AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 9(5): 10478–10493.
DOI:10.3934/math.2024512
Received: 19 January 2024
Revised: 27 February 2024
Accepted: 01 March 2024
Published: 18 March 2024

Research article

Uncertain logistic regression models

Jinling Gao[1] **and Zengtai Gong**[2,*]

[1] School of Information Engineering and Artificial Intelligence, Lanzhou University of Finance and Economics, Lanzhou 730010, China.

[2] College of Mathematics and Statistics, Northwest Normal University, Lanzhou 730070, China

* **Correspondence:** Email: zt-gong@163.com, gongzt@nwnu.edu.cn; Tel: +8613993196400.

Abstract: Logistic regression is a generalized nonlinear regression analysis model and is often used for data mining, automatic disease diagnosis, economic prediction, and other fields. In this paper, we first aimed to introduce the concept of uncertain logistic regression based on the uncertainty theory, as well as investigating the likelihood function in the sense of uncertain measure to represent the likelihood of unknown parameters.

Keywords: uncertain logistic regression; maximum likelihood estimation; uncertainty theory
Mathematics Subject Classification: 03E72, 08A72, 26E50

1. Introduction

Since the interval-valued dates with the human uncertain were frequently used in real life, the concept of uncertainty theory was initiated by Liu [5]. In order to rationally deal with belief degrees of an uncertain event, he gave the exact definition of uncertain measure, which is a set function satisfying normality, monotonicity, self-duality, countable subadditivity, and product axioms. In recent years, many authors began to study and develop the uncertain variables, such as Liu [9, 10], Wang et al. [13], Wen et al. [14], Zhao [20], and Das et al. [1]. Ye and Liu [19] employed uncertain hypothesis tests to determine whether an uncertain differential equation is suitable for an observed date. With the development of uncertainty theory, uncertain variable had been applied to many fields, including statistics by Yao [16], Bayesain statistics by Ding and Zhang [2], and regression analysis by Ding and Zhang [3]. Apparently, uncertain regression analysis is a significant mathematic tool to estimate the relationship between explanatory variables and response variables when the imprecise observations are regarded as uncertain variables. The regression models include linear regression, nonlinear regression, and nonparametric regression.

For that matter, Yao and Liu [17] investigated the principle of least squares, which estimate the unknown parameters of uncertain linear regression models. Song and Fu [12] discussed the uncertain linear regression model with multiple response variables. Next, Lio and Liu [8] proposed the method of maximum likelihood estimation to estimate the parameter of uncertain linear regression models. The least squares estimation was used for an uncertain moving average model by Yang and Ni [18]. For the nonlinear regression, Fang et al. [4] proposed the Johnson-Schumacher growth model by the least squares method.

The logistic regression is the standard approach, which analyzes the binary and categorical date applied to many areas, such as epidemiologic and biomedical studies. The traditional logistic regression analysis estimates the relationships between the explanatory variables and response variables on the basis of the assumption that the samples of these variables are precisely observed. However, since the observation of the samples are imprecise, we need to estimate the logistic regression model, which is based on the relationships between the variables with imprecisely observed samples. To sum up, we propose the uncertain logistic regression. Compared with the uncertain linear regression models, the uncertain logistic regression model can better fit the classification problems.

We study the uncertain logistic regression models. The rest of this paper is organized as follows: In Section 2, we recall some basic concepts of measure and distribution of uncertain variables. Then, we introduce the concept of uncertain logistic distribution in Section 3. The estimations of uncertain logistic regressions of binary classification and its maximum likelihood estimation are presented in Section 4. We research the uncertain cumulative logistic regression and derive the maximum likelihood in Section 5. Finally, some conclusions are given in Section 6.

2. Preliminaries

In this section, we review some basic concepts and theorems of uncertainty theory, as they will be used throughout paper.

Definition 2.1. *[5] Let $\mathcal{L}$ be a σ-algebra on a nonempty set Γ. A set function $\mathcal{M} : \mathcal{L} \to [0, 1]$ is called an uncertain measure if it satisfies the following axioms:*

Axiom 1: (Normality Axiom) $\mathcal{M}\{\Gamma\} = 1$ for the universal sat Γ.
Axiom 2: (Duality Axiom) $\mathcal{M}\{\Lambda\} + \mathcal{M}\{\Lambda^c\} = 1$ for any event Λ.
Axiom 3: (Subadditivity Axiom) For every countable sequence of events $\Lambda_1, \Lambda_2, ...$, we have

$$\mathcal{M}\{\bigcup_{i=1}^{\infty} \Lambda_i\} \leq \sum_{i=1}^{\infty} \mathcal{M}\{\Lambda_i\}.$$

Axitom 4: (Product Axiom) Let $(\Gamma_k, \mathcal{L}_k, \mathcal{M}_k)$ be uncertain space for $k = 1, 2, ...$ The product uncertain measure $\mathcal{M}$ is an uncertain measure satisfying

$$\mathcal{M}\{\prod_{k=1}^{\infty}\} = \bigwedge_{k=1}^{\infty} \mathcal{M}\{\Lambda_k\}.$$

Definition 2.2. *[5] An uncertain variable ξ is a measurable function from an uncertainty space $(\Gamma, \mathcal{L}, \mathcal{M})$ to the set of real numbers, i.e., for any Borel set B of real numbers, the set*

$$\{\xi \in B\} = \{\gamma \in \Gamma | \xi(\gamma) \in B\}$$

in an event.

Definition 2.3. *[5] The uncertainty distribution of an uncertain variable ξ is defined by*

$$\Phi(x) = \mathcal{M}\{\xi \leq x\}.$$

for any $x \in \mathbb{R}$.

Definition 2.4. *[7] An uncertainty distribution $\Phi(x)$ is said to be regular if it is a continuous and strictly increasing function with respect to x at which $0 < \Phi(x) < 1$, and*

$$\lim_{x \to -\infty} \Phi(x) = 0, \ \lim_{x \to +\infty} \Phi(x) = 1.$$

Definition 2.5. *[7] Let ξ be an uncertain variable with regular uncertainty distribution $\Phi(x)$. Then the inverse function $\Phi^{-1}(\alpha)$ is called the inverse uncertainty distribution of ξ.*

Theorem 2.1. *(Sufficient and Necessary Condition) [7] A function $\Phi^{-1}(\alpha) : (0, 1) \to \Re$ is an inverse uncertainty distribution if and only if it is a continuous and strictly increasing function with respect to α.*

Definition 2.6. *[7] The events $\Lambda_1, \Lambda_2, ...\Lambda_n$ are said to be independent if*

$$\mathcal{M}\{\bigcap_{i=1}^{n} \Lambda_i^*\} = \bigwedge_{i=1}^{n} \mathcal{M}\{\Lambda_i^*\}, \tag{2.1}$$

where Λ_i^* are arbitrarily chosen from $\{\Lambda_i, \Lambda_i^c, \Gamma\}$, $i = 1, 2, ...n$, respectively and Γ is the sure event.

Theorem 2.2. *[7] Let $\xi_1, \xi_2, ..., \xi_n$ be independent uncertain variables with regular uncertainty distributions $\Phi_1, \Phi_2, ..., \Phi_n$, respectively. If $f(\xi_1, \xi_2, ..., \xi_n)$ is a continuous and strictly increasing function, then ξ has an uncertainty distribution*

$$\Psi(x) = \sup_{f(x_1, x_2, ..., x_n) = x} \min_{1 \leq i \leq n} \Phi_i(x_i). \tag{2.2}$$

Definition 2.7. *[5] Let ξ be an uncertain variable. Then the expected value of ξ is defined by*

$$E[\xi] = \int_0^{+\infty} \mathcal{M}\{\xi \geq x\} dx - \int_{-\infty}^0 \mathcal{M}\{\xi \leq x\} dx, \tag{2.3}$$

provided that at least one of the two integrals is finite.

Theorem 2.3. *[7] Let ξ be an uncertain variable with uncertainty distribution Φ. Then,*

$$E[\xi] = \int_{-\infty}^{+\infty} x d\Phi(x). \tag{2.4}$$

Definition 2.8. *[5] Let ξ be an uncertain variable with finite expected value e. Then, the variance of ξ is*

$$V[\xi] = E[(\xi - e)^2]. \tag{2.5}$$

3. Uncertain logistic distribution

In this section, we introduce the uncertain logistic variable and its distribution.

Definition 3.1. *An uncertain variable ξ on $(\Gamma, \mathcal{L}, \mathcal{M})$ is called obeying logistic distribution if its distribution function takes the form*

$$\Lambda(x) = (1 + \mathrm{e}^{\frac{-(x-\mu)}{\sigma}})^{-1}, x \in \mathbb{R}. \tag{3.1}$$

Definition 3.2. *An uncertain logistic distribution is standard if its uncertain distribution takes the form*

$$\Lambda(x) = (1 + \mathrm{e}^{-x})^{-1}, x \in \mathbb{R}. \tag{3.2}$$

Theorem 3.1. *[7] Let ξ be an uncertain variable with uncertainty distribution Φ. Then,*

$$E[\xi] = \int_0^1 \Phi^{-1}(\alpha) d\alpha. \tag{3.3}$$

Theorem 3.2. *[15] Let ξ be an uncertain variable with uncertainty distribution Φ and finite expected value m. Then,*

$$V[\xi] = \int_0^1 (\Phi^{-1}(\alpha) - m)^2 d\alpha. \tag{3.4}$$

Theorem 3.3. *Let ξ be a uncertain logistic variable on the uncertain space $(\Gamma, \mathcal{L}, \mathcal{M})$. Then the expected value and variance of ξ are*

$$E[\xi] = \mu, \tag{3.5}$$

and

$$Var[\xi] = \frac{\pi^2\sigma^2}{3}. \tag{3.6}$$

Proof.

$$E[\xi] = \int_0^1 \sigma[\ln(\alpha) - \ln(1-\alpha)] + \mu d\alpha.$$

Note

$$\begin{aligned} &\int_0^1 \ln(\alpha) - \ln(1-\alpha) d\alpha \\ &= \int_0^1 \ln(\alpha) d\alpha - (-1)\int_0^1 \ln(1-\alpha) d\alpha \\ &= \int_0^1 \ln(\alpha) d\alpha - \int_0^1 \ln(\beta) d\beta = 0. \end{aligned}$$

Therefore, the expected value of ξ is μ.

$$Var[\xi] = \int_0^1 (\Phi^{-1}(\alpha) - m)^2 d\alpha$$

$$
\begin{aligned}
&= \int_0^1 (\sigma[\ln(\alpha) - \ln(1-\alpha)])^2 d\alpha \\
&= \sigma^2[\int_0^1 \ln^2(\alpha)d\alpha - 2\int_0^1 \ln(\alpha)ln(1-\alpha)d\alpha + \int_0^1 \ln^2(1-\alpha)d\alpha] \\
&= 2 + (-1)\int_0^1 \ln^2(\alpha)d\alpha - 2\int_0^1 \ln(\alpha)\ln(1-\alpha)d\alpha + \int_0^1 \ln^2(1-\alpha)d\alpha \\
&= \sigma^2[4 - 2\times(2-\frac{\pi^2}{6})] = \frac{\pi^2\sigma^2}{3}.
\end{aligned}
$$

Theorem 3.4. *[11] Let ξ be an uncertain variable with regular uncertainty distribution Φ, and let k be a positive integer. Then the k-th moment of ξ is*

$$E[\xi^k] = \int_0^1 (\Phi^{-1}(\alpha))^k d\alpha. \tag{3.7}$$

Theorem 3.5. *Let ξ be the standard uncertain logistic distribution. Then k-th moment of ξ is*

$$E[\xi^k] = \begin{cases} 0, & k = 2m+1, \\ \sum_{i=0}^{2m}\binom{2m}{i}(-1)^{2m-i}\int_0^1 (\ln\alpha)^i(\ln(1-\alpha))^{2m-i}d\alpha, & k = 2m. \end{cases} \tag{3.8}$$

Proof. According to (3.7), we have

$$
\begin{aligned}
E[\xi^k] &= \int_0^1 [\ln(\alpha) - \ln(1-\alpha)]^k d\alpha \\
&= \int_0^1 \sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}(\ln\alpha)^i(\ln(1-\alpha))^{k-i}d\alpha \\
&= \sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}\int_0^1 (\ln\alpha)^i(\ln(1-\alpha))^{k-i}d\alpha.
\end{aligned}
$$

We notice when $k = 2m+1$, then by changing variable, $\beta = 1-\alpha$

$$
\begin{aligned}
&\sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}\int_0^1 (\ln\alpha)^i(\ln(1-\alpha))^{k-i}d\alpha \\
&= \sum_{i=0}^{2m+1}\binom{2m+1}{i}(-1)^{(2m+1)-i}\int_0^1 (\ln\alpha)^i(\ln(1-\alpha))^{(2m+1)-i}d\alpha \\
&= \sum_{i=0}^{2m+1}\binom{2m+1}{(2m+1)-i}(-1)(-1)^i\int_0^1 (\ln\beta)^{(2m+1)-i}(\ln(1-\beta))^i d\beta \\
&= -\int_0^1 [\ln(\beta) - \ln(1-\beta)]^k d\beta.
\end{aligned}
$$

Thus,

$$E[\xi^{2m+1}] = 0.$$

Theorem 3.6. *[20] Let ξ_1 and ξ_2 be two regular uncertain variables with distributions Ξ_1 and Ξ_2 and finite expected values $E[\xi_1]$ and $E[\xi_2]$. Then the covariance of Ξ_1 and Ξ_2 is*

$$Cov[\xi_1, \xi_2] = \int_0^1 \{\Xi_1^{-1}(\alpha) - E[\xi_1]\}\{\Xi_2^{-1}(\alpha) - E[\xi_2]\}d\alpha. \tag{3.9}$$

Theorem 3.7. *Let ξ_1 and ξ_2 be uncertain logistic variable with the uncertain distribution*

$$\Lambda(x) = (1 + \mathrm{e}^{\frac{-(x-\mu_1)}{\sigma_1}})^{-1}, x \in \mathbb{R},$$

and

$$\Lambda(y) = (1 + \mathrm{e}^{\frac{-(y-\mu_2)}{\sigma_2}})^{-1}, y \in \mathbb{R}.$$

Then, the covariance of uncertain variables is

$$Cov[\xi_1, \xi_2] = \frac{\pi^2\sigma_1\sigma_2}{3}. \tag{3.10}$$

Proof. Since the expected values of ξ_1 and ξ_2 are μ_1 and μ_2, then,

$$\begin{aligned} Cov[\xi_1, \xi_2] &= \int_0^1 \{\sigma_1[\ln(\alpha) - \ln(1-\alpha)]\}\{\sigma_2[\ln(\alpha) - \ln(1-\alpha)]\}d\alpha \\ &= \sigma_1\sigma_2 \int_0^1 [ln(\alpha) - ln(1-\alpha)]^2 d\alpha \\ &= \frac{\pi^2\sigma_1\sigma_2}{3}. \end{aligned}$$

Definition 3.3. *[6] Suppose that ξ is an uncertain set with membership function Ξ. Then its entropy is defined by*

$$H[\xi] = \int_{-\infty}^{+\infty} S(\Xi(x))dx, \tag{3.11}$$

where $S(t) = -t\ln t - (1-t)\ln(1-t)$.

Theorem 3.8. *Let ξ be a uncertain logistic variable with the uncertain distribution*

$$\Lambda(x) = (1 + \mathrm{e}^{\frac{-(x-\mu)}{\sigma}})^{-1}, x \in \mathbb{R}. \tag{3.12}$$

Then, the entropy is expressed by σ.

Proof. Note that

$$\begin{aligned} H[\xi] &= \int_0^1 [-t\ln t - (1-t)\ln(1-t)]d(\sigma(\ln(\alpha) - \ln(1-\alpha)) + \mu) \\ &= -\sigma \int_0^1 -\frac{\ln t}{1-t} - \frac{\ln(1-t)}{t}dt \\ &= \sigma. \end{aligned}$$

4. Uncertain binary logistic regression

4.1. The definition of model

Let $x = (x_1, x_2, ..., x_p)$ be a vector of crisp explanatory variables, and $z = (z_1, z_2, ...z_p)$ be a uncertain response variable. We set a threshold value C (such as $(C = 0)$). The event occurs when $z_i > C$. Then, we have

$$y_i = \begin{cases} 1, & z_i > 0, \\ 0, & z_i \le 0. \end{cases} \tag{4.1}$$

Let z_i and x_i be linear relationship, i.e.,

$$z_i = a + bx_i + \varepsilon_i. \tag{4.2}$$

Then, the uncertain logistic regression model of binary classification is

$$\mathcal{M}\{y_i = 1\} = \frac{1}{1 + \exp -(a + \sum_{j=1}^{k} b_j x_{ij})}, \tag{4.3}$$

where a, b are vectors of unknown parameters, and ε_i is a disturbance term. Next let ε_i obey the standard uncertain logistic distribution, then,

$$\mathcal{M}\{y_i = 1\} = \mathcal{M}\{z_i > 0\}$$

$$= \mathcal{M}\{a + bx_i + \varepsilon_i > 0\}$$

$$= \mathcal{M}\{\varepsilon_i > -(a + bx_i)\}.$$

Due to the symmetry of standard uncertain logistic distribution, the previous formula can be written

$$\mathcal{M}\{\varepsilon_i \le (a + bx_i)\} = (1 + \mathrm{e}^{-(a+bx_i)})^{-1}. \tag{4.4}$$

The (4.3) is defined as uncertain logistic regression model. According to (4.3), we have

$$\frac{\mathcal{M}\{z_i > 0\}}{\mathcal{M}\{z_i \le 0\}} = \mathrm{e}^{a+bx_i}. \tag{4.5}$$

Then, we take the logarithm of (4.5), and obtain

$$\ln(\frac{\mathcal{M}\{z_i > 0\}}{\mathcal{M}\{z_i \le 0\}}) = a + bx_i. \tag{4.6}$$

If there are k explanatory variables, then,

$$\ln(\frac{\mathcal{M}\{z_i > 0\}}{\mathcal{M}\{z_i \le 0\}}) = a + \sum_{j=1}^{k} b_j x_{ij}. \tag{4.7}$$

4.2. Likelihood function

Let $z_1, z_2, ..., z_n$, which are corresponding to $y_1, y_2, ..., y_n$ with uncertainty distribution F be iid samples, and let $z_1, z_2, ..., z_n$ have observing values $c_1, c_2, ..., c_n$, respectively. Then, the likelihood of taking value at β can be represented by

$$\lim_{\Delta z \to 0} \mathcal{M}\{\bigcap_{i=1}^{n}(c_i - \frac{\Delta z}{2} < z_i \le c_i + \frac{\Delta z}{2})|\beta\}$$

$$= \lim_{\Delta z \to 0} \bigwedge_{i=1}^{n} \mathcal{M}\{(c_i - \frac{\Delta z}{2} < z_i \le c_i + \frac{\Delta z}{2})|\beta\},$$

where β is a set of unknown parameters, and Δz is a small number.

Theorem 4.1. *Suppose $z_1, z_2, ..., z_n$ which are corresponding to $y_1, y_2, ..., y_n$ with uncertainty distribution F are iid samples and β is a set of unknown parameters. Given that $c_1, c_2, ..., c_n$ are observed, then we have*

$$L(\beta|c_1, c_2, ..., c_n) = \bigwedge_{i=1}^{n} [\exp(a + \sum_{j=1}^{k} b_j x_{ij})]^{y_i} \frac{1}{1 + \exp(a + \sum_{j=1}^{k} b_j x_{ij})}. \tag{4.8}$$

Proof. The likelihood function that we obtain is

$$\begin{aligned} &L(\beta|c_1, c_2, ..., c_n) \\ &= \lim_{\Delta z \to 0} \bigwedge_{i=1}^{n} \mathcal{M}\{(c_i - \frac{\Delta z}{2} < z_i \le c_i + \frac{\Delta z}{2})|\beta\} \\ &= \bigwedge_{i=1}^{n} \mathcal{M}\{z_i > 0\}^{y_i} \mathcal{M}\{z_i \le 0\}^{1-y_i} \\ &= \bigwedge_{i=1}^{n} [\frac{1}{1 + \exp -(a + \sum_{j=1}^{k} b_j x_{ij})}]^{y_i} [\frac{\exp -(a + \sum_{j=1}^{k} b_j x_{ij})}{1 + \exp -(a + \sum_{j=1}^{k} b_j x_{ij})}]^{1-y_i} \\ &= \bigwedge_{i=1}^{n} [\exp(a + \sum_{j=1}^{k} b_j x_{ij})]^{y_i} \frac{1}{1 + \exp(a + \sum_{j=1}^{k} b_j x_{ij})}. \end{aligned}$$

The theorem is proved.

Next, we obtain the parameter value through letting the likelihood function reach maximum point. Since the likelihood function is

$$L(\beta|c_1, c_2, ..., c_n) = \bigwedge_{i=1}^{n} [\exp(a + \sum_{j=1}^{k} b_j x_{ij})]^{y_i} \frac{1}{1 + \exp(a + \sum_{j=1}^{k} b_j x_{ij})}.$$

Then, the maximum likelihood estimator $\widehat{\beta}$ solves the maximization problem

$$\max L(\beta|c_1, c_2, ..., c_n).$$

4.3. Binary classification example

Example 4.1. *Let* $(x_{i1}, x_{i2}, x_{i3}, z_i), i = 1, 2, ..., 24$ *be a set of observed values (see Table 1), and the uncertain binary logistic regression model is*

$$z = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \epsilon,$$

where ϵ is a standard uncertain logistic distribution. Then we have

$$\ln \frac{\mathcal{M}\{y = 1\}}{1 - \mathcal{M}\{y = 1\}} = a + \sum_{j=1}^{3} b_j x_j,$$

and

$$\mathcal{M}\{y = 1\} = \frac{1}{1 + \exp -(a + \sum_{j=1}^{3} b_j x_j)},$$

where $y = 0$ when $z \leq 30$ and $y = 1$ when $z > 30$ according to personal experimental. Then, we solve the maximum problem

$$\max \bigwedge_{i=1}^{24} [\exp(a + \sum_{j=1}^{3} b_j x_{ij})]^{y_i} \frac{1}{1 + \exp(a + \sum_{j=1}^{3} b_j x_{ij})}$$

for obtaining $\widehat{\beta}$. Hence, the uncertain logistic regression model is

$$\mathcal{M}\{y = 1\} = \frac{1}{1 + \exp -(85.799 + 8.456x_1 - 3.776x_2 + 2.525x_3)}.$$

Table 1. Observed values of explanatory variables and response variable.

i	1	2	3	4	5	6	7	8
x_{i1}	11	13	15	21	22	7	8	19
x_{i2}	27	27	28	11	17	22	27	15
x_{i3}	10	21	19	34	32	13	26	31
z_i	23	30	29	37	31	27	29	31
i	9	10	11	12	13	14	15	16
x_{i1}	4	22	5	6	4	6	6	13
x_{i2}	21	9	16	23	22	7	8	12
x_{i3}	6	48	26	12	10	21	19	34
z_i	20	43	28	22	23	30	29	37
i	17	18	19	20	21	22	23	24
x_{i1}	12	17	12	4	7	8	5	6
x_{i2}	16	13	11	25	26	18	17	24
x_{i3}	24	36	38	20	8	41	36	24
z_i	33	34	32	24	24	38	28	26

5. Uncertain cumulative logistic regression

5.1. Uncertain cumulative logistic regression

Next we generalized uncertain logistic regression model of binary classification. Let $j = 1, 2, ..., J$ be order variables, then the uncertain cumulative logistic regression can be represented by

$$z_i = a + \sum_{k=1}^{n} b_k x_{ik} + \varepsilon_i,$$

where ε_i is a disturbance term with standard uncertain logistic distribution, Suppose y_i have J kinds of situation, then,

$$y_i = 1, z_i \le \eta_1,$$

$$y_i = 2, \eta_1 < z_i \le \eta_2,$$

$$......$$

$$y_i = J, z_i > \eta_{J-1},$$

where η_j is the boundary point, and $\eta_1 < \eta_2 < ... < \eta_{J-1}$. Then,

$$\begin{aligned}
\mathcal{M}\{y \le j\} &= \mathcal{M}\{z \le \eta_j\} \\
&= F(\eta_j) \\
&= \mathcal{M}\{a + \sum_{k=1}^{K} b_k x_k + \varepsilon \le \eta_j\} \\
&= \mathcal{M}\{\varepsilon \le \eta_j - (a + \sum_{k=1}^{K} b_k x_k)\} \\
&= \Lambda(\eta_j - (a + \sum_{k=1}^{K} b_k x_k)) \\
&= \frac{1}{1 + \exp[-(\eta_j - (a + \sum_{k=1}^{K} b_k x_k))]} \\
&= \frac{\exp[\eta_j - (a + \sum_{k=1}^{K} b_k x_k)]}{1 + \exp[\eta_j - (a + \sum_{k=1}^{K} b_k x_k)]},
\end{aligned}$$

where F is the uncertain distribution of z. Let $\alpha_j = \eta_j - a$,

$$\ln \frac{\mathcal{M}\{y \le j\}}{1 - \mathcal{M}\{y \le j\}} = \alpha_j + \sum_{k=1}^{K} b_k x_k.$$

5.2. Likelihood function of uncertain cumulative logistic regression

Let $z_1, z_2, ..., z_n$ which are corresponding to $y_1, y_2, ..., y_n$ with uncertainty distribution F be iid samples, and β be a set of unknown parameters. Furthermore, let $z_1, z_2, ..., z_n$ have observing values

$c_1, c_2, ..., c_n$, respectively. Then, the likelihood of taking value at β can be represented by

$$\mathcal{M}\{\bigcap_{i=1}^{n}(z_i = c_i)|\beta\}.$$

From the independent of $y_1, y_2, ..., y_n$, we get

$$= \bigwedge_{i=1}^{n} \mathcal{M}\{(z_i = c_i)|\beta\}$$

$$= \bigwedge_{i=1}^{n} \prod_{j=1}^{J} \mathcal{M}\{(y_i = j)\}^{y_{ij}},$$

where

$$y_{ij} = \begin{cases} 1, & \eta_{j-1} < z_i \leq \eta_j, \\ 0, & others. \end{cases}$$

On the other hand, from the subadditivity of uncertain measure, for each i, we obtain

$$\begin{aligned} &\mathcal{M}\{j - \frac{\Delta}{2} < y_i \leq j + \frac{\Delta}{2}\} \\ &\geq \mathcal{M}\{y_i \leq j + \frac{\Delta}{2}\} - \mathcal{M}\{y_i \leq j - \frac{\Delta}{2}\} \\ &= F(j + \frac{\Delta}{2}) - F(j - \frac{\Delta}{2}). \end{aligned}$$

Hence, we have

$$\begin{aligned} &\bigwedge_{i=1}^{n} \prod_{j=1}^{J} [\lim_{\Delta \to 0} \frac{\mathcal{M}\{j - \frac{\Delta}{2} < y_i \leq j + \frac{\Delta}{2}\}}{\Delta}]^{y_{ij}} \\ &\geq \bigwedge_{i=1}^{n} \prod_{j=1}^{J} [\lim_{\Delta \to 0} \frac{\mathcal{M}\{y_i \leq j + \frac{\Delta}{2}\} - \mathcal{M}\{y_i \leq j - \frac{\Delta}{2}\}}{\Delta}]^{y_{ij}} \\ &= \bigwedge_{i=1}^{n} \prod_{j=1}^{J} [\lim_{\Delta \to 0} \frac{F(\eta_j + \frac{\Delta}{2}) - F(\eta_j - \frac{\Delta}{2})}{\Delta}]^{y_{ij}}. \end{aligned}$$

Next, we define a likelihood function of uncertain cumulative logistic regression with respect to a set of unknown parameters β as follows.

Definition 5.1. *Let $z_1, z_2, ..., z_n$ which are corresponding to $y_1, y_2, ..., y_n$ with uncertainty distribution F be iid samples and β be a set of unknown parameters. Suppose $z_1, z_2, ..., z_n$ have observing values $c_1, c_2, ..., c_n$, respectively. Then the likelihood function of taking value at β can be represented by*

$$L(\beta) = \bigwedge_{i=1}^{n} \prod_{j=1}^{J} [\lim_{\Delta \to 0} \frac{F(\eta_j + \frac{\Delta}{2}) - F(\eta_j - \frac{\Delta}{2})}{\Delta}]^{y_{ij}}. \tag{5.1}$$

Theorem 5.1. *Suppose $z_1, z_2, ..., z_n$ which are corresponding to $y_1, y_2, ..., y_n$ with uncertainty distribution F are iid samples and β is a set of unknown parameters. Let $z_1, z_2, ..., z_n$ have observing values $c_1, c_2, ..., c_n$, respectively. If F is differentiable at $c_1, c_2, ..., c_n$. Then the likelihood function of taking value at β can be represented by*

$$L(\beta) = \bigwedge_{i=1}^{n} \prod_{j=1}^{J} \{\frac{\exp[\eta_j - (a + \sum_{k=1}^{K} b_k x_k)]}{\{1 + \exp[\eta_j - (a + \sum_{k=1}^{K} b_k x_k)]\}^2}\}^{y_{ij}}. \tag{5.2}$$

Proof. Due to $F(c|\beta)$ is differentiable at $c_1, c_2, ..., c_n$, we have

$$\begin{aligned} L(\beta) &= \bigwedge_{i=1}^{n} \prod_{j=1}^{J} [\lim_{\Delta \to 0} \frac{F(\eta_j + \frac{\Delta}{2}) - F(\eta_j - \frac{\Delta}{2})}{\Delta}]^{y_{ij}} \\ &= \bigwedge_{i=1}^{n} \prod_{j=1}^{J} F'(\eta_j)^{y_{ij}} \\ &= \bigwedge_{i=1}^{n} \prod_{j=1}^{J} \{\frac{\exp[\eta_j - (a + \sum_{k=1}^{K} b_k x_k)]}{\{1 + \exp[\eta_j - (a + \sum_{k=1}^{K} b_k x_k)]\}^2}\}^{y_{ij}}. \end{aligned}$$

Next, we obtain the parameter value through letting the likelihood function reach maximum point. Since the likelihood function is

$$L(\beta) = \bigwedge_{i=1}^{n} \prod_{j=1}^{J} \{\frac{\exp[\eta_j - (a + \sum_{k=1}^{K} b_k x_k)]}{\{1 + \exp[\eta_j - (a + \sum_{k=1}^{K} b_k x_k)]\}^2}\}^{y_{ij}},$$

so, we have

$$\ln L(\beta) = \bigwedge_{i=1}^{n} \sum_{j=1}^{J} y_{ij}\{\eta_j - (a + \sum_{k=1}^{K} b_k x_k) - 2\ln\{1 + \exp[\eta_j - (a + \sum_{k=1}^{K} b_k x_k)]\}\}.$$

Then, the maximum likelihood estimator $\widehat{\beta}$ solves the maximization problem

$$\max \ln L(\beta|c_1, c_2, ..., c_n).$$

5.3. Cumulative logistic regression example

Example 5.1. *Let $(x_{i1}, x_{i2}, x_{i3}, z_i), i = 1, 2, ..., 24$ be a set of observed values (see Table 2), and the uncertain multivariate logistic regression model is*

$$z = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \epsilon,$$

where ϵ is a standard uncertain logistic distribution. Then, we have

$$\ln \frac{\mathcal{M}\{y = 1\}}{1 - \mathcal{M}\{y = 1\}} = \ln \frac{\mathcal{M}\{y = 1\}}{\mathcal{M}\{y = 2\} + \mathcal{M}\{y = 3\}} = a + \sum_{j=1}^{3} b_j x_j,$$

and

$$\mathcal{M}\{y=1\} = \frac{\exp[30-(a+\sum_{k=1}^{3} b_k x_k)]}{1+\exp[30-(a+\sum_{k=1}^{3} b_k x_k)]}$$

$$\mathcal{M}\{y=2\} = \frac{\exp[40-(a+\sum_{k=1}^{3} b_k x_k)]-\exp[30-(a+\sum_{k=1}^{3} b_k x_k)]}{\{1+\exp[40-(a+\sum_{k=1}^{3} b_k x_k)]\}\{1+\exp[30-(a+\sum_{k=1}^{3} b_k x_k)]\}}$$

$$\mathcal{M}\{y=3\} = \frac{1}{1+\exp[40-(a+\sum_{k=1}^{3} b_k x_k)]},$$

where

$$y = \begin{cases} 1, & z \le 30, \\ 2, & 30 < z \le 40, \\ 3, & z > 40, \end{cases}$$

according to personal experimental. Then, we solve the maximum problem

$$\max \bigwedge_{i=1}^{32} \sum_{j=1}^{3} y_{ij}\{\eta_j - (a + \sum_{k=1}^{3} b_k x_k) - 2\ln\{1+\exp[\eta_j - (a+\sum_{k=1}^{3} b_k x_k)]\}\}$$

for obtaining $\widehat{\beta}$.

Hence, the uncertain logistic regression model is

$$\mathcal{M}\{y=1\} = \frac{\exp[30-(9.305-20.341x_1+1.877x_2-0.639x_3)]}{1+\exp[30-(9.305-20.341x_1+1.877x_2-0.639x_3)]},$$

$$\mathcal{M}\{y=3\} = \frac{1}{1+\exp[40-(9.305-20.341x_1+1.877x_2-0.639x_3)]},$$

$$\mathcal{M}\{y=2\} = 1-\mathcal{M}\{y=1\}-\mathcal{M}\{y=3\}.$$

Table 2. Observed values of explanatory variables and response variable.

i	1	2	3	4	5	6	7	8
x_{i1}	8	4	15	16	24	3	7	25
x_{i2}	1	1	3	4	2	3	2	1
x_{i3}	2	1	2	1	1	1	2	2
z_i	15	23	34	45	43	41	47	21
i	9	10	11	12	13	14	15	16
x_{i1}	3	9	25	17	3	7	3	15
x_{i2}	3	3	4	4	3	3	4	3
x_{i3}	1	2	1	2	2	1	1	1
z_i	22	34	20	42	41	23	24	45
i	17	18	19	20	21	22	23	24
x_{i1}	17	4	4	23	25	26	18	3
x_{i2}	1	3	2	1	3	4	2	1
x_{i3}	2	2	2	1	1	1	2	2
z_i	17	36	22	37	38	31	32	25
i	25	26	27	28	29	30	31	32
x_{i1}	3	14	29	30	9	9	4	31
x_{i2}	1	1	3	3	4	3	2	4
x_{i3}	1	1	2	1	1	2	2	1
z_i	20	23	43	22	43	45	20	43

6. Conclusions

In this paper, we first recall some basic concepts of measure and distributions of uncertain variables. Then, we introduce the concept of uncertain logistic distribution. In addition, we estimate the uncertain logistic regression of binary classification and presented its maximum likelihood estimation. Finally, we generalize the binary classification to the uncertain cumulative logistic regression and derive the maximum likelihood. Through the above work, we found that uncertain logistic regression models are more suitable to deal with the classification problems in practice than uncertain linear regression models. Moreover, the output results of uncertain logistic models have probabilistic significance, and the classification results can be interpreted by probability. However, due to the limitations of uncertainty theorem, we can build only the order multi-classification model, and cannot establish the disorder multi-classification model. In the future, we plan to apply our results to the field of uncertain statistics research.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

The authors would like to thank the anonymous referees and the editor. This work is supported by the National Natural Science Foundation of China (12061607).

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. B. Das, B. C. Tripathy, P. Debnath, B. Bhattacharya, Characterization of statistical convergence of complex uncertain double sequence, *Anal. Math. Phys.*, **10** (2020), 71. https://doi.org/10.1007/s13324-020-00419-7
2. J. Ding, Z. Zhang, Bayesian statistical models with uncertainty variables, *J. Intell. Fuzzy Syst.*, **39** (2020), 1109–1117. https://doi.org/10.3233/JIFS-192014
3. J. H. Ding, Z. Q. Zhang, Statistical inference on unceryain nonparametric regression model, *Fuzzy Optim. Decis. Making*, **20** (2021), 451–469. https://doi.org/10.1007/s10700-021-09353-0
4. L. Fang, S. Liu, Z. Huang, Uncertain Johnson-Schumacher growth model with imprecise observations and k-fold cross-validation test, *Soft Computing*, **24** (2020), 2715–2720. https://doi.org/10.1007/s00500-019-04090-4
5. B. D. Liu, *Uncertainty theory*, 2 Eds., Berlin: Springer-Verlag, 2007. https://doi.org/10.1007/978-3-662-44354-5
6. B. D. Liu, Uncertain entailment and modus ponens in the framework of uncertain logic, *Journal of Uncertain Systems*, **3** (2009), 243–251.
7. B. D. Liu, *Uncertainty theory: A branch of mathematics for modeling human uncertainty*, Berlin: Springgger-Verlag, 2010. https://doi.org/10.1007/978-3-642-13959-8
8. W. C. Lio, B. D. Liu, Uncertain maximum likelihood estimation with application to uncertain regression analysis, *Soft Comput.*, **24** (2020), 9351–9360. https://doi.org/10.1007/s00500-020-04951-3
9. Z. Liu, L. F. Jia, Cross-validation for the uncertain Chapman-Richards growth model with imprecise observations, *Int. J. Uncertain. Fuzziness Knowl. Based Syst.*, **28** (2020), 769–783. https://doi.org/10.1142/s0218488520500336
10. Z. Liu, Y. Yang, Least absolute deviations eatimation for uncertain regression with imprecise observations, *Fuzzy Optim. Decis. Making*, **19** (2020), 33–52. https://doi.org/10.1007/s10700-019-09312-w
11. Y. H. Sheng, S. Kar, Some results of moments of uncertain variable through inverse uncertainty distribution, *Fuzzy Optim. Decis. Making*, **14** (2015), 57–76. https://doi.org/10.1007/s10700-014-9193-1
12. Y. L. Song, Z. F. Fu, Uncertain multivariable regression model, *Soft Comput.*, **22** (2018), 5861–5866. https://doi.org/10.1007/s00500-018-3324-5

13. X. S. Wang, Z. C. Gao, H. Y. Guo, Delphi method for estimating uncertainty distributions, *International Journaal on Information*, **15** (2012), 449–460.
14. M. L. Wen, Q. Y. Zhang, R. Kang, Y. Yang, Some new ranking criteria in data envelopment analysis under uncertain environment, *Comput. Ind. Eng.*, **110** (2017), 498–504. https://doi.org/10.1016/j.cie.2017.05.034
15. K. Yao, A formula to calculate the variance of uncertain variable, *Soft Comput.*, **19** (2015), 2947–2953. https://doi.org/10.1007/s00500-014-1457-8
16. K. Yao, Uncertain statistical inference models with imprecise observations, *IEEE T. Fuzzy Syst.*, **26** (2018), 409–415. https://doi.dog/10.1109/TFUZZ.2017.2666846
17. K. Yao, B. D. Liu, Uncertain regression analysis: An approach for imprecise observations, *Soft Comput.*, **22** (2018), 5579–5582. https://doi.org/10.1007/s00500-017-2521-y
18. X. Yang, Y. Ni, Least-squares estimation for uncertain moving average model, *Commun. Stat.-Theor. M.*, **50** (2021), 4134–4143. https://doi.org/10.1080/03610926.2020.1713373
19. T. Q. Ye, B. D. Liu, Uncertain hypothesis test for uncertain differential equations, *Fuzzy Optim. Decis. Making*, **22** (2023), 195–211. https://doi.org/10.1007/s10700-022-09389-w
20. M. X. Zhao, Y. H. Liu, D. A. Ralescu, J. Zhou, The covariance of uncertain variables: definition and calculation formulae, *Fuzzy Optim. Decis. Making*, **17** (2018), 211–232. https://doi.org/10.1007/s10700-017-9270-3

AIMS Press