International Journal of Applied Mathematics & Statistics,
Int. J. Appl. Math. Stat.; Vol. 27; Issue No. 3; Year 2012,
ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)

Note on the Unbiased Estimation of a Function of the Parameter of the Geometric Distribution

Tamás Lengyel[1]

[1]Mathematics Department
Occidental College
Los Angeles, CA90041, USA
lengyel@oxy.edu

ABSTRACT

Kolmogorov studied the problem of whether a function of the parameter p of the Bernoulli distribution $Bernoulli[p]$ has an unbiased estimator based on a sample $X_1, X_2, \ldots, X_n$ of size n and proved that exactly the polynomial functions of degree at most n can be estimated. For the geometric distribution $Geometric[p]$, we prove that exactly the functions that are analytic at $p = 1$ have unbiased estimators and present the best estimators.

Keywords: estimability, minimum variance unbiased estimator, hypergeometric functions, reliability theory.

2000 Mathematics Subject Classification: 62F10.

1 Introduction

Typically, to estimate certain unknown parameters, we search for unbiased estimators with small mean squared error. Sometimes a biased estimator may have a smaller mean squared error and thus, suggesting a search for "best" estimators among both biased and unbiased estimators. On the other hand, the study of unbiased estimators of a parameter or functions of the parameter has lead to many important results (e.g., Cramér-Rao inequality, Rao-Blackwell and Lehmann-Scheffé theorems) that help in comparing such estimators and constructing best unbiased estimators. For this reason, unbiased estimators are often favored over biased ones. (Kolmogorov, 1950) studied the problem of whether a function $g(p)$ of the parameter p of the Bernoulli distribution $Bernoulli[p]$ has an unbiased estimator, i.e., $g(p)$ is *estimable* based on a sample $X_1, X_2, \ldots, X_n$ of size n. He proved that exactly the polynomial functions of degree at most n can be estimated. With the notation $(x)_k = x(x-1)\ldots(x-k+1)$ and $(x)_k = 0$ if $x < k$ for the falling factorial and $S = \sum_{i=1}^{n} X_i$ for the sample sum, the statistic

$$\frac{(S)_k}{(n)_k} = \frac{\binom{S}{k}}{\binom{n}{k}} \tag{1.1}$$

provides us with an unbiased estimator of $p^k, 0 \le k \le n$ (Voinov and Nikulin, 1993, Appendix A24., No. 13), in fact, the only unbiased estimator for p^k in the case of the Bernoulli distribution.

In a similar fashion, we get that

$$\sum_{k=0}^{n} a_k \frac{(S)_k}{(n)_k}$$

is the only unbiased, hence the best or minimum variance unbiased estimator (MVUE) of $\sum_{k=0}^{n} a_k p^k$ (Voinov and Nikulin, 1993, Appendix A24., No. 14). (Note that most MVUEs of this note can be found in (Voinov and Nikulin, 1993).) A simple application of Theorem 2.1 below immediately provides the above MVUE and also guarantees that no other function which is analytic at $p = 0$ is estimable.

In this note, we discuss the similar problem of estimating functions of the parameter p of the geometric distribution $Geometric[p]$ which is defined as the probability distribution $P(X = k) = p(1-p)^{k-1}, k = 1, 2, \ldots,$ of the number X of Bernoulli trials needed to get the first success. Sometimes this distribution is referred to as the shifted geometric distribution. The geometric distribution is a common discrete distribution in modeling the life time of a device in reliability theory. Typically, we search for the maximum likelihood estimator and MVUE for the reliability and failure rate functions, however, for a general function it has not been known if an MVUE let alone an unbiased estimator exists. In Section 2 we answer the question of estimability. In the last section we outline some applications in which unbiased estimators of some function of the parameter p are preferable.

2 From estimations by hypergeometric functions to estimations by power series

For the geometric distribution, different hypergeometric functions of the sample sum can be used to obtain unbiased estimators of $p^k, -\infty < k \le n-1$. In fact, we have that

$$E\left(\frac{\binom{S-k-1}{n-k-1}}{\binom{S-1}{n-1}}\right) = p^k, \tag{2.1}$$

see (Voinov and Nikulin, 1993, Appendix A25., Nos. 13 and 16, with $\theta = 1-p$ and a translation in the value of X_i) for $0 \le k \le n-1$ and (Haldane, 1945) for $k = 1$ and $n \ge 2$. The case with $k \ge n$ is covered by the Remark after Theorem 2.2.

A general tool to obtain MVUEs is to apply the Lehmann-Scheffé theorem. After finding an unbiased estimator T for $g(p)$ and a sufficient and complete estimator S for p, we can use the Lehmann-Scheffé theorem to obtain the MVUE for $g(p)$. For example, for $0 \le k \le n-1$, we set $g(p) = p^k$, the indicator variable $T = I_{X_1 X_2 \ldots X_k = 1}$, and S to be the sample sum. Then $E(T) = p^k$ and

$$\begin{aligned} E(T\,|\,S = N) &= \frac{P(X_1 = X_2 = \cdots = X_k = 1, \sum_{i=1}^{n} X_i = N)}{P(\sum_{i=1}^{n} X_i = N)} &= \frac{p^k P(\sum_{i=k+1}^{n} X_i = N-k)}{P(\sum_{i=1}^{n} X_i = N)} \\ &= \frac{p^k \binom{N-k-1}{n-k-1} p^{n-k} (1-p)^{N-n}}{\binom{N-1}{n-1} p^n (1-p)^{N-n}} &= \frac{\binom{N-k-1}{n-k-1}}{\binom{N-1}{n-1}}, \end{aligned}$$

that is,

$$E(T\,|\,S) = \frac{\binom{S-k-1}{n-k-1}}{\binom{S-1}{n-1}}.$$

Note, however, that unless there is a trivial candidate, finding an unbiased estimator T for a general distribution might be a difficult task, and this is the focus of the monograph by (Voinov and Nikulin, 1993). In order to prove (2.1) for $1/p^k, k \geq 0$, we observe that $\binom{S+k-1}{k}/\binom{n+k-1}{k} = \binom{S+k-1}{n+k-1}/\binom{S-1}{n-1}$, and

$$\begin{aligned} E\left(\frac{\binom{S+k-1}{k}}{\binom{n+k-1}{k}}\right) &= \sum_{N=n}^{\infty}\prod_{i=0}^{k-1}\frac{N+k-1-i}{n+k-1-i}\binom{N-1}{n-1}p^n(1-p)^{N-n} \\ &= \sum_{N=n}^{\infty}\binom{N+k-1}{n+k-1}p^n(1-p)^{N-n} \\ &= \sum_{N=0}^{\infty}\binom{N+n+k-1}{N}p^n(1-p)^N = \frac{p^n}{(1-(1-p))^{n+k}} = \frac{1}{p^k}. \end{aligned}$$

Again, this statistic is a function of the sufficient and complete statistic S for p; thus, it is MVUE for $1/p^k$. One might wonder if $1/(1-p)$ has an MVUE and hopelessly sort through some candidates, based on the ad hoc applications of hypergeometric identities, until recognizing that the answer relies on a powerful general result of (Patil, 1963, Theorem 4) on the estimability of the function $g(p)$ and its generalizations, see (Voinov and Nikulin, 1993). If an MVUE exist for a function of the parameter then we say that the function is MVU estimable. Our main result states that exactly the functions that are analytic at $p = 1$ are MVU (and simply) estimable, and it is given in Theorem 2.8. Let $W(d(\theta))$ denote the index set $\{k \mid a_k \neq 0\}$ of the nonzero coefficients of the power series $d(\theta) = \sum_{k=0}^{\infty} a_k\theta^k$ around 0. We have

Theorem 2.1 (Patil)**.** *Let X follow the generalized power series distribution (GPSD) given by $P(X = x) = p(x;\theta) = a(x)\,\theta^x/f(\theta), x = 0, 1, 2\ldots$, with $f(\theta) = \sum_{x=0}^{\infty} a(x)\theta^x$, and $g(\theta)$ be a function of θ such that $g(\theta)f_n(\theta)$, with $f_n(\theta) = (f(\theta))^n$, admits a powers series expansion in θ. The necessary and sufficient condition for $g(\theta)$ to be MVU estimable on the basis of a random sample of size n from this GPSD is that $W(g(\theta)f_n(\theta)) \subseteq W(f_n(\theta))$. Also, whenever it exists, the MVUE for $g(\theta)$ is given by $c(z,n)/b(z,n)$ for $z \in W(g(\theta)f_n(\theta))$ and it is 0, otherwise, where $c(z,n)$ is the coefficient of θ^z in the expansion of $g(\theta)f_n(\theta)$, z is the sample sum, and $b(z,n)$ is the coefficient of θ^z in $f_n(\theta) = \sum b(z,n)\theta^z$.*

Now we can derive the general solution for the geometric distribution.

Theorem 2.2. *Let $h(\theta)$ be an arbitrary function such that $h(\theta)(\frac{\theta}{1-\theta})^n$ admits a power series expansion in θ and assume that the function $g(p)$ can be written as the function h of $1-p$, i.e., $g(p) = h(\theta)$ by using the substitution $\theta = 1-p$. Then $g(p)$ is estimable on the basis of the sample $X_1, X_2, \ldots, X_n\sim$ $Geometric[p]$ exactly if $g(p)$ is analytic about 1. With $g(p) = \sum_{k=0}^{\infty} a_k(1-p)^k$, the MVUE for $g(p)$ is*

$$\frac{1}{\binom{S-1}{n-1}}\sum_{k=0}^{\infty} a_k\binom{S-k-1}{n-1} = \frac{1}{\binom{S-1}{n-1}}\sum_{k=0}^{S-n} a_k\binom{S-k-1}{n-1}. \tag{2.2}$$

Proof. According to Theorem 2.1, for the geometric distribution we have $a(x) = 1$ for $x \geq 1$, $\theta = 1-p$ and $f(\theta) = (1-p)/p = \theta/(1-\theta)$. Thus, $f_n(\theta) = (\theta/(1-\theta))^n = \sum_{k=0}^{\infty}\binom{n+k-1}{k}\theta^{n+k}$, $b(z,n) = \binom{z-1}{n-1}$, $W(f_n(\theta)) = \{n, n+1, \ldots\}$, and we want to test whether

$$W(h(\theta)f_n(\theta)) \subseteq \{n, n+1, \ldots\}. \tag{2.3}$$

Clearly, the relation (2.3) holds if and only if $h(\theta)$ is analytic at $\theta = 0$, i.e., if $g(p)$ is analytic at $p = 1$.

To obtain the MVUE for $g(p) = \sum_{k=0}^{\infty} a_k(1-p)^k$, we first set $g(p) = (1-p)^k = \theta^k$ with any integer exponent $k \geq 0$ and get that $h(\theta)f_n(\theta) = \theta^{n+k}/(1-\theta)^n = \theta^{n+k}\sum_{t=0}^{\infty}\binom{n+t-1}{t}\theta^t$, and thus, by Theorem 2.1 the MVUE is $\binom{S-k-1}{n-1}/\binom{S-1}{n-1}$ if $S = z \in W(h(\theta)f_n(\theta)) = \{n+k, n+k+1, \dots\}$ and 0, otherwise (which is indirectly incorporated in the first case as the binomial coefficient in the numerator becomes zero). □

Remark. We can easily obtain the relation (2.1) by this theorem. For $g(p) = p^k$, i.e., $h(\theta) = (1-\theta)^k$ with any integer exponent $k \leq n-1$, the MVUE is $\binom{S-k-1}{n-k-1}/\binom{S-1}{n-1}$ for $S \geq n$ and thus, $S = z \in W(h(\theta)f_n(\theta)) = \{n, n+1, \dots\}$. For $g(p) = p^k$ with $k \geq n$, we have a slightly different form. For example, if $k = n$ then $W(h(\theta)f_n(\theta)) = \{n\}$ and the MVUE is 1 if $S = n$ and 0, otherwise. In particular, if $k = n = 1$ then, as it can be easily seen, the only unbiased estimator (thus MVUE) of $p = 1-\theta$ is 1 if $S = 1$ and 0, otherwise. In general, if $k = n+m, m \geq 0$, then $W(h(\theta)f_n(\theta)) = \{n, n+1, \dots, n+m\}$, and we get that the MVUE of p^{n+m} is $\binom{m}{S-n}(-1)^{S-n}/\binom{S-1}{n-1}$ if $S \leq n+m = k$.

We note that a GPSD is of exponential type and thus, according to the factorization theorem, the sample sum is a complete sufficient statistic for θ. By Theorem 2.2 we can conclude that $g(p) = 1/(1-p) = 1/\theta = h(\theta)$ is not MVU (and in the case of a GPSD, simply not) estimable and $W(h(\theta)f_n(\theta)) = \{n-1, n, \dots\}$. The assumption on $h(\theta)(\frac{\theta}{1-\theta})^n$ restricts the applicability of Theorem 2.2, however, Corollary 2.3 still easily follows. Although the above result does not seem to appear in the literature, we mention that (Patil, 1963) observed that $h(\theta) = 1/f(\theta)$ is MVU estimable if and only if $0 \in W(f(\theta))$. Here we have $f(\theta) = \theta/(1-\theta)$; thus, $h(\theta) = -1+1/\theta$, i.e., $1/(1-p)$ is not MVU estimable.

Corollary 2.3. *Assume that the meromorphic function $g(p)$ has a pole of order n or less at $p = 1$. Then $g(p)$ is not estimable on the basis of a sample of size n from the distribution $Geometric[p]$.*

We note an obvious lemma and its consequence, Corollary 2.5 which covers the higher order poles, too.

Lemma 2.4. *If $g(p)$ is estimable for a sample of size n then it is estimable for every sample size exceeding n.*

Corollary 2.5. *Assume that the meromorphic function $g(p)$ has a pole of any order $k \geq 1$ at $p = 1$. Then in the case of the distribution $Geometric[p]$, $g(p)$ is not estimable for any sample size.*

Proof. For sample size n, the case with $1 \leq k \leq n$ is taken care of by Corollary 2.3 so assume that $k > n$ and that $g(p)$ is estimable. By Lemma 2.4, $g(p)$ is also estimable based on the larger sample size k. However, by Corollary 2.3 it is not possible, and we have a contradiction. □

A generalization of Theorem 2.1 for modified power series distributions appears in (Voinov and Nikulin, 1993, p. 210 (see p. 221, too)) that we restate here for GPSDs.

Theorem 2.6. *Let X follow the GPSD given by $P(X = x) = p(x;\theta) = a(x)\,\theta^x/f(\theta), x = 0, 1, 2\dots$, with $f(\theta) = \sum_{x=0}^{\infty} a(x)\theta^x$ and $f_n(\theta) = (f(\theta))^n$. The necessary and sufficient condition for $g(\theta)$ to be MVU estimable on the basis of a random sample of size n from this GPSD* *is that $g(\theta)f_n(\theta)$ admits a powers series expansion in θ and $W(g(\theta)f_n(\theta)) \subseteq W(f_n(\theta))$. Also, whenever it exists, the MVUE for $g(\theta)$ is given by $c(z,n)/b(z,n)$ for $z \in W(g(\theta)f_n(\theta))$ and it is 0, otherwise, where $c(z,n)$ is the coefficient of θ^z in the expansion of $g(\theta)f_n(\theta)$, z is the sample sum, and $b(z,n)$ is the coefficient of θ^z in $f_n(\theta) = \sum b(z,n)\theta^z$.*

It follows in general, that a necessary condition for $g(\theta)$ to be MVU estimable is that it is analytic at $\theta = 0$. This leads to the generalization of Corollary 2.5.

Corollary 2.7. *Let $g(p) = 1/(1-p)^\alpha$ be with any real $\alpha > 0$. In the case of the distribution $Geometric[p]$, $g(p)$ is not estimable for any sample size.*

We get our main result in

Theorem 2.8. *Regardless of the sample size, the real function $g(p)$ is estimable on the basis of a sample from the distribution $Geometric[p]$ exactly if $g(p)$ is analytic at $p = 1$. The MVUE is given in (2.2) in Theorem 2.2.*

The result of Theorem 2.2 can be generalized to left-truncated geometric distributions with known or unknown shift parameters. The proof of Theorem 2.2 also suggests Theorem 2.9 which is of independent interest, cf. (Voinov and Nikulin, 1993, Appendix A25., No. 16).

Theorem 2.9. *The MVUE for $g(p) = p^k(1-p)^r$ with $k \leq n-1$ and $r \geq 0$ integers is*

$$\frac{\binom{S-r-k-1}{n-k-1}}{\binom{S-1}{n-1}}$$

if $S \geq n + r$.

3 Applications

An important application, both historically and statistically, of an unbiased estimator for the parameter $g(p) = p$ of the geometric distribution comes from the estimation of small population frequencies. (Haldane, 1945) considered the following paradigm. If p is the frequency of some attribute of a population, $q = 1 - p$, and $X_i \sim Bernoulli[p], i = 1, 2, \ldots, n$, is the indicator variable of the presence of the attribute in the ith sample from the population, then the standard deviation of the sample mean S/n is $\sqrt{pq/n}$ with S having a binomial distribution with parameters n and p. For small values of p it is unsatisfactory to have an error which is proportional to $\sqrt{p}$.

Haldane opted for a different approach with $X_i \sim Geometric[p]$ being the number of trials it takes to observe an occurrence of the attribute. This so called inverse sampling procedure continues until the nth occurrence, and gives now S a negative binomial distribution with parameters n and p. As was mentioned in Section 2, the unbiased estimator $(n-1)/(S-1)$ for p can be used, and it has an error which is approximately $p\sqrt{q/(n-2)}$ for any $n \geq 3$ provided n is kept constant, i.e., of order of magnitude p when p is small.

For larger values of p and for practical reasons, one might want to find an estimator which has approximately constant standard deviation for all possible values of the frequency rate p and carries information about p so that we can calculate an estimator of p from it. The delta method (cf. (Oehlert, 1992) and (Schwarz, 2008)) is a suitable choice for such problems and other problems involving functions of the parameters. Although it provides an improvement over the standard error of Haldane's approach in the sense that it will be independent of p, it results only in an approximately unbiased estimator of some function $h(p)$ of p, to be determined later in the process, rather than an unbiased estimator of p itself.

Estimating functions of the parameter of the geometric distribution is often required in quality control where the life time X of a certain component can be modeled by the geometric distribution $Geometric[p]$ provided that measurements are taken in discrete time. It is a common problem to derive estimators of the reliability and failure (or hazard) rate functions of individual components in a multi-component series system based on masked system life test data. For example, (Xie, Gaudoin and Bracquemond, 2002) defines the hazard rate function, in terms of the reliability function, as $-\ln(1-p)$. According to Theorem 2.2 this function of p does not have an unbiased estimator.

Another typical application of unbiased estimators of a function of some parameter is when seeking an unbiased estimator of the distribution function $P(X \le k)$ of the random variable X (cf. (Guenther, 1978)). This problem is related to the estimation of the reliability function in quality control. In the case of the geometric distribution, this reduces to the estimation of $P(X \le k) = 1 - P(X > k) = 1 - (1-p)^k$. For example, for $k = 2$ the MVUE is

$$1 - \frac{(S-n)(S-n-1)}{(S-1)(S-2)}, \text{ if } S \ge n+2$$

(cf. Theorem 2.9) and 1, if $S = n$ or $n+1$, as given by (2.2) of Theorem 2.2, and it has standard deviation approximately $2p/\sqrt{n-2}$ for any small p. This estimator is preferable over the biased estimators $1-(1-n/S)^2$ and $1-(1-(n-1)/(S-1))^2, n \ge 2$, constructed by naively substituting the MLE and the MVUE for p, respectively. In fact, the former one overestimates the expected value by approximately $2p/(n-1)$ while the latter one underestimates it by approximately $p^2/(n-2)$ with approximate standard deviations $2pn/((n-1)\sqrt{n-2})$ and $2p/\sqrt{n-2}$. Also, the former one is in agreement with (Patil, 1962) in which the amount of bias in the maximum likelihood estimation (MLE) of a differentiable function of the parameter is determined for GPSDs. Note that the method of moments leads to the MLE for these distributions.

By Theorem 2.8, we can also construct the MVUE for the moment generating function $M_X(t)$ for any particular value t where it is analytic. In general, this might help in calculating the Chernoff bound on tail probabilities. Of course, for the geometric distribution the above approach directly estimates the tail probabilities.

Acknowledgment

The author wishes to thank András Krámli and Gregory P. Tollisen for making helpful suggestions and comments.

References

Guenther, W. C. 1978. Some easily found minimum variance unbiased estimators, *Amer. Stat.* **32**: 29–34.

Haldane, J. B. S. 1945. On a method of estimating frequencies, *Biometrika* **33**: 222–225.

Kolmogorov, A. N. 1950. Unbiased estimators, *Izv. Aka. Nauk SSSR, Ser. Mat.* **14**: 303–326 (in Russian).

Oehlert, G. W. 1992. A note on the delta method, *Amer. Stat.* **46**: 27–29.

Patil, G. P. 1962. Maximum likelihood estimation for generalized power series distributions and its application to a truncated binomial distribution, *Biometrika* **49**: 227–237.

Patil, G. P. 1963. Minimum variance unbiased estimation and certain problems of additive number theory, *Ann. Math. Stat.* **34**: 1050–1056.

Schwarz, W. 2008. *40 Puzzles and Problems in Probability and Mathematical Statistics, Problem Books in Mathematics*, Springer, New York.

Voinov, G. V. and Nikulin, M. S. 1993. *Unbiased Estimators and Their Applications*, Vol. 1: Univariate Case, Kluiwer, Dordrecht/Boston/London.

Xie, M., Gaudoin, O. and Bracquemond, C. 2002. Redefining failure rate function for discrete distributions, *Int. J. Reliab. Quality Saf. Eng.* **9**(3): 275–286.