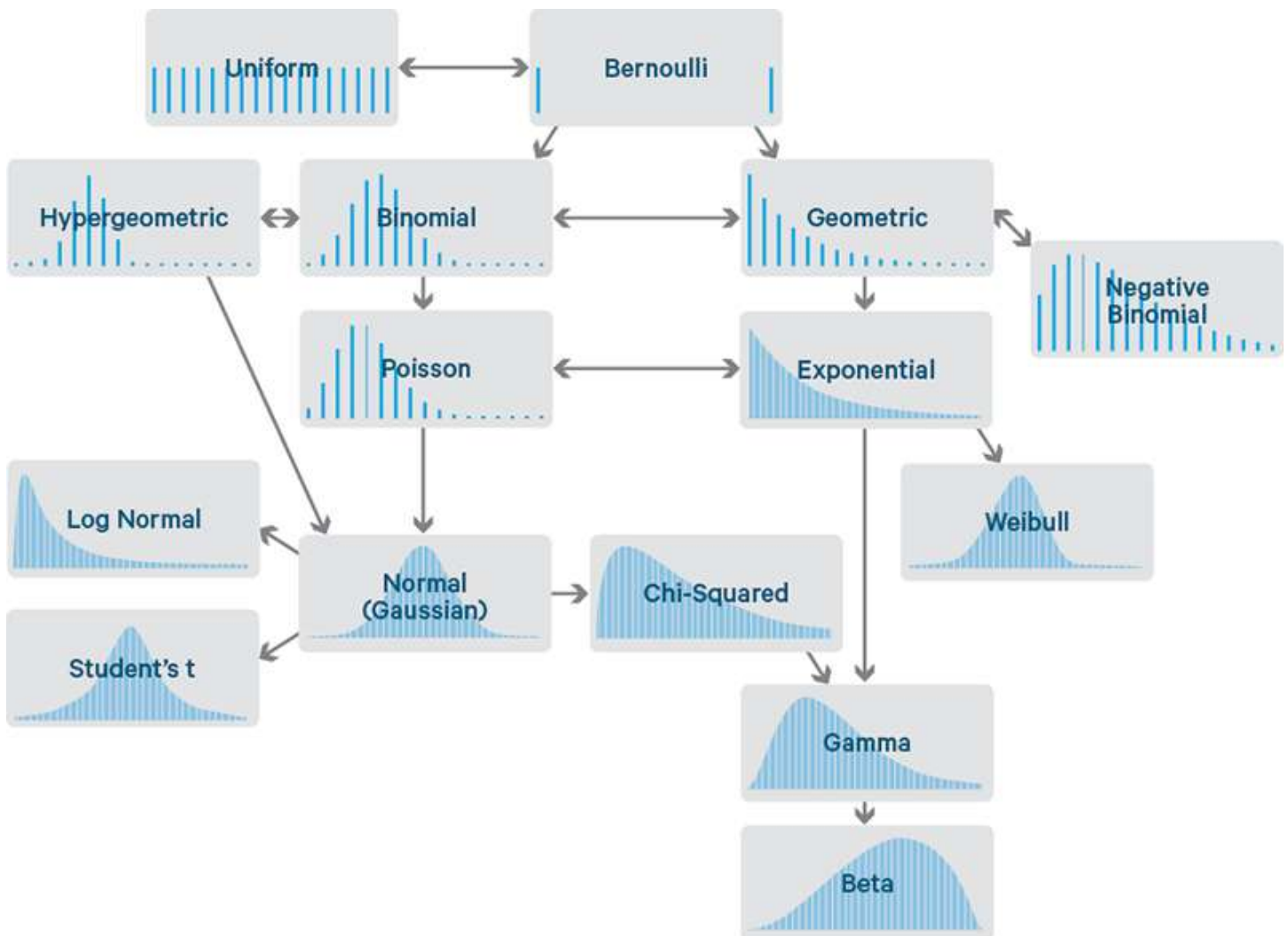


Probability density functions for common distributions



Kom ihåg newdata vs data Krav för LDA: data ska vara normalfördelad i varje klass och ha samma varians i båda klasserna. (Multivariate normality) Linear Discriminant Analysis. Dock kan LDA vara en bra metod för att klassificera.

För att göra tree:s så behöver man göra om target till factors om den är non numeric. Använd: `as.factor`

När man gör general error estimation (out of sample error) så använder man både validation och training data och sen testar på test-data

"Model to give to the user" : train on all data (train + valid + test), estimated error is the generalization error

`Knn.cv` för crossvalidation `knn`, library

Assumption Naive Bayes: You assume strong independence of the features

Logistic regression: `glm(formula, data=data, family = "binomial")`

Link funktioner (canonical link): logit : $\ln(p/(1-p))$ eller som in R $\log(p/(1-p))$ Log: $\ln(p)$ eller $\log(p)$