Estimating the Mean of a Poisson Population From a Sample Set

Given: y_i, i = 1 to N samples from a population believed to have a Poisson distribution

Estimate: the population mean M_p (and thus also its variance V_p)

The standard estimator for a Poisson population mean based on a sample is the unweighted sample mean $\bar{y}$; this is a maximum-likelihood unbiased estimator. The uncertainty of the sample mean, expressed as a variance, is the sample variance V_s divided by N.

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$V_s = \frac{1}{N}\sum_{i=1}^{N} (y_i - \bar{y})^2$$

$$\sigma_{\bar{y}}^2 = \frac{V_s}{N}$$

Since the parent population is Poisson, its mean and variance are equal, and so both $\bar{y}$ and the sample variance multiplied by $N/(N$-1) are unbiased estimators of the population variance V_p; this estimator will de denoted with a circumflex:

$$\hat{V}_p = \frac{N}{N-1}V_s = \frac{1}{N-1}\sum_{i=1}^{N} (y_i - \bar{y})^2$$

The sample variance V_s in the asymptotically Gaussian limit is related to chi-square according to

$$V_s = \frac{V_p}{N}\chi_{N-1}^2 = \frac{M_p}{N}\chi_{N-1}^2$$

Where M_p has been substituted for V_p because the parent population is assumed to be Poisson-distributed. Since V_s is proportional to a chi-square random variable with N-1 degrees of freedom, its *own* variance is proportional to 2(N-1). The factor scaling the chi-square random variable is M_p/N, so the square of this scales the variance of V_s. Replacing M_p with its estimator $\bar{y}$, the uncertainty variances of V_s and V_p are therefore approximately

$$\sigma_{V_s}^2 \approx \left(\frac{\bar{y}}{N}\right)^2 \big(2(N-1)\big) = \frac{2(N-1)\bar{y}^2}{N^2}$$

$$\sigma_{\hat{V}_p}^2 = \left(\frac{N}{N-1}\right)^2 \sigma_{V_s}^2 \approx \frac{2\bar{y}^2}{N-1}$$

We will investigate averaging these two estimators of the population mean with inverse-variance weights to obtain a better estimate of the population mean M_p. Since the Poisson population has positive skewness $1/\sqrt{M_p}$, the sample mean and variance are correlated according to the general rule

$$\mathrm{cov}(\bar{y}, V_s) = \frac{N-1}{N^2}\mu_3$$

where μ_3 is the third central moment of the parent population. Since the skewness $S = \mu_3/\sigma^3$, and for a Poisson distribution $S = 1/\sqrt{M_p}$ and $\sigma^3 = M_p^{3/2}$, we have $\mu_3 = M_p$, and

$$\mathrm{cov}(\bar{y}, V_s) = \frac{(N-1)M_p}{N^2} \approx \frac{(N-1)\bar{y}}{N^2}$$

Now we take the sample mean and sample estimate of the population variance as two measurements, z_1 and z_2, of the population mean; these and the corresponding error covariance matrix are

$$z_1 = \bar{y}$$
$$z_2 = \frac{N}{N-1}V_s$$
$$\Omega \approx \begin{pmatrix} \dfrac{V_s}{N} & \dfrac{(N-1)\bar{y}}{N^2} \\ \dfrac{(N-1)\bar{y}}{N^2} & \dfrac{2\bar{y}^2}{N-1} \end{pmatrix}$$

We apply standard chi-square minimization to evaluate a scalar model $z = M_p$ using these two measurements with correlated errors. The weight matrix is the inverse of the error covariance matrix:

$$W = \Omega^{-1} \approx \frac{1}{\dfrac{2\bar{y}^2 V_s}{N(N-1)} - \dfrac{(N-1)^2\bar{y}^2}{N^4}} \begin{pmatrix} \dfrac{2\bar{y}^2}{N-1} & \dfrac{(1-N)\bar{y}}{N^2} \\ \dfrac{(1-N)\bar{y}}{N^2} & \dfrac{V_s}{N} \end{pmatrix}$$
$$\equiv \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$$

The chi-square to be minimized is defined as follows.

$$u_1 \equiv z_1 - M_p\,,\; u_2 \equiv z_2 - M_p$$
$$\chi^2 = w_{11}u_1^2 + 2w_{12}u_1u_2 + w_{22}u_2^2$$

Setting the derivative of chi-square with respect to M_p equal to zero yields the following estimator for

M_p, which we will indicate with a caret.

$$a \equiv w_{11} + 2w_{12} + w_{22}$$
$$b \equiv w_{11}z_1 + w_{12}(z_1 + z_2) + w_{22}z_2$$
$$\hat{M}_p = \frac{b}{a}$$
$$\sigma_{\hat{M}_p} = \sqrt{\frac{1}{a}}$$

Since one feature of Gaussian estimation is the fact that the uncertainty does not depend on the measured values, we can get a good idea of how much we gain by the averaging of two estimators for the population mean compared to the single usual estimator, the sample mean. We assume that the sample mean and variance are not drastically incorrect and substitute the population mean and the population variance times (N-1)/N for them, respectively, in the expression for W above.

$$W \approx \frac{1}{\frac{2M_p^3}{(N-1)^2} - \frac{(N-1)^2 M_p^2}{N^4}} \begin{pmatrix} \frac{2M_p^2}{N-1} & \frac{(1-N)M_p}{N^2} \\ \frac{(1-N)M_p}{N^2} & \frac{M_p}{N-1} \end{pmatrix}$$
$$\approx \frac{1}{M_p(2M_pN^4 - N^4 + 4N^3 - 6N^2 + 4N - 1)} \begin{pmatrix} 2M_pN^4(N-1) & -N^2(N-1)^3 \\ -N^2(N-1)^3 & N^4(N-1) \end{pmatrix}$$
$$\approx \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$$

Now we can approximate the uncertainty of the estimator for the population mean:

$$\sigma_{\hat{M}_p} \approx \sqrt{\frac{M_p(2M_pN^4 - N^4 + 4N^3 - 6N^2 + 4N - 1)}{N^2(N-1)(2N^2M_p - N^2 + 4N - 2)}}$$

This can be compared to the uncertainty of the sample mean, which is the usual estimator for the population mean all by itself, for any M_p and N. The result is that there is essentially no useful gain from using the sample-based estimate of the population variance as another estimator of the Poisson population mean. This can be seen in the following table.

M_p	N	$\sigma_{\bar{y}}$	$\sigma_{\hat{V}_p}$	$co-\sigma_{\bar{y}\hat{V}_p}$	$\sigma_{\hat{M}_p}$
10	5	1.58114	7.07107	1.26491	1.57593
10	10	1.05409	4.71045	0.94868	1.05311
100	5	5.00000	70.71068	4.00000	4.99838
100	10	3.33333	47.14045	3.00000	3.33303
100	25	2.04124	28.86751	1.95959	2.04121
1000	10	10.54093	471.40452	9.48683	10.54083
1000	50	4.51754	202.03051	4.42719	4.51754
10000	5	50.00000	7071.06781	40.00000	49.99984
10000	10	33.33333	4714.04521	30.00000	33.33330
10000	50	14.28571	2020.03051	14.00000	14.28571