

# 电力系统负荷预报误差的概率 密度函数建模

赵希人 李大为 李国斌 郝长泉

(哈尔滨船舶工程学院自控系, 150001)

## 摘 要

本文依据数理统计理论, 提出对电力系统负荷预报误差的概率密度函数的建模方法, 给出建模公式并证明其渐近无偏性及均方收敛性。该方法已成功地用于黑龙江省电力系统负荷预报误差分析。

**关键词:** 概率密度函数, 建模, 预报误差, 电力系统负荷, 检验。

## 一、引 言

电力系统负荷预报是现代电力科学中的重要研究课题, 而预报误差的统计分析又是为预报的可靠性提供了直接的理论依据及数据结果。因此, 预报误差统计分析是电力系统负荷预报中的一项重要内容。

由于电力系统负荷变化是一非平稳过程, 所以, 通常采用分解建模及预报方法<sup>[1-3]</sup>。设  $\{y(n), n = 1, 2, \dots, N\}$  是电力系统负荷序列, 可以认为  $y(n)$  由三项组成, 即

$$y(n) = f(n) + P(n) + X(n), \quad n = 1, 2, \dots, N. \quad (1)$$

其中  $f(n)$  为趋势项,  $P(n)$  为周期项,  $X(n)$  为平稳随机项。利用分解建模及预报方法<sup>[3]</sup>, 可以得到  $y(n)$  的预报  $\hat{y}(n)$ 。为了给出预报误差, 通常是利用简单的统计方法来计算后验预报误差的均方差  $\sigma_y$ ,

$$\sigma_y = \sqrt{\frac{1}{N} \sum_{n=1}^N [y(n) - \hat{y}(n)]^2}, \quad (2)$$

这里, 值得研究的问题是, 用(2)式进行计算所得的  $\sigma_y$  的概率含义是什么? 能否表征预报精度。作者认为, 只有将负荷预报误差  $\tilde{y}(n) = y(n) - \hat{y}(n)$  的概率密度函数建模求出, 才能圆满地回答上述问题, 从而可给出在一定概率意义下的预报置信区间。

关于概率密度函数的建模及其应用, 国外已有许多报导<sup>[4-6]</sup>, 但在国内尚属少见。

本文所要解决的问题是, 依据预报后验误差序列  $\{\tilde{y}(n) = y(n) - \hat{y}(n), n = 1, 2, \dots, N\}$ , 对误差  $\tilde{y}(n)$  的概率密度函数  $f_N(y)$  进行建模, 给出建模公式并讨论其渐近无偏性及均方收敛性。

## 二、概率密度函数建模的理论依据

为了讨论密度函数的建模方法,首先引进核函数定义.

**定义 1.** 设函数  $K(x)$  是实数域  $R$  上的概率密度函数且满足

$$1) \sup_x K(x) \leq M, M \text{ 为常数}, x \in R, \quad (3)$$

$$\text{并且 } |x|K(x) \rightarrow 0, |x| \rightarrow \infty. \quad (4)$$

$$2) K(-x) = K(x), \int_{-\infty}^{+\infty} x^2 K(x) dx < \infty, x \in R, \quad (5)$$

3)  $K(x)$  的特征函数  $\hat{K}(u)$  是绝对可积的,则称  $K(x)$  是核函数.  
可以验证,下述函数:

$$K_1(x) = \begin{cases} 1/2, & |x| \leq 1, \\ 0, & |x| > 1, \end{cases} \quad K_2(x) = \begin{cases} 1 - |x|, & |x| \leq 1, \\ 0, & |x| > 1, \end{cases} \quad (6), (7)$$

$$K_3(x) = (2\pi)^{-\frac{1}{2}} \exp\{-x^2/2\}, \quad (8)$$

均是核函数.

**引理 1.** 设  $K(x)$  是实数域  $R$  上的函数,且

$$|K(z)| \leq M, z \in R, M \text{ 为常数}, \quad (9)$$

$$\int_{-\infty}^{+\infty} |K(z)| dz < \infty, \quad (10)$$

$$|z||K(z)| \rightarrow 0, |z| \rightarrow \infty, \quad (11)$$

$g(x)$  是定义在  $R$  上的函数,且

$$\int_{-\infty}^{+\infty} |g(z)| dz < \infty, \quad (12)$$

现定义

$$g_n(x) = \frac{1}{h_n} \int_{-\infty}^{+\infty} K(z/h_n) g(x-z) dz, \quad (13)$$

则当  $n \rightarrow \infty, 0 < h_n \rightarrow 0$  时,如果  $g(x)$  连续,必有

$$\lim_{n \rightarrow \infty} g_n(x) = g(x) \int_{-\infty}^{+\infty} K(z) dz, \quad (14)$$

并且若  $g(x)$  一致连续,(14)式的收敛也是一致的.

证明. 对任意  $\delta > 0$ , 易知

$$\left| g_n(x) - g(x) \int_{-\infty}^{+\infty} K(z) dz \right| \quad (15)$$

$$= \left| \frac{1}{h_n} \int_{-\infty}^{+\infty} K(z/h_n) g(x-z) dz - g(x) \int_{-\infty}^{+\infty} K(z) dz \right|$$

$$\leq \left| \int_{|z| \leq \delta} \frac{1}{h_n} K(z/h_n) g(x-z) dz - \int_{|z| \leq \delta/h_n} g(x) K(z) dz \right|$$

$$+ \left| \int_{|z| > \delta} \frac{1}{h_n} K(z/h_n) g(x-z) dz \right| + \left| \int_{|z| > \delta/h_n} g(x) K(z) dz \right|. \quad (16)$$

因  $g(x)$  连续并考虑(10)式,于是不等号右第一项为

$$\begin{aligned}
& \left| \int_{|z|<\delta} \frac{1}{h_n} K(z/h_n) g(x-z) dz - \int_{|z|\leq\delta/h_n} g(x) K(z) dz \right| \\
&= \left| \int_{|z|\leq\delta/h_n} K(z) g(x-h_n z) dz - \int_{|z|\leq\delta/h_n} g(x) K(z) dz \right| \\
&\leq \int_{|z|\leq\delta/h_n} |g(x-h_n z) - g(x)| |K(z)| dz \\
&\leq \sup_{|z|\leq\delta} |g(x-z) - g(x)| \int_{-\infty}^{+\infty} |K(z)| dz < \varepsilon, \quad \forall \delta > 0. \tag{17}
\end{aligned}$$

又因  $\delta/h_n \rightarrow \infty, n \rightarrow \infty$ , 并考虑(11),(12)式, 则(16)式不等号右边第二项为

$$\begin{aligned}
& \left| \int_{|z|>\delta} \frac{1}{h_n} K(z/h_n) g(x-z) dz \right| \\
&\leq \int_{|z|>\delta} \frac{1}{\delta} \left| \frac{z}{h_n} K(z/h_n) \right| |g(x-z)| dz \\
&\leq \frac{1}{\delta} \sup_{|z|\geq\delta/h_n} |z| |K(z)| \int_{-\infty}^{+\infty} |g(z)| dz \rightarrow 0, \quad n \rightarrow \infty. \tag{18}
\end{aligned}$$

最后, 由  $g(x)$  连续并考虑到(10)式, 则(16)式不等号右边第三项为

$$\left| \int_{|z|>\delta/h_n} g(x) K(z) dz \right| \leq |g(x)| \int_{|z|>\delta/h_n} |K(z)| dz \rightarrow 0, \quad n \rightarrow \infty. \tag{19}$$

将以上三式代入(16)式立得(14)式. 进一步, 若  $g(x)$  一致连续, 则由(17)式可知, (14)式的收敛是一致的. 引理证毕.

**定义 2.** 设  $K(\cdot)$  是  $R$  中的核函数,  $(x_1, x_2, \dots, x_n)$  是母体  $X$  的简单子样, 则称

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-x_i}{h_n}\right) \tag{20}$$

是母体  $X$  的概率密度函数  $f(x)$  的估计.

密度函数估计  $\hat{f}_n(x)$  有如下性质:

**定理 1.** 如果母体  $X$  的密度函数  $f(x)$  连续, 则当  $n \rightarrow \infty, 0 < h_n \rightarrow 0$  时, 有

$$E\hat{f}_n(x) = f(x), \quad n \rightarrow \infty, \tag{21}$$

即  $\hat{f}_n(x)$  是  $f(x)$  的渐近无偏估计. 进一步, 若  $f(x)$  一致连续, 则  $\hat{f}_n(x)$  是  $f(x)$  的一致渐近无偏估计.

证明. 由于  $(x_1, x_2, \dots, x_n)$  是母体  $X$  的简单子样, 故由引理 1 可得

$$\begin{aligned}
E[\hat{f}_n(x)] &= E\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-x_i}{h_n}\right)\right] = \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K\left(\frac{x-x_i}{h_n}\right) f(x_i) dx_i \\
&= \int_{-\infty}^{+\infty} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) f(z) dz = f(x) \int_{-\infty}^{+\infty} K(z) dz, \quad n \rightarrow \infty, \\
&= f(x), \quad n \rightarrow \infty, \tag{证毕.}
\end{aligned}$$

**定理 2.** 如果母体  $X$  的密度函数  $f(x)$  连续, 则当  $n \rightarrow \infty, 0 < h_n \rightarrow 0, nh_n \rightarrow \infty$  时有

$$E[\hat{f}_n(x) - f(x)]^2 = 0, \quad n \rightarrow \infty, \tag{22}$$

即  $\hat{f}_n(x)$  均方收敛于  $f(x)$ . 进一步, 若  $f(x)$  一致连续, 则  $\hat{f}_n(x)$  一致均方收敛于  $f(x)$ .

证明. 由  $(x_1, x_2, \dots, x_n)$  是  $X$  的简单子样及引理 1 可推得

$$\begin{aligned}
 E[\hat{f}_n(x) - f(x)]^2 &\triangleq D[\hat{f}_n(x)] \\
 &= \frac{1}{n^2} \sum_{i=1}^n D \left[ \frac{1}{h_n} K \left( \frac{x - x_i}{h_n} \right) \right] = \frac{1}{n} D \left[ \frac{1}{h_n} K \left( \frac{x - x_i}{h_n} \right) \right] \\
 &\leq \frac{1}{n} E \left[ \frac{1}{h_n} K \left( \frac{x - x_i}{h_n} \right) \right]^2 = \frac{1}{nh_n} \int_{-\infty}^{+\infty} \frac{1}{h_n} K^2 \left( \frac{x - z}{h_n} \right) f(z) dz \\
 &= \frac{1}{nh_n} f(x) \int_{-\infty}^{+\infty} K^2(z) dz \rightarrow 0, \quad n \rightarrow \infty, \quad nh_n \rightarrow \infty.
 \end{aligned}$$

证毕。

### 三、小时负荷预报误差的密度函数建模

利用上面介绍的方法，作者对黑龙江省电力系统小时负荷预报误差的概率密度函数进行了建模。其中，小时负荷预报是采用分解建模预报法<sup>[3]</sup>，后验误差数据取自 90 年 4 月 14 日至 5 月 24 日，共计 40 天的数据。

具体的建模步骤如下：

1) 密度函数估计。取 40 天的第  $j$  ( $j = 1, 2, \dots, 24$ ) 点钟的误差数据，利用定义 2 中(20)式，可计算出第  $j$  点钟的预报误差密度函数  $\hat{f}_{nj}(x)$  并画出曲线。

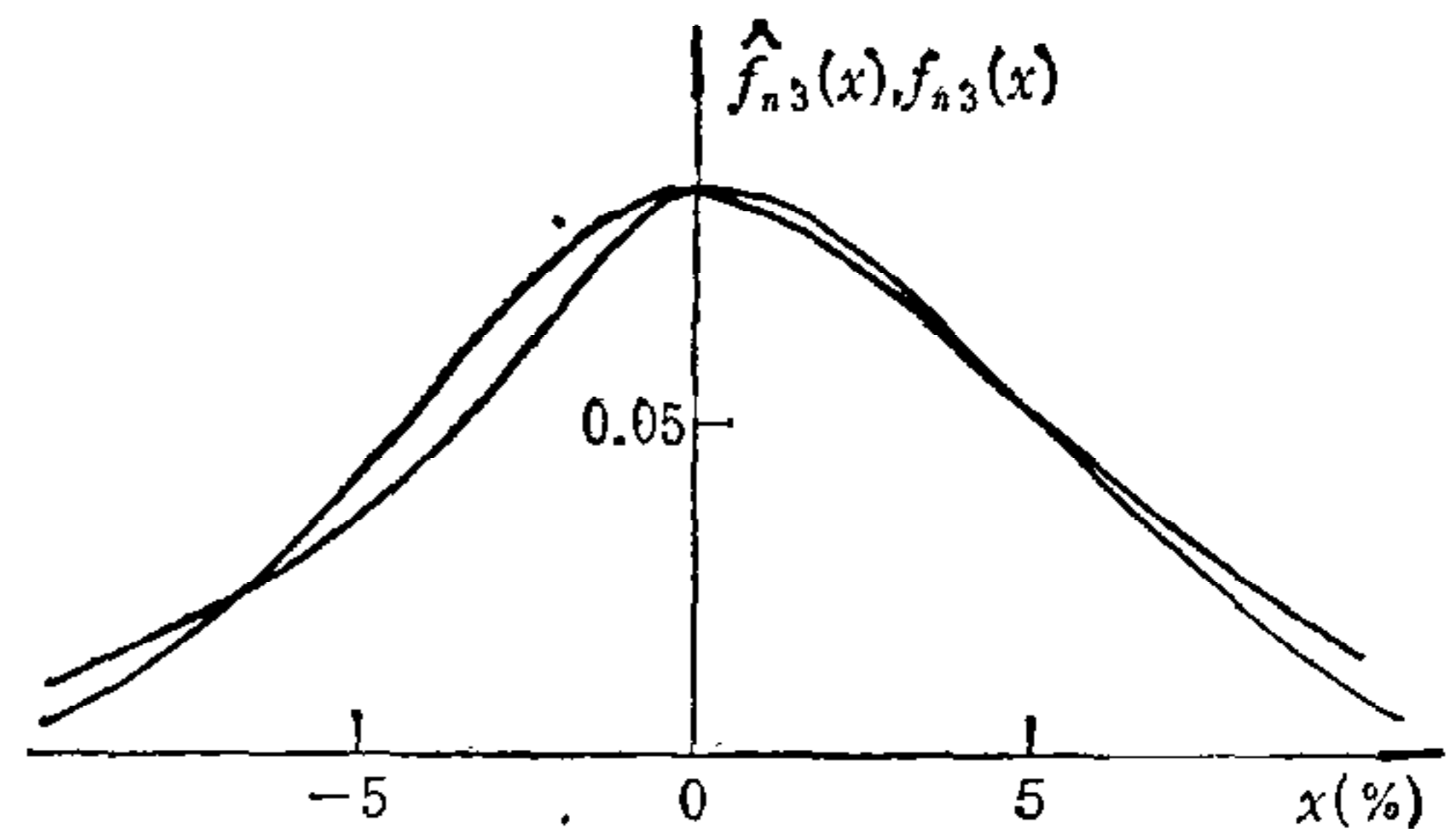


图 1 三点钟的密度估计  $\hat{f}_{n3}(x)$  及正态拟合  $f_{n3}(x)$

$$\hat{f}_{nj}(x) = \frac{1}{nh_n} \sum_{i=1}^n K \left( \frac{x - x_{ij}}{h_n} \right), \tag{23}$$

其中  $n = 40$ ,  $h_n = 1.8 \sim 2$ ,  $K(\cdot)$  取正态核函数。

例如，4 月 14 日至 5 月 24 日三点钟的预报误差密度函数曲线  $\hat{f}_{n3}(x)$  示于图 1。

2) 密度函数的正态拟合。利用(23)式对 24 小时的预报误差密度函数分别进行估计并画出曲线后，可以看出，这些曲线的形状基本接近于正态分布曲线，这说明利用分解建模预报法所得到的预报误差基本服从正态分布，于是可用正态曲线来拟合  $\hat{f}_{nj}(x)$ ,  $j = 1, 2, \dots, 24$ 。由函数  $\hat{f}_{nj}(x)$  知，当  $x = \bar{x}_j$  时函数出现最大并记作  $\hat{f}_{nj\max} \triangleq \hat{f}_{nj}(\bar{x}_j)$ ，这样，若取正态拟合密度函数为

$$f_{nj}(x) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp \left\{ -\frac{(x - a_j)^2}{2\sigma_j^2} \right\} \triangleq N(a_j, \sigma_j^2). \tag{24}$$

则由

$$a_j = \bar{x}_j, \quad \frac{1}{\sqrt{2\pi}\sigma_j} = \hat{f}_{nj\max} \tag{25}, (26)$$

可近似估计出  $a_j$  及  $\sigma_j$ 。

例如，4 月 14 日至 5 月 24 日三点钟的预报误差密度函数估计  $\hat{f}_{n3}(x)$  的正态拟合曲线  $f_{n3}(x)$  示于图 1，可以看出两者是比较接近的。

3) 统计检验。通过以上计算所得到的正态拟合函数  $f_{nj}(x)$ ,  $j = 1, 2, \dots, 24$  是否

可以接受,应采用皮尔逊定理进行  $\chi^2$  检验.

例如,三点钟的正态拟合函数是  $f_{n3}(x) = N(0.42, 4.5^2)$ . 现假设  $H_0: f_{n3}(x) = N(0.42, 4.5^2)$  并取水平  $\alpha = 0.05$ . 为检验  $H_0$  是否可以接受,取 150 天的三点钟预报误差数据并将数轴划分为 8 个子区间. 运用皮尔逊  $\chi^2$  检验的计算结果由表 1 给出.

表 1 对三点钟  $H_0$  检验计算结果

子区间(%)	$(-\infty, -8.58]$	$(-8.58, -5.58]$	$(-5.58, -2.58]$	$(-2.58, 0.42]$
$n_i$	4	11	20	40
$nP_i$	4.5	9.75	23.25	37.5
$(n_i - nP_i)^2/nP_i$	0.06	0.16	0.45	0.17
子区间(%)	$(0.42, 3.42]$	$(3.42, 6.42]$	$(6.42, 9.42]$	$(9.42, +\infty)$
$n_i$	37	24	8	6
$nP_i$	37.5	23.25	9.75	4.5
$(n_i - nP_i)^2/nP_i$	0.01	0.02	0.31	0.5

因为  $\sum_{i=1}^8 (n_i - nP_i)^2/nP_i = 1.68 < \chi^2(7, 0.05) = 14.07$ , 所以可以接受  $H_0$ , 或者说

$f_{n3}(x) = N(0.42, 4.5^2)$  与三点钟预报误差的实际分布密度函数无显著差异.

按照以上三步作法,对 24 小时的小时负荷预报误差的概率密度函数进行了估计、拟合和检验,现将结果列于表 2.

表 2 小时负荷预报误差的密度函数建模结果

小时	按(2)式 计算 $\sigma_j\%$	按(23)式 建模 $\sigma_j\%$	按(23)式 建模 $a_j\%$	密度函数	小时	按(2)式 计算 $\sigma\%$	按(23)式 建模 $\sigma\%$	按(23)式 建模 $a_j\%$	密函数度
1	3.1	2.91	0.41	$N(0.41, 2.91^2)$	13	3.2	3.11	-0.21	$N(-0.21, 3.11^2)$
2	3.5	3.34	0.08	$N(0.08, 3.34^2)$	14	3.9	3.64	-0.18	$N(-0.18, 3.64^2)$
3	5.1	4.50	0.47	$N(0.42, 4.5^2)$	15	3.5	3.27	-0.10	$N(-0.1, 3.27^2)$
4	3.6	3.41	-0.19	$N(-0.19, 3.41^2)$	16	4.0	3.74	0.14	$N(0.14, 3.74^2)$
5	4.9	4.44	0.14	$N(0.14, 4.44^2)$	17	5.2	4.64	0.28	$N(0.28, 4.64^2)$
6	4.5	4.12	0.21	$N(0.21, 4.12^2)$	18	3.9	3.71	-0.15	$N(-0.15, 3.71^2)$
7	4.9	4.56	-0.36	$N(-0.36, 4.56^2)$	19	4.7	4.58	-0.39	$N(-0.39, 4.58^2)$
8	4.2	4.03	-0.42	$N(-0.42, 4.03^2)$	20	5.3	4.83	-0.18	$N(-0.18, 4.83^2)$
9	4.3	3.98	-0.34	$N(-0.34, 3.98^2)$	21	5.2	4.79	-0.06	$N(-0.06, 4.79^2)$
10	4.0	3.82	0.16	$N(0.16, 3.82^2)$	22	4.9	4.62	-0.10	$N(-0.10, 4.62^2)$
11	2.7	2.86	0.10	$N(0.1, 2.86^2)$	23	4.7	4.21	0.29	$N(0.29, 4.21^2)$
12	2.4	2.38	0.07	$N(0.07, 2.38^2)$	24	3.7	3.52	0.32	$N(0.32, 3.52^2)$

由以上分析计算可得如下结论:

1) 按本文(20)式对电力系统负荷预报误差的概率密度函数进行估计,然后正态拟合,最后进行统计检验,从而得到预报误差的概率密度函数建模,这种方法是可行的.

2) 电力系统小时负荷预报误差服从正态分布,这一结论从统计意义上讲是可以接受的.

3) 按上述方法建模所得的预报均方误差  $\sigma_j\%$  ( $j = 1, 2, \dots, 24$ ) 是有明确概率意义的。

4) 通常按(2)式计算所得的预报均方误差, 虽略偏大, 但与建模均方误差比较接近, 因此仍有实用参考意义。

## 四、讨 论

在本文的方法中, 值得讨论的是, 是否存在最优核函数以及如何选取核函数。所谓最优核函数是指密度估计  $\hat{f}_n(x)$  的均方差  $E[\hat{f}_n(x) - f(x)]^2$  达到最小, 事实上由

$$\begin{aligned} E[\hat{f}_n(x) - f(x)]^2 &= \frac{1}{n} E \left[ \frac{1}{h_n} K \left( \frac{x - x_i}{h_n} \right) \right]^2 - \frac{1}{n} \left[ E \frac{1}{h_n} K \left( \frac{x - x_i}{h_n} \right) \right]^2 \\ &= \frac{1}{nh_n} f(x) \int_{-\infty}^{+\infty} K^2(z) dz - \frac{1}{n} f^2(x) \end{aligned} \quad (27)$$

可知, 只须使  $L = \int_{-\infty}^{+\infty} K^2(z) dz$  达到最小即可。于是, 在定义 1 所述的条件下, 利用变分法可求出最优核函数  $K_0(x)$  为<sup>[4]</sup>

$$K_0(x) = \begin{cases} \frac{3}{4\sqrt{5}} \left( 1 - \frac{x^2}{5} \right), & |x| \leq \sqrt{5}, \\ 0, & |x| > \sqrt{5}. \end{cases} \quad (28)$$

在实际应用中, 并不采用最优核  $K_0(x)$ , 而是采用正态核函数, 这是因为利用计算机进行计算方便, 其性能也接近最优核, 伊凡乃钦尼可夫给出了如下计算结果<sup>[7]</sup>:

$$R \triangleq \frac{\int_{-\infty}^{+\infty} K^2(z) dz}{\int_{-\infty}^{+\infty} K_0^2(z) dz} = 1.05.$$

对于满足定义 1 的其他核函数, 其  $R$  值仍在 1.01~1.10 之间<sup>[7]</sup>。这样一来, 利用核函数对密度函数估计时, 可不必取最优核  $K_0(x)$ , 只须选取某一较方便的核函数即可。

## 参 考 文 献

- [1] 刘晨晖, 电力系统负荷预报理论与方法, 哈尔滨工业大学出版社, 1987.
- [2] Hagan, M.T., Behr, S. M., The Time Series Approach to Short Term Load Forecasting, *IEEE Trans, On Power Systems*, PWRS-2 (1987), (3).
- [3] 赵希人、王晓陵、郑焱、马宏颖, 电力系统负荷的分解建模及预报方法, *自动化学报*, 17(1991), (6), 713—720.
- [4] B. L. S. Prakasa Rao, *Nonparametric functional Estimation*, 1983.
- [5] Rao. C. R. *Linear Statistical Inference and Its Applications*, 1973.
- [6] Silverman, B.W., Density estimation: Are theoretical results useful in practice. In "Asymptotic Theory of Statistical Tests and Estimation", Academic Press, New York, 179—203.
- [7] Epanechnikov, V. A., Nonparametric estimation of a multi—dimensional Probability density, *Theory Probab. Appl.* 14, 153—158.
- [8] 复旦大学编, 概率论, 第二册数理统计, 人民教育出版社, 1982.

## PROBABILITY DENSITY FUNCTION MODELING FOR POWER SYSTEM LOAD FORECASTING ERRORS

ZHAO XIREN LI DAWEI LI GUOBIN HAO CHANGQUAN

(Dept. of Automatic Control, Harbin Shipbuilding Engineering Institute, 150001)

### ABSTRACT

According to the principles of statistics, a modeling method of probability density functions for power system load forecasting errors is presented in this paper. The modeling formulae are given and the asymptotic unbiased and the mean square convergence are derived. This method has been successfully applied to the analysis of HLJ power system load forecasting errors.

**Key words:** Probability density function; modeling; forecasting errors; power system load; test.

**赵希人** 现任哈尔滨船舶工程学院教授。其他情况见本刊第17卷第6期。



**李大为** 1956年生于哈尔滨。1986年在哈尔滨船舶工程学院获硕士学位。现任自控系副主任，主要从事现代控制理论及应用方面的科研、教学工作。研究方向是船舶自适应控制；随机控制理论；系统建模及参数辨识。发表论文多篇。



**李国斌** 1937年生于哈尔滨，1960年毕业于哈尔滨军事工程学院。现任哈尔滨船舶工程学院自控系主任，教授。长期从事现代控制理论及应用的科研、教学工作。研究方向是船舶姿态估计与控制，系统建模及参数估计，曾多次获国家级、部级成果奖并出版专著。

**郝长泉** 1960年生于吉林市，1990年在哈尔滨船舶工程学院获硕士学位，主要从事系统建模与参数估计、随机信号处理等方面科研工作。