The efficiencies of the root-mean-square and power-divergence statistics for testing goodness-of-fit

by Jiayang Gao

A Summer Undergraduate Research Experience at NYU

2011

1. Introduction

This article compares the efficiencies of the root-mean-square and power-divergence statistics in goodness-of-fit testing, giving advice on which statistics are the most appropriate for various distributions. The word "efficiency" here is defined to be the number of draws required to reach simultaneously a confidence level of 99% and a rejection rate of 99% for a distribution that differs from the model – a statistic is more efficient when it requires fewer draws. For introductions to goodness-of-fit testing and the standard power-divergence statistics, see, for example, [1], [2], or [3]. For an introduction to the root-mean-square statistic as used for goodness-of-fit testing, see [4]. For a summary of our recommendations, see the concluding section.

Power-divergence statistics are a family of statistics defined by the following equation

$$I_\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{k=1}^{n} mY_k \left[\left(\frac{Y_k}{P_k} \right)^\lambda - 1 \right]; \quad \lambda \in R,$$

where n is the number of bins, m is the number of draws, Y_k is the proportion of experimentally observed draws falling in the k^{th} bin, and P_k is the probability that a draw falls in the k^{th} bin according to the model distribution. I_0 and I_{-1} are defined by $\lim_{\lambda \to 0} I_\lambda$ and $\lim_{\lambda \to -1} I_\lambda$, respectively.

Among the family of power divergence statistics, some are particularly well known. For example, I_1 is Pearson's chi-square, I_0 is the log–likelihood-ratio, $I_{0.5}$ is the Freeman-Tukey statistic, and I_{-1} is the modified log–likelihood-ratio. In this article we also compare these classic power-divergence statistics with the ones generated by other "λ"s.

The root-mean-square statistic is $\sqrt{\sum m(Y_k - P_k)^2}$.

2. Root-mean-square statistic vs. power-divergence statistics

As we can see from the following examples, the differences of efficiencies among power-divergence statistics are relatively small compared to the difference between the root-mean-

square and power-divergence statistics. In this section, we compare the efficiencies of the root-mean-square statistic and the power-divergence statistics.

The statistics used in this section are the root-mean-square and the power-divergence statistics for λ= -0.9, -0.5 (Freeman-Tukey), 0 (log–likelihood-ratio), 0.5, and 1 (chi-square).

We generated the plots using Gnuplot and Fortran 77 as described in [4].

2.1 An example where the root-mean-square is more efficient than the power-divergence statistics

Example 1:

Let the model distribution be

$$p(1)=p(2)=\frac{1}{4}$$

$$p(k)=\frac{1}{2(n-2)} \quad \text{for } k=3, 4, \dots, n-1, n$$

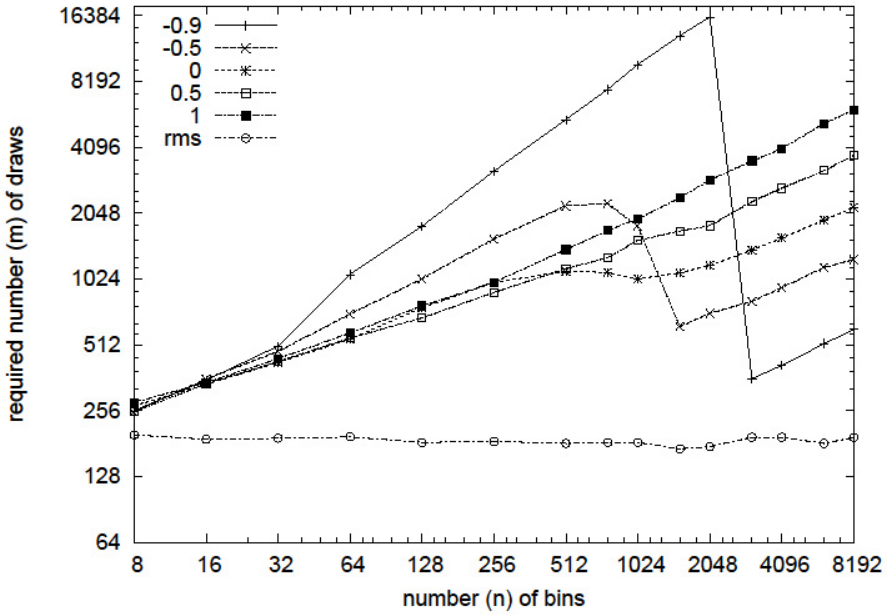
and consider m i.i.d draws from the distribution

$$q(1)=\frac{3}{8}$$

$$q(2)=\frac{1}{8}$$

$$q(k)=\frac{1}{2(n-2)} \quad \text{for } k=3, 4, \dots, n-1, n$$

The following figure plots the number of draws (m) required for each statistic to distinguish the actual distribution of the draws from the model distribution. The graph illustrates that the root-mean-square requires a constant 190 draws for any number of bins, while the power-diverence statistics require 36% more draws when the number of bins (n) is 8. The graph also shows that the number of draws required for the power-divergence statistics tends to increase as n increases (except for the sudden drop when n=3072 with λ= -0.9, which will be discussed later). Moreover, the most efficient power-divergence statistic (λ=0) requires 513% more draws than the root-mean-square when n=512, and the most efficient (λ= -0.9) requires 212% more draws than the root-mean-square when n=8192.



In the example above, the root-mean-square is much more efficient than the power-divergence statistics. Other examples in which the root-mean-square is superior can be found in section 6.2, 6.4, and 6.5 of [4].

2.2 An example where the root-mean-square is less efficient than the power-divergence statistics

Example 2:

Let the model distribution be

$$p(1)= \frac{1}{2}$$

$$p(2)=0$$

$$p(k)= \frac{1}{2(n-2)} \quad \text{for } k=3, 4, \dots , n-1, n$$

and consider m i.i.d draws from the distribution

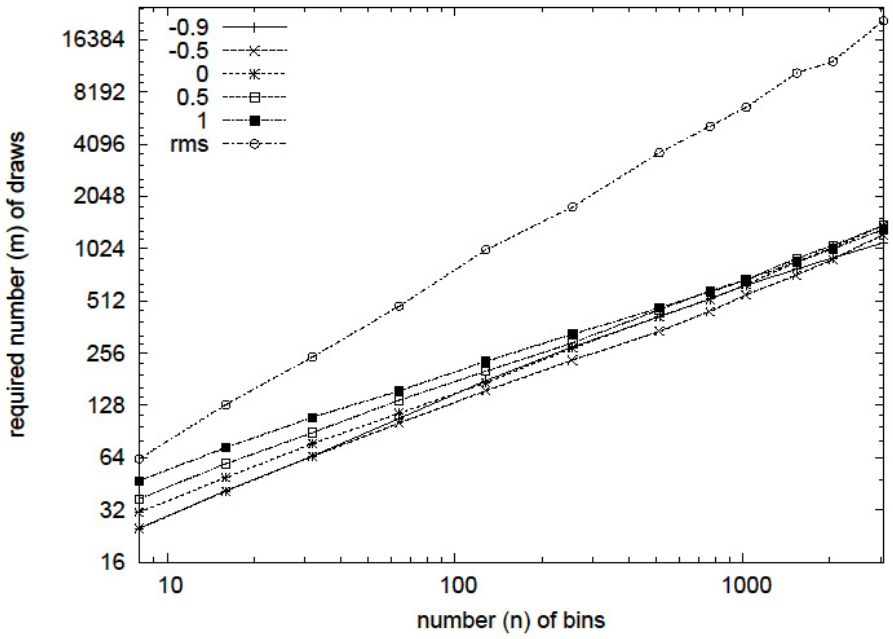
$$q(1)= \frac{1}{2}$$

$$q(2)=0$$

$$q(k)= \frac{1}{2(n/2-1)} = \frac{1}{n-2} \text{ for } k=3, 4, \ldots, \frac{n}{2}, \frac{n}{2}+1$$

$$q(k)=0 \text{ for } k= \frac{n}{2}+2, \frac{n}{2}+3, \ldots, n-1, n$$

The following figure plots the number of draws (m) required for each statistic to distinguish the actual distribution from the model distribution. The graph shows that the power-divergence statistic is more efficient than the root-mean-square, and the required number of draws for both the root-mean-square and the power-divergence statistic increase as the number of bins (n) increases, as does the difference between the power-divergence statistic and the root-mean-square. When n=8, the root-mean-square requires 153% more draws than the most efficient of all power-divergence statistics tested (the most efficient that we tested corresponds to λ= -0.7), and 34% more than the least efficient (λ=1). When n=3072, the root-mean-square requires 1961% more draws than the most efficient (λ=-0.7) of all power-divergence statistics tested, and 1405% more than the least efficient (λ=0).



2.3 Some remarks comparing the root-mean-square and power-divergence statistics

Remark 1: When the probability P_k in the model distribution is small relative to $\max_j P_j$, the chi-square statistic tends to be more sensitive to discrepancies in the k^{th} bin than the root-mean-square statistic. (Recall that P_k is the probability of the k^{th} bin in the model distribution)

Justification of remark 1:

The root-mean-square statistic is $\sqrt{\sum m(Y_k - P_k)^2}$,

where n is the number of bins, m is the number of draws, Y_k is the proportion of experimentally observed draws falling in the k^{th} bin, and P_k is the probability that a draw falls in the k^{th} bin according to the model distribution.

The confidence level determined by the root-mean-square statistic or by its square is the same, because the square root is monotonically increasing. Therefore we can look at the square of the root-mean-square instead, which is:

$$X = m \sum (Y_k - P_k)^2.$$

Also, the chi-square statistic is:

$$\chi 2 = m \sum_{k=1}^{n} \frac{(Y_k - P_k)^2}{P_k}.$$

Notice that

$$\frac{\text{The } k^{th} \text{ summand of } \chi 2}{\text{The } k^{th} \text{ summand of } X} = \frac{\frac{(Y_k - P_k)^2}{P_k}}{(Y_k - P_k)^2} = \frac{1}{P_k}.$$

The k^{th} summand of $\chi 2$ is much, much larger than the k^{th} summand of X for very small P_k. ■

Remark 2: When both P_k and Q_k are small relative to $\max_j P_j$ or $\max_j Q_j$, the Freeman-Tukey statistic tends to be more sensitive to discrepancies in the k^{th} bin than the root-mean-square statistic. (Recall that P_k is the probability of the k^{th} bin in the model distribution, and Q_k is the probability of the k^{th} bin in the actual distribution of the draws.)

Justification of remark 2:

The root-mean-square statistic is $\sqrt{\sum m(Y_k - P_k)^2}$,

where n is the number of bins, m is the number of draws, Y_k is the proportion of experimentally observed draws falling in the k^{th} bin, and P_k is the probability that a draw falls in the k^{th} bin according to the model distribution.

The confidence level determined by the root-mean-square statistic or by its square is the same, because the square root is monotonically increasing. Therefore we can look at the square of the root-mean-square instead, which is:

$$X = m \sum (Y_k - P_k)^2.$$

Also, the Freeman-Tukey statistic is:

$$F = 4m\sum_{k=1}^{n}(\sqrt{Y_k} - \sqrt{P_k})^2.$$

Notice that

$$\frac{\text{The } k^{th} \text{ summand of F}}{\text{The } k^{th} \text{ summand of X}} = \frac{4(\sqrt{Y_k}-\sqrt{P_k})^2}{(Y_k-P_k)^2} = \frac{4}{(\sqrt{Y_k}+\sqrt{P_k})^2}.$$

According to the central limit theorem,

$$Y_k \approx Q_k \pm \frac{\sqrt{(1-Q_k)Q_k}}{\sqrt{m}}.$$

The k^{th} summand of F is much, much larger than the k^{th} summand of X when both P_k and Q_k are very small. ■

Remark 1 and remark 2 above compare the root-mean-square with two special power-divergence-statistics. For more general power-divergence statistics, we have the following remark.

Remark 3: The root-mean-square is not very sensitive to relative discrepancies between the model and actual distributions when the absolute differences are small. That is, the root-mean-square tends to require many more draws when the relative discrepancy $|1 - \frac{P_k}{Q_k}|$ is sufficiently larger than $\min_j \{|1 - \frac{P_j}{Q_j}|\}$ for some k, while the absolute difference $|P_k - Q_k|$ is sufficiently less than $\min\{\frac{P_k}{Q_k}, \frac{Q_k}{P_k}\}$ for every k (example 2 illustrates this).

Heuristic justification of remark 3:

The root-mean-square statistic is $\sqrt{\sum m(Y_k - P_k)^2}$,

where n is the number of bins, m is the number of draws, Y_k is the proportion of experimentally observed draws falling in the k^{th} bin, and P_k is the probability that a draw falls in the k^{th} bin according to the model distribution.

The confidence level determined by the root-mean-square statistic or by its square is the same, because the root-mean-square statistic is monotonically increasing. Therefore we can look at its square instead, which is:

$$X=m\sum(Y_k - P_k)^2.$$

Also, the power-divergence statistic is

$$I_\lambda = \frac{2m}{\lambda(\lambda+1)} \sum_{k=1}^{n} Y_k \left[\left(\frac{Y_k}{P_k} \right)^\lambda - 1 \right].$$

Assume without loss of generality $P_k > Q_k$.

Let us write

$$P_k - Q_k = 10^{-a} \text{ and } \frac{P_k}{Q_k} = b$$

where a is a large positive real number, $1 << b << 10^a$.

Then

$$P_k = \frac{b}{b-1} 10^{-a}, \text{ and } Q_k = \frac{1}{b-1} 10^{-a}$$

According to the central limit theorem,

$$Y_k \approx Q_k + \frac{\sqrt{(1-Q_k)Q_k}}{\sqrt{m}}.$$

Because $Q_k = P_k/b$, and $P_k < 1$, $Q_k \in (0, \frac{1}{b})$

$$\frac{1}{b-1} 10^{-a} + \frac{10^{-\frac{a}{2}}}{\sqrt{bm}} \lesssim Y_k \lesssim \frac{1}{b-1} 10^{-a} + \frac{10^{-\frac{a}{2}}}{\sqrt{m(b-1)}}$$

Let $\sqrt{mb} = 10^c$, where c is large positive real number.

Look at the k^{th} summand in both root-mean-square statistic and power-divergence statistics.

The k^{th} summand of X:

$$X(k) = (Y_k - P_k)^2 \lesssim [-10^{-a} + \frac{10^{-\frac{a}{2}}}{\sqrt{m(b-1)}}]^2 < (10^{-\frac{a}{2}-c})^2 = 10^{-a-2c}$$

The k^{th} summand of I_λ:

$$I_\lambda(k) = \frac{2}{\lambda(\lambda+1)} Y_k \left[\left(\frac{Y_k}{P_k} \right)^\lambda - 1 \right] \gtrsim \left(\frac{1}{b-1} 10^{-a} + \frac{(10^{-\frac{a}{2}})}{\sqrt{mb}} \right) \left[\left(\frac{1}{b} + \frac{\sqrt{b-1}}{b\sqrt{m}} 10^{\frac{a}{2}} \right)^\lambda - 1 \right] \frac{2}{\lambda(\lambda+1)}$$
$$> O(10^{-\frac{a}{2}-c})$$

Therefore, $\frac{I_\lambda(k)}{X(k)} > O(10^{\frac{a}{2}+c}) = \sqrt{mb}\, O(10^{\frac{a}{2}})$ **(1)**

$$\Rightarrow \frac{I_\lambda(k)}{X(k)} > \sqrt{m\frac{P_k}{Q_k}\frac{1}{P_k-Q_k}} >> \sqrt{m\frac{P_k}{Q_k}\frac{1}{\frac{Q_k}{P_k}}} = \sqrt{m(\frac{P_k}{Q_k})^2} \qquad (2)$$

We conclude that the k^{th} summand of I_λ (power-divergence statistics) is much larger than the k^{th} summand of X (root-mean-square statistic) when $\frac{P_k}{Q_k}$ is sufficiently larger than 1 and the absolute difference $|P_k - Q_k|$ is sufficiently less than $\min\{\frac{P_k}{Q_k}, \frac{Q_k}{P_k}\}$. This means the power-divergence statistics are much more sensitive to relative discrepancy than the root-mean-square statistic when absolute difference is small. ■

2.4 Further explanation of remark 3 using examples

Example 3:

Let the model distribution be

$$p(1)= \frac{1}{2}$$

$$p(2)=0$$

$$p(k)= \frac{0.99}{2(n/2-1)} \text{ for } k=3, 4, \dots, \frac{n}{2}, \frac{n}{2}+1$$

$$p(k)= \frac{0.01}{2(n/2-1)} \text{ for } k= \frac{n}{2}+2, \frac{n}{2}+3, \dots, n-1, n$$

and consider m i.i.d draws from the distribution

$$q(1)= \frac{1}{2}$$

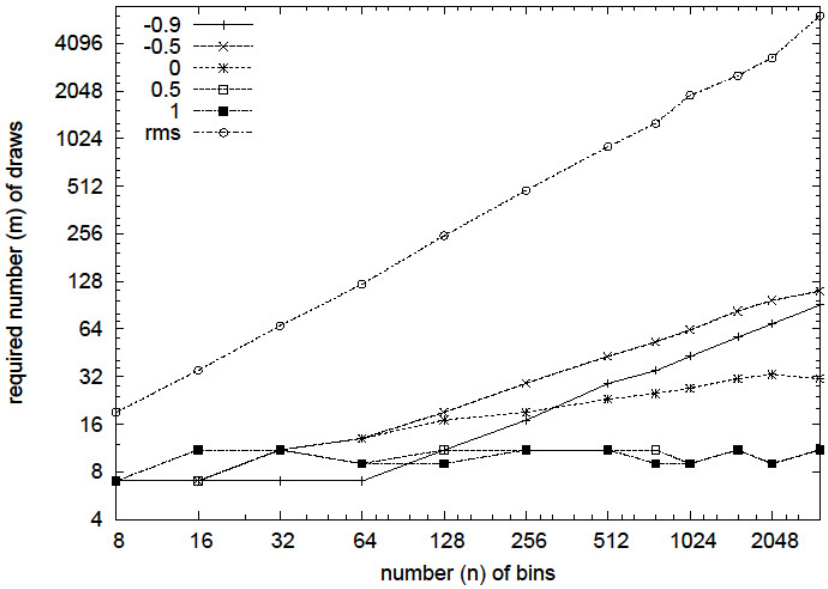
$$q(2)=0$$

$$q(k)= \frac{0.01}{2(n/2-1)} \text{ for } k=3, 4, \dots, \frac{n}{2}, \frac{n}{2}+1$$

$$q(k)= \frac{0.99}{2(n/2-1)} \text{ for } k= \frac{n}{2}+2, \frac{n}{2}+3, \dots, n-1, n$$

This is another example where $\frac{P_k}{Q_k}$ is sufficiently far from 1 for some k, while $|P_k - Q_k|$ is sufficiently less than $\min\{\frac{P_k}{Q_k}, \frac{Q_k}{P_k}\}$ for every k.

The following figure plots the number of draws (m) required for each statistic to distinguish the actual distribution of the draws from the model distribution. As can be expected from remark 3, power-divergence statistics are much more efficient than the root-mean-square statistic. When n=8, the root-mean-square requires 171% more draws than all power-divergence statistics tested. When n=3072, the root-mean-square requires 55764% more draws than the most efficient (λ= 1) of all power-divergence statistics tested, and 4520% more than the least efficient (λ= -0.7).



The superiority of the power-divergence statistics over the root-mean-square in example 3 is more significant than that in example 2. Comparing example 3 with example 2, we get the following remark:

Remark 4: When $\frac{P_k}{Q_k}$ is sufficiently far from 1 for some k, while $|P_k - Q_k|$ is sufficiently less than $\min\{\frac{P_k}{Q_k}, \frac{Q_k}{P_k}\}$ for all k, the superiority of the power-divergence statistics over the root-mean-square increases as the relative discrepancy between P_k and Q_k increases.

This remark follows directly from equation (1) in the justification of remark 3. Because

$$\frac{I_\lambda(k)}{X(k)} > O(10^{\frac{a}{2}+c}) = \sqrt{mb}\, O(10^{\frac{a}{2}}),$$

as b gets larger, less number of draws m is required.

Example 4:

Let the model distribution be

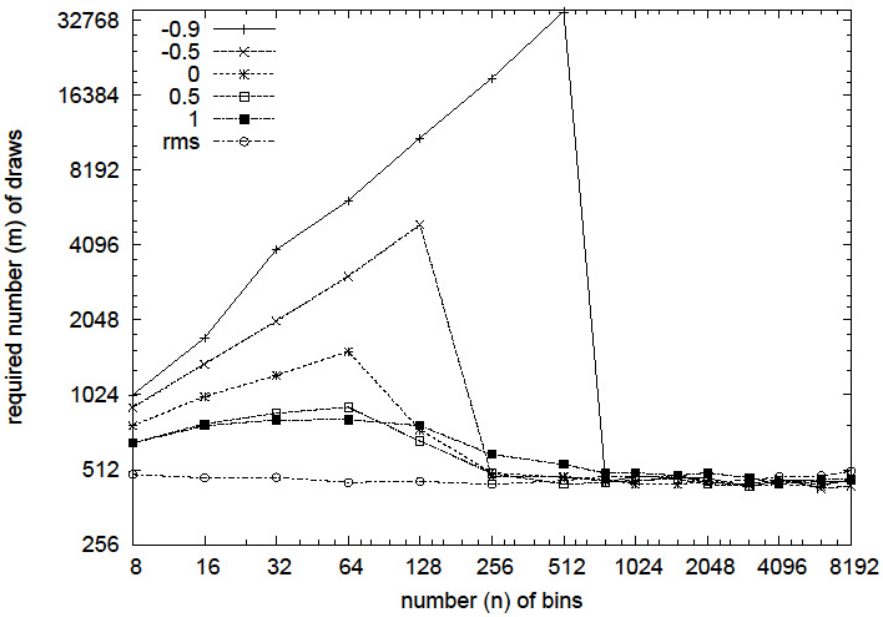
$$p(1)= \frac{15}{16}$$

$$p(k)= \frac{1}{16(n-1)} \quad \text{for } k=2, 4, \dots, n-1, n$$

and consider m i.i.d draws from the distribution

$$q(1)= \frac{7}{8}$$

$$q(k)= \frac{1}{8(n-1)} \quad \text{for } k=2, 4, \dots, n-1, n$$



The figure above plots the number of draws (m) required for each statistic to distinguish the actual distribution of the draws from the model distribution.

The graph illustrates that the root-mean-square statistic is more efficient than the power-divergence statistics, even though as n becomes larger, the difference becomes negligible. This will be discussed later. When n=8, the most efficient power-divergence statistic (λ =1) requires 34% more draws than the root-mean-square, and when n=64, the most efficient power-divergence statistic (λ =1) requires 79% more draws.

In example 4, $|P_k - Q_k|$ is small for all k except for k=1, and $\frac{P_k}{Q_k}$ is sufficiently far from 1 for all k. The only condition of remark 3 this example fails to meet is that in this example, $|P_1 - Q_1| = \frac{1}{16}$ is not small enough.

Therefore, we can make the following remarks:

Remark 5: As long as there is at least one bin for which the absolute difference between the actual and model probabilities is not small enough, the root-mean-square will be sensitive to that bin and thus will be competitive with and possibly outperform the power-divergence statistics.

Remark 6: In Example 4, the difference between the root-mean-square and the power-divergence statistics becomes negligible as n becomes large. As n becomes larger, the relative discrepancy in the tail (k=2, 3,…,n) remains the same while the absolute difference becomes smaller and a larger fraction of the bins fall in the tail. Thus, as n becomes larger, it is easier (requires fewer draws) for the power-divergence statistics to sense the difference.

3. Power-divergence statistics generated by different λ's

In example 1 and example 3 above, not only are the efficiencies of the root-mean-square and power-divergence statistics different, but so are the efficiencies of power-divergence statistics generated by different λ's. In this section we briefly compare the efficiencies of different power-divergence statistics.

The statistics used in this section are the power-divergence statistics for λ= -0.9, -0.75, -0.5 (Freeman-Tukey), -0.25, 0 (log-likelihood-ratio), 0.5, 1 (chi-square).

We generated the plots using Gnuplot and Fortran 77 as described in [4].

Example 5 (same distribution as used in example 1):

Let the model distribution be

$$p(1)=p(2)=\frac{1}{4}$$

$$p(k)=\frac{1}{2(n-2)} \quad \text{for } k=3, 4, \dots, n-1, n$$

and consider m i.i.d draws from the distribution

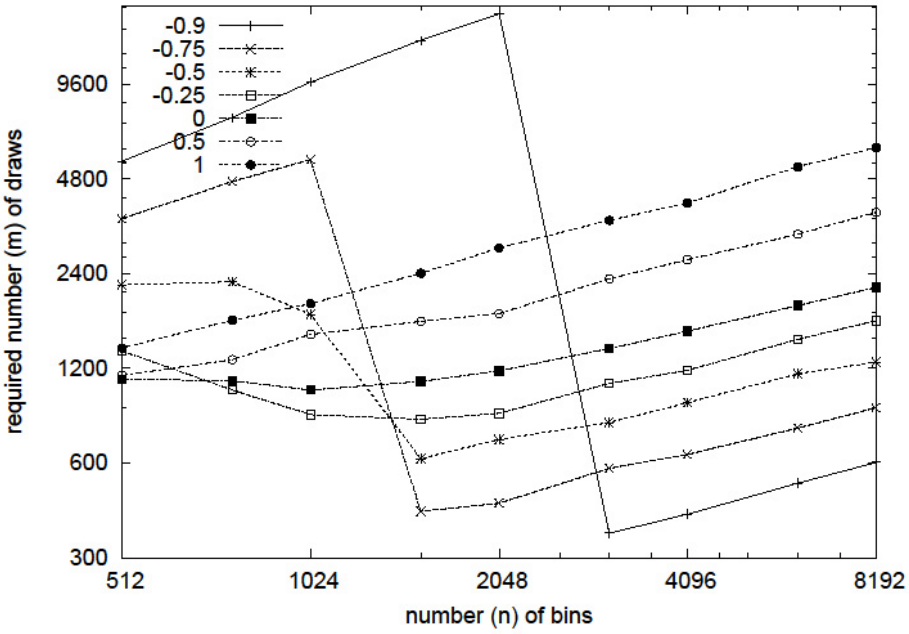
$$q(1)=\frac{3}{8}$$

$$q(2)=\frac{1}{8}$$

$$q(k)=\frac{1}{2(n-2)} \quad \text{for } k=3, 4, \ldots, n-1, n$$

The following figure plots the number of draws (m) required for each statistic to distinguish the actual distribution of the draws from the model distribution. The graph shows that

(1) When λ is positive, the number of draws (m) required increases as the number of bins (n) increases.
(2) When λ is negative, the number of draws (m) required first increases as n increases, then suddenly drops as n becomes large.
(3) The parameter λ for the most efficient power-divergence statistic is *nonpositive* for all numbers of bins (n) tested. When n=512, chi-square (λ=1) requires 36% more draws than the most efficient power-divergence statistic (λ=0). When n=8192, chi-square (λ=1) requires 901% more draws than the most efficient power-divergence statistic (λ= -0.9).



Example 6 (same distribution as used in example 2):

Let the model distribution be

$$p(1)= \frac{1}{2}$$

$$p(2)=0$$

$$p(k)= \frac{1}{2(n-2)} \quad \text{for } k=3, 4, \ldots, n-1, n$$

and consider m i.i.d draws from the distribution

$$q(1)= \frac{1}{2}$$

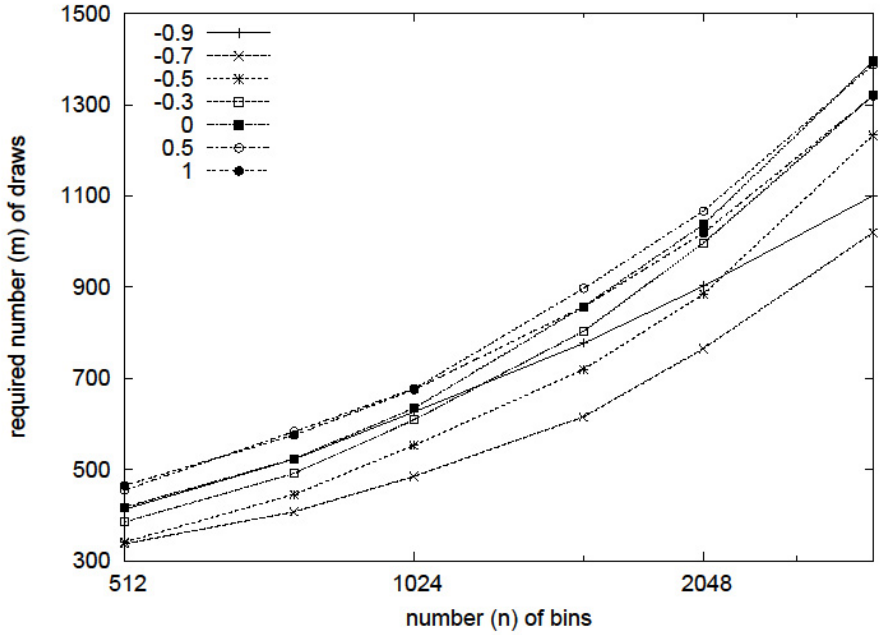
$$q(2)=0$$

$$q(k)= \frac{1}{2(n/2-1)} = \frac{1}{n-2} \text{ for } k=3, 4, \ldots, \frac{n}{2}, \frac{n}{2}+1$$

$$q(k)=0 \text{ for } k= \frac{n}{2}+2, \frac{n}{2}+3, \ldots, n-1, n$$

The following figure plots the number of draws (m) required for each statistic to distinguish the actual distribution of the draws from the model distribution. The graph shows that

(1) All power-divergence statistics for the negative λ's tested require fewer draws than the ones generated by positive λ's.
(2) The most efficient power-divergence statistic is $I_{0.7}$ for all numbers of bins (n) tested. When n=512, chi-square (λ=1) requires 38% more draws than the most efficient power-divergence statistic (λ= -0.7). When n=8192, chi-square (λ=1) requires 29% more draws than the most efficient power-divergence statistic (λ= -0.7).



Example 7 (same distribution as used in example 3):

Let the model distribution be

$$p(1)= \frac{1}{2}$$

$$p(2)=0$$

$$p(k)= \frac{0.99}{2(n/2-1)} \quad \text{for } k=3, 4, \dots, \frac{n}{2}, \frac{n}{2}+1$$

$$p(k)= \frac{0.01}{2(n/2-1)} \quad \text{for } k= \frac{n}{2}+2, \frac{n}{2}+3, \dots, n-1, n$$

and consider m i.i.d draws from the distribution

$$q(1)= \frac{1}{2}$$

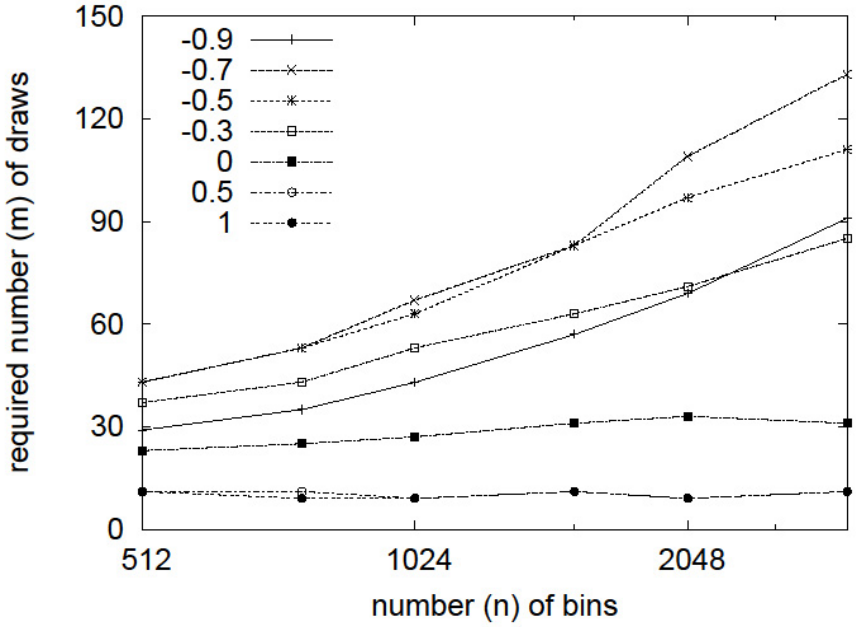
$$q(2)=0$$

$$q(k)= \frac{0.01}{2(n/2-1)} \quad \text{for } k=3, 4, \dots, \frac{n}{2}, \frac{n}{2}+1$$

$$q(k)=\frac{0.99}{2(n/2-1)} \quad \text{for } k=\frac{n}{2}+2\ ,\ \frac{n}{2}+3,\ \ldots,\ n-1,\ n$$

The following figure plots the number of draws (m) required for each statistic to distinguish the actual distribution of the draws from the model distribution. The graph shows that

(1) All statistics for the positive λ's tested require fewer draws than the ones for negative λ's.
(2) The most efficient power-divergence statistic is chi-square (λ=1) for all number of bins (n) tested. This means the classic chi-square is more efficient than the other power-divergence statistics tested. When n=512, power divergence statistic for λ=-0.7 requires 291% more draws than chi-square. When n=8192, power divergence statistic for λ=-0.7 requires 1109% more draws than chi-square.



Remark 7:

When the number of bins n is large, the power-divergence statistics for $-1<\lambda\leq 0$ tend to have the characteristics of both chi-square statistic and the root-mean-square statistic. In other words, for n large, the power-divergence statistics for nonpositive λ's seem to draw a compromise between the chi-square statistic and the root-mean-square statistic. Furthermore, the power-divergence statistic is more like the root-mean-square statistic as λ becomes more negative.

Several of the previous examples illustrate Remark 7: in example 1 the root-mean-square is much more efficient than chi-square; in example 5 with the same distribution, the λ for the most efficient power-divergence statistic is -0.9 for n large, which is not surprising since the power-divergence statistic is most similar to the root-mean-square statistic when λ=-.9, at least among those values for λ considered in the plot.

In example 3, the root-mean-square is much less efficient than chi-square; in example 7 with the same distribution, the most efficient power-divergence statistic is chi-square ($\lambda = 1$) — all other λ's behave more like the root-mean-square.

In example 2, the root-mean-square is less efficient than chi-square, but the difference is not as large as in example 3. In example 6, which concerns the same distribution as example 2, the λ for the most efficient power-divergence statistic is -0.7 for n large; this statistic can draw on the advantages of both the root-mean-square and chi-square.

And in example 4, where the root-mean-square is more efficient than the power-divergence statistics, the differences become negligible when n becomes large.

Remark 8:

For complicated model distributions, we recommend using both the root-mean-square and the power-divergence statistics for $-1<\lambda\leq 0$.

An example of a distribution for which Remark 7 is apropos is the following.

Example 8:

Let the model distribution be

$$p(1)=\frac{4}{10}$$

$$p(2)=\frac{1}{10}$$

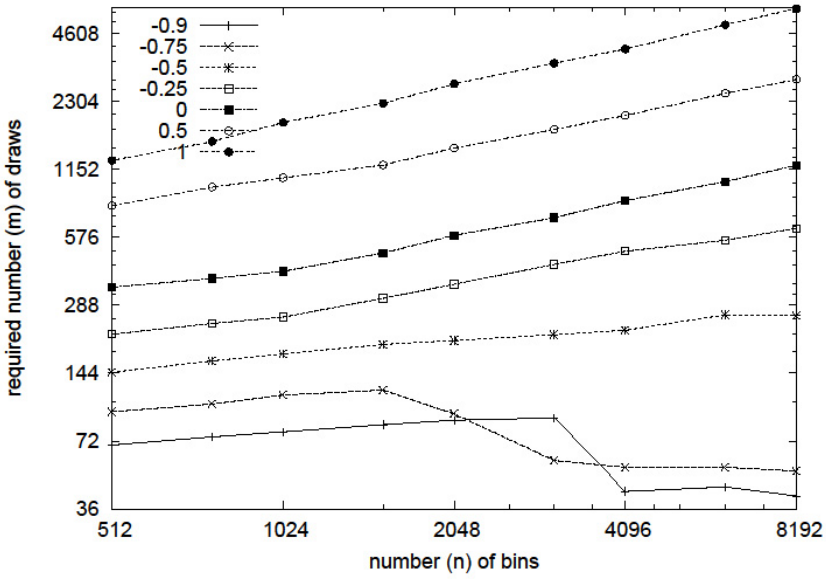
$$p(k)=\frac{1}{2(n-2)} \quad \text{for } k=3, 4, \ldots, n-1, n$$

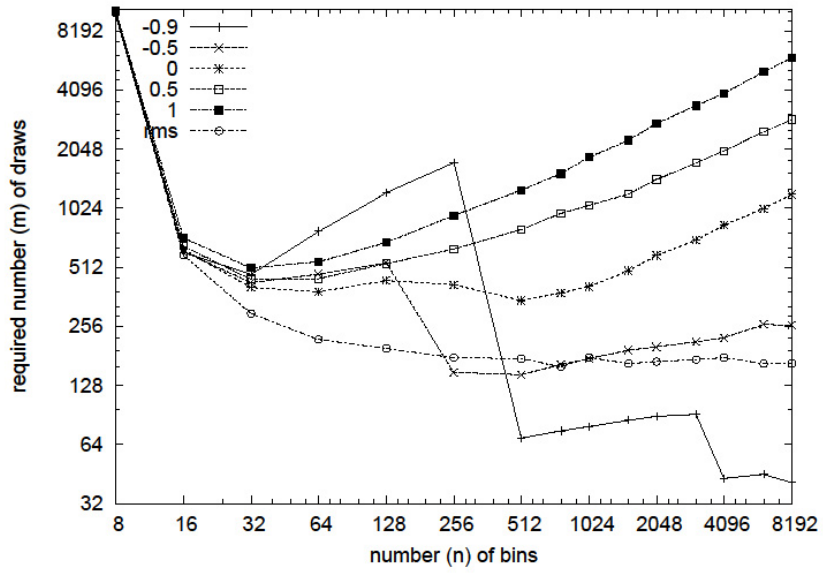
and consider m i.i.d draws from the distribution

$$q(1)=\frac{1}{2}-\frac{1}{2(n-2)}$$

$$q(k)= \frac{1}{2(n-2)} \quad \text{for } k=2, 3, \ldots, n-1, n$$

The two graphs below show that the root-mean-square is the most efficient when n is small, and the power-divergence statistic for λ= -0.9 is the most efficient when n is large.





4. Conclusion

In this article, we compared the efficiencies of the root-mean-square and the power-divergence statistics, and we can draw the following conclusions:

(1) The root-mean-square statistic is not sensitive to relative discrepancies between actual and model distributions in bins with small absolute discrepancies. Therefore, when the absolute difference between the actual and model distributions is small for all bins, and the relative discrepancy is large for some bins, we recommend using the power-divergence statistics.
(2) In contrast, when an absolute difference is large and the relative discrepancies are small, we recommend using the root-mean-square.
(3) For distributions that do not satisfy the criteria of either (1) or (2), we recommend using both the root-mean-square and the power-divergence statistics for $-1<\lambda\leq 0$.

Acknowledgements

I would like to thank my faculty sponsor, Mark Tygert.

References

[1] C. R. Rao, "Karl Pearson chi-square test: The dawn of statistical inference," in: C. Huber-Carol, N. Balakrishnan, M. S. Nikulin, M. Mesbah (Eds.), *Goodness-of-Fit Tests and Model Validity*, Birkhauser, Boston, pp. 9–24, 2002.

[2] M. G. Kendall, A. Stuart, K. Ord, S. Arnold, *Kendall's Advanced Theory of Statistics*, vol. 2A, Wiley, 6th edition, 2009.

[3] S. R. S. Varadhan, M. Levandowsky, N. Rubin, *Mathematical Statistics*, Lecture Notes Series, Courant Institute of Mathematical Sciences, NYU, New York, 1974.

[4] W. Perkins, M. Tygert, and R. Ward, "Computing the confidence levels for a root-mean-square test of goodness-of-fit," Applied Mathematics and Computation, vol. 217, no. 22, pp. 9072–9084, 2011.