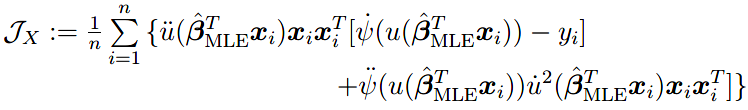
**OSMAC for GLM**

Subsampling with replacement & hard pooling pilot <> (can improve by Poisson sampling & efficient pooling)

**GLM setting:**

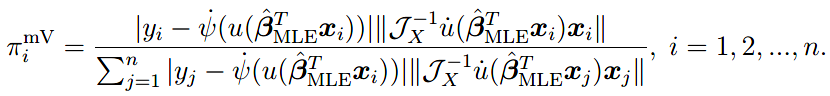
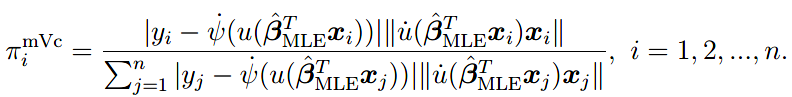


**Observed information:**



# Optimal Subsampling

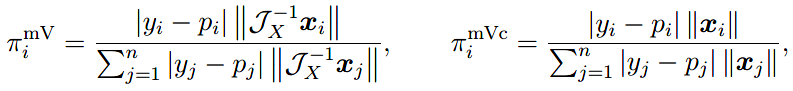
1. General GLM:

Observed information can be replaced by empirical information:



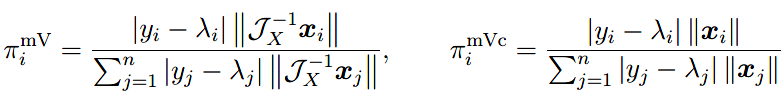
1. Logistic regression:







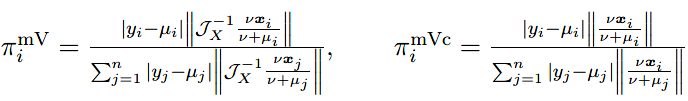
1. Poisson regression:







1. Negative-binomial regression (NBR):







# Implementation Issues

1. Subsample size: proportion to p ()
2. Issues for weighted estimators <can solved by unweighted estimators with bias correction>
   1. Small sub-prob 🡪 large weight 🡪 sensitive
   2. Setting threshold (truncation)

