

# PHƯƠNG PHÁP BOOTSTRAP VÀ ỨNG DỤNG TRONG BÀI TOÁN ƯỚC LƯỢNG

Nguyễn Thị Quyên<sup>1</sup>, Nguyễn Ngân Hà<sup>2</sup>, Đoàn Ngọc Diệu Linh<sup>3</sup>

<sup>1</sup>*Khoa Toán và KHTN*

<sup>2</sup>*Trường THCS Kiến Phúc, Kiến Thụy, Hải Phòng*

<sup>3</sup>*Trường THCS Thanh Quang, Thanh Hà, Hải Dương*

<sup>1</sup>*Email: quyennnt@dhhp.edu.vn*

*Ngày nhận bài: 21/11/2024*

*Ngày PB đánh giá: 26/11/2024*

*Ngày duyệt đăng: 13/01/2025*

**Tóm tắt:** Trong bài báo này, chúng tôi giới thiệu phương pháp tái tạo mẫu Bootstrap là phương pháp tạo mẫu mới từ một mẫu gốc. Trong trường hợp phân bố của biến ngẫu nhiên gốc hoàn toàn chưa biết, cùng với việc tái tạo mẫu Bootstrap, sử dụng kết quả của Luật số lớn, một số cách ước lượng tham số tổng thể được đề xuất: khoảng tin cậy chuẩn, khoảng tin cậy Student hóa; việc đánh giá độ chệch của ước lượng cũng được nêu ra ở đây. Tuy nhiên, trong trường hợp mẫu gốc nhỏ, nhờ việc tái tạo mẫu nhiều lần, chúng tôi đề xuất phương pháp ước lượng sử dụng phân vị - dựa trực tiếp vào các giá trị ước lượng thu được từ các mẫu Bootstrap. Bằng việc sử dụng dụng phần mềm R để mô phỏng cho trường hợp mẫu nhỏ, chúng tôi nhận thấy việc ước lượng bằng phương pháp sử dụng phân vị dựa trên mẫu tái tạo Bootstrap khá hiệu quả, các khoảng tin cậy gần sát với khoảng tin cậy thực.

**Từ khóa:** tái tạo mẫu, Bootstrap, phân bố gốc, ước lượng, phần mềm R.

## BOOTSTRAP METHOD AND ITS APPLICATIONS FOR ESTIMATION PROBLEMS

**Abstract:** In this paper, we introduce the Bootstrap method which generates a new sample from an original one. In the case that the distribution of the original random variable is completely unknown, along with using the Bootstrap method and the results of the Law of Large Numbers, several estimation methods for population parameters are proposed, such as standard confidence intervals, studentized confidence intervals; the assessment of the bias of the estimators is also presented. However, in the case of a small original sample, depending on repeated resampling, we propose an estimation method that uses quantiles derived from the estimators obtained through Bootstrap samples. By using R software to simulate small sample, we found that the quantile-

based estimation method using Bootstrap resampling is quite effective and the confidence intervals closely approximate the real ones.

**Keywords:** Resampling, Bootstrap, population distribution, estimate, R software.

## 1. Giới thiệu

Trong thống kê truyền thống, để kiểm định hoặc ước lượng một tham số  $\theta$  của tổng thể, ta thường dựa vào các giả định về biến ngẫu nhiên gốc để tìm ra phân bố của một thống kê nào đó chứa tham số  $\theta$ . Tuy nhiên không phải lúc nào giả định về biến ngẫu nhiên gốc cũng được thỏa mãn. Khi đó, người ta lập lại mẫu nhiều lần để ước lượng tham số  $\theta$  và sai số tiêu chuẩn của nó. Song việc lập lại mẫu nhiều lần cũng không dễ thực hiện, trên thực tế, có nhiều trường hợp, người ta chỉ có một mẫu, mẫu này gọi là mẫu gốc. Để lấy được nhiều mẫu, người ta thực hiện việc tái lấy mẫu từ mẫu gốc đó. Việc làm như trên gọi là lấy mẫu Bootstrap.

Xét một mẫu ngẫu nhiên  $(X_1, X_2, \dots, X_n)$  lấy từ tổng thể có phân bố  $F$ , chứa tham số  $\theta = t(F)$ , có ước lượng điểm là  $\hat{\theta} = f(X_1, X_2, \dots, X_n)$ . Khi đó phân bố xác suất của  $\hat{\theta}$  sẽ phụ thuộc vào  $n$  và  $F$ . Khi họ phân bố  $F$  đã biết, việc tạo ra nhiều mẫu có thể thực hiện được bằng cách mô phỏng số liệu nhiều lần. Từ đó có thể chứng tỏ rằng việc ước lượng tham số  $\theta$  và sai số tiêu chuẩn có thể thực hiện được thuận lợi và đáng tin cậy.

Trong trường hợp họ phân bố  $F$  hoàn toàn chưa biết, ta sẽ tạo ra mẫu bằng

cách mô phỏng từ phân bố thực nghiệm  $\hat{F} = F_n$  - là hàm phân bố xác suất rời rạc mà mỗi quan sát  $x_1, x_2, \dots, x_n$  đều có xác suất xảy ra bằng  $1/n$ . Như vậy, hàm phân bố  $\hat{F}$  xác định từ bảng xác suất của biến ngẫu nhiên  $X$ :

$X$	$x_1$	$x_2$	$\dots$	$x_n$
$P$	$1/n$	$1/n$	$\dots$	$1/n$

Để ước lượng tham số  $\theta = t(F)$ , ta sử dụng phương pháp đơn giản để ước lượng tham số từ mẫu được xác định như sau [2, 3].

**Định nghĩa 1.** Đối với hàm phân bố  $F$ , ước lượng rập khuôn (plug-in estimate) cho tham số  $\theta = t(F)$  là  $\hat{\theta} = t(\hat{F})$ .

Để thuận tiện, ước lượng sai số tiêu chuẩn của ước lượng cho tham số  $\theta = t(F)$  bất kỳ dựa trên mẫu quan sát  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  được kí hiệu là  $\hat{\theta} = s(\mathbf{x})$ .

**Định nghĩa 2.** Mẫu tái tạo Bootstrap (mẫu Bootstrap) là mẫu ngẫu nhiên cỡ  $n$  được lấy ngẫu nhiên có hoàn lại các phần tử từ mẫu gốc  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , kí hiệu là  $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ .

Như vậy, mẫu tái tạo Bootstrap là mẫu được lấy từ biến ngẫu nhiên gốc  $X$  có các giá trị  $x_1, x_2, \dots, x_n$  và phân bố  $\hat{F}$ .

Khi đó, từ mẫu Bootstrap  $\mathbf{x}^*$ , ước lượng cho tham số  $\theta$  được tính bởi  $\hat{\theta}^* = s(\mathbf{x}^*)$  - được gọi là ước lượng Bootstrap.

## 2. Tổng quan nghiên cứu

Để đảm bảo các tiêu chuẩn ước lượng tham số tổng thể, trong [3] đã chỉ ra độ chệch của ước lượng  $\hat{\theta} = s(\mathbf{x})$  cho tham số  $\theta$  được xác định bởi biểu thức:

$$\text{bias}_F = \text{bias}_F(\hat{\theta}, \theta) = E_F[s(\mathbf{x})] - t(F)$$

Vì độ chệch trên phụ thuộc vào phân bố  $F$  nên ta có thể ước lượng bởi

$$\text{bias}_{\hat{F}} = E_{\hat{F}}[s(\mathbf{x}^*)] - t(\hat{F}).$$

Bằng việc sử dụng mẫu Bootstrap, ước lượng  $E_{\hat{F}}[s(\mathbf{x}^*)]$  và sai số tiêu chuẩn cho  $\text{se}_{\hat{F}}(s(\mathbf{x}^*))$ , được tiến hành theo thủ tục sau:

*Bước 1.* Lấy một mẫu ngẫu nhiên có giá trị quan sát  $\mathbf{x}$  từ tổng thể có phân bố  $F$ , coi đó là mẫu gốc.

*Bước 2.* Sinh ra  $B$  mẫu Bootstrap  $\mathbf{x}^*(1), \mathbf{x}^*(2), \dots, \mathbf{x}^*(B)$  từ mẫu gốc  $\mathbf{x}$ .

*Bước 3.* Với mẫu thứ  $b$  (ký hiệu là  $\mathbf{x}^*(b)$ ) tính ước lượng  $\hat{\theta}^*(b) = s(\mathbf{x}^*(b))$ .

*Bước 4.* Tính trung bình, sai số tiêu chuẩn mẫu  $\text{se}_B$  của  $B$  giá trị  $\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(B)$ :

$$\bar{\theta}^* = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B},$$

$$\text{se}_B = \sqrt{\frac{\sum_{b=1}^B [\hat{\theta}^*(b) - \bar{\theta}^*]^2}{B-1}}.$$

Khi đó  $E_{\hat{F}}[s(\mathbf{x}^*)]$  được xấp xỉ bởi

$$\bar{\theta}^* = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B} \text{ và độ chệch được ước}$$

lượng là  $\text{bias}_B = \bar{\theta}^* - t(\hat{F})$ . Các ước lượng dựa vào mẫu Bootstrap được gọi là ước lượng Bootstrap.

**Định nghĩa 3.** Giới hạn của  $\text{se}_B$  khi  $B \rightarrow \infty$  là ước lượng Bootstrap lý tưởng của  $\text{se}_F(\hat{\theta})$ , tức là:

$$\lim_{B \rightarrow \infty} \text{se}_B = \text{se}_{\hat{F}} = \text{se}_{\hat{F}}(\theta^*).$$

Như vậy, giá trị  $\text{se}_B$  sẽ tiến đến giá trị  $\text{se}_{\hat{F}}$  khi  $B \rightarrow \infty$ , hay là sai số tiêu chuẩn thực nghiệm tiến tới sai số tiêu chuẩn tổng thể khi số lần tái lấy mẫu lớn. “Tổng thể” ở đây ta hiểu là tập các giá trị  $\hat{\theta}^*$ . Cụ thể, với mẫu gốc gồm  $n$  phần tử khác nhau thì số mẫu Bootstrap phân biệt

thực sự là  $C_{2n-1}^n$ , lúc đó “tổng thể” ở đây sẽ gồm  $C_{2n-1}^n$  phần tử.

Người ta chỉ được rằng:  $E_F se_B^2 = E_F se_\infty^2$  và  $V_F se_B^2 \geq V_F se_\infty^2$ .

Trong [1] đã chỉ ra số lượng mẫu Bootstrap B cần thiết như sau:

- Ngay khi số mẫu Bootstrap nhỏ,  $B = 25$  cũng đã đủ để ước lượng  $se_F(\hat{\theta})$ , với  $B = 50$  thường là đủ tốt.

- Rất hiếm khi ta phải dùng đến một lượng lớn mẫu Bootstrap ( $B < 200$ ). Trừ khi tìm khoảng tin cậy thì cần số lần tái tạo mẫu lớn ( $B = 1000$ ).

Để đánh giá độ chệch ta xét căn bậc hai của sai số bình phương trung bình của ước lượng  $\hat{\theta}$  cho  $\theta$  là  $\sqrt{E_F[\hat{\theta} - \theta]^2}$ . Ta có,

$$\begin{aligned} & \sqrt{E_F[\hat{\theta} - \theta]^2} \\ &= \sqrt{E_F[(\hat{\theta} - E\hat{\theta}) + (E\hat{\theta} - \theta)]^2} \\ &= \sqrt{se_F^2(\hat{\theta}) + bias_F^2(\hat{\theta}, \theta)} \\ &= se_F(\hat{\theta}) \cdot \sqrt{1 + \left(\frac{bias_F}{se_F}\right)^2} \\ &= se_F(\hat{\theta}) \left[1 + \left(\frac{bias_F}{se_F}\right)^2\right]^{1/2}. \end{aligned}$$

$$\text{Nếu } \left|\frac{bias_F}{se_F}\right| \leq 0.25$$

thì  $\sqrt{E_F[\hat{\theta} - \theta]^2} \leq 1.03 se_F(\hat{\theta})$ .

Ước lượng Bootstrap lý tưởng cho độ chệch là  $bias_\infty = bias_{\hat{F}}$ . Để xác định ước lượng độ chệch  $bias_B$ , ta đã ước lượng  $E_{\hat{F}}(\hat{\theta}^*)$  bởi  $\bar{\theta}^*$ , theo kết quả của Luật số lớn,

$$\begin{aligned} 0.95 &\simeq P_{\hat{F}} \left\{ \left| \bar{\theta}^* - E_{\hat{F}}(\hat{\theta}^*) \right| < 2 \cdot \frac{se_B}{\sqrt{B}} \right\} \\ &= P_{\hat{F}} \left\{ \left| (\bar{\theta}^* - \hat{\theta}) + (E_{\hat{F}}(\hat{\theta}^*) - \hat{\theta}) \right| < 2 \cdot \frac{se_B}{\sqrt{B}} \right\} \\ &= P_{\hat{F}} \left\{ \left| bias_B - bias_\infty \right| < 2 \cdot \frac{se_B}{\sqrt{B}} \right\}. \end{aligned}$$

### 3. Kết quả nghiên cứu

Giả sử ta có một mẫu quan sát  $\mathbf{x}$  lấy từ tổng thể phân bố F (chưa biết) chứa tham số  $\theta = t(F)$  chưa biết, tức là  $F \rightarrow \mathbf{x} = (x_1, x_2, \dots, x_n)$ . Ước lượng rập khuôn tương ứng  $\hat{\theta} = t(\hat{F})$  là một ước lượng điểm của nó và ước lượng cho sai số tiêu chuẩn của  $\hat{\theta}$  là  $se$ . Với độ tin cậy  $1 - \alpha$ , trong [1], khi kích thước mẫu đủ lớn, bằng việc sử dụng kết quả của Luật số lớn, ước lượng cho  $\theta$  được xây dựng bằng phương pháp khoảng tin cậy xấp xỉ chuẩn, khoảng tin cậy Student hóa. Trong trường hợp kích thước mẫu gốc

nhỏ, chúng tôi đề xuất phương pháp ước lượng phân vị.

### 3.1. Khoảng tin cậy xấp xỉ chuẩn

Khi kích thước mẫu  $n$  lớn thì phân bố của ước lượng  $\hat{\theta}$  xấp xỉ chuẩn, tức là:

$$Z = \frac{\hat{\theta} - E_F(\hat{\theta})}{se} \approx N(0,1)$$

$$\text{hay } Z = \frac{\hat{\theta} - \theta - \text{bias}_F}{se} \approx N(0,1) .$$

Vì độ chệch  $\text{bias}_F$  và  $se$  chưa biết nên ta có thể xấp xỉ bởi ước lượng Bootstrap. Khi đó, với việc dùng  $B$  mẫu Bootstrap theo thủ tục ở trên, ta thu được khoảng tin cậy cho  $\theta$  :

$$\left( \begin{array}{l} \hat{\theta} - \text{bias}_B - u_{\alpha/2} \cdot \text{se}_B; \\ \hat{\theta} - \text{bias}_B + u_{\alpha/2} \cdot \text{se}_B \end{array} \right) \quad (2.1)$$

trong đó  $u_{\alpha/2}$  là giá trị tới hạn mức  $\alpha/2$  của phân bố chuẩn tắc.

Rõ ràng, khoảng tin cậy trên chỉ mang tính xấp xỉ bởi ta đã xấp xỉ phân bố của  $Z$  dựa trên luật số lớn và xấp xỉ độ chệch thông qua ước lượng Bootstrap của nó.

Trong trường hợp kích thước mẫu không đủ lớn thì ta có thể sử dụng một phân bố tốt hơn cho  $Z$ , đó là phân bố Student, cụ thể,

$$Z = \frac{\hat{\theta} - E_F(\hat{\theta})}{se} \approx T^{(n-1)} .$$

Tương tự cách xây dựng như trên, khoảng tin cậy được xấp xỉ

$$\left( \begin{array}{l} \hat{\theta} - \text{bias}_B - t_{\alpha/2}^{(n-1)} \text{se}_B; \\ \hat{\theta} - \text{bias}_B + t_{\alpha/2}^{(n-1)} \text{se}_B \end{array} \right) \quad (2.2)$$

**Nhận xét.** Khi  $\hat{\theta}$  là ước lượng không chệch cho  $\theta$  thì (2.1) và (2.2) được viết lại là

$$\left( \hat{\theta} - u_{\alpha/2} \text{se}_B; \hat{\theta} + u_{\alpha/2} \text{se}_B \right)$$

$$\text{và } \left( \hat{\theta} - t_{\alpha/2}^{(n-1)} \text{se}_B; \hat{\theta} + t_{\alpha/2}^{(n-1)} \text{se}_B \right)$$

### 3.2. Khoảng tin cậy Student hóa

Đây là phương pháp xây dựng khoảng tin cậy dựa vào ước lượng phân bố của  $Z$ , với việc tìm các phân vị của phân bố Student dựa trực tiếp vào mẫu Bootstrap. Thủ tục tiến hành như sau

*Bước 1.* Sinh ra  $B$  mẫu Bootstrap từ mẫu gốc.

*Bước 2.* Từ mỗi mẫu tính giá trị

$$Z^*(b) = \frac{\hat{\theta}^*(b) - \hat{\theta}}{se^*(b)}, \text{ với } \hat{\theta}^*(b), se^*(b)$$

là giá trị và sai số của  $\hat{\theta}$  ở mẫu Bootstrap thứ  $b$ .

*Bước 3.* Phân vị mức  $\alpha$  của  $Z^*(b)$  được ước lượng bởi giá trị  $\hat{t}_{(\alpha)}$  sao cho

$$\frac{\sum_{b=1}^B I_{(Z^*(b) \leq \hat{t}_{(\alpha)})}}{B} = \alpha .$$

Khi đó, khoảng tin cậy Student hóa là  $\left( \hat{\theta} - \hat{t}_{(1-\alpha/2)} se; \hat{\theta} - \hat{t}_{(\alpha/2)} se \right)$ .

### 3.3. Khoảng tin cậy phân vị

Việc xây dựng khoảng tin cậy ở 3.1 và 3.2 vẫn được dựa trên Luật số lớn, ở đó phân bố được xấp xỉ bởi phân bố chuẩn và phân bố Student. Việc làm này chưa cho thấy hiệu quả của phương pháp Bootstrap. Thay cho việc dựa vào phân bố của  $Z = (\hat{\theta} - \theta) / se$ , ta nghiên cứu trực tiếp phân bố Bootstrap của  $\hat{\theta}$ .

Giả sử  $\hat{\theta}$  là một ước lượng điểm của  $\theta$  và  $se$  là sai số tiêu chuẩn của nó. Việc xây dựng khoảng tin cậy dựa vào phân vị mức  $\alpha / 2$  và  $1 - \alpha / 2$ . Xét một mẫu Bootstrap  $\mathbf{x}^*$  được sinh từ phân bố  $\hat{F}$  và có ước lượng Bootstrap  $\hat{\theta}^* = s(\mathbf{x}^*)$ . Giả sử  $\hat{G}$  là phân bố tích lũy của  $\hat{\theta}^*$ . Khi đó khoảng tin cậy phân vị mức  $1 - \alpha$  được xác định bởi phân vị mức  $\alpha / 2$  và  $1 - \alpha / 2$  của  $\hat{G}$  là  $(\hat{\theta}_{lo}, \hat{\theta}_{up}) = (\hat{G}^{-1}(\alpha / 2), \hat{G}^{-1}(1 - \alpha / 2))$ , từ việc xác định  $\hat{G}^{-1}(\alpha / 2) = \hat{\theta}^*(\alpha / 2)$  là phân vị mức  $\alpha / 2$  của phân bố Bootstrap, ta có thể viết khoảng tin cậy phân vị là

$$(\hat{\theta}_{lo}, \hat{\theta}_{up}) = (\hat{\theta}^*(\alpha / 2), \hat{\theta}^*(1 - \alpha / 2)).$$

Để ước lượng được các phân vị trên, ta tạo ra  $B$  mẫu Bootstrap độc lập  $(\mathbf{x}^{*1}, \dots, \mathbf{x}^{*B})$  và tính các ước lượng

Bootstrap  $\hat{\theta}^*(b) = s(\mathbf{x}^{*b})$ . Khi đó xác định  $\hat{\theta}_B^{*(\alpha)}$  là giá trị thứ  $B \cdot \alpha$  trong  $B$  giá trị sắp xếp có thứ tự của các  $\hat{\theta}^*(b)$ , khoảng tin cậy phân vị mức  $1 - \alpha$  xấp xỉ là  $(\hat{\theta}_{lo}, \hat{\theta}_{up}) = (\hat{\theta}_B^*(\alpha / 2), \hat{\theta}_B^*(1 - \alpha / 2))$ .

### 3.4. Nghiên cứu mô phỏng

Theo định lý giới hạn trung tâm khi  $n$  đủ lớn, thì biểu đồ Bootstrap có dáng điệu của phân bố chuẩn, tuy nhiên khi kích thước mẫu nhỏ thì nó không có dáng điệu chuẩn. Khi đó, ước lượng bằng khoảng tin cậy phân vị là một giải pháp. Để xem xét các khoảng tin cậy trên sẽ khác nhau thế nào và tính hiệu quả của ước lượng phân vị, ta sẽ tiến hành mô phỏng- thực hiện bằng phần mềm R - là ngôn ngữ lập trình thống kê sáng lập bởi nhóm tác giả "R Development Core Team" tại đại học Auckland, do Ross Ihaka và Robert Gentleman khởi xướng và phát triển. R trở thành ngôn ngữ rất phổ biến trong thống kê, xử lý số liệu, do mã nguồn mở và việc cài đặt đơn giản từ web <http://r.can-project.org>.

Cụ thể, với mẫu nhỏ, ta mô phỏng để tạo ra một mẫu gốc kích thước  $n = 10$  từ phân bố chuẩn hóa  $N(0,1)$  và ước lượng tham số  $\theta = \mu$ . Khi đó, ước lượng dựa trên mẫu gốc là  $\bar{x}$  và giá trị chân thực là 0. Code R cho việc tái tạo  $B$  mẫu tính toán thực hiện như sau:

```

> dx=function(B)
+ {x=rnorm(10)
+ print(x)
+ {for (i in 1: B)
+ {s=sample(x,10, replace=TRUE)
+ m[i]=mean(s)}
+ print(mean(m))}
+ print(sd(m))
+ a=quantile(m, 0.025)
+ b=quantile(m, 0.975)
+ print(a)
+ print(b)}

```

Kết quả trả về:

```

[1] 0.2009353
[1] 0.2811288
2.5%
-0.3494263
97.5%
0.7243752

```

Dựa trên việc tái tạo  $B = 1000$  mẫu Bootstrap, ta thu được khoảng tin cậy phân vị 95% cho  $\theta$  là  $(-0.3494; 0.72437)$ . Ngoài ra, nhờ việc tính được  $\hat{\theta}^* = 0.2009$  và sai số tiêu chuẩn là  $se = 0.2811$  nên nếu dùng khoảng tin cậy dựa vào phân bố chuẩn, ta được khoảng tin cậy tương ứng là  $(-0.35; 0.7518)$ . Rõ ràng, hai khoảng tin cậy là khá tương đồng.

Để tiếp tục, ta so sánh khoảng tin cậy phân vị với khoảng tin cậy Student hóa với Code trong R

```

> St=function(B)
+ {x=rnorm(10)
+ print(mean(x))
+ print(sd(x))
+ {for (i in 1: B)
+ {s=sample(x,10, replace=TRUE)
+ m[i]=(mean(s)-mean(x))/sd(s)}
+ a=quantile(m, 0.025)
+ b=quantile(m, 0.975)
+ print(a)
+ print(b)}}}

```

Kết quả trả về

```

[1] 0.5170109
[1] 0.8552643
2.5%
-0.915104
97.5%
0.5731135

```

Dựa trên việc tái tạo mẫu với  $B = 1000$ , khoảng tin cậy 95% khi đó là  $(0.02688; 1.2996)$ . Trong trường hợp này, rõ ràng kết quả trả về không tốt bằng phương pháp tái tạo mẫu Bootstrap.

#### 4. Kết luận

Với trường hợp mẫu lớn, bằng việc sử dụng Luật số lớn, các phân bố giới hạn đã được chỉ rõ, nhưng điều này không còn đúng với mẫu nhỏ. Bằng việc sử dụng R để mô phỏng ước lượng trong trường hợp mẫu nhỏ, chúng ta thu được kết luận khá ưu việt về việc sử dụng phương pháp tái tạo mẫu Bootstrap để ước lượng bằng

phân vị. Về mặt lý thuyết, phương pháp này có điều kiện áp dụng khá rộng rãi khi không cần yêu cầu về phân bố của biến ngẫu nhiên gốc; trong trường hợp mẫu nhỏ, nó cho kết quả khá tương đồng khi biết phân bố của biến ngẫu nhiên gốc.

#### **TÀI LIỆU THAM KHẢO**

1. A.C. Davison, D.V.Hinkley, (2006), *Bootstrap Methods and their*

*Application (8th ed.)*, Cambridge Series in Statistical and Probabilistic Mathematics.

2. Bradley Efron, Robert J.Tibshirani (1993), *An introduction to the Bootstrap*, Chapman Hall.

3. Michael R. Chernick. (2007), *Bootstrap methods: A giude to practitioners and researchers*, Willey.