

Teachers' corner

# A rich learning lesson using the Poisson distribution

Mohammad Fraiwan Al-Saleh

*Department of Maths and Physics, Qatar University, Qatar*

Received 29 December 2006; accepted 25 January 2007

---

## Abstract

In this note, we explore the rich information about inference that the Poisson distribution has. The source of this information is mainly the fact that the mean and variance of this distribution are equal.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Poisson distribution; Complete sufficient statistic; Minimum variance unbiased estimator

---

## 1. Introduction

The Poisson probability distribution is believed to be one of the three most important distributions, the other two being the binomial and the normal distribution. The mean,  $\mu$ , and variance,  $\sigma^2$ , are usually the main features of a given distribution. The mean is a measure of central tendency, while the variance is a measure of the dispersion, spread or variability of a distribution. If  $X$  is binomial with parameters  $n$ , a positive integer, and  $p$ ,  $0 < p < 1$ , denoted by  $b(n, p)$ , then  $\mu = np$  and  $\sigma^2 = np(1 - p)$ ; clearly,  $\mu > \sigma^2$ . If  $X$  is geometric with parameter  $p$ ,  $0 < p < 1$ , denoted by  $g(n, p)$ , then  $\mu = \frac{1-p}{p}$  and  $\sigma^2 = \frac{1-p}{p^2}$ ; clearly,  $\mu < \sigma^2$ . Finally if  $X$  is a Poisson random variable with parameter  $\lambda$ , denoted by  $P(\lambda)$ , then  $\mu = \sigma^2 = \lambda$ . The equality of the mean and variance of the Poisson distribution make it a very rich example in inference. The Poisson example, if used properly in *classrooms*, can give a deep intuitive understanding of some of the ideas in statistical inference. In the next section, we discuss some of these interesting results.

## 2. The Poisson example

Assume that  $X \sim P(\lambda)$ . Then for  $x = 0, 1, 2, \dots$ ,

$$f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (1)$$

---

*E-mail address:* [malsaleh@qu.edu.qa](mailto:malsaleh@qu.edu.qa).

and

$$\mu = E(X) = \sigma^2 = \text{Var}(X) = \lambda. \tag{2}$$

Let  $X_1, X_2, \dots, X_n$  be an independent and identically distributed (iid) sample from  $P(\lambda)$ . The sample mean ( $\bar{X}$ ) and sample variance ( $S^2$ ) are given respectively by

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \& \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \tag{3}$$

It is well known that  $\bar{X}$  and  $S^2$  are always unbiased estimators for  $\mu$  and  $\sigma^2$ , respectively. Hence in our case,  $E(\bar{X}) = E(S^2) = \lambda$ . Since  $P(\lambda)$  is a member of the regular exponential family, it follows that  $T = \bar{X}$  is a complete sufficient statistic for  $\lambda$ . Since  $\bar{X}$  is also unbiased, it follows by the Lehmann–Scheffe theorem that  $\bar{X}$  is the unique minimum variance unbiased estimator (MVUE) of  $\lambda$ . Now,  $S^2$  is also an unbiased estimator of  $\lambda$ , and  $S^2$  is not a function of  $T$  (we cannot obtain the value of  $S^2$  if we are given only the value of  $T$ ). Thus,  $S^2$  is not the MVUE of  $\lambda$ . Therefore,

$$\text{Var}(S^2) > \text{Var}(\bar{X}) = \frac{\lambda}{n}. \tag{4}$$

In accordance with the Rao–Blackwell theorem (see Hogg and Craig [1] or Casella and Berger [2]),  $\phi(\bar{X}) = E(S^2|\bar{X})$  is also an unbiased estimator of  $\lambda$ . Therefore, by the uniqueness of the MVUE,

$$E(S^2|\bar{X}) = \bar{X}, \tag{5}$$

which implies that  $S^2$  and  $\bar{X}$  are dependent.

Now,  $E(S^2\bar{X}) = E(E(S^2\bar{X}|\bar{X})) = E(\bar{X}E(S^2|\bar{X})) = E(\bar{X}^2)$ . Thus,  $E(S^2\bar{X}) = \frac{\lambda}{n} + \lambda^2$ , which implies that

$$\text{Cov}(S^2, \bar{X}) = \frac{\lambda}{n}, \tag{6}$$

i.e.  $S^2$  and  $\bar{X}$  are positively correlated, their covariance is decreasing with  $n$ ; and goes to zero as  $n \rightarrow \infty$ . Furthermore, the correlation between  $S^2$  and  $\bar{X}$ ,  $\rho$ , is given by

$$\rho = \frac{\text{Cov}(S^2, \bar{X})}{\sqrt{\text{Var}(S^2)\text{Var}(\bar{X})}} = \sqrt{\frac{\text{Var}(\bar{X})}{\text{Var}(S^2)}}. \tag{7}$$

It can be shown that  $\text{Var}(S^2) = \frac{\lambda}{n} + \frac{2\lambda^2}{n-1}$  (Casella and Berger [2]). Thus,

$$\rho = \sqrt{\left(\frac{1}{1 + \frac{2\lambda n}{n-1}}\right)}$$

which is always positive and goes to  $\sqrt{\frac{1}{1+2\lambda}}$  as  $n \rightarrow \infty$ .

$$E\left(\frac{S^2}{\bar{X}}\right) = E\left(E\left(\frac{S^2}{\bar{X}} \mid \bar{X}\right)\right) = E\left(\frac{1}{\bar{X}} E(S^2 \mid \bar{X})\right) = 1, \text{ i.e.}$$

$$E\left(\frac{S^2}{\bar{X}}\right) = \frac{E(S^2)}{E(\bar{X})} = 1. \quad (8)$$

On the other hand,  $E\left(\frac{\bar{X}}{S^2}\right) = E\left(E\left(\frac{\bar{X}}{S^2} \mid \bar{X}\right)\right) = E\left(\bar{X} E\left(\frac{1}{S^2} \mid \bar{X}\right)\right) > E\left(\bar{X} \left(\frac{1}{E(S^2)}\right)\right)$  (by Jensen's inequality). Thus,

$$E\left(\frac{\bar{X}}{S^2}\right) > 1. \quad (9)$$

Consider next  $\psi(S^2) = E(\bar{X} \mid S^2)$ . This is again an unbiased function, i.e.  $E(\psi(S^2)) = E(\bar{X}) = \lambda$ . Now,

$$\text{Var}(\bar{X}) = \text{Var}(E(\bar{X} \mid S^2)) + E(\text{Var}(\bar{X} \mid S^2)) \quad (10)$$

implies

$$\text{Var}(\bar{X}) > \text{Var}(\psi(S^2)). \quad (11)$$

It seems at first glance that (11) is in contradiction with the fact that  $\bar{X}$  is the unique MVUE of  $\lambda$ , since no unbiased estimator of  $\lambda$  can have a variance smaller than the variance of  $\bar{X}$ . It follows that  $\psi(S^2) = E(\bar{X} \mid S^2)$  is not an estimator of  $\lambda$ . In other words,  $\psi(S^2)$  is an unbiased function of  $S^2$  and  $\lambda$ , i.e.  $\psi(S^2)$  is not free of the parameter  $\lambda$ , so it is not an estimator. Note also that, since  $S^2$  is not a complete statistic,  $E(\bullet \mid S^2)$  is not guaranteed to be free of  $\lambda$ .

Since  $E(\bar{X}(\bar{X} - S^2)) = E(E(\bar{X}(\bar{X} - S^2) \mid \bar{X})) = 0$ , it follows that  $\text{Cov}(\bar{X}, (\bar{X} - S^2)) = 0$ . Therefore,  $\bar{X}$  and  $(\bar{X} - S^2)$  are uncorrelated. In a similar manner it can be easily deduced that  $S^2$  and  $(S^2 - \bar{X})$  are correlated. So,  $(\bar{X} - S^2)$  is uncorrelated with  $\bar{X}$  but correlated with  $S^2$ . From this, one may conclude that

$$\text{Var}(S^2 - \bar{X}) = \text{Var}(S^2) - \text{Var}(\bar{X}), \quad (12)$$

and using (7), we have

$$\frac{\text{Var}(S^2 - \bar{X})}{\text{Var}(S^2)} = 1 - \rho^2 = 1 - \frac{\text{Var}(\bar{X})}{\text{Var}(S^2)} \quad (13)$$

$$= 1 - \frac{1}{\text{eff}(\bar{X}; S^2)}. \quad (14)$$

Thus

$$\text{eff}(\bar{X}; S^2) = \frac{1}{\rho^2} = 1 + \frac{2\lambda n}{n-1}. \quad (15)$$

This quantity is called by some authors the *relative saving* of using  $\bar{X}$  instead of  $S^2$  as an estimator of  $\lambda$ .

Using the fact that  $E\left(\frac{S^2}{\bar{X}}\right) = 1$ , which was shown above, one can deduce that

$$E\left(\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i} - \frac{n-1}{n}\right) = \lambda, \quad (16)$$

and hence  $\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i} - \frac{n-1}{n}$  is another unbiased estimator of  $\lambda$ .

### 3. Concluding remarks

The Poisson distribution has some distinct properties that make it a rich learning example. The main property is that of the equality of its mean and variance, which implies that both the sample mean and variance are unbiased estimators of  $\lambda$ . Using the fact that the sample mean is the unique MVUE, several results have been derived. These results, when put together, though well known, give a unified example that can be given to students in a mathematical statistics course. This example gives students a deep understanding of many important statistical concepts such as *sufficiency*, *completeness*, *MVUE*, etc.

### References

- [1] R. Hogg, A. Craig, Introduction to Mathematical Statistics, Macmillan publishing Co., Inc., New York, 1978.
- [2] G. Casella, R. Berger, Statistical Inference, Wadsworth & Brooks/Cole Advanced Book & Software, California, 1990.