Preferred citation style

Hess, S. and K.W. Axhausen (2005) Distributional assumptions in the representation of random taste heterogeneity, 5th STRC, Ascona, March 2005.

Distributional assumptions in the representation of random taste heterogeneity

Stephane Hess
Imperial College London & RAND Europe

Kay Axhausen
ETH Zurich

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Imperial College London

Mixed Logit

- No a priori constraints on choice of f(β|Ω)

$$P_{ni} = \int_{\beta} \left(\frac{e^{f(\beta, X_{ni})}}{\sum_{j=1}^{I} e^{f(\beta, X_{nj})}} f(\beta|\Omega) \right) d\beta$$

- No closed-form solution for choice-probabilities
  - ➔ Need simulation in estimation and application

Issues for the a-priori distribution

- Generally no a priori knowledge about true distribution
- Limited set of distributions available in standard estimation packages
- Most researchers/practitioners use *Normal*
- Limited success with using alternative distributions

Current choices

- Normal
- Log-normal
- Triangular
- Sb

Effect of poor distributional assumptions

- Unbounded:
  - risk of wrong conclusions wrt existence of counter-intuitively signed coefficients and related trade-offs (e.g. VTTS)
  - risk especially great with symmetrical distributions
- Strictly bounded:
  - can prevent data or model specification issues to manifest themselves

Effect of poor distributional assumptions

- Symmetry
  - Over/underestimation of the tails of the distributions
  - Bias in the mean recovery with skewed underlying distributions
- Consideration of mass points

Framework

- Recovery of true distribution
  - Normal
  - Truncated Normal
  - Censored Normal
  - Normal with Mass
  - Lognormal
  - S_B

- Four different settings per distribution (24 exp)

- 8,000 independent draws per experiment (ExpertFit used for analysis)

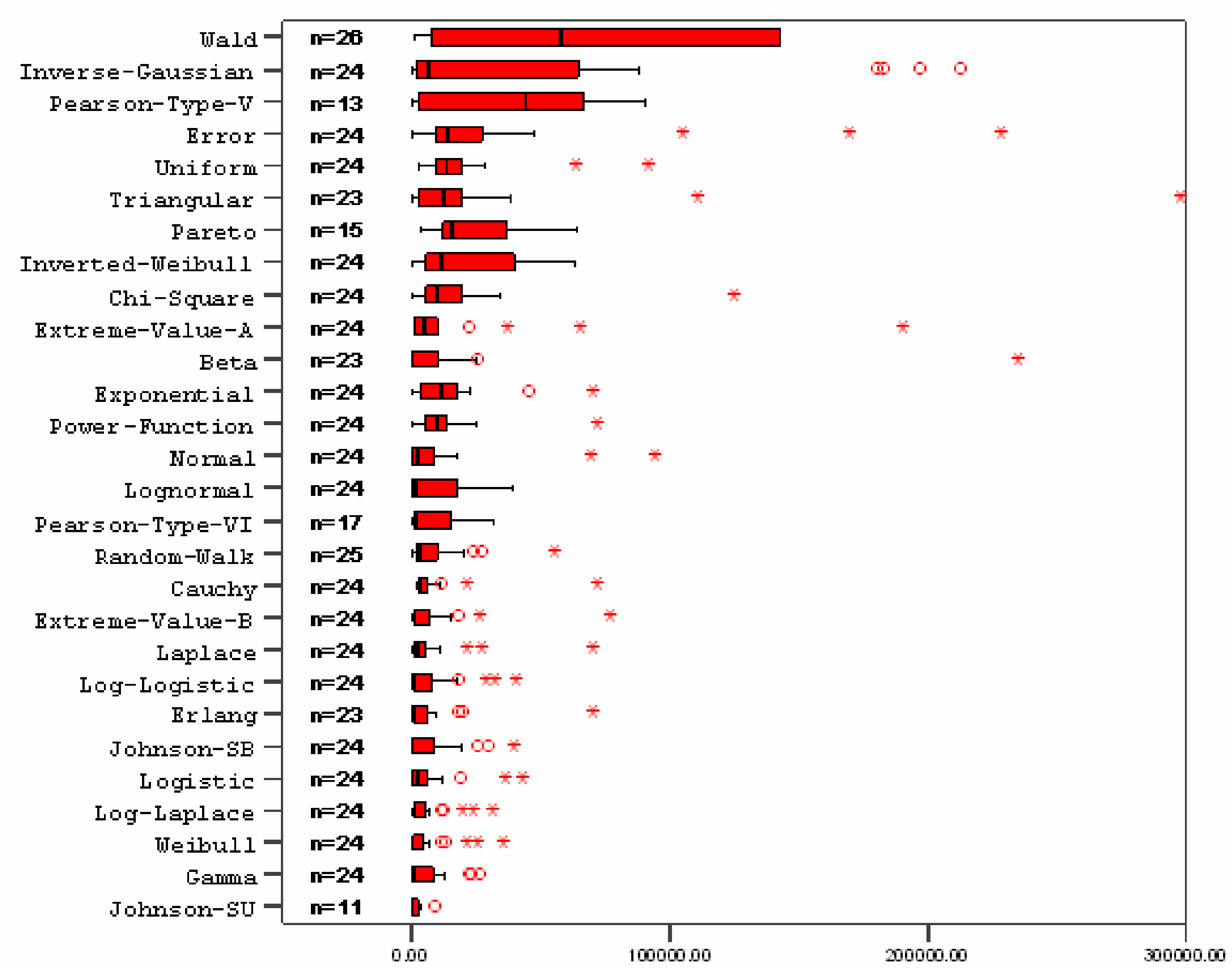
Distributions

- Unbounded (*Normal*, S_u, Cauchy, Error, Exp. Power, extreme value A&B, Laplace, Student‘s t, Logistic)
- Bounded on one side (*Log normal*, Chi2, Erlang, F, Gamma, Inverse Gaussion & Weibull, Log-Laplace, Log-logistic, Pareta, Pearson V&VI, Random walk, Wald, Weibull)
- Bounded on both sides (*Triangular, Uniform,* S_b, Beta, Power function)

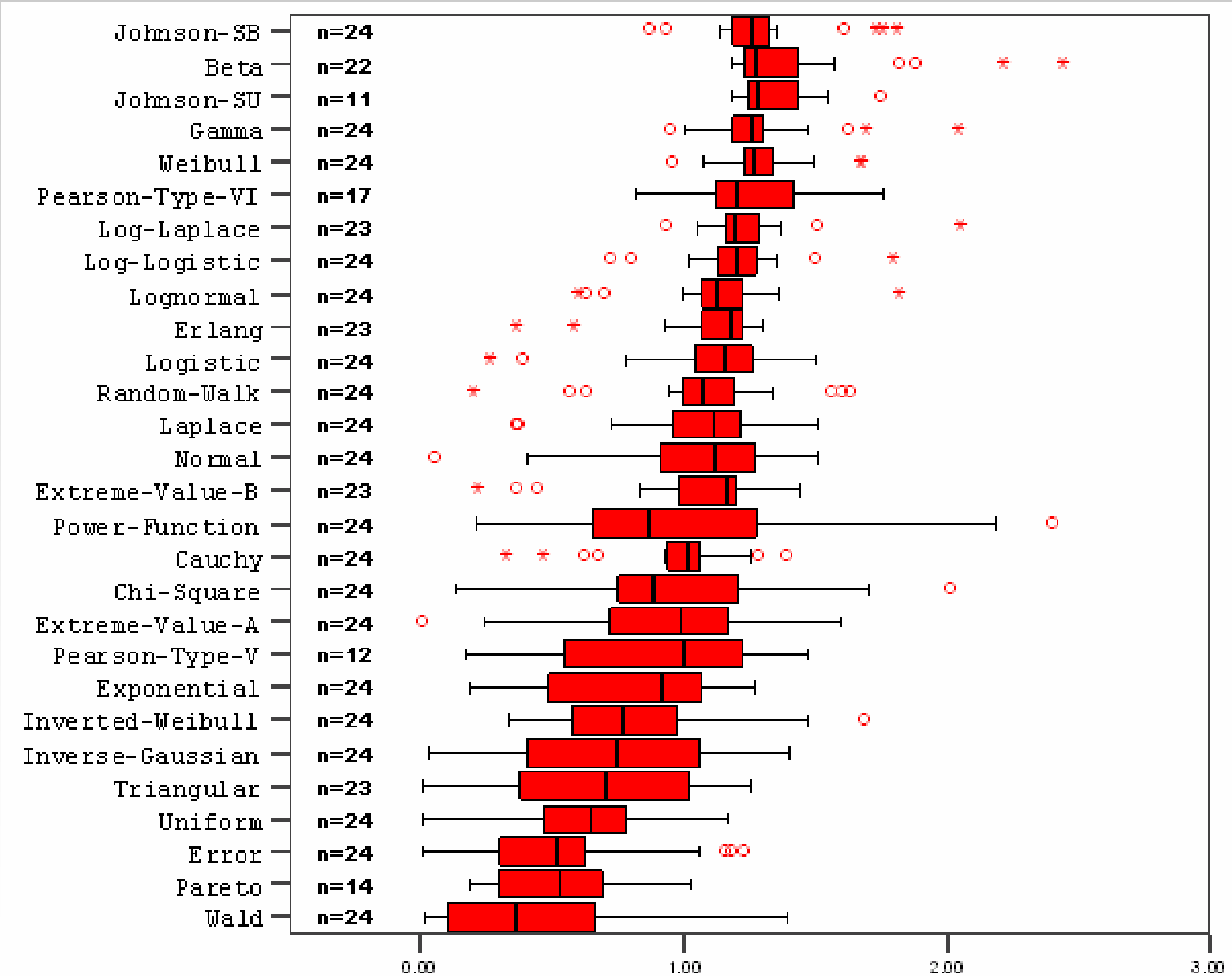
Comparison Criteria

- Equal chi-squared test
- Recovery of true mean & variance
- Likelihood function of observed values
- Recovery of lower & upper octiles
- Difference in density functions, mean & max
- Weighted sum of differences across lower & upper octiles, quartiles and median

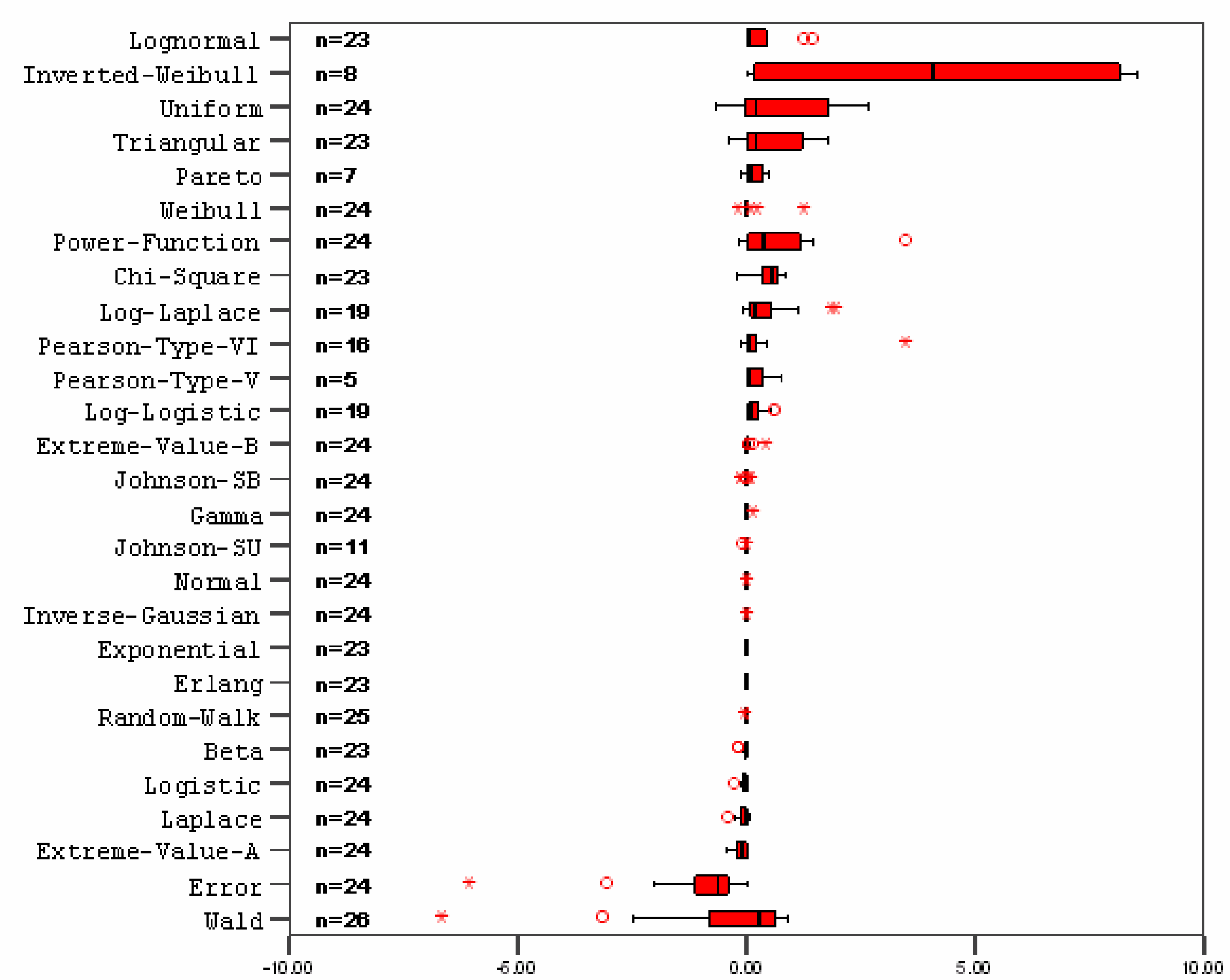
Equal probability chi-square test



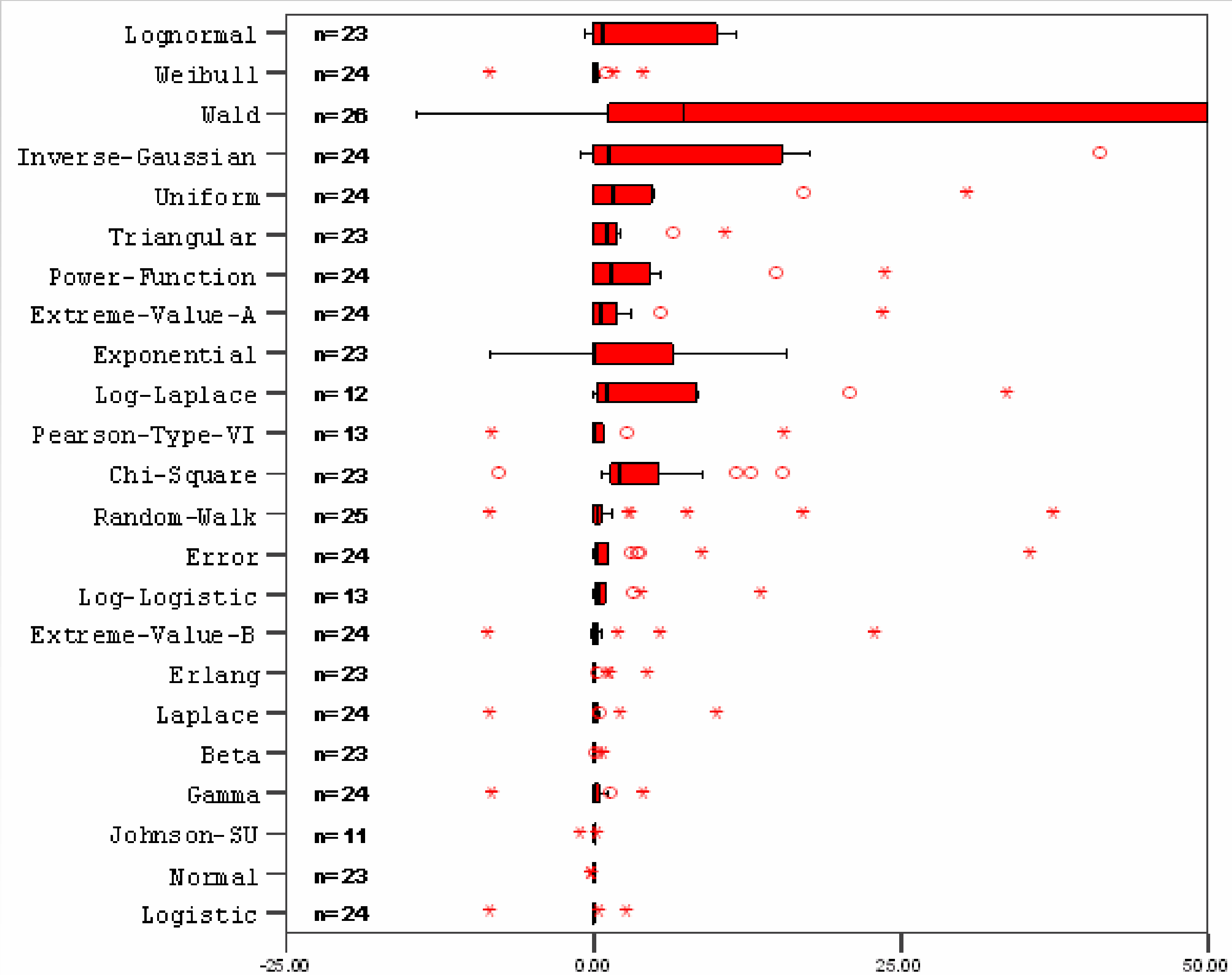
Relative likelihood per observation



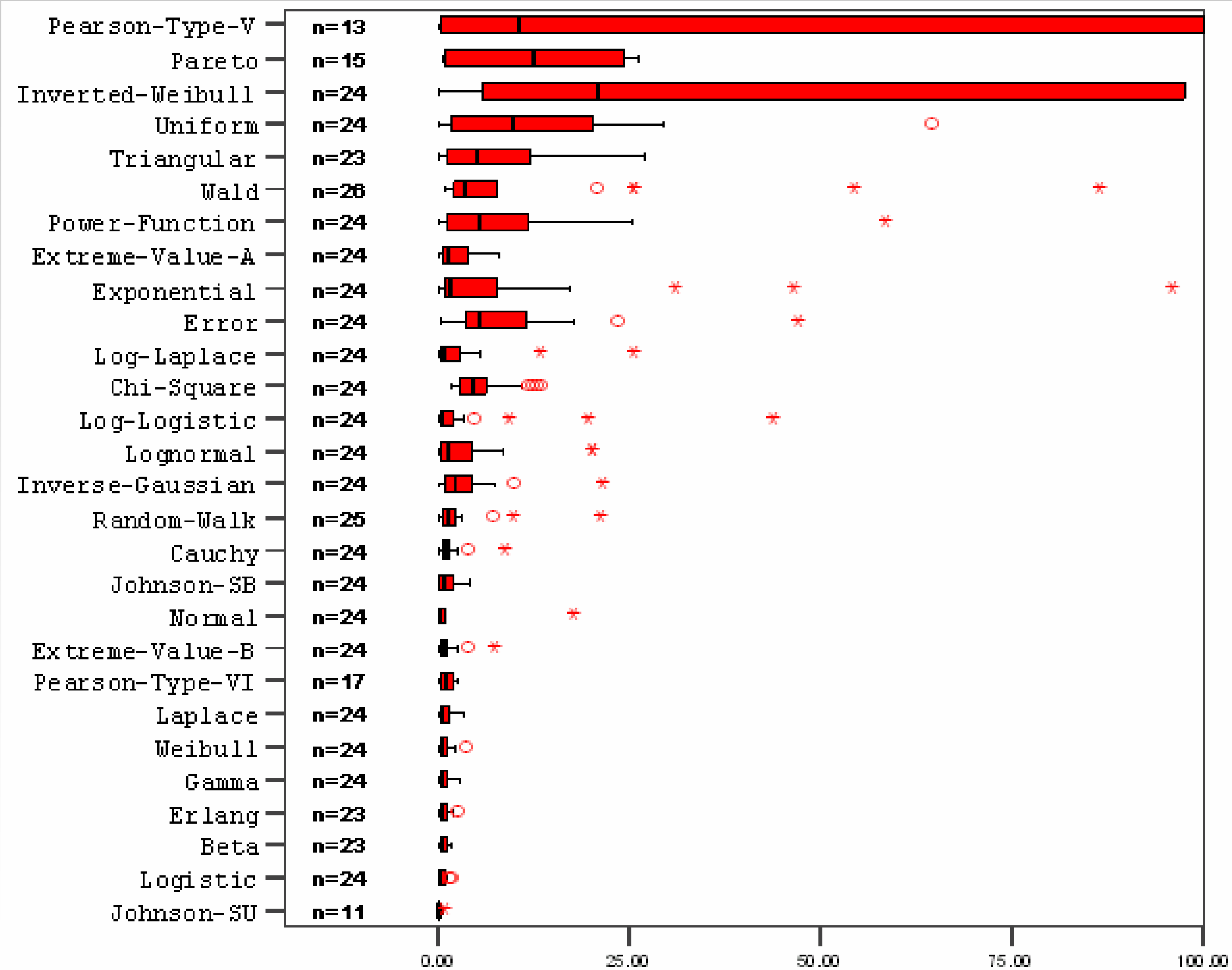
Recovery of mean



Recovery of variance



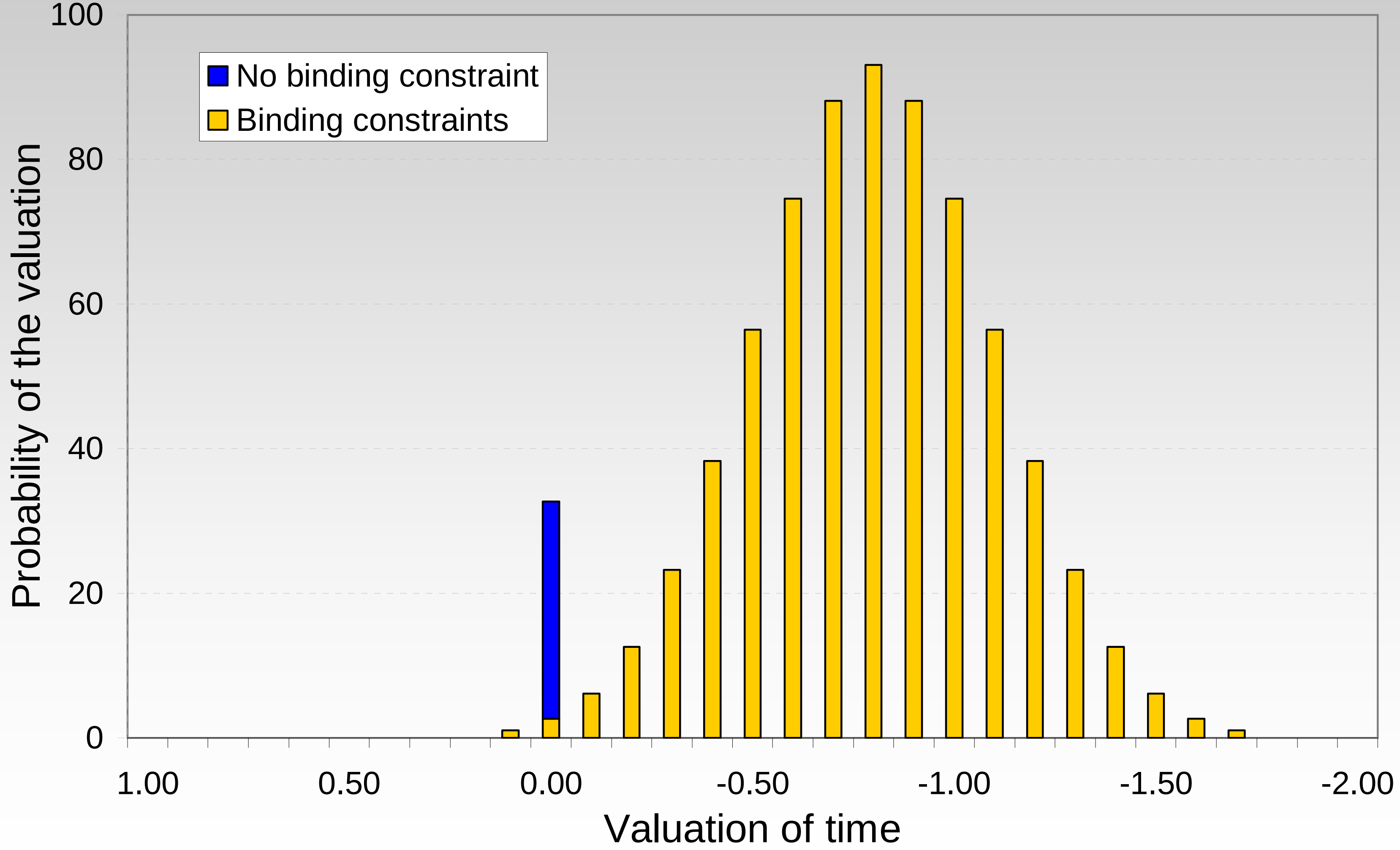
Weighted difference, 5 points



Conclusions

- Group of well performing distributions (Su, Sb, Beta, Gamma, Weibull)
- Normal not too bad, but a priori difficulties
- Software availability
- Still, what to do .....

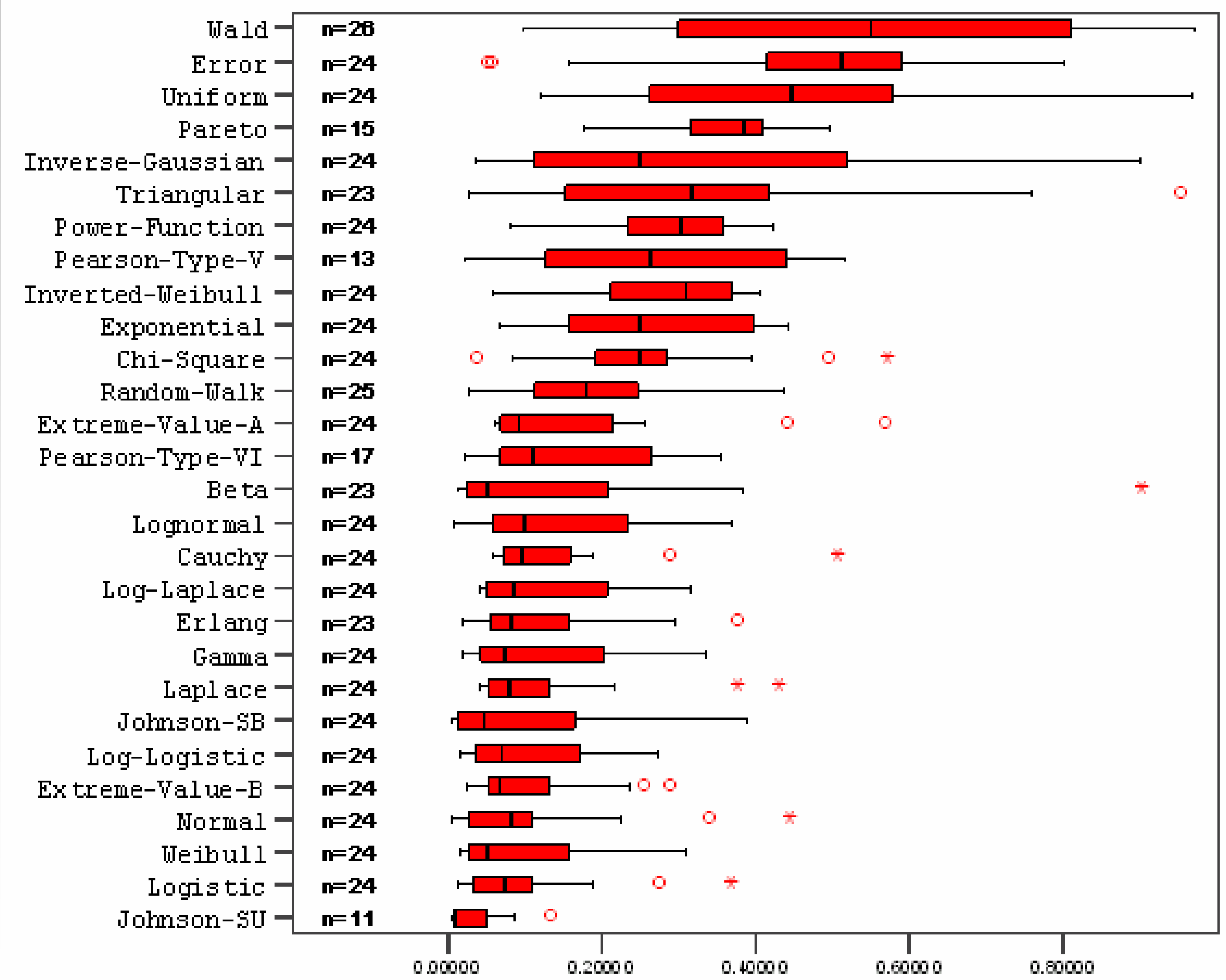
In this case ?



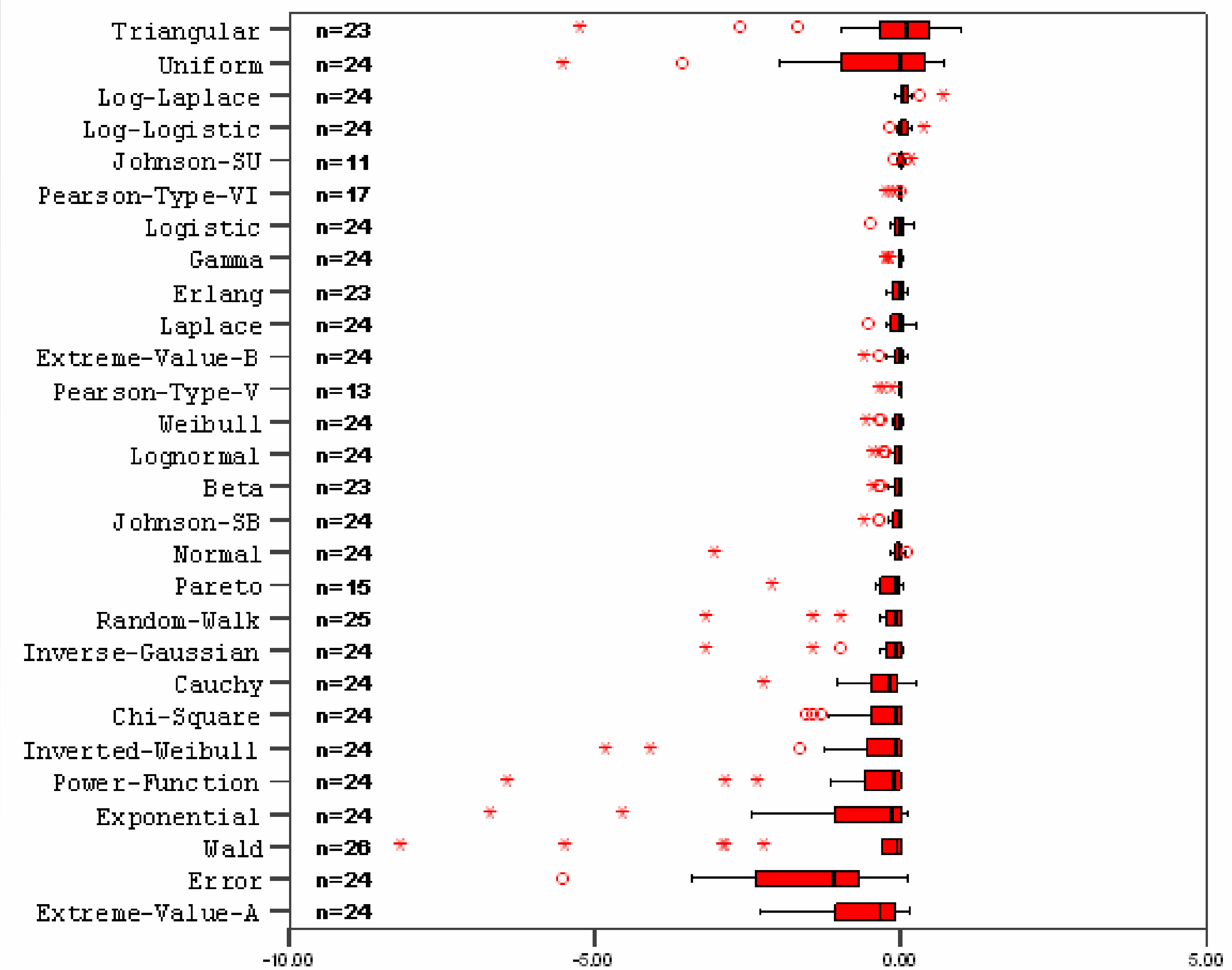
No binding constraint
Binding constraints
100
80
60
40
20
0
Probability of the valuation
1.00
0.50
0.00
-0.50
-1.00
-1.50
-2.00
Valuation of time

Appendices

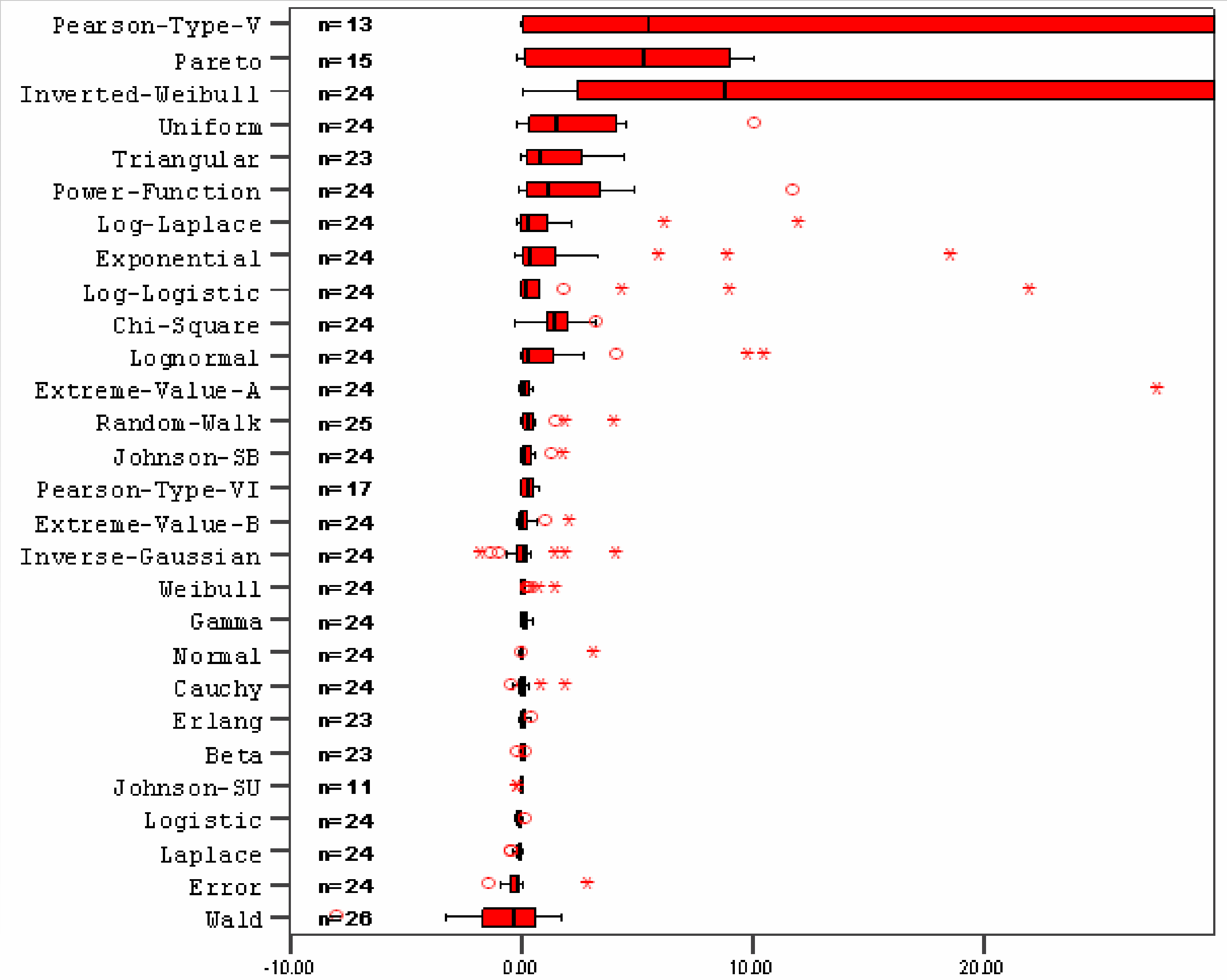
Density function maximum difference



Difference in lower octile



Difference in upper octile



Density function mean difference