

## CN-informed GLM

$$K \sim NB(\text{mean} = \mu_{ij}, \text{dispersion} = \alpha_i)$$

$$\mu_{ij} = O_{ij} q_{ij}$$

$$O_{ij} = s_j + \log\left(\frac{CN_{ij}}{2}\right)$$

$$\log(q_{ij}) = \beta_{i0} X_{j0} + \beta_{i1} X_{j1}$$

$K_{ij}$  = mRNA counts

$CN_{ij}$  = copy number integers (0,1,2,...,5)

$X_j = [1, X_{j1}]$  = covariate vector (design matrix, including intercept)

$\beta_i = [\beta_{i0}, \beta_{i1}]$  = regression coefficient vector (base level, effect size)

$s_j$  = library size factor

$O_{ij}$  = offset

*Statistical testing (Wald test):*

$$H_0 : \beta_{1i} = 0 \text{ vs } H_1 : \beta_{1i} \neq 0$$

```
[56]: # Classical GLM
      irls_solver(counts, size_factors, design_matrix, disp)

Beta parameters: [7.42819011 1.85418179]
Estimated mean: [ 1967.99858081 1828.10357256 1935.38075053 2009.81021647
 1830.70609429 1980.23685096 1961.79232929 1308.2826297
 1977.05650339 1533.27021047 9650.78978847 11394.07347687
 9731.54634261 10071.64775884 8459.58717055 10729.67835561
11572.035241 10725.47900749 8576.10056945 11369.34391072]
```

```
[37]: # CN normalized GLM
      irls_solver(counts, size_factors, design_matrix, cnv, disp)

Beta parameters: [7.42812371 0.27144301]
Estimated mean: [1967.88474121 1827.99902134 1935.26907659 2009.69360077 1830.60137027
 1980.12219879 1961.67890175 1308.21259218 1976.94206237 1533.18523481
 3512.31212302 4362.94223821 2893.87612598 4834.04187735 3267.6413555
 5263.95537222 4801.94880507 3733.0512585 4186.58948289 4357.86283204]
```