

Possion Regression

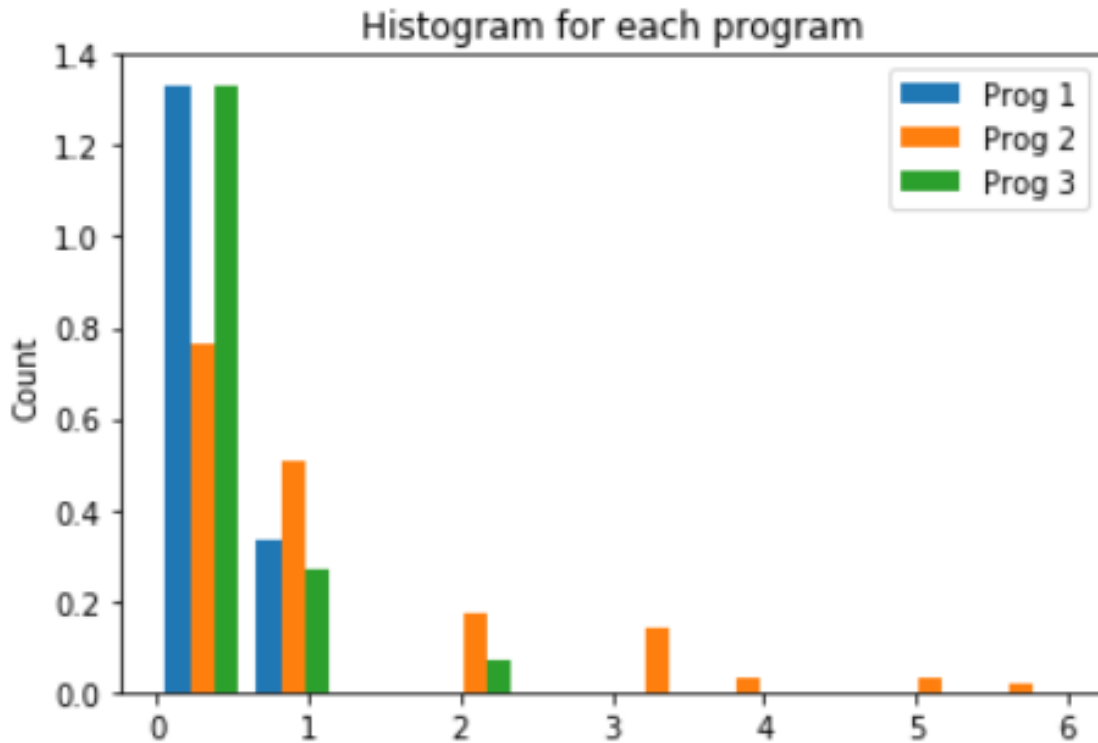
数据集主要来自参考教程:

<https://github.com/mahat/PoissonRegression/blob/master/SimplePoissonRegression.py>

数据集大概是根据学生目前的项目以及书写成绩得到其获奖次数。(原链接也没有说明)

1. 根据函数 show_hist():

可以得到如下图表, 看得出来 Y 服从泊松分布



2. 采用统计包计算得到以下的结果:

Poisson Regression Results						
Dep. Variable:	y	No. Observations:	200			
Model:	Poisson	Df Residuals:	196			
Method:	MLE	Df Model:	3			
Date:	Wed, 17 Jan 2018	Pseudo R-squ.:	0.2118			
Time:	21:03:29	Log-Likelihood:	-182.75			
converged:	True	LL-Null:	-231.86			
		LLR p-value:	3.747e-21			
	coef	std err	z	P> z	[0.025	0.975]
x1	0.0702	0.011	6.619	0.000	0.049	0.091
x2	-0.3698	0.441	-0.838	0.402	-1.234	0.495
x3	0.7140	0.320	2.231	0.026	0.087	1.341
const	-4.8773	0.628	-7.764	0.000	-6.109	-3.646

3. 自己写的 poisson regression:

```
: opt_solution.x
: fun: 135.10516063788083
hess_inv: <4x4 LbfgsInvHessProduct with dtype=float64>
jac: array([ 1.48059721e-04,  6.68338417e-07,  2.12977541e-06,
            1.89003131e-06])
message: b'CONVERGENCE: REL_REDUCTION_OF_F_<=_FACTR*EPSMCH'
nfev: 42
nit: 31
status: 0
success: True
x: array([ 0.0701524 , -0.36980908,  0.71404998, -4.87731541])
```

结果求得的参数与模型几乎一样

具体思路：

似然函数估计:

Poisson Likelihood

$$\log \mathcal{L}(\lambda|y) = \sum_{i=1}^N y_i \log \lambda_i - \lambda_i$$

Substitute the systematic component, $\lambda_i = \exp(\mathbf{x}_i\beta)$

$$\log \mathcal{L}(\beta|y, \mathbf{X}) = \sum_{i=1}^N y_i \mathbf{x}_i \beta - \exp(\mathbf{x}_i \beta)$$

梯度推导如下:

$$\nabla \log L(\beta|y, \mathbf{X}) = \sum_{i=1}^N y_i \mathbf{x}_i - \exp(\mathbf{x}_i \beta) \mathbf{x}_i$$

根据以上公式，通过 **scipy** 提供的优化函数，得到最优解。