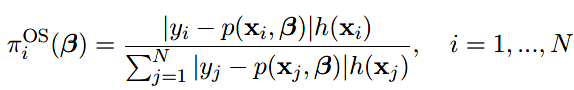
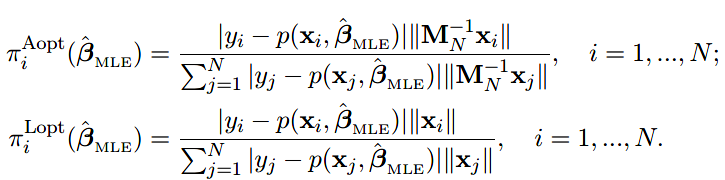
**Unweighted OSMAC + Poisson Sampling + pilot pooling**

# OSMAC Review & Weighted Estimators

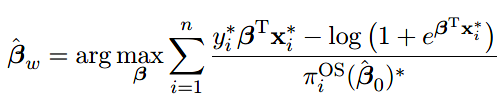
**SSP for logistic regression**

1. General form: 
2. A-optimal & L-optimal:



**Weighted Estimator vs. Unweighted Estimator**

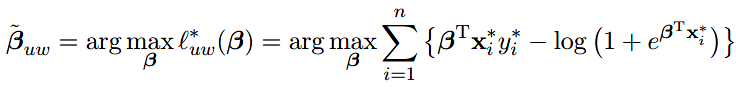
1. Weighted:



* 1. Unbiased
  2. Large 🡪 more information 🡪 smaller weight 🡪 reduce contribution 🡪inefficient
  3. Small 🡪 larger weight🡪if include, inflate Var-Cov <**solve by truncation in OSMAC\_GLM**>

1. Un-weighted:
   1. Biased 🡪 need correction
   2. More efficient, smaller Var-Cov

# Unbiased Estimator (unbiased)


1. is asymptotically unbiased (-consistency)
2. More efficient than weighted estimator (Var-Cov matrix smaller)

# Poisson Subsampling

**Sampling with replacement:**

1. Iid conditional on full data, but may not independent unconditionally
2. Generate all subsampling probabilities at once & store (need large RAM)
3. With replacement 🡪 1 pt, multiple times 🡪 inefficient

**Poisson subsampling:**

1. Algorithm:
   1. Generate 
   2. Include pt if 
2. Independent unconditionally & conditionally
3. No need to generate all subsampling probs at once
4. 1 pt, only 1 time at most 🡪 efficient

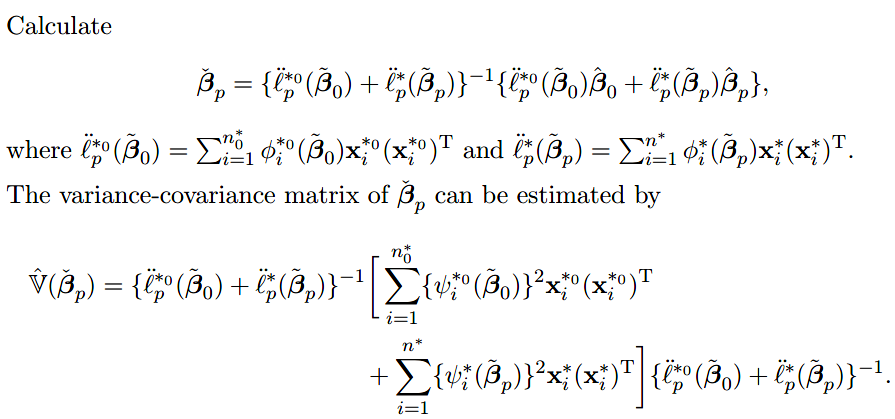
Var-Cov matrix smaller than sub-sampling with replacement

1. Limitation: subsample size, random. Expectation of n

# Pooling the Pilot Result

Aggregation procedure in the divide-and-conquer method <more computationally efficient>

Take Poisson subsampling for example:

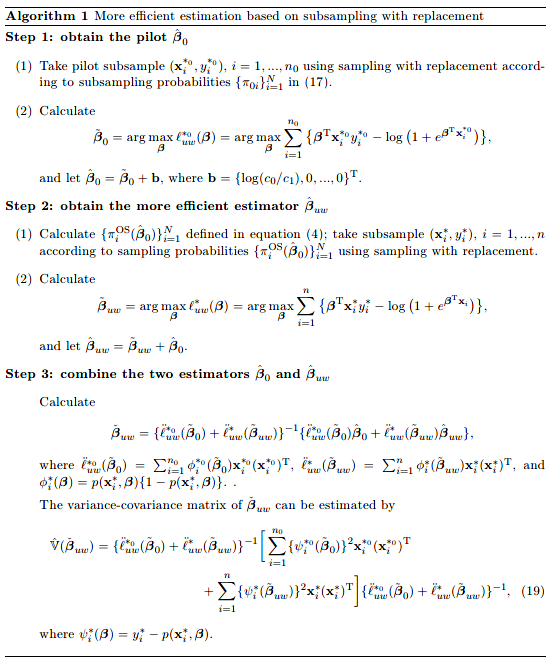


If subsample size small (n/N << 1), can approximate by

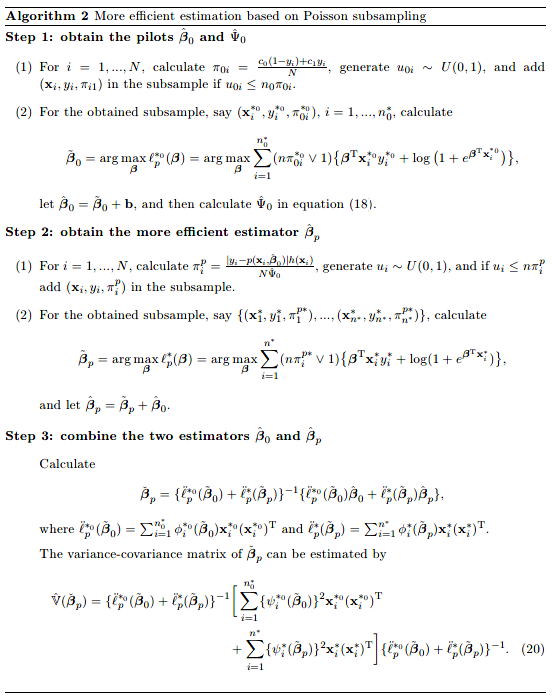


# Algorithms: unweighted (subsampling with replacement/ Poisson)

**Subsampling with replacement**



**Poisson sampling:**



# Misspecification

Pilot Estimate Misspecification <model correct>:

1. still consistent
2. Poisson subsampling still more efficient

Model Misspecification <pilot estimators consistent>:

1. Subsampling with replacement:
   1. critical to have a good pilot estimator
   2. inflate Var-Cov by a small pilot sample size
2. Poisson subsampling: weaker on   🡪 converge w/o specifying rate