Chapter 2. Loss Functions

♠ Certain Standard Loss Functions

• Squared-Error Loss

The loss function

$$L(\theta, a) = (\theta - a)^2$$

is called *squared-error loss.*

• Reasons for Considering Squared-Error Loss

(i) If $\delta(\boldsymbol{X})$ is an unbiased estimator of θ, then the risk function

$$R(\theta, \delta) = E^{\boldsymbol{X}}[L(\theta, \delta(\boldsymbol{X}))]$$
$$= E^{\boldsymbol{X}}[(\theta - \delta(\boldsymbol{X}))^2] = \mathrm{Var}(\delta(\boldsymbol{X})).$$

(ii) It has a close relationship to classical least squares theory.

(iii) The use of squared-error loss makes the calculations relatively straightforward and simple.

- **Limitation of Squared-Error Loss**

(i) It is unbounded.

(ii) Large errors are penalized much too severely.

- **Generalization of Squared-Error Loss**

A generalization of squared-error loss is the *weighted squared-error loss*, which is defined as

$$L(\theta, a) = w(\theta)(\theta - a)^2.$$

The weighted squared-error loss has the attractive feature of allowing the square error, $(\theta - a)^2$, to be weighted by a function of θ.

• **Quadratic Loss**

If $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ is a vector to be estimated by $\boldsymbol{a}(a_1, a_2, \ldots, a_p)'$, and Q is a $p \times p$ positive definite matrix, then

$$L(\boldsymbol{\theta}, \boldsymbol{a}) = \sum_{i=1}^{p} (\boldsymbol{\theta} - \boldsymbol{a})' Q (\boldsymbol{\theta} - \boldsymbol{a})$$

is called *quadratic loss*. When Q is diagonal, this reduces to

$$L(\boldsymbol{\theta}, \boldsymbol{a}) = \sum_{i=1}^{p} q_i (\theta_i - a_i)^2,$$

and is a natural extension of squared-error loss to the multivariate situation.

• **Linear Loss**

The linear loss is defined by

$$L(\theta, a) = \begin{cases} K_0(\theta - a) & \text{if } \theta - a \geq 0, \\ K_1(a - \theta) & \text{if } \theta - a < 0. \end{cases}$$

The constants K_0 and K_1 can be chosen to reflect the relative importance of underestimation and overestimation. If overestimation is a more serious problem than underestimation, we choose $K_1 > K_0$.

If $K_0 = K_1 = K$, then

$$L(\theta, a) = K|\theta - a|,$$

which is termed as the **absolute error loss**.

If K_0 and K_1, or K, are the functions of θ, the loss will be called *weighted linear loss*.

- **"0-1" Loss**

In the two-action decision problem (of which hypothesis testing is an example), it is typically the case that a_0 is "correct" if $\theta \in \Theta_0$, and a_1 is correct if $\theta_1 \in \Theta_1$. (Thus could correspond to testing H_0: $\theta \in \Theta_0$ versus H_1: $\theta \in \Theta_1$. The loss

$$L(\theta, a_i) = \begin{cases} 0 & \text{if } \theta \in \Theta_i, \\ 1 & \text{if } \theta \in \Theta_j, \end{cases} \quad (j \neq i),$$

is called "0-1" *loss*. The corresponding loss matrix is given by

	Θ_0	Θ_1
a_0	0	1
a_1	1	0

In a hypothesis testing problem, the risk function of a decision rule (or test) $\delta(\boldsymbol{x})$ is

$$\begin{aligned} R(\theta, \delta) &= E_\theta[L(\theta, \delta(\boldsymbol{X}))] \\ &= 1 \times P_\theta(\delta(\boldsymbol{X}) \text{ is the incorrect decision}). \end{aligned}$$

Case 1: $\theta \in \Theta_0$

$$\begin{aligned} R(\theta, \delta) &= P(\text{incorrect decision} | \theta \in \Theta_0) \\ &= P(\text{reject } H_0 | \theta \in \Theta_0) \\ &= P(\text{Type I error}), \end{aligned}$$

which is the probability of false rejection.

Case 2: $\theta \in \Theta_1$

$$\begin{aligned} R(\theta, \delta) &= P(\text{incorrect decision} | \theta \in \Theta_1) \\ &= P(\text{accept } H_0 | \theta \in \Theta_1) \\ &= P(\text{Type II error}), \end{aligned}$$

which is the probability of false acceptance.

Suppose $a_0 = \{\text{accept } H_0\}$ and $a_1 = \{\text{reject } H_0\}$. From the conditional perspective, the Bayes expected loss is

$$\rho(\pi^*, a_i) = \int L(\theta, a_i) dF^{\pi^*}(\theta) = P^{\pi^*}(\theta \in \Theta_j)$$
$$= 1 - P^{\pi^*}(\theta \in \Theta_i), \quad j \neq i.$$

In practice, "0-1" loss will rarely be a good approximation to the true loss. More realistic losses are

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
a_0	0	k_0
a_1	k_1	0

or

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
a_0	0	$k_0(\theta)$
a_1	$k_1(\theta)$	0

♠ Loss for Confidence Procedure

Let $C(\boldsymbol{x}) \subset \Theta$ denote a confidence set for θ. Consider the loss function

$$L(\theta, C(\boldsymbol{x})) = 1 - I_{C(\boldsymbol{x})}(\theta) = \begin{cases} 1 & \text{if } \theta \neq C(\boldsymbol{x}), \\ 0 & \text{if } \theta \in C(\boldsymbol{x}). \end{cases}$$

Then

$$R(\theta, C) = E_\theta[1 - I_{C(\boldsymbol{X})}(\theta)] = 1 - P_\theta(C(\boldsymbol{X}) \text{ contains } \theta),$$

which is one minus the frequentist coverage probability.

From the conditional perspective, the Bayes expected loss is

$$\rho(\pi^*, C(\boldsymbol{x})) = E^{\pi^*}[1 - I_{C(\boldsymbol{x})}(\theta)] = 1 - P^{\pi^*}(\theta \in C(\boldsymbol{x})),$$

which is one minus the actual (subjective) probability that θ is in the specific set $C(\boldsymbol{x})$.

♠ Invariant Loss

• Location

Let θ denote a location parameter. Then, the loss

$$L(\theta, a) = (\theta - a)^2$$

is invariant with respect to location changes. More specifically, if we let

$$\theta' = \theta + b \text{ and } \delta'(\boldsymbol{x}) = \delta(\boldsymbol{x}) + b,$$

then

$$L(\theta', \delta') = L(\theta, \delta).$$

A more general location invariant loss is

$$L(\theta, a) = h(\theta - a),$$

where h is any nonnegative function of $\theta - a$.

If we take

$$h(\theta - a) = e^{c(\theta - a)} - c(\theta - a) = 1,$$

where c is a constant, the resulting loss is called the **Linex loss**.

• **Scale**

Let $\theta > 0$ is a scalar parameter. Consider either

$$L(\theta, a) = (\theta/a - 1)^2$$

or

$$L(\theta, a) = \theta/a - \ln(\theta/a) - 1,$$

where $a > 0$. Then $L(\theta, a)$ is scale-invariant, since for $\delta'(\boldsymbol{x}) = \delta(\boldsymbol{x})/c$ and $a' = a/c$,

$$L(\theta', \delta') = L(\theta, \delta).$$

A more general scale-invariant loss can be defined by

$$L(\theta, a) = g(\theta/a),$$

where g is any nonnegative function of θ/a.

A question: Is $\theta/a - \ln(\theta/a) - 1 \geq 0$?

♠ Entropy Loss

The entropy loss is defined as

$$L(\theta, a) = E_\theta \left[\ln \left(\frac{f(X|a)}{f(X|\theta)} \right) \right].$$

• Example 2.1: $N(\theta, 1)$ Distribution

Assume $X \sim N(\theta, 1)$. We have

$$f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$$

and

$$\ln f(x|\theta) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2}(x-\theta)^2.$$

Then the entropy loss is

$$\begin{aligned} L(\theta, a) =& E_\theta \left[-\frac{1}{2}(X-\theta)^2 + \frac{1}{2}(X-a)^2 \right] \\ =& -\frac{1}{2} + \frac{1}{2} + \frac{1}{2}(\theta - a)^2 = \frac{1}{2}(\theta - a)^2, \end{aligned}$$

which is a squared-error loss multiplied by a constant.

• **Example 2.2: Exponential Distribution**

Assume $X \sim \mathcal{E}(\theta)$ with density

$$f(x|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}.$$

Then

$$\ln f(x|\theta) = -\ln \theta - \frac{x}{\theta}$$

and the entropy loss is

$$\begin{aligned} L(\theta, a) =& E_\theta \left[-\ln \theta - \frac{X}{\theta} + \ln(a) + \frac{X}{a} \right] \\ =& \frac{\theta}{a} - \ln\left(\frac{\theta}{a}\right) - 1. \end{aligned}$$

It can be shown that (i) $L(\theta, a) \geq 0$; (ii) $L(\theta, a)$ is convex in a; and (iii) the minimum value of $L(\theta, a)$ is 0, which is attained at $a = \theta$.

• **Example 2.3: Poisson Distribution**

Assume $X \sim \mathcal{P}(\theta)$ with density

$$f(x|\theta) = \frac{\theta^x}{x!} e^{-\theta}.$$

Then

$$\ln f(x|\theta) = x \ln(\theta) - \ln(x!) - \theta$$

and the entropy loss is

$$\begin{aligned} L(\theta, a) =& E_\theta \left[X \ln(\theta) - \ln(X!) - \theta - X \ln(a) + \ln(X!) + a \right] \\ =& \theta \ln(\theta) - \theta - \theta \ln(a) + a \\ =& \theta \left[\frac{a}{\theta} - \ln\left(\frac{a}{\theta}\right) - 1 \right], \end{aligned}$$

which is very similar to the loss for the exponential distribution.

• **Example 2.4: $N(\theta, \sigma^2)$ Distribution**

Assume $X \sim N(\theta, \sigma^2)$ with density

$$f(x|\theta, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}.$$

Then

$$\ln f(x|\theta, \sigma) = -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{(x-\theta)^2}{2\sigma^2}.$$

Let $\boldsymbol{a} = (a, b)'$. The entropy loss is

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{a}) =& E_{\boldsymbol{\theta}}\left[-\ln(\sigma) - \frac{(X-\theta)^2}{2\sigma^2} + \ln(b) + \frac{(X-a)^2}{2b^2}\right] \\ =& -\ln(\sigma) - \frac{1}{2} + \ln(b) + \frac{\sigma^2 + (\theta - a)^2}{2b^2} \\ =& \frac{(\theta-a)^2}{2b^2} + \frac{1}{2}\left[\frac{\sigma^2}{b^2} - \ln\left(\frac{\sigma^2}{b^2}\right) - 1\right]. \end{aligned}$$

$$\begin{array}{cc} \Downarrow & \Downarrow \\ \text{location} & \text{scale} \end{array}$$

Clearly, $L(\boldsymbol{\theta}, \boldsymbol{a}) \geq 0$.

♠ Multivariate Loss Functions

Assume that $\boldsymbol{\theta}$ is a p-dimensional vector and Σ is a $p \times p$ positive definite matrix.

For example, $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \Sigma)$.

Then, the standard multivariate loss functions include

(a) $L(\boldsymbol{\theta}, \boldsymbol{a}) = (\boldsymbol{\theta} - \boldsymbol{a})' \Sigma^{-1} (\boldsymbol{\theta} - \boldsymbol{a})$ for location (Σ known).

(b) $L(\boldsymbol{\theta}, \boldsymbol{a}) = \sum_{i=1}^{p} \left(\frac{a_i}{\theta_i} - \ln \frac{a_i}{\theta_i} - 1 \right)$ for scale.

(c) $L(\Sigma, \widehat{\Sigma}) = tr(\Sigma - \widehat{\Sigma})^2$.

(d) $L(\Sigma, \widehat{\Sigma}) = tr(\Sigma \widehat{\Sigma}^{-1} - I)^2$.

(e) $L(\Sigma, \widehat{\Sigma}) = tr(\Sigma \widehat{\Sigma}^{-1}) - \ln(tr(\Sigma \widehat{\Sigma}^{-1})) - p$. (Stein's loss)

(f) $L(\Sigma, \widehat{\Sigma}) = \dfrac{\max \lambda_j}{\max \hat{\lambda}_i}$,

where λ_j and $\hat{\lambda}_i$ are the eigenvalues of Σ and $\widehat{\Sigma}$.

♠ For Predictive Problems

Assume Z is a future observation with density $g(z|\theta)$, which is independent of X.

Let $L^*(z,a)$ denote the loss involving prediction of Z. The loss for a predictive problem is defined by

$$L(\theta,a) = E_\theta^{\boldsymbol{Z}}[L^*(Z,a)] = \int L^*(z,a)g(z|\theta)dz.$$

• **Example 2.5 (Example 5 on page 67)**:

Suppose Z is $N(\theta,\sigma^2)$, and that it is desired to estimate Z under the squared-error loss

$$L^*(z,a) = (z-a)^2.$$

Then

$$\begin{aligned} L(\theta,a) =& E_\theta^{\boldsymbol{Z}}[L^*(Z,a)] = E_\theta^{\boldsymbol{Z}}[(Z-\theta+\theta-a)^2] \\ =& E_\theta^{\boldsymbol{Z}}[(Z-\theta)^2] + E_\theta^{\boldsymbol{Z}}[\theta-a)^2] \\ =& \sigma^2 + (\theta-a)^2. \end{aligned}$$

Thus, working with $L(\theta,a)$ is equivalent to working with squared-error loss for θ.

♠ Balanced Loss Function

Let $\boldsymbol{x} = (x_1, \ldots, x_n)'$ be an observation vector which satisfies

$$\boldsymbol{x} = \boldsymbol{\theta} + \boldsymbol{u},$$

where $\boldsymbol{\theta} \in \Theta$ a p-dimensional subspace of R^n, and $\boldsymbol{u}$ is an error vector such that $E[\boldsymbol{u}] = 0$ and $\mathrm{Var}(\boldsymbol{u}) = \sigma^2 I$.

Zellner (1994) proposed the *balanced loss function* as a means of incorporating both goodness of fit and precision of estimation in the evaluation of an estimator (an action or a decision rule).

A general form of the *balanced loss function* for the above setup is

$$L_w(\boldsymbol{\theta}, \boldsymbol{a}) = w(\boldsymbol{x} - \boldsymbol{a})'(\boldsymbol{x} - \boldsymbol{a}) + (1 - w)(\boldsymbol{\theta} - \boldsymbol{a})'(\boldsymbol{\theta} - \boldsymbol{a}),$$

where $\boldsymbol{a} = (a_1, \ldots, a_n)'$ is an action, $0 \leq w \leq 1$ is the relative weight given to the goodness-of-fit portion of the loss and $1 - w$ is the relative weight given to the precision of action portion.

Note that the squared-error loss is a special case of the balanced loss by taking $w = 0$.

If $\pi^*(\boldsymbol{\theta})$ is the believed probability distribution of $\boldsymbol{\theta}$ at the time of decision making, the *Bayesian expected balanced loss* of the action $\boldsymbol{a}$ is

$$\rho_w(\pi^*, \boldsymbol{a}) = E^{\pi^*}[L_w(\boldsymbol{\theta}, \boldsymbol{a})] = \int_\Theta L_w(\boldsymbol{\theta}, \boldsymbol{a}) dF^{\pi^*}(\boldsymbol{\theta}),$$

where $F^{\pi^*}(\boldsymbol{\theta})$ is the (joint) cdf corresponding to $\pi^*(\boldsymbol{\theta})$.

After some algebra, we obtain

$$\rho_w(\pi^*, \boldsymbol{a}) = \sum_{i=1}^{n} \Big\{ w(x_i - a_i)^2 + (1-w)\Big[\mathrm{Var}^{\pi^*}(\theta_i) + (E^{\pi^*}[\theta_i] - a_i)^2\Big] \Big\}$$

Thus, the Bayes action is

$$\boldsymbol{a}_w^{\pi^*} = w\boldsymbol{x} + (1-w)E^{\pi^*}[\boldsymbol{\theta}],$$

and the minimum value of $\rho_w(\pi^*, \boldsymbol{a})$ is

$$\rho_w(\pi^*, \boldsymbol{a}_w^{\pi^*}) = \sum_{i=1}^{n} \Big\{ w(1-w)(x_i - E^{\pi^*}[\theta_i])^2 + (1-w)\mathrm{Var}^{\pi^*}(\theta_i) \Big\}.$$

The references for Balanced Loss include

Dey, D.K., Ghosh, M., and Strawderman, W.E. (1999). On Estimation with Balanced Loss Functions. *Statistus & Probability Letters, 45*, 97-101.

Zellner, A. (1994). Bayesian and non-Bayesian Estimation Using Balanced Loss Functions. In *Statistica Decision Theory and Related Topics*, Eds. S.S,. Gupta and J.O. Berger. Springer-Verlag, pp. 371-390.

♠ L-Measure

Consider an experiment that yields the data $\boldsymbol{x} = (x_1, \ldots, x_n)'$. Denote the joint sampling density of the x_i's by $f(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of indexing parameters. Let $\boldsymbol{z} = (z_1, \ldots, z_n)'$ denote future values of a replicate experiment. That is, $\boldsymbol{z}$ is a future response vector with the same sampling density as $\boldsymbol{x}|\boldsymbol{\theta}$.

Let $\boldsymbol{a} = (a_1, \ldots, a_n)'$. We define the loss function

$$L(\boldsymbol{\theta}, \boldsymbol{a}) = E^{\boldsymbol{Z}}_{\boldsymbol{\theta}} \left[(\boldsymbol{Z} - \boldsymbol{a})'(\boldsymbol{Z} - \boldsymbol{a}) \right] + k(\boldsymbol{x} - \boldsymbol{a})'(\boldsymbol{x} - \boldsymbol{a}),$$

where $k > 0$. Then

$$\begin{aligned} & L(\boldsymbol{\theta}, \boldsymbol{a}) \\ = & \sum_{i=1}^{n} \left\{ \mathrm{Var}(Z_i|\boldsymbol{\theta}) + (E[Z_i|\boldsymbol{\theta}] - a_i)^2 + k(x_i - a_i)^2 \right\}. \end{aligned}$$

If $\pi^*(\boldsymbol{\theta})$ is the believed probability distribution of $\boldsymbol{\theta}$ at the time of decision making, the *Bayesian expected loss* of the action $\boldsymbol{a}$ is

$$\rho(\pi^*, \boldsymbol{a}) = E^{\pi^*}[L(\boldsymbol{\theta}, \boldsymbol{a})] = \int_{\Theta} L(\boldsymbol{\theta}, \boldsymbol{a}) dF^{\pi^*}(\boldsymbol{\theta}),$$

where $F^{\pi^*}(\boldsymbol{\theta})$ is the (joint) cdf corresponding to $\pi^*(\boldsymbol{\theta})$.

After some algebra, we obtain,

$$\rho(\pi^*, \boldsymbol{a}) = \sum_{i=1}^{n} \Bigg\{ E^{\pi^*}\Big[\mathrm{Var}(Z_i|\boldsymbol{\theta})\Big] + \mathrm{Var}^{\pi^*}(E[Z_i|\boldsymbol{\theta}]) + (\mu_i - a_i)^2 + k(x_i - a_i)^2 \Bigg\},$$

where $\mu_i = E^{\pi^*}(E[Z_i|\boldsymbol{\theta}])$. Minimizing $\rho(\pi^*, \boldsymbol{a})$ gives the Bayes action:

$$\boldsymbol{a}^{\pi^*} = (1 - \nu)\boldsymbol{\mu} + \nu \boldsymbol{x},$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ and $\nu = \frac{k}{1+k}$.

Upon substitution in $\rho(\pi^*, \boldsymbol{a})$, we obtain the *L-measure*

$$L(\pi^*) = \sum_{i=1}^{n} \left\{ E^{\pi^*}\left[\mathrm{Var}(Z_i|\boldsymbol{\theta})\right] + \mathrm{Var}^{\pi^*}\left(E[Z_i|\boldsymbol{\theta}]\right) \right\} + \nu \sum_{i=1}^{n} (\mu_i - y_i)^2.$$

Note: It is common to choose π^* to be the posterior distribution of $\boldsymbol{\theta}$ given the data $\boldsymbol{x}$ and in this case,

$$E^{\pi^*}\left[\mathrm{Var}(Z_i|\boldsymbol{\theta})\right] + \mathrm{Var}^{\pi^*}\left(E[Z_i|\boldsymbol{\theta}]\right) = \mathrm{Var}(Z_i|\boldsymbol{x}),$$

which is the variance of Z_i with respect to the posterior predictive distribution. Thus, the L-measure reduces to

$$L(\pi^*) = \sum_{i=1}^{n} \mathrm{Var}(Z_i|\boldsymbol{x}) + \nu \sum_{i=1}^{n} (\mu_i - y_i)^2.$$

Thus we see that $L(\pi^*)$ has the appealing decomposition as a sum involving the predictive variances plus the squared "bias" terms, $(\mu_i - a_i^{\pi^*})^2$ and where ν is a weight for the second bias component.

The references for L-measure include

Gelfand, A., and Ghosh, S. (1998), "Model Choice: A Minimum Posterior Predictive Loss Approach," *Biometrika, 85*, 1-13.

Ibrahim, J.G., Chen, M.-H., and Sinha, D. (2001). Criterion Based Methods for Bayesian Model Assessment. *Statistica Sinica, 11*, 419-443.

♠ Deviance Loss Function

We assume an exponential family model for x_i of the form

$$f(x_i|\theta_i, \phi) = h(x_i, \phi) \exp\{\tau_i[x_i\theta_i - b(\theta_i)]/\phi\}.$$

Hence,

$$E[X_i|\theta_i, \phi] = b'(\theta_i)$$

and

$$\mathrm{Var}(X_i|\theta_i, \phi) = (\phi/\tau_i)b''(\theta_i).$$

Since b' is strictly increasing, ${b'}^{-1}(\cdot)$ exists and is strictly increasing. We denote it by $\theta(\cdot)$.

Let

$$\begin{aligned} & L(x_i, a_i) \\ =& 2\phi \ln\left(\frac{f(x_i|\theta(x_i), \phi)}{f(x_i|\theta(a_i), \phi)}\right) \\ =& 2\tau_i\Big\{x_i[\theta(x_i) - \theta(a_i)] - [b(\theta(x_i)) - b(\theta(a_i))]\Big\}. \end{aligned}$$

Let $\boldsymbol{z} = (z_1, \ldots, z_n)'$ denote future values of a replicate experiment. That is, $\boldsymbol{z}$ is a future response vector with the same sampling density as $\boldsymbol{x}|\boldsymbol{\theta}$. The *deviance loss* is defined as

$$L(\boldsymbol{\theta}, \boldsymbol{a}) = \sum_{i=1}^{n} \Big\{ E_{\boldsymbol{\theta}} \left[L(Z_i, a_i) \right] + k L(x_i, a_i) \Big\}.$$

• **Example 2.5: Poisson Distribution**

For a Poisson distribution, we have

$$f(x_i|\theta_i, \phi) = h(x_i, \phi) \exp \Big\{ x_i \theta_i - b(\theta_i) \Big\},$$

where $\phi = 1$, $b(\theta_i) = e^{\theta_i}$, and $h(x_i, \phi) = \frac{1}{x_i!}$. Hence, $b'(\theta_i) = e^{\theta_i}$ and $\theta(\cdot) = \log(\cdot)$. Therefore,

$$\begin{aligned} L(x_i, a_i) =& 2 \Big\{ x_i [\theta(x_i) - \theta(a_i)] - [b(\theta(x_i)) - b(\theta(a_i))] \Big\} \\ =& 2 \Big\{ x_i \ln(x_i/a_i) - (x_i - a_i) \Big\}. \end{aligned}$$

When $x_i = 0$, Gelfand and Ghosh (1998, Biometrika) suggest continuity corrections. That is, we use $L(x_i + \frac{1}{2}, a_i + \frac{1}{2})$ instead of $L(x_i, a_i)$.