International Journal of Difference Equations (IJDE).
ISSN 0973-6069, Volume 16, Number 1, (2021). 169-178

https://dx.doi.org/10.37622/IJDE/16.1.2021.169-178

Efficiency Sample From Some Sampling Distributions

Mohammed Elmoudier Mohammed Gobar
Mathematics Department, Faculty of Science and Arts,
Al-Baha University, Kingdom of Saudi Arabia.

Abstract

This paper introduces a method to evaluate the efficiency of a random sample from some sampling distributions. The Shannon entropy and A-entropy are considered to calculate the efficiency of such samples. The obtained results are discussed. An example is given to illustrate the computations of efficiency of random sample from the geometric, binomial and Poisson distributions.
Keywords: Shannon entropy; A-entropy; sup-entropy; efficiency of a random sample; relative efficiency; amount of information.

2010 AMS subject classifications: 94A17.

1. INTRODUCTION

Sample size determination is an important statistical factor of planning in many fields of studies in statistics. Various criteria may be used to determine sample size [1], [2], [3], [13], [14] and [16]. Measures of information appear in several contexts in probability and statistics, Awad and Al Sarie [5] gave a survey of these measures and some of them are used by author's researches [8], [9], [10], [11] and [12] when the sample size is selected for approximating some sampling distributions by using information measures. This paper aims to derive some results which are analogue to those in [3]. Moreover, the paper suggests a method that evaluates the efficiency of a random sample from different discrete distributions and determine the sample size in order to obtain a sample of a given efficiency. The amount of information in the data is measured by the methods as in [4], [15]. The suggested procedure is tested in cases of geometric, binomial and Poisson distributions.

Let X be a random variable in a given probability space $(\Omega, \mathcal{F}, P)$ with probability density function (p.d.f)$f(X;\theta)$, the Shannon introduced a measure for the amount of information in a random variable X known as " Shannon entropy", which is defined by:

$$H(X) = -E(logf(X;\theta) \tag{1.1}$$

where E denotes for expectation with respect to the probability measure P.
In [4] an extension of the entropy H(X) defined by Shannon is defined as following:

$$H(X) = -E(log\frac{f(X;\theta}{\delta}) \tag{1.2}$$

where $\delta = sup_x f(x;\theta)$ which is called A-entropy (known as the sup-entropy).

Definition 1 *Let I be an entropy measure and $f(x,y) = f_1(x)f_2(y)$ where $f(x,y)$ is the joint density of the random vector (X,Y) then*
I is said to be additive if $I(f) = I(f_1) + I(f_2)$,
I is said to be subadditive if $I(f) \leq I(f_1) + I(f_2)$

Theorem 1.1 *If $S_n = \{X_1, X_2, ..., X_n\}$ are independent and identically distributed (i.i.d) random variables of size n from the distribution $f(x_i;\theta)$, then:*
$H(S_n) = nH(X)$ *and*
$A(S_n) = nA(X)$.
Proof:
Note that the joint p.d.f of $X_1, X_2, ..., X_n$ is
$f(x_1, x_2, ..., x_n) = \prod_{j=1}^{n} f(x_j)$
then
$H(S_n) = -E[log(f(x_1, x_2, ..., x_n))] = -\sum_{j=1}^{n} E[log(f(x_j))] = nH(X)$.
Similarly,
$(\frac{f(x_1}{\delta})(\frac{f(x_2}{\delta})...(\frac{f(x_n}{\delta}) = \prod_{j=1}^{n} \frac{f(x_j}{\delta})$
then
$A(S_n) = -E[log\frac{f(x_1}{\delta})(\frac{f(x_2}{\delta})...(\frac{f(x_n}{\delta})] = -\sum_{j=1}^{n} E[log\frac{f(x_j}{\delta}] = nA(X)$.
It is known that the Shannon entropy is additive and subadditive [7].
By additivity of $H(S_n)$ and $A(S_n)$ then the efficiency of such random sample S_n is given from it's distributions. To see this let $S_n = \{X_1, ..., X_n\}$ are i.i.d then the Shannon entropy for a random sample S_n is given by:

$$H(S_n) = -\sum_{i=1}^{n} E[log(f(X_i,\theta)], \tag{1.3}$$

and the A-entropy is given by:

$$A(S_n) = -\sum_{j=1}^{n} E[log\frac{f(X_j, \theta}{\delta}] \tag{1.4}$$

where $\delta = sup_{x_i} f(x_i; \theta)$

Definition 2 *Let* X *be a random variable whose taking values in the set* $C := Y_1, Y_2, ..., Y_m$. *Assume that we have a random sample* S_n *of size* n, *and let* $\{x_1, x_2, ..., x_k\}$ *be it's values with frequencies* $\{r_1, r_2, ..., r_k\}$, *respectively, where* $k \leq m$. *Now the relative frequency of* x_i *is:*

$$r(x_i) = \frac{r_i}{n}, i = 1, 2, ..., k \tag{1.5}$$

Hence the estimated Shannon entropy in the sample S_n *is denoted by*

$$HE(S_n) = -\Sigma_{i=1}^{k} r(x_i) log(r(x_i)) \tag{1.6}$$

and the estimated A-entropy in the sample S_n *is denoted by*

$$AE(S_n) = -\Sigma_{i=1}^{k} r(x_i) log\frac{r(x_i)}{\delta} \tag{1.7}$$

where $\delta = max_{x_i} r(x_i)$

The Shannon entropy in X is defined by Equation(1.3)and so the estimated relative efficiency in the sample S_n is given by:

$$eff = \frac{HE(S_n)}{H(S_n)} \tag{1.8}$$

where eff is Shannon relative efficiency.
Also the A-entropy in X is defined by Equation (1.4) and so the estimated relative efficiency in the sample S_n is given by:

$$eff^* = \frac{AE(S_n)}{A(S_n)}. \tag{1.9}$$

where eff^* is Awad relative efficiency. It is known that the amount of entropy in a given sample size is measured by Equations (1.8) and (1.9) are less than or equal to $(1-\beta)100$

2. MEASURES OF ENTROPY FROM DISTRIBUTIONS

In this section Shannon and A-entropies are derived for a given discrete distributions.

2.1. CaseI:Geometric distribution

Let $Y : G(\theta)$ with p.d.f

$$f_1(y;\theta) = \theta(1-\theta)^{y-1}, y = 1, 2, , ...andzerootherwise \tag{2.1}$$

Let $\delta_1 = sup_y f_1(y;\theta)$. To calculate Shannon and A-entropies.
Since $log(f_1(y;\theta)) = log\theta + (y-1)log(1-\theta)$,
hence

$$H_G(Y) = -E(logf_1(y;\theta)) = -\frac{\theta log\theta + (1-\theta)log(1-\theta)}{\theta} \tag{2.2}$$

The A-entropy for the $Y : G(\theta)$ from Equation (1.2) is given by:

$$A_G(Y) = H_G(Y) + log\delta_1 \tag{2.3}$$

Let $Y_1, Y_2, ..., Y_n$ be a random sample from a Geometric distribution with p.d.f given by Equation (2.1), let $S_n = -\Sigma_{i=1}^n Y_i$, then by the Theorem1 the Shannon entropy for S_n is given by:

$$H(S_n) = nH_G(Y) \tag{2.4}$$

and the A-entropy for S_n is:

$$A(S_n) = nA_G(Y) \tag{2.5}$$

2.2. CaseII:Binomial distribution

Let $Y_1, Y_2, ..., Y_n$ be a random sample from a Bernoulli distribution with p.d.f
$f(y;\theta) = \theta^y(1-\theta)^{n-y}$ if $y \in 0, 1$ and zero otherwise.
Let $X = \Sigma_{i=1}^n Y_i$; notice that $X : B(n,\theta)$ with p.d.f
$f_2(x;\theta) = \frac{n!}{(n-x)!x!}\theta^x(1-\theta)^{n-x}$ if $x \in 0, 1, 2, ..., n$ and zero otherwise
Let $\delta_2 = sup_x f_2(x;\theta) = f(\lceil\theta(n+1)\rceil;\theta)$, $n \geq 2$
whenever $\lceil k \rceil$ is the greatest integer less than or equal to k.
then the Shannon entropy is

$$\begin{aligned} H_B(X) &= -E(log(f_2(x;\theta))) \\ &= -E(log(x!(n-x)!)) + log(n!) + n\theta log\theta + n(1-\theta)log(1-\theta)) \end{aligned} \tag{2.6}$$

Then A-entropy for the X is given by:

$$A_B(X) = H_B(X) + log\delta_2 \tag{2.7}$$

where $\theta^\wedge = x$.

2.3. CaseIII:Poisson distribution

Let $Y_1, Y_2, ..., Y_n$ be a random sample from a Poisson distribution with p.d.f
$f(y;\theta) = \frac{e^{-\theta}\theta^y}{y!}$ if $y \in 0, 1, 2, ...$ and zero otherwise.
Let $X = \sum_{i=1}^{n} Y_i$,that is for large n and small θ ,$X \approx$ Poisson distribution with $n\theta$ [6].The approximate p.d.f of X is given by
$f_3(x;\theta) = \frac{e^{-n\theta}(n\theta)^x}{x!}$ if $x \in 0, 1, 2, ...$ and zero otherwise
notice that:
$\delta_3 = sup_x f_3(x;\theta) = f_3(\lceil n\theta \rceil;\theta), n \geq 2$
and
$\theta^\wedge = \frac{X}{n}$.
from which $logf_3(x;\theta) = -n\theta + xlogn\theta - logx!$,
Then the Shannon entropy is:

$$H_P(X) = -E(log(f_3(x;\theta))) = n\theta(1 - logn\theta) + E(logx!) \tag{2.8}$$

It can be shown that the A-entropy is given by:

$$A_P(X) = E(logx!) - log(\lceil n\theta \rceil!) - (n\theta - \lceil n\theta \rceil)log(n\theta). \tag{2.9}$$

3. RESULTS

Throughout this section, the Matlab 8.1 program have been used to perform the computations.The estimate of the Shannon entropy,A-entropy and the efficiency in the sample S_n are given by Equations (1.8)and(1.9)in each Case I ,II and III.

3.1. Example

a)To illustrate the concepts,as in section 2.1 CaseI, 80 random samples are simulated from geometric distributions:$G(\theta)$. For each $\theta = 0.2 : 0.1 : 0.9$ and for each $n = 10 : 10 : 100$, in each case the efficiency eff and eff^* of each simulated sample is computed in Tables 1 and 2 from Equations (2.4) and (2.5) respectively.

Table 1: Estimated efficiency of sample of size n from geometric distribution with parameter θ by Shannon entropy

n \ θ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	0.6110	0.6158	0.6732	0.4046	0.5957	0.6764	0.4331	0.0000
20	0.6703	0.5730	0.5969	0.7797	0.5877	0.6764	0.4331	0.9752
30	0.7083	0.6423	0.6708	0.6934	0.8065	0.8589	0.9533	0.6646
40	0.7612	0.7449	0.6854	0.7749	0.7933	0.7900	0.7103	0.6146
50	0.6823	0.8251	0.6899	0.9437	0.8343	0.5672	0.7117	0.0000
60	0.7584	0.7479	0.7534	0.7286	0.9050	0.7709	0.8947	0.6618
70	0.7604	0.8045	0.7548	0.8120	0.7004	0.6708	0.7456	0.4082
80	0.7519	0.7259	0.8199	0.7344	0.7971	0.7880	0.6778	0.6146
90	0.8401	0.8602	0.7475	0.7341	0.8236	0.6806	0.7567	0.9553
100	0.8281	0.7898	0.8273	0.8976	0.8586	0.8195	0.8789	0.9652

Table 2: Estimated efficiency of sample of size n from geometric distribution with parameter θ by A-entropy

n \ θ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	0.1637	0.3404	0.1788	0.3805	0.2870	0.2518	0.4064	0.6083
20	0.3451	0.2565	0.2633	0.3392	0.3564	0.4076	0.4160	0.7203
30	0.2614	0.3376	0.2346	0.4216	0.5289	0.5710	0.5622	0.7426
40	0.3321	0.4539	0.2460	0.4651	0.5221	0.5281	0.5486	0.5217
50	0.2538	0.4862	0.5133	0.3726	0.5486	0.5154	0.6120	0.9295
60	0.3243	0.4005	0.4819	0.4951	0.4365	0.5907	0.6135	0.5532
70	0.2706	0.3450	0.3963	0.5608	0.6174	0.5781	0.5744	0.8372
80	0.2521	0.3328	0.3507	0.4764	0.5317	0.6281	0.5517	0.8795
90	0.3072	0.3960	0.3936	0.4360	0.4181	0.5409	0.5494	0.7397
100	0.2479	0.3341	0.4474	0.4851	0.4567	0.5539	0.5999	0.5013

b) For estimating the efficiency, as in section 2.2 Case II also 80 random samples are simulated from binomial distributions, $X : B(n, \theta)$. For each $\theta = 0.2 : 0.1 : 0.9$ and for each $n = 10 : 10 : 100$. In each case the efficiency eff and eff^* of each simulated sample is computed in Tables 3 and 4 respectively from the Equations (2.6) and (2.7).

Table 3: Estimated efficiency of sample of size n from binomial distribution with parameter θ by Shannon entropy

n \ θ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	0.4073	0.4733	0.5597	0.6521	0.6077	0.6834	0.7785	0.6945
20	0.6092	0.6946	0.6746	0.7408	0.8231	0.7698	0.7483	0.8421
30	0.5960	0.6613	0.7339	0.7334	0.8026	0.8982	0.9599	0.7911
40	0.6456	0.7057	0.7633	0.8217	0.8554	0.9848	0.9553	0.8220
50	0.6515	0.6708	0.7771	0.8301	0.9836	0.9139	0.9653	0.9551
60	0.6544	0.6847	0.7311	0.8446	0.9026	0.9199	0.9953	0.8018
70	0.6505	0.6938	0.7715	0.9062	0.9474	0.9560	0.9534	0.9297
80	0.6752	0.7181	0.7680	0.8155	0.9041	0.9902	0.9927	0.9815
90	0.6603	0.7264	0.7922	0.8540	0.9330	0.9919	0.9907	0.9370
100	0.6595	0.7331	0.7832	0.8839	0.9230	0.9911	0.9904	0.9084

Table 4: Estimated efficiency of sample of size n from binomial distribution with parameter θ by A-entropy

n \ θ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
10	0.1979	0.1951	0.1000	0.1077	0.2978	0.7094	0.9830	0.1569
20	0.3519	0.3135	0.2282	0.2200	0.3508	0.9646	0.6733	0.1395
30	0.3321	0.1488	0.1361	0.3019	0.2660	0.6755	0.7425	0.1740
40	0.4159	0.4870	0.3313	0.2406	0.3608	0.7391	0.5980	0.1499
50	0.4028	0.4256	0.2986	0.2583	0.3702	0.9342	0.9365	0.1403
60	0.3255	0.2642	0.2170	0.2463	0.3043	0.6762	0.7232	0.1953
70	0.3125	0.3641	0.2605	0.2342	0.2705	0.8284	0.9497	0.1739
80	0.3455	0.3803	0.2564	0.2607	0.3565	0.6122	0.9991	0.1804
90	0.2396	0.3047	0.2214	0.2532	0.2610	0.7851	0.9938	0.1743
100	0.3498	0.3603	0.2600	0.3097	0.3505	0.6255	0.7963	0.1493

c) The Poisson distribution $X : P(\theta)$ is used for simulated 80 random samples as in section 2.3 Case III, for each $\theta = 6 : 1 : 13$ and for each $n = 10 : 10 : 100$. In each case the efficiency eff and eff^* of each simulated sample is computed in Tables 5 and 6 respectively from the Equations (2.8) and (2.9). Table 5. Estimated efficiency of sample of size n from Poisson distribution with parameter ? by Shannon entropy.

Table 5: Estimated efficiency of sample of size n from Poisson distribution with parameter θ by Shannon entropy

n \ θ	6	7	8	9	10	11	12	13
10	0.6949	0.6950	0.5700	0.7672	0.6225	0.6424	0.7672	0.7673
20	0.8425	0.7375	0.8262	0.7887	0.8674	0.7900	0.8266	0.8787
30	0.8103	0.7787	0.8321	0.8410	0.8685	0.9196	0.8794	0.7862
40	0.7715	0.8552	0.8046	0.8222	0.8870	0.9101	0.8425	0.9410
50	0.8721	0.8649	0.8535	0.8964	0.8553	0.8731	0.9382	0.9743
60	0.8066	0.9135	0.8781	0.8854	0.9652	0.9676	0.9306	0.9664
70	0.7718	0.8562	0.9251	0.9256	0.9448	0.9877	0.9284	0.9465
80	0.8205	0.8599	0.8545	0.8963	0.9327	0.9637	0.9305	0.9824
90	0.8401	0.8864	0.8743	0.9323	0.9601	0.9474	0.9066	0.9590
100	0.8033	0.9360	0.9336	0.8946	0.9169	0.9428	0.9841	0.9989

Table 6: Estimated efficiency of sample of size n from Poisson distribution with parameter θ by A-entropy

n \ θ	6	7	8	9	10	11	12	13
10	0.4530	0.7562	0.6845	0.7855	0.4465	0.6046	0.3455	0.5623
20	0.6928	0.9027	0.8004	0.8857	0.7098	0.5003	0.7089	0.5867
30	0.9992	0.9294	0.5332	0.9940	0.7733	0.5780	0.8140	0.6902
40	0.5653	0.7833	0.5771	0.9759	0.7160	0.9094	0.8569	0.9113
50	0.6678	0.6197	0.9025	0.8160	0.9757	0.7775	0.9007	0.6946
60	0.6239	0.6793	0.6916	0.8603	0.9464	0.8957	0.8362	0.8231
70	0.7765	0.6358	0.7023	0.7731	0.9113	0.7219	0.7805	0.9144
80	0.6529	0.5911	0.7682	0.7720	0.8446	0.6312	0.6041	0.8502
90	0.4900	0.8213	0.8412	0.8171	0.7598	0.7442	0.7130	0.7734
100	0.7157	0.9224	0.7485	0.8633	0.7386	0.7219	0.7201	0.6014

4. CONCLUSIONS

The calculations of the estimated relative efficiency of entropies given in [4], [15] are derived for the cases where the distribution of sample size that selected are geometric, binomial and Poisson. These calculations that given by using Tables1 to 6 in example1 shows that:

(a) $0 \leq eff \leq 1$ and $0 \leq eff^* \leq 1$.

(b) The estimated relative efficiency increases with sample size n in both Shannon and Awad sup-entropies.

(c) The estimated relative efficiency which is calculated from Shannon and Awad sup-entropies are closed to each other especially for Poisson distribution.

More precisely one may take the sample size n as follow:

1. (1) For the case of geometric distribution, calculations shows that when size of the sample $n = 50$, for $p.d.f = f(x, \theta), \theta = 0.5$ and hence $eff \approx 94$ by Table 1.
2. For the binomial distributions case, we considered the size $n = 80$ and $\theta = 0.7$, then $eff \approx 99$ by Table 3.
3. For the case of Poisson distribution,$n = 40$ and $\theta = 9$ and hence $eff^* \approx 99$ by Table 6.

ACKNOWLEDGMENTS

The author would like to thank Professor I.H.Elsanousi for his comments,suggestions and recommendations.

REFERENCES

[1] C.J. Adcock,1997,The Choice of sample size and the maximum expected utility-Comments on the paper by Lindley. The Statistician Vol.46 (2):155-162.

[2] C.J. Adcock,1997,Sample size determination: A Review. The Statistician Vol.46 (2): 261-283.

[3] A.M.Awad,1996. Entropy of a partition and efficiency of a sample. Abhath Al-Yarmouk pure sci. and eng. Vol. 5: 9- 30.

[4] A. M.Awad,1987,Statistical information measures, Dirasat (Science), XIV 12: 7-20.

[5] A. M.Awad,and T.Al Sarie,1996. Cluster of entropy measures based on binomial distribution. Abhath Al-Yarmouk, Vol.5: 31-69.

[6] M. M.Azzam ,And A.M.Awad,1996,Entropy measures and some distribution Approximations. Microelectron Reliab. Vol. 36(10):1569-1580.

[7] A.Feinstein,1958. Foundation of information theory. McGraw-Hill: N. Y.

[8] M. E.Gobar,E.B.Bashier,and A.M.Awad,2014.Information measures and some distribution approximations. Journal of advances in mathematics Vol.6 (3) :1008-1014.

[9] M.E.Gobar,and E.B.Bashier,2014.Akaike information criterion and some distribution approximations. International J.of Math. Sci. and Engg. Appl. (IJMSEA) Vol.8 No II: 1-9.

[10] M.E.Gobar,E.B.Bashier,and A.M.Awad,2014. The best sample size in approximating some sampling distributions. International J. of Math. Sci. and Engg. Applec. (IJMSEA). Vol.8 No3: 179-192.

[11] M.E.Gobar,and E.B.Bashier,2015.Estimating sample size to approximating some sampling distributions by information measures. Journal of progressive research in mathematics Vol. 3(3):203-215.

[12] M.E.Gobar,1999. Optimal sample size in approximating some sampling Distributions. M.Sc. Thesis .University of Jordan: Jordan.

[13] L.Joseph,1997. Interval-based versus decision theoretic criteria for the choice of sample size. The Statistician Vol.46 (2): 145-149.

[14] D.V.Lindley,1997. The Choice of sample size. The Statistician Vol.46 (2): 129-138.

[15] E.E.Shannon,1948.A mathematical theory of communication. Bulletin of systems Technology J. 27: 379-423 and 623-659.

[16] S.A.Singh, and Masuku, B.M. 2014. Sampling techniques and determination of sample size in applied statistics research. An overview. Introduction journal of Economics. commerce and management Vol.11(11): 1-22.