Vol. 35 (2015)
No. 4

数 学 杂 志
J. of Math. (PRC)

APPLICATION OF THE PARETO POSITIVE STABLE DISTRIBUTION IN INSURANCE CLAIM

XUAN Hai-yan [1], BAO Hai-ming [2], SHI Yong-xia [3]

(*1.College of Economics and Mangement, Lanzhou University of Technology, Lanzhou 730050, China*)

(*2.Lanzhou Jincheng Branch, China Construction Bank, Lanzhou 730030, China*)

(*3.College of Science, Lanzhou University of Technology, Lanzhou 730050, China*)

Abstract: In this paper, we study the application of the Pareto positive stable distribution in insurance. The parameter estimates of Pareto positive stable distribution, normal distribution and Pareto distribution are obtained using the method of maximum likelihood estimates. By Akaike information criterion, it is indicated that the Pareto positive stable distribution can fit the insurance data well.

Keywords: Pareto positive stable distribution; parameter estimation; insurance claim; Akaike information criterion

2010 MR Subject Classification: 62P20; 91G10

Document code: A **Article ID:** 0255-7797(2015)04-0889-09

1 Introduction

Insurance company claim is an important factor in its development. Insurance companies commonly use exponential distribution, lognormal distribution or Pareto distribution to fit claims data and control risk. In the insurance claim model, making premium to insurance company and resisting the risk, Pareto distribution model is of guiding significance. It is suitable for fitting large claims data. From the historical data, insurance claim often shows high positive bias. On the distribution, it shows fat tail shape. But Pareto distribution has a heavy-tailed charcteristic. So when simulating these data, Pareto distribution is popular with the scholars.

The Pareto positive stable (PPS) distribution was firstly proposed by Sarabia and Prieto in their thesis in 2009 [1]. They explained the reason why the Pareto positive stable distribution is used to model losses in insurance. For instance, the Pareto positive stable distribution easily fit and have a simple quantile expression. It makes the Monte Carlo simulation simple. For the risk value, an analytical expression is provided. Ortobelli et al. [2] proposed some stable Paretian models for optimal portfolio selection and for quantifying

* **Received date:** 2014-09-01 **Accepted date:** 2014-12-24

Foundation item: Supported by National Natural Science Foundation of China (11261031); the Natural Science Foundation of Inner Mongolia Autonomous Region (2013MS0101).

Biography: Xuan Haiyan(1973–), female, korean, born at Wangqing, Jilin, associate professor, major in applied probability statistics.

the risk of a given portfolio. Guillen et al. [3] proposed using the Pareto positive stable distribution simulation insurance data and studied it.

This paper firstly introduces Pareto positive stable distribution, its the probability density function and the quantile function. Then we give the Pareto positive stable distribution moment estimation, regression estimation, and maximum likelihood estimation. Then, for the randomly generated data using maximum likelihood estimation method, we do estimation for the parameters of the Pareto positive stable distribution, the normal distribution and the Pareto distribution. And their parameters are compared. By AIC information criterion [4], we get the Pareto positive stable distribution can better fit the data in insurance claims. Therefore, the Pareto positive stable distribution can better analyze insurance claims data.

2 The Pareto Positive Stable Distribution

The Pareto positive stable distribution is given by

$$\begin{aligned} F(x;\lambda,\nu,\sigma) &= P(X \le x) \\ &= \begin{cases} 1-\exp\{-\lambda[\log(x/\sigma)]^v\}, & x \ge \sigma, \\ 0, & x < \sigma, \end{cases} \end{aligned} \tag{2.1}$$

where $\lambda, \sigma, \nu > 0$. Note that $\lambda, \nu > 0$ are shape parameters and σ is a scale parameter.

Derivating to the Pareto positive stable cumulative distribution function, we can obtain the probability density function (*pdf*) of it:

$$f(x;\lambda,\nu,\sigma) = \begin{cases} \dfrac{\lambda\nu[\log(x/\sigma)]^{\nu-1}}{x}\exp\{-\lambda[\log(x/\sigma)]^v\}, & x \ge \sigma, \\ 0, & x < \sigma. \end{cases} \tag{2.2}$$

The Pareto positive stable distribution have a two-fold origin. One is the classical Pareto distribution [5–6]. Its cumulative distribution function (*cdf*) is

$$\begin{aligned} F(x) &= P(X \le x) \\ &= \begin{cases} 1-(\dfrac{x}{\sigma})^{-\alpha}, x \ge \sigma > 0, \\ 0, \quad x < \sigma, \end{cases} \end{aligned}$$

where $\alpha > 0$ is a shape parameter and σ is a scale parameter, which represents the smallest value in the sample. Let $\alpha = \lambda\nu$, then we can obtain the cumulative distribution function of the PPS distribution. The other is from a simple transformation of the classical Weibull distribution [7–8]. Let Z be a classical Weibull distribution with cumulative distribution function (*cdf*)

$$F_Z(z) = 1-\exp(-z^\nu),\ z > 0,$$

where $\nu > 0$. Then, the new random variable

$$X = \sigma\exp(\lambda^{-1/\nu}Z)$$

follow the Pareto positive stable distribution, denoted by $X \sim PPS(\lambda, \sigma, \nu)$, where $\sigma, \lambda > 0$.

Figure 1, Figure 2, Figure 3 and Figure 4 are the probability density function of the PPS distribution with different parameters.

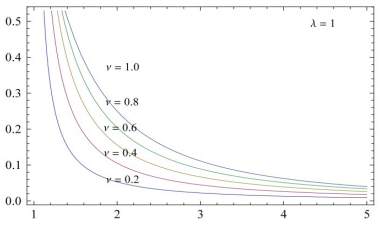

Figure 1: The *pdf* of PPS distribution when $\sigma = 1$, $\lambda = 1$ and $\nu = 0.2, 0.4, 0.6, 0.8, 1$

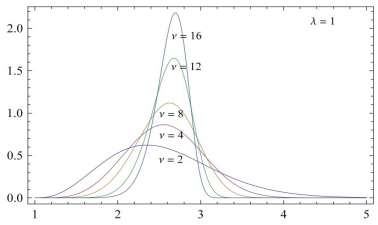

Figure 2: The *pdf* of PPS distribution when $\sigma = 1$, $\lambda = 1$ and $\nu = 2, 4, 8, 12, 16$

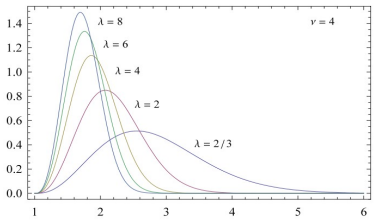

Figure 3: The *pdf* of PPS distribution when $\sigma = 1$, $\nu = 4$ and $\lambda = 2/3, 2, 4, 6, 8$

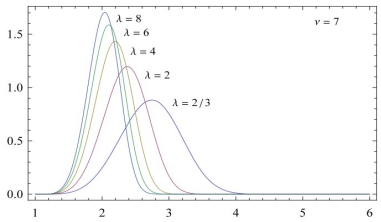

Figure 4: The *pdf* of PPS distribution when $\sigma = 1$, $\nu = 7$ and $\lambda = 2/3, 2, 4, 6, 8$

The quantile function of the Pareto positive stable distribution can be easily obtained. Let $p = F(x)$, then

$$p = 1 - \exp\{-\lambda[\log(x/\sigma)]^{v}\},$$

where $x \geq \sigma$, and then

$$-\frac{1}{\lambda}\log(1-p) = [\log(\frac{x}{\sigma})]^{\nu},$$

we can obtain

$$Q(p) = F^{-1}(p) = \sigma \exp\{[-\frac{1}{\lambda}\log(1-p)]^{1/\nu}\},\ 0 < p < 1.$$

3 Parameter Estimation

Let $x_1, x_2, \cdots, x_n$ be a sample of size n drawn from a Pareto positive stable distribution. We assume that parameter σ is the smallest sample value. Then, we introduce three estimation methods of Pareto positive stable distribution: moments estimates, regression estimates and maximum likelihood estimates. We define the random variable $Z = \log(X/\sigma)$ and its observed values is $z_i = \log(x_i/\sigma), i = 1, 2, \cdots, n$.

3.1 Moments Estimates

The r-order origin moments of the random variable z is

$$E(Z^r) = E[(\log \frac{X}{\sigma})^r] = \int_{\sigma}^{\infty} z^r f(x)dx$$
$$= \lambda^{-r/\nu} \int_0^{\infty} e^{-\lambda z^{\nu}} \cdot (\lambda z^{\nu})^{\frac{r}{\nu}} \cdot d(\lambda z^{\nu}) = \lambda^{-r/\nu}\Gamma(1+\frac{r}{\nu}),\ r>0,$$

where $\Gamma(x) = \int_0^{\infty} e^{-t}t^{x-1}dt$, $\Gamma(1+\frac{r}{\nu}) = \int_0^{\infty} e^{-t}t^{\frac{r}{\nu}}dt$.

Note that $E(Z) = \bar{z} = \lambda^{-1/\nu}\Gamma(1+\frac{1}{\nu})$, $E(Z^2) = \lambda^{-2/\nu}\Gamma(1+\frac{2}{\nu})$, and

$$s_z^2 = EZ^2 - [E(Z)]^2 = \lambda^{-2/\nu}[\Gamma(1+\frac{2}{\nu}) - \Gamma(1+\frac{1}{\nu})^2],$$

thus

$$\frac{(\bar{z})^2}{s_z^2} = \frac{\Gamma(1+\frac{1}{\nu})^2}{\Gamma(1+\frac{2}{\nu}) - \Gamma(1+\frac{1}{\nu})^2}, \tag{3.1}$$

where $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$ and $s_z^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2$ are mean and variance of sample to random variable Z respectively. We solve the estimator of ν from the formula (3.1), since $E(Z) = \bar{z} = \lambda^{-1/\nu}\Gamma(1+\frac{1}{\nu})$, we obtain the estimator of λ:

$$\hat{\lambda} = [\frac{\bar{z}}{\Gamma(1+\frac{1}{\hat{\nu}})}]^{-\hat{\nu}}. \tag{3.2}$$

3.2 Regression Estimates

From expression (2.1), taking logarithms twice in $1 - F(x)$, we get

$$\log[-\log(1-F(x))] = \log\lambda + \nu\log[\log(x/\sigma)]. \tag{3.3}$$

If σ is know, it is a linear relation in $\log[\log(x/\sigma)]$. Let $a = \log\lambda$, $X = \log[\log(x/\sigma)]$, $b = \nu$, and $y_i = \log[-\log(1-F_n(x_i))]$, then the residual sum of squares (RSS):

$$RSS = \sum_{i=1}^{n}[y_i - a - bX_i]^2.$$

Taking partial derivative for RSS we get

$$\begin{cases} na + (\sum_{i=1}^{n} X_i)b = \sum_{i=1}^{n} y_i, \\ (\sum_{i=1}^{n} X_i)a + (\sum_{i=1}^{n} X_i^2)b = \sum_{i=1}^{n} X_i y_i. \end{cases}$$

Because X_i are not all equal, the coefficient determinant

$$\begin{vmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{vmatrix} = n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2 = n\sum_{i=1}^{n}(X_i - \overline{X})^2 \neq 0,$$

hence, equations have a unique solution. The estimators of b, a are

$$\hat{b} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(y_i - \overline{y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2},$$
$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{b}{n}\sum_{i=1}^{n} X_i = \overline{y} - \hat{b}\overline{X},$$

where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $X_i = \log z_i$, then

$$\hat{\nu} = \frac{\sum_{i=1}^{n}(\log z_i - \overline{\log z})(y_i - \overline{y})}{\sum_{i=1}^{n}(\log z_i - \overline{\log z})^2}, \tag{3.4}$$

and

$$\hat{\lambda} = \exp\{\overline{y} - \hat{\nu}\overline{\log z}\}. \tag{3.5}$$

3.2 Maximum Likelihood Estimates

From expression (2.2) of the probability density function (*pdf*) of the PPS distribution, the log-likelihood function of PPS distribution is given by

$$\begin{aligned}\log \ell(\lambda, \nu) &= \sum_{i=1}^{n} \log f(x_i) \\ &= n\log\lambda + n\log\nu + (\nu - 1)\sum_{i=1}^{n}\log z_i - \lambda\sum_{i=1}^{n} z_i^{\nu} - \sum_{i=1}^{n}\log x_i.\end{aligned}$$

Taking partial derivative with respect to λ and ν we obtain the equations

$$\begin{cases} \dfrac{\partial \log \ell}{\partial \lambda} = \dfrac{n}{\lambda} - \sum_{i=1}^{n} z_i^{\nu} = 0, \\ \dfrac{\partial \log \ell}{\partial \nu} = \dfrac{n}{\nu} + \sum_{i=1}^{n}\log z_i - \lambda\sum_{i=1}^{n} z_i^{\nu}\log z_i = 0. \end{cases}$$

We solve λ in the first equation and put it into the second equation. We obtain the equation in ν,

$$\frac{1}{\nu} + \frac{1}{n}\sum_{i=1}^{n}\log z_i - \frac{\sum_{i=1}^{n} z_i^{\nu}\log z_i}{\sum_{i=1}^{n} z_i^{\nu}} = 0 \tag{3.6}$$

and solve the estimator $\widehat{\nu}$. Then we put it into the first equation and so obtain the estimator of λ:

$$\hat{\lambda} = [\frac{1}{n}\sum_{i=1}^{n} z_i^{\hat{\nu}}]^{-1}. \tag{3.7}$$

4 Application in Insurance Claim

We consider the data on motor insurance claims of a major insurance company. A sample of 518 randomly generates by MATLAB between the minimum and maximum claims. For each claim i, we observe X_1 (cost of property damage) and X_2 (cost of medical expenses). Unit of data is thousand of yuan. The basic numerical characteristics of X_1 and X_2 can be seen in Table 1.

Table 1: The Basic Numerical Characteristics of X_1 And X_2

	Mean	Std.Dev	Skewness	Kurtosis	Min	Max
X_1	201.6638	228.2939	1.2744	3.3027	0.1074	819.5100
X_2	20.5363	21.9714	0.9908	2.5048	0.0070	71.2090

In order to test the adequacy of the Pareto positive stable distribution for data, we estimate the parameters of the PPS distribution by the date X_1 and X_2 with the maximum likelihood estimation method. The parameter values see Table 2.

Table 2: The Parameter Estimation of The PPS Distribution

	PPS	
	$\widehat{\nu}$	$\widehat{\lambda}$
X_1	6.1579	2.7222×10^{-6}
X_2	6.3388	1.3268×10^{-6}

Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ and $Z_{(1)}, Z_{(2)}, \cdots, Z_{(n)}$ be the order statistics of $X_1$and X_2 respectively. $F_n(x_{(i)}) = \frac{i}{n+1}$ is the empirical cumulative distribution function of the sample, then $1 - F_n(x_{(i)})$ corresponds to the rank of the ith data divided by $n+1$. For the two sets of data, we take logarithm. The horizontal axis represents the natural logarithm of size of the size of the sample observation value and the vertical axis represents the logarithm of the samples' rank. Then fitting, the abscissa is $\log(x)$ and the ordinate is $\log[(n+1)(1-F(x))]$, as showing in Figure 5 and Figure 6.

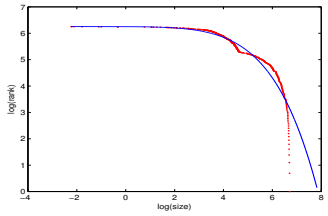

Figure 5: The fitted plot in log-log scale for X_1 data set

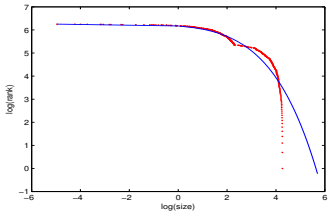

Figure 6: The fitted plot in log-log scale for X_2 data set

By equation (3.3), then

$$\log[-\log(1 - F_n(x_i))] = \log\lambda + \nu \log[\log(x_i/\sigma)]. \tag{4.1}$$

Hence, if $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ follow the Pareto positive stable distribution, we infer the double log-log scatter plots demonstrate linear features and slope is positive. For the two sets of data, we take logarithm twice. The horizontal axis represents $\log[\log(\frac{size}{\sigma})]$ and the vertical axis represents $\log[-\log(\frac{rank}{n+1})]$. Then fitting, the abscissa is $\log[\log(\frac{x}{\sigma})]$ and the ordinate is $\log[-\log(1-F(x)]$, as showing in Figure 7 and Figure 8.

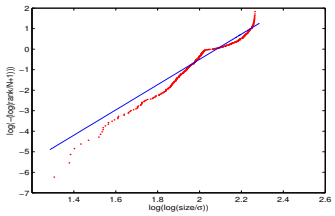

Figure 7: The fitted plot in double log-log scale for X_1 data set

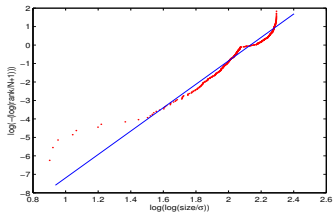

Figure 8: The fitted plot in double log-log scale for X_2 data set

Form Figure 5 and Figure 6, we find as long as a deviation appears in the large data fitting, this is just corresponding to huge claims. Form Figure 7 and Figure 8, we find that both plots are clearly linear, which supports the assumption of the Pareto positive stable distribution for both sets of data. Therefore, these two sets of data fit well.

Table 3: The *cdf* and *pdf* of Different Distribution

Distribution	$F(x)$	$f(x)$
PPS	$1-\exp\{-\lambda[\log(x/\sigma)]^{\nu}\}$, $x \geq \sigma > 0$	$\frac{\lambda\nu[\log(x/\sigma)]^{\nu-1}}{x}\exp\{-\lambda[\log(x/\sigma)]^{\nu}\}$, $x \geq \sigma > 0$
normal	$\Phi(\frac{x-\mu}{\sigma})$, $x \in R$	$\frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\}$, $x \in R$
Pareto	$1-(\frac{x}{\sigma})^{-\alpha}$, $x \geq \sigma > 0$	$\frac{\alpha\sigma^{\alpha}}{x^{\alpha+1}}$, $x \geq \sigma > 0$

In order to illustrate the advantages of the Pareto positive stable distribution for fitting insurance claims data, we compare the Pareto positive stable distribution with the normal distribution and the Pareto distribution. Table 3 is the expression of probability density functions and the cumulative distribution functions of different distributions.

All the parameters of three distributions are got by maximum likelihood estimation. The parameters' estimated results are shown in Table 4.

Table 4: The Parameters Estimation of Different Distribution

	PPS分布		Pareto	normal	
	$\widehat{\nu}$	$\widehat{\lambda}$	$\widehat{\alpha}$	$\widehat{\mu}$	$\widehat{\sigma}$
X_1	6.1579	2.7222×10^{-6}	0.1340	201.6638	228.2939
X_2	6.3388	1.3268×10^{-6}	0.1270	20.5363	21.9714

We select the preferred model by using Akaike information criterion (AIC). Akaike information criterion is defined as

$$AIC = 2(s - \log \ell),$$

where s is the number of parameters and $\log \ell$ is the log-likelihood function. Akaike information criterion shows that the preferred model is the one with the lowest AIC value. If the AIC value of the PPS distribution is smaller than the AIC value of the normal and Pareto distribution, indicating the PPS distribution can fit date better than the normal distribution and Pareto distribution. Importing X_1 and X_2 into MATLAB to calculate, we can see the AIC (normal)–$AIC(PPS)$ and AIC(Pareto)–$AIC(PPS)$ as a function of the sample sequence number N to plot. If the difference is positive, which means that the AIC of the PPS distribution is smaller, it shows the fitting effect of the PPS distribution is better.

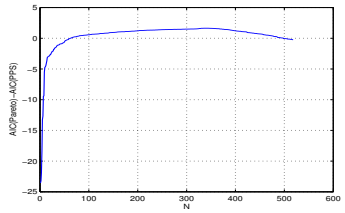

Figure 9: AIC(Pareto)-$AIC(PPS)$ as a function of N for X_1 data set

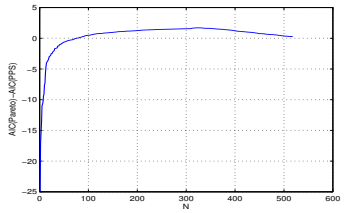

Figure 10: AIC(Pareto)-$AIC(PPS)$ as a function of N for X_2 data set

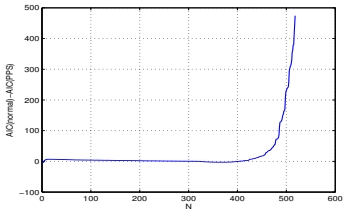

Figure 11: AIC(normal)-$AIC(PPS)$ as a function of N for X_1 data set

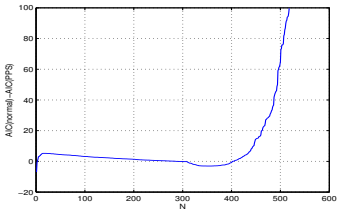

Figure 12: AIC(normal)-$AIC(PPS)$ as a function of N for X_2 data set

Looking at Figure 9, Figure 10, Figure 11 and Figure 12, regardless of the data X_1 or X_2, the vast majority difference of AIC value is positive, so it shows the PPS distribution is better than the normal distribution and Pareto distribution for fitting insurance claims data.

5 Conclusions

In the insurance claims, there exists many small and large claims. Some insurance claims data are with relatively thick tail, for example motor vehicle insurance. For this insurance data, using the Pareto positive stable distribution to fit, it will get better fitting effect. The Pareto positive stable distribution has the simple expression of probability density function and quantile function. Its parameters estimate can be obtained by moments estimates, regression estimates and maximum likelihood estimates. On the basis of parameter estimates, comparing with other distributions, the Pareto positive stable distribution can fit better

insurance claims data. Therefore, in the insurance industry, using Pareto positive stable distribution in the analysis of insurance claims data has a better application.

References

[1] Sarabia J M, Prieto F. The Pareto-positive stable distribution: a new descriptive method for city size data[J]. Physica A: Stat. Mech. Appl., 2009, 388(19): 4179–4191.

[2] Ortobelli S, Rachev S T, Fabozzi F J. Risk management and dynamic portfolio selection with stable Paretian distribution[J]. Journal of Empirical Finance, 2010, 17(2): 195–211.

[3] Guillen M, Prieto F, Sarabia J M. Modelling losses and locating the tail with the Pareto Positive Stable distribution[J]. Insurance: Math. Econ., 2011, 49(3): 454–461.

[4] Akaike H. A new look at the statistical model identification[J]. IEEE Trans. Automatic Control, 1974, 19(6): 716–723.

[5] Yao Hui, Dai Yong, Xie Lin. Pareto-geometric distribution[J]. J. Mathematics, 2012, 32(2): 339–351.

[6] Arnold B C. Pareto distributions[M]. Fairland, Maryland: International Co-operative Publishing House, 1983.

[7] Balakrishnan N, Nevzorov V B. A primer on statistical distributions[M]. New York: John Wiley, 2003.

[8] Castillo E, Hadi A S, Balakrishnan N, Sarabia J M. Extreme value and related models with applications in engineering and science[M]. New York: John Wiley, 2004.

Pareto严格稳定分布在保险理赔中的应用

玄海燕[1], 包海明[2], 史永侠[3]

(1.兰州理工大学经济管理学院, 甘肃 兰州 730050)

(2.中国建设银行兰州金城支行, 甘肃 兰州 730030)

(3.兰州理工大学理学院, 甘肃 兰州 730050)

摘要: 本文研究了Pareto严格稳定分布在保险中的应用. 利用极大似然估计的方法得到了Pareto严格稳定分布, 正态分布和Pareto分布的参数估计. 根据信息准则, 表明Pareto严格稳定分布能够较好地拟合保险数据.

关键词: Pareto严格稳定分布; 参数估计; 保险理赔; 信息准则

MR(2010)**主题分类号**: 62P20; 91G10 **中图分类号**: O213