Statistical Properties of Population Stability Index

Bilal Yurdakul[1] and Joshua Naranjo[2]

[1] Stress Testing Analytics Group,USAA
[2] Dept. of Statistics, Western Michigan University

Abstract. The population stability index (PSI) is a widely used statistic that measures how much a variable has shifted over time. PSI is also used to measure applicability of a model by measuring shift in either independent or dependent variables. Consequently, a high PSI may alert the business to a change in the characteristics of a population that requires investigation and possibly a model update. PSI is commonly used among banks. Since banks are heavily regulated, an unsuitable use of a model may result in additional model risk. Therefore incorrect use of PSI may bring unexpected risks for the institution. However, there are not many studies about the statistical properties of PSI. Presently, the following rule of thumb is being used: PSI<0.10 means "little shift", .10<PSI<.25 means "moderate shift", and PSI>0.25 means "significant shift, action required". These benchmarks are used without reference to statistical Type I or Type II error rates. We fill the gap by providing statistical properties of PSI and some recommendations.

Keywords: population stability · divergence · risk · model validation · model monitoring

1 Introduction

PSI is a measure of population stability between two population samples. Suppose we are interested in comparing the distribution of credit scores for a base year, say 2017, and a target year, say 2019. Did the distribution of credit scores remain the same or did it change? PSI is calculated by classifying the credit scores into B bins, and comparing the multinomial frequencies over the two years. For example, credit scores are classified into seven bins in Table 1, with the highest scores in Grade A and lowest scores in Grade G. The table is based on data freely available from Lending Club. The bin boundaries were determined by Lending Club.

Let N be the sample size for base population and M be the sample size for target population. Let $\hat{p}_i = n_i/N$ be the relative frequency of the i^{th} grade for the base year, and let $\hat{q}_i = m_i/M$ be the relative frequency of the i^{th} grade for the target year. The population stability index is computed as

$$\text{PSI} = \sum_{i=1}^{B} (\hat{p}_i - \hat{q}_i)(\ln \hat{p}_i - \ln \hat{q}_i) \tag{1}$$

Grade (i)	Base (B) ($\hat{p}_i$)	Target (T) ($\hat{q}_i$)	B-T	ln(B)-ln(T)	Product
A	0.253	0.177	0.077	0.357	0.027
B	0.302	0.262	0.040	0.142	0.006
C	0.204	0.285	-0.081	-0.334	0.027
D	0.134	0.158	-0.024	-0.165	0.004
E	0.072	0.088	-0.016	-0.201	0.003
F	0.026	0.025	0.001	0.039	0.000
G	0.008	0.006	0.002	0.288	0.001
Total	1.000	1.000			PSI=.068

Table 1. PSI Calculation for Distribution of Credit Scores

In general,base year data refers to the development dataset for model validation and PSI gives a measure of the distribution changes between current and development data. A high PSI indicates a large shift which then makes the model questionable for use. Most of the papers about population stability index are blog posts, studies about its use in the industry (Siddiqi, 2016),(Pruitt, 2010). There is a patent issued for a machine which does the calculation of PSI (Liu et al., 2009). In addition to these publications, there are also papers and manuals by governmental regulatory bodies. Although these publications do not directly discuss PSI, they include mentions about measuring stability of population as an ongoing monitoring requirement (FDIC, 2007), (OCC, 2016), (FED, 2011), (FDIC, 2007). Even though PSI is a widely used tool to measure population stability, its distributional properties are not well known. In this paper, we discuss the statistical properties of PSI. Knowing the properties of PSI allows derivation of, among other things, benchmarks with known rates of false significance.

2 Statistical properties of PSI

Let $\mathbf{X} = (X_1, X_2, \ldots, X_B)$ be a *multinomial* random variable with parameters n and $\mathbf{p}$, where $\mathbf{p} = (p_1, p_2, \ldots, p_B)$ and $p_i > 0$ for $i = 1, \ldots, B$. In our motivating example, n is the total number of base-year credit scores, and X_i is the number that fall in the i^{th} grade. Similarly, let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_B)$ be a multinomial random variable with parameters m and $\mathbf{q} = (q_1, q_2, \ldots, q_B)$ with $q_i > 0$ for $i = 1, \ldots, B$. The following theorem provides a polynomial approximation of PSI.

Theorem 1. *Let $\hat{p}_i = X_i/n$ where $(X_1, \ldots, X_B)$ is multinomial with parameters n and $(p_1, \ldots, p_B)$. Let $\hat{q}_i = Y_i/n$ where $(Y_1, \ldots, Y_B)$ is multinomial with parameters m and $(q_1, \ldots, q_B)$. Then*

$$PSI = PSI^* + O_p(n^{-3/2}) + O_p(m^{-3/2}) \tag{2}$$

where $PSI^* = \sum_{i=1}^{B} (\hat{p}_i - \hat{q}_i) \left[\ln p_i - \ln q_i + \frac{\hat{p}_i - p_i}{p_i} - \frac{\hat{q}_i - q_i}{q_i} - \frac{(\hat{p}_i - p_i)^2}{2p_i^2} + \frac{(\hat{q}_i - q_i)^2}{2q_i^2}\right]$

Proof. By Taylor's expansion, $\ln(\hat{p}_i) = \ln(p_i) + \frac{\hat{p}_i - p_i}{p_i} - \frac{(\hat{p}_i - p_i)^2}{2p_i^2} + R_n$ where R_n consists of higher order terms. Now $\sqrt{n}(\hat{p}_i - p_i)$ is asymptotically normal so $(\hat{p}_i - p_i)$ is $O_p(n^{-1/2})$. Thus $(\hat{p}_i - p_i)^3$ and higher order terms are $O_p(n^{-3/2})$. Expanding $\ln(\hat{q}_i)$ similarly, and substituting into $(\ln \hat{p}_i - \ln \hat{q}_i)$, the result follows.

The polynomial approximation simplifies considerably when the two multinomials have equal probability vectors.

Theorem 2. *If $p_i = q_i,\ i = 1, \ldots, B$ and the assumptions of Theorem 1 hold, then*

$$PSI = \sum_{i=1}^{B} \frac{(\hat{p}_i - \hat{q}_i)^2}{p_i} + o_p(n^{-1}) + o_p(m^{-1}) \tag{3}$$

Proof. Note that $\ln p_i - \ln q_i = 0$. Secondly $\frac{\hat{p}_i - p_i}{p_i} - \frac{\hat{q}_i - q_i}{q_i} = \frac{\hat{p}_i - \hat{q}_i}{p_i}$. Finally, $(\hat{p}_i - \hat{q}_i)\frac{(\hat{p}_i - p_i)^2}{2p_i^2} = [(\hat{p}_i - p_i) - (\hat{q}_i - q_i)]\frac{(\hat{p}_i - p_i)^2}{2p_i^2} = o_p(n^{-1})$ since $(\hat{p}_i - p_i)$ and $(\hat{q}_i - q_i)$ are $o_p(1)$, and $(\hat{p}_i - p_i)^2$ is $O_p(n^{-1})$.

Let $\mathrm{PSI}^{**} = \sum_{i=1}^{B} \frac{(\hat{p}_i - \hat{q}_i)^2}{p_i}$. We will derive the properties of this quadratic approximation. First, we write it in matrix form, i.e.

$$\mathrm{PSI}^{**} = \sum_{i=1}^{B} \frac{(\hat{p}_i - \hat{q}_i)^2}{p_i} = V'AV \tag{4}$$

where

$$V = \begin{bmatrix} \hat{p}_1 - \hat{q}_1 \\ \hat{p}_2 - \hat{q}_2 \\ \vdots \\ \hat{p}_B - \hat{q}_B \end{bmatrix} \text{ and } A = \begin{bmatrix} \frac{1}{p_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{p_2} & & 0 \\ \vdots & & & \\ 0 & 0 & & \frac{1}{p_B} \end{bmatrix}$$

Theorem 3. *Under the assumptions of Theorem 2, $\left(\frac{1}{n} + \frac{1}{m}\right)^{-1} PSI^{**}$ has an approximate χ^2 distribution with $B - 1$ degrees of freedom.*

Proof. Note that $\left(\frac{1}{n} + \frac{1}{m}\right)^{-1} \mathrm{PSI}^{**} = V'A^*V$ where $A^* = \left(\frac{1}{n} + \frac{1}{m}\right)^{-1} A$. It is known from multinomial theory that $(\hat{p}_1, \ldots, \hat{p}_B)$ is asymptotically multivariate normal, and hence V is approximately normal for large n and m (see e.g. Agresti, 2007). To show that $V'A^*V$ has a chi-square distribution, a sufficient condition is that $A^*\Sigma$ is idempotent, where Σ is the variance-covariance matrix of V (see e.g. Stapleton 2009). By independence of $\hat{p}_i$ and $\hat{q}_i$, $\mathrm{var}(\hat{p}_i - \hat{q}_i) = \frac{p_i(1-p_i)}{n} + \frac{q_i(1-q_i)}{m} = p_i(1-p_i)\left(\frac{1}{n} + \frac{1}{m}\right)$ since $p_i = q_i$. Also, $\mathrm{cov}(\hat{p}_i - \hat{q}_i, \hat{p}_j - \hat{q}_j) = \mathrm{cov}(\hat{p}_i, \hat{p}_j) - \mathrm{cov}(\hat{p}_i, \hat{q}_j) - \mathrm{cov}(\hat{q}_i, \hat{p}_j) + \mathrm{cov}(\hat{q}_i, \hat{q}_j)$ which equals $\frac{-p_i p_j}{n} - 0 + 0 + \frac{-q_i q_j}{m}$, or $-p_i p_j\left(\frac{1}{n} + \frac{1}{m}\right)$. (For covariance between multinomial frequencies, see e.g. Cochran, 1977). Then

$$A^*\Sigma = \begin{bmatrix} \frac{1}{p_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{p_2} & & 0 \\ \vdots & & & \\ 0 & 0 & & \frac{1}{p_B} \end{bmatrix} \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & \ldots & -p_1 p_n \\ -p_2 p_1 & p_2(1-p_2) & & -p_2 p_B \\ \vdots & & & \\ -p_n p_1 & -p_n p_2 & & p_B(1-p_B) \end{bmatrix}$$

which can be shown to be idempotent (Yurdakul, 2018). Finally, the degrees of freedom is

$$\mathrm{tr}(A^*\Sigma) = \mathrm{tr}\begin{bmatrix} 1-p_1 & -p_2 & \dots & -p_B \\ -p_1 & 1-p_2 & & -p_b \\ \vdots & & & \\ -p_1 & -p_2 & & 1-p_B \end{bmatrix} = B-1$$

Theorems 1-3 suggest that PSI has a distribution that is approximated by $\left(\frac{1}{n}+\frac{1}{m}\right)$ times a χ^2 random variable with $B-1$ degrees of freedom. We ran simulations to confirm this. For example, we used the *rmultinom()* function in R to generate multinomial counts of $n = 400$ observations falling into $B = 10$ categories with equal probabilities $p_i = .10$ for $i = 1, \dots, 10$. The second sample was generated with $m = 400$ and similar probabilities $q_i = .10$ for all i. The PSI value (1) for comparing the two samples was calculated. Figure 1 contains the histogram of 10,000 simulated values of PSI. The density of $\left(\frac{1}{n}+\frac{1}{m}\right)\chi^2_9$ is overlayed for comparison.

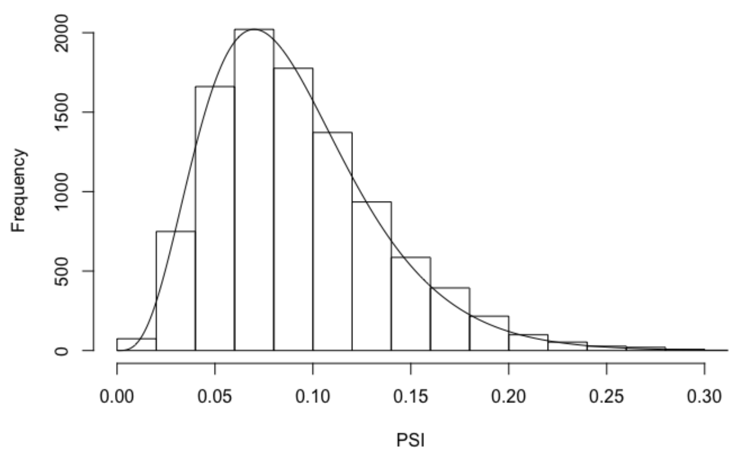

Fig. 1. Histogram of 10,000 simulated PSI values and overlay of chi-square density

3 Benchmarks for PSI

PSI is used to measure how much a distribution has shifted between two populations or over time. A typical rule of thumb for the extent to which a distribution has shifted is the following: PSI $< 0.10 \Rightarrow$'Little change', PSI between

$0.10 - 0.25 \Rightarrow$'Moderate change', and PSI $> 0.25 \Rightarrow$ 'Significant change'. However, the statistical properties of these benchmarks are unknown. For example, how frequently will PSI exceed .10 when in truth there has been no population shift? The asymptotic distribution of PSI allows us to use benchmarks that control false significance rate. Consider formally testing

$$H_0 : (p_1, \ldots, p_B) = (q_1, \ldots, q_B) \text{ versus } H_1 : p_i \neq q_i \text{ for at least one } i$$

A test with *false rejection rate* α would reject H_0 and declare a change if

$$\text{PSI} > \left(\frac{1}{n} + \frac{1}{m}\right) \chi^2_{\alpha,B-1} \tag{5}$$

where $\chi^2_{\alpha,B-1}$ is the upper α percentile of the χ^2 distribution with $B-1$ degrees of freedom. Table 2 provides values of the benchmark (5) for $B = 10$ and $\alpha = .05$. The benchmarks change considerably with n and m.

	m=100	200	400	600	800	1000
n=100	0.338	0.254	0.211	0.197	0.190	0.186
200	0.254	0.169	0.127	0.113	0.106	0.102
400	0.211	0.127	0.085	0.070	0.063	0.059
600	0.197	0.113	0.070	0.056	0.049	0.045
800	0.190	0.106	0.063	0.049	0.042	0.038
1000	0.186	0.102	0.059	0.045	0.038	0.034

Table 2. PSI Benchmarks for $B = 10$ and $\alpha = .05$

When $B = 10$, the rule of thumb PSI $> .25$ seems reasonable for sample sizes n and m between 100 and 200, but is too conservative for larger sample sizes. The benchmark also changes with the number of bins B. Table 3 provides the benchmark values (5) for $B = 20$ and $\alpha = .05$.

	m=100	200	400	600	800	1000
n=100	0.603	0.452	0.377	0.352	0.339	0.332
200	0.452	0.301	0.226	0.201	0.188	0.181
400	0.377	0.226	0.151	0.126	0.113	0.106
600	0.352	0.201	0.126	0.100	0.088	0.080
800	0.339	0.188	0.113	0.088	0.075	0.068
1000	0.332	0.181	0.106	0.080	0.068	0.060

Table 3. PSI Benchmarks for $B = 20$ and $\alpha = .05$

4 Conclusion

PSI is used to measure divergence of frequency distributions between two samples or in time. In practice, PSI is used to measure the divergence between development data set of a particular model and the current population the model is used on. It is possible to measure this divergence based on dependent variable or independent variables in the model. In short, it is used as an applicability test for the model to the current population and also used as a validation tool for the model to ensure the development data set is similar to the current population. In its current use in the industry, PSI is compared to rule-of-thumb benchmarks, that have no known properties. In this paper, we provide some guidance by deriving an approximate distribution of PSI. Users may calculate benchmarks depending on the desired false significance rate α. Besides α, the benchmarks depend on the number of bins B, and the base and target sample sizes n and m.

References

1. Agresti (2007), *An Introduction to Categorical Data Analysis, 2nd. Ed*, John Wiley & Sons.
2. Cochran (1977), *Sampling Techniques*, Wiley.
3. FED (2011), Supervisory guidance on model risk management contents `http://www.federalreserve.gov/supervisionreg/srletters/sr1107a1.pdf`
4. FDIC (2007), Scoring and Modeling VIII. SCORING AND MODELING `http://www.fdic.gov/regulations/examinations/credit_card/pdf_version/ch8.pdf`
5. OCC (2016), Sensitivity to Market Risk Installment Lending. `http://www.occ.gov/publications/publications-by-type/comptrollershandbook/installment-lending/pub-ch-installment-lending.pdf`
6. Stapleton, J.H. (2009), *Linear statistical models*, John Wiley & Sons.
7. Yurdakul (2018), Statistical Properties of Population Stability Index `https://scholarworks.wmich.edu/dissertations/3208/`