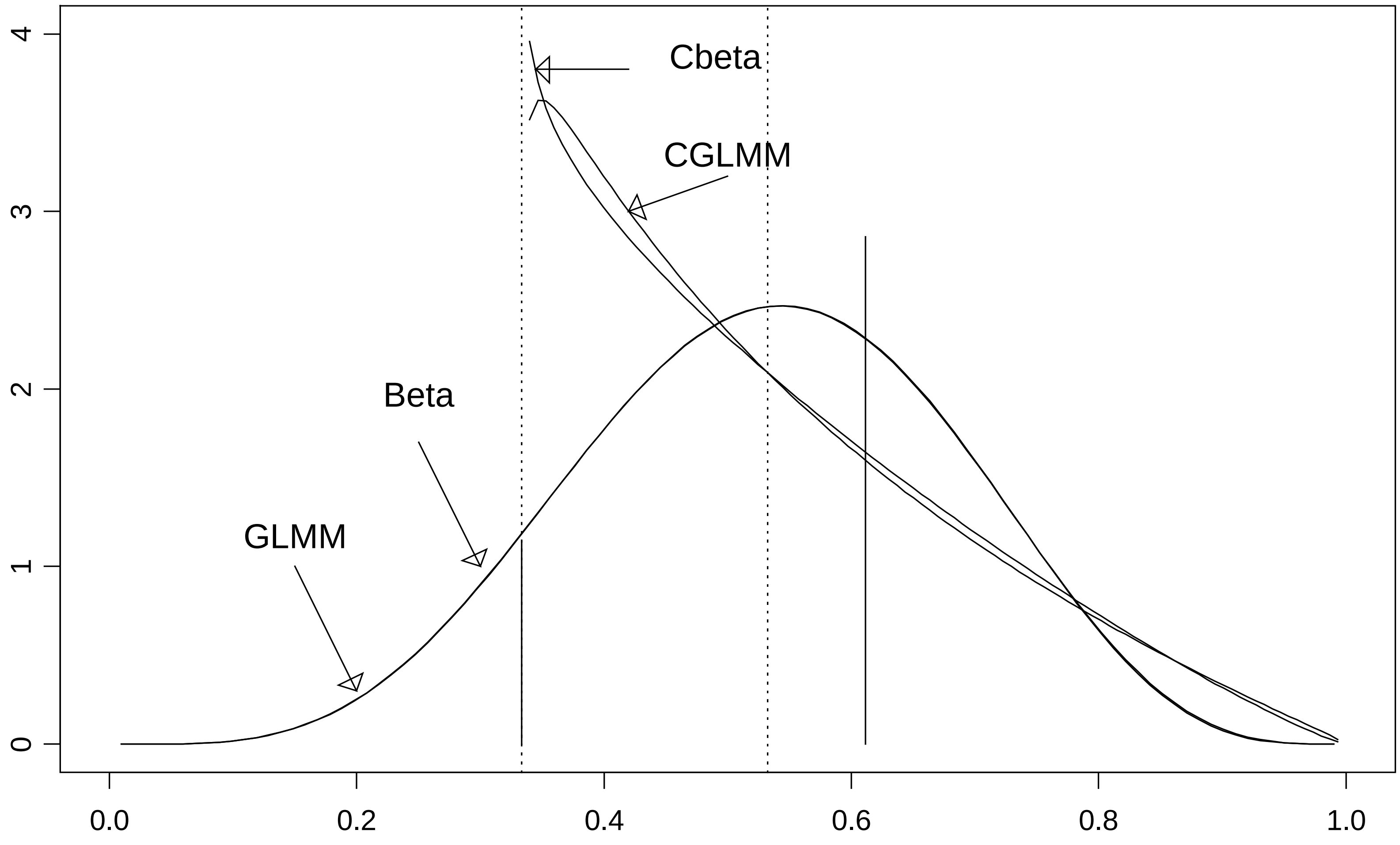
Model comparison



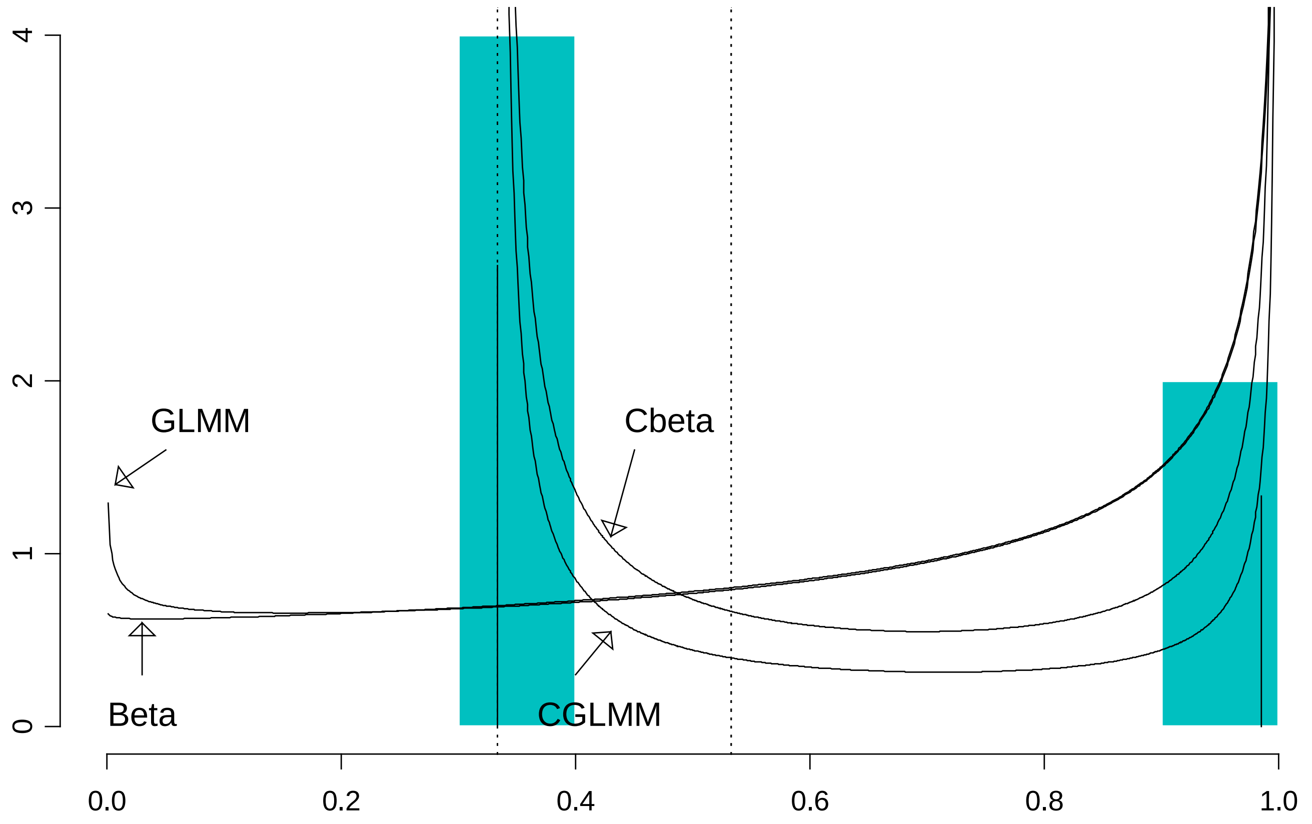
Data: n=23, k=12



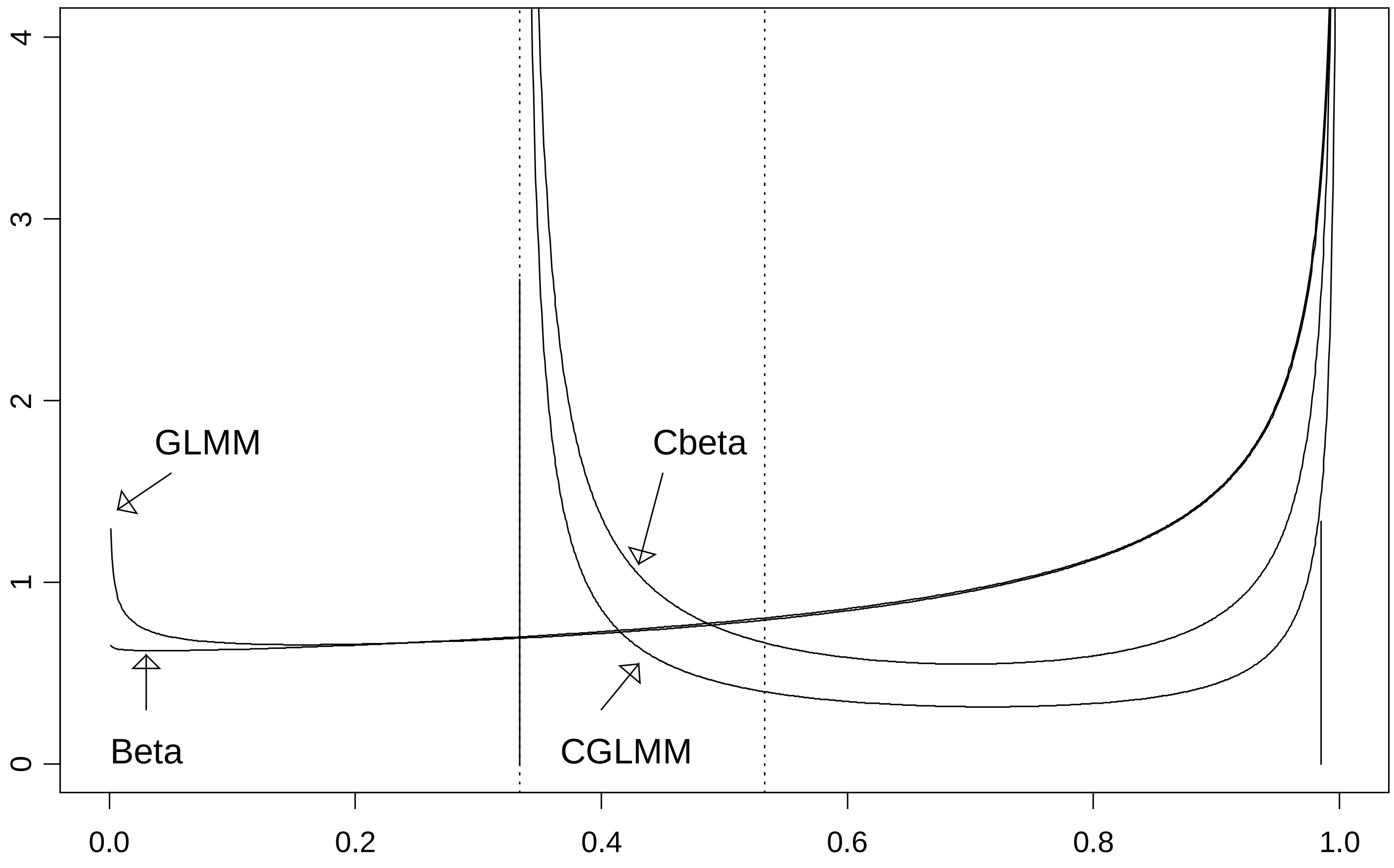
Model comparison



Data: n=6, k=100



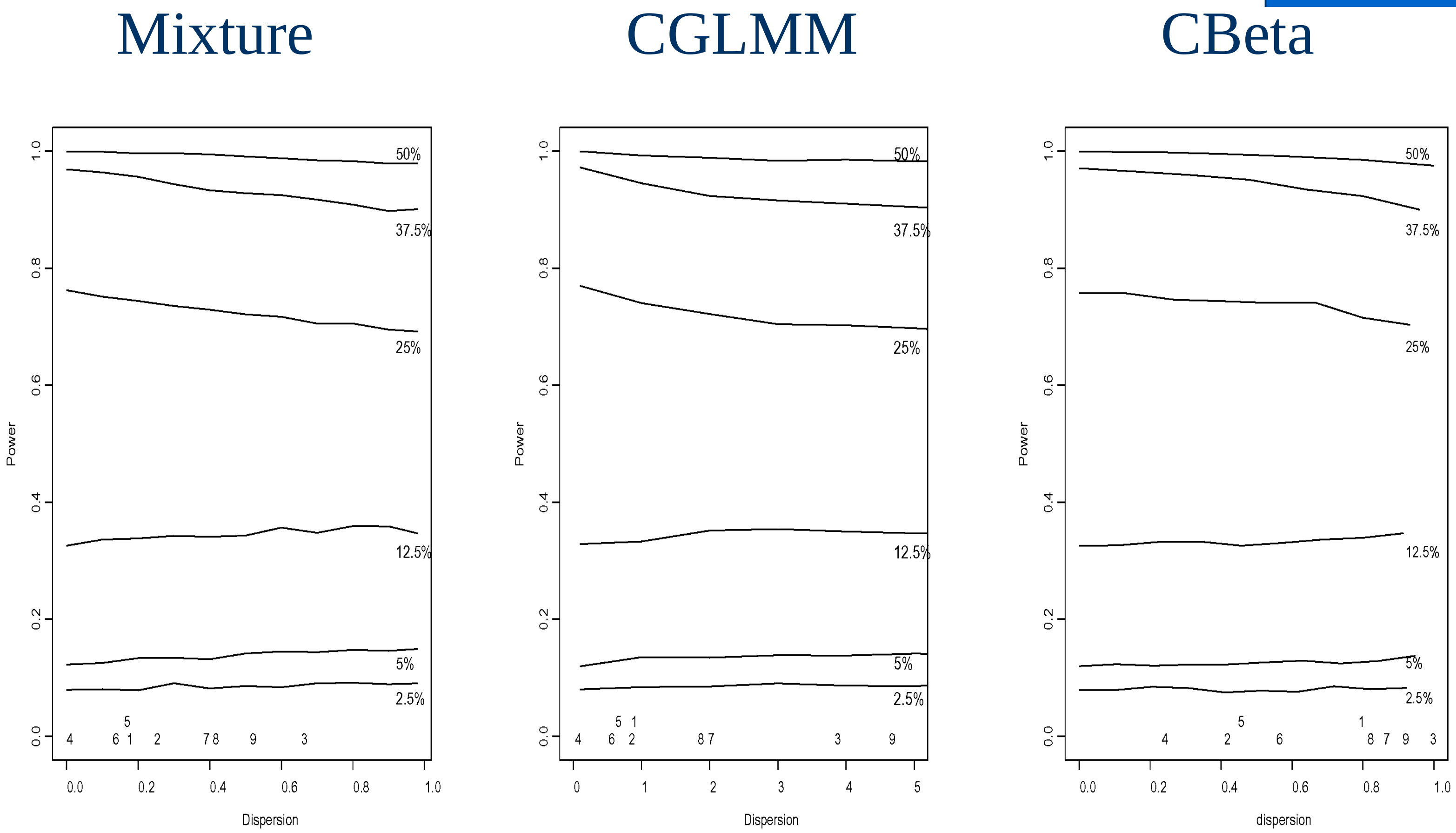
Model comparison



Calculation of power

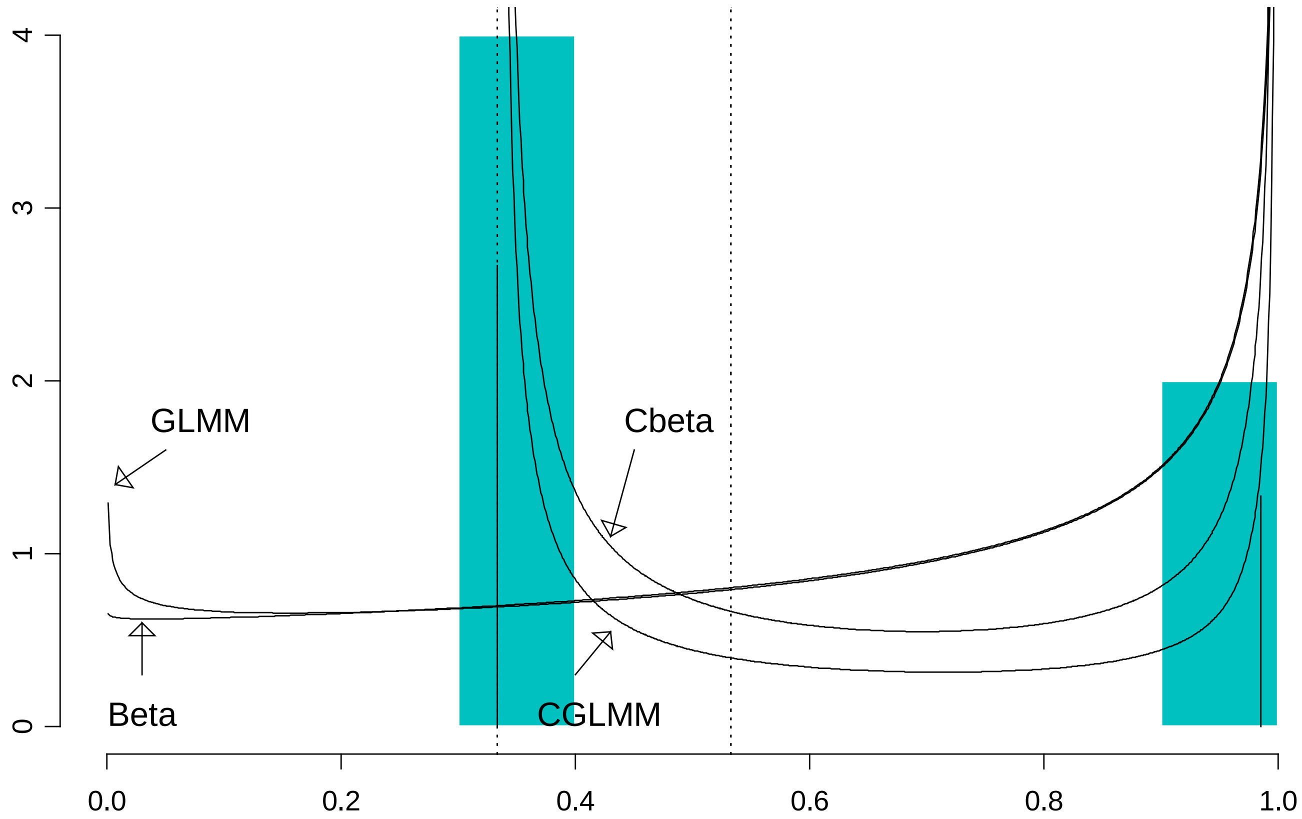
- Calculation of $P(X=x_{critical})$
- Done by Monte Carlo methods:
  - Simulate X (total number of correct answers)
  - Count how often X is larger than or equal to the critical value
- Easy if software has built in functions: normal, beta, binomial. (e.g. Splus/R/SAS)

Power, n=12, k=4



Mixture
CGLMM
CBeta
Power
Dispersion
dispersion
50%
37.5%
25%
12.5%
5%
2.5%
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5

Extreme data: n=6, k=100



The common limit model

- All 3 models converge to the same limit/extreme situation:
- All individuals are either 100% discriminators or non-discriminators (guessers)

The common limit model

For each assessor: $X_i \sim \text{bin}(k, p_i)$

The individual probabilities are randomly distributed:

$$p_i = \begin{cases} \frac{1}{3}, & \text{with probability } 1-\gamma \\ 1, & \text{with probability } \gamma \end{cases}$$

Monte Carlo for the common limit model

- Fix the effect size
- Simulate n binomial(1, ?)s
- For each outcome of 1 set x_i =k
- For each outcome of 0 simulate a binomial(k,1/3)
- Count the number of times X becomes as large as the critical value.

Limits of power for triangle test

Level 5%, 37.5% effect

n:	k=1	k=2	k=3	k=4	k=5
12 .	40%	70%	81%	90%	91%
38 39	88% 92%				
48	97%				

Summary

- There is no big difference between the different ”complicated models” to handle replications
- The loss of power by substituting assessors by replications is remarkable small
- Given the panel size, a few replications increase the power considerably
- Tables of limit power is given for some situations – a simple Monte Carlo method for other.

Some additional insights

- Computationally, the mixture model is the easiest to handle:
  - The EM-algorithm easy to implement AND gives the option of ”fuzzy clustering” of the individual assessors
- More powerfull test than the independent binomial exist!