УДК 519.2

Ю.Г. Дмитриев, С.В. Скрипин

О КОМБИНИРОВАННОЙ ОЦЕНКЕ ВЕРОЯТНОСТИ БЕЗОТКАЗНОЙ РАБОТЫ ПО ПОЛНОЙ ВЫБОРКЕ

Строятся статистические оценки для вероятности безотказной работы объекта в виде взвешенной суммы непараметрической оценки вероятности и заданной параметрической оценки вероятности. Находятся асимптотические распределения предлагаемых оценок. Проводится сравнение точности оценок при конечном объеме наблюдений путем имитационного моделирования.

Ключевые слова: *случайная наработка, вероятность безотказной работы, непараметрическая оценка, комбинированная оценка вероятности, асимптотическое распределение.*

В теории надежности при анализе и статистическом оценивании вероятности безотказной работы (ВБР) объекта используются те или иные вероятностно-статистические модели случайной наработки элемента X до первого отказа [1]. Выбор модели обусловлен наличием априорной информации о виде функции распределения F случайной наработки X. Функция распределения может принадлежать как непараметрическому семейству $\boldsymbol{F}$, так и параметрическому семейству $\boldsymbol{G} = \{G(t;\theta),\theta\subset\Theta\}$ распределений наработки. Для новых объектов эта функция распределения, как правило, полностью неизвестна и поэтому для ее статистической оценки применяют непараметрическую оценку, построенную по результатам испытаний. Однако опыт исследователя позволяет ему выдвинуть некоторые априорные догадки о принадлежности F некоторому параметрическому семейству. Эта догадка может быть как верной, так и неверной. Возникает желание использовать имеющиеся знания о возможной параметризации распределения в непараметрических оценках с целью улучшения их свойств. В данной работе предлагается подход к построению комбинированных оценок в виде взвешенной суммы непараметрической оценки вероятности и заданной параметрической оценки вероятности. Обсуждаются проблемы, возникающие при реализации таких оценок на практике, анализируются асимптотические свойства, а также свойства при конечном объеме наблюдений путем имитационного моделирования.

1. Постановка задачи

Пусть $F(t) = P\{X \le t\}$, $t > 0$, – функция распределения случайной наработки X. Запишем ВБР объекта, достигшего возраста t, в виде

$$J(x;t,F) = \frac{1-F(t+x)}{1-F(t)} = \frac{P_F(A_{t+x})}{P_F(A_t)}, \qquad (1)$$

где событие $A_{t+x} = \{X>t+x\}$, $A_t = X>t$.

Имеется полная простая выборка $X_1,\ldots, X_n$ объема n случайной наработки X с $F\in\boldsymbol{F}$. Непараметрическую оценку ВБР (1) возьмем в виде

$$\hat{J}(x;t) = \frac{1-F_n(x+t)}{1-F_n(t)} = \frac{\hat{P}(B)}{\hat{P}(A)}, \qquad (2)$$

где $F_n(t) = n^{-1}\sum_{i=1}^{n} c\,(t - X_i)$, $c(t) = \{0\colon t<0,\ 1\colon t\geq 0\}$ – эмпирическая функция распределения, $\hat{P}\,(\cdot)$ – эмпирические вероятности, $B = A_{t+x}$, $A = A_t$.

Пусть имеется предположение, что $F\in\boldsymbol{G}$. В этом случае ВБР запишется в виде

$$J(x;t,G) = \frac{1-G(t+x,\theta)}{1-G(t,\theta)} = \frac{P_\theta(A_{t+x})}{P_\theta(A_t)} = \Psi(x;t,\theta), \theta\in\Theta. \quad (3)$$

Назовем величину $\Psi(x;t,\theta)$ априорной догадкой. Задача состоит в построении оценки ВБР, учитывающей непараметрическую оценку (2) и априорную догадку (3) совместно.

Рассмотрим разложение непараметрической оценки в окрестности истинных вероятностей $P(B)$ и $P(A)$ по формуле Тейлора с остаточным членом $R_n(t, x)$ в форме Лагранжа для произвольных, но фиксированных t, x:

$$\hat{J}(x;t) = \frac{P_F(A_{t+x})}{P_F(A_t)} + \frac{1}{P(A)}(\hat{P}(B) - P(B)) - \frac{P(B)}{P^2(A)}(\hat{P}(A) - P(A)) + R_n(t,x). \quad (4)$$

Обозначим главную часть в (4) через

$$\hat{J}_g(x;t) = J(x;t,F) + \frac{1}{P(A)}(\hat{P}(B) - P(B)) - \frac{P(B)}{P^2(A)}(\hat{P}(A) - P(A)). \quad (5)$$

Из (5) следует, что математическое ожидание и дисперсия главной части равны соответственно

$$M_F\hat{J}_g(x;t) = J(x;t,F),\ \ D_F\hat{J}_g(x;t) = n^{-1}\sigma_F^2,$$

где $$\sigma_F^2 = \sigma_F^2(x;t) = \frac{P(B)(P(A)-P(B))}{P^3(A)} = \frac{[1-F(t+x)][F(t+x)-F(t)]}{(1-F(t))^3}. \quad (6)$$

Кроме того, из (5) вытекает, что при увеличении объема наблюдений n в силу центральной предельной теоремы последовательность

$$\sqrt{n}(\hat{J}_g(x;t) - J(x;t,F)) \sim N\left(0,\sigma_F^2\right),$$

т.е. имеет асимптотически нормальное распределение с нулевым математическим ожиданием и дисперсией, определяемой формулой (6). Поскольку вторые производные в остаточном члене $R_n(t, x)$ являются непрерывными функциями относительно $P(B)$ и $P(A)$, то в силу теоремы непрерывности (см. [2], глава 6) последовательность $\sqrt{n}\ R_n(t, x)$ слабо сходится к нулю. Отсюда следует, что и непараметрическая оценка (2) также имеет асимптотически нормальное распределение, т.е. $\sqrt{n}\ (\hat{J}(x;t) - J(x;t,F)) \sim N(0,\sigma^2_F)$.

2. Комбинированная оценка

Совместное использование непараметрической оценки (2) и априорной догадки (3) осуществим с помощью комбинированной оценки следующего вида:

$$\hat{J}_\lambda(x;t) = (1-\lambda)\hat{J}(x,t) + \lambda\Psi(x;t,\theta)) = \hat{J}(x,t) - \lambda(\hat{J}(x;t) - \Psi(x;t,\theta)), \quad (7)$$

где λ – коэффициент взвешивания, выбираемый из заданного критерия качества, x, t, θ произвольные, но фиксированы. Поскольку определяющим элементом в асимптотическом поведении оценки (2) является ее главная часть, то выберем коэффициент λ из условия минимума среднеквадратической ошибки (СКО) главной части

$$S_F^2(x;t,\lambda)=M_F[\hat{J}_g(x;t)-\lambda(\hat{J}_g(x;t)-\Psi(x;t,\theta)]^2\,. \qquad (8)$$

Минимизация выражения (8) по λ приводит к оптимальному значению коэффициента

$$\lambda_o=\lambda_o(x;t,F)=\frac{\sigma_F^2}{\sigma_F^2+n\Delta_F^2}=\left(1+n\frac{\Delta_F^2}{\sigma_F^2}\right)^{-1}, \qquad (9)$$

где $\sigma^2{}_F$ определяется формулой (6), а $\Delta_F = J(x;t,F) - \Psi(x;t,\theta)$ – величина смещения (отклонения) априорной догадки от истинного значения ВБР. Минимум СКО главной части удовлетворяет соотношению

$$nS_F^2(x;t,\lambda_o)=\sigma_F^2(1-\lambda_o)=\sigma_F^2\left(\frac{n\Delta_F^2}{\sigma_F^2+n\Delta_F^2}\right), \qquad (10)$$

Как показывают выражения (9) и (10), оптимальное значение λ_o изменятся в пределах $0 < \lambda_o \le 1$, а выигрыш в точности оценивания комбинированной оценки по сравнению с непараметрической оценкой характеризуется отношением

$$V=\frac{S_F^2(x;t,\lambda_o)}{D_F\hat{J}_g(x;t)}=(1-\lambda_o)=\left(\frac{n\Delta_F^2}{\sigma_F^2+n\Delta_F^2}\right). \qquad (11)$$

Формула (11) показывает, что чем V ближе к нулю, тем сильнее влияние априорной догадки на точность оценивания ВБР, и чем ближе V к единице, тем слабее это влияние. При $\Delta_F = 0$ имеем $\lambda_o = 1$, $V = 0$, и в качестве оценки ВБР следует взять априорную догадку $\Psi(x;t,\theta)$. При $\Delta_F \neq 0$, что обычно и бывает на практике, $\lambda_o < 1$ и с ростом объема наблюдений ($n\to\infty$) $\lambda_o\to 0$, влияние априорной догадки уменьшается, выигрыш в точности оценивания убывает, $V\to 1$.

3. Адаптивные оценки

Оптимальное значение λ_o зависит от функции распределения F случайной наработки и, как правило, неизвестно. Это обстоятельство затрудняет применение комбинированной оценки (7) на практике и приводит к поиску тех или иных оценок $\hat{\lambda}_0$ для λ_o, при которых выигрыш в точности еще сохраняется. Комбинированные оценки (7) с $\hat{\lambda}_0$ назовем адаптивными. В этом разделе мы рассмотрим некоторые из возможных оценок для λ_o и приведем свойства получающихся при этом адаптивных оценок

$$\hat{J}(x;t,\hat{\lambda}_0)=\hat{J}(x;t)-\hat{\lambda}_0(\hat{J}(x;t)-\Psi(x,t,\theta))\,. \qquad (12)$$

1. Первая оценка λ_o. Подставим в формулу (9) вместо неизвестной F эмпирическую функцию распределения F_n (или вместо неизвестных вероятностей P – их эмпирические оценки $\hat{P}$). Получим оценку

$$\hat{\lambda}_{01}=\left(1+n\hat{\Delta}^2/\hat{\sigma}^2\right)^{-1}, \qquad (13)$$

где

$$\hat{\Delta}=\hat{\Delta}(x;t)=\hat{J}(x;t)-\Psi(x;t,\theta)=\frac{\hat{P}(B)}{\hat{P}(A)}-\Psi(x;t,\theta)\,,$$

$$\hat{\sigma}^2=\hat{P}(B)(\hat{P}(A)-\hat{P}(B))/\hat{P}^3(A)\,.$$

$$\frac{\hat{\Delta}^2}{\hat{\sigma}^2} = \frac{(\hat{P}(B) - \Psi\hat{P}(A))^2 \hat{P}(A)}{\hat{P}(B)(\hat{P}(A) - \hat{P}(B))}. \tag{14}$$

Пусть $\Delta_F \neq 0$. Рассмотрим разность

$$\sqrt{n}(\hat{J}(x;t,\hat{\lambda}_{01}) - J(x;t)) = \sqrt{n}(\hat{J}(x;t) - J(x;t)) - \sqrt{n}\hat{\lambda}_{01}\hat{\Delta}. \tag{15}$$

Поскольку (13) и отношение (14) являются непрерывными функциями от $\hat{P}(A), \hat{P}(B)$, то в силу состоятельности эмпирических вероятностей, по теореме непрерывности (см. [2], глава 6) следует, что для каждых фиксированных x, t при $n\to\infty$

$$\frac{\hat{\Delta}^2}{\hat{\sigma}^2} \xrightarrow{p} \frac{\Delta_F^2}{\sigma_F^2},\ \hat{\lambda}_{01} \xrightarrow{p} 0,\ \sqrt{n}\hat{\lambda}_{01}\hat{\Delta} \xrightarrow{p} 0.$$

Отсюда из (15) вытекает, что при $\Delta_F \neq 0$ асимптотические распределения адаптивной оценки $\hat{J}(x;t,\hat{\lambda}_{01})$ и непараметрической $\hat{J}(x;t)$ одинаковы и совпадают, как установлено выше, с нормальным законом $N(0,\sigma^2_F)$.

Пусть $\Delta_F = 0$. Тогда при $n\to\infty$ в силу центральной предельной теоремы и теоремы непрерывности ([2], глава 6) случайная последовательность $\eta_n = \sqrt{n}\hat{\Delta}/\hat{\sigma}$ стремится по распределению к стандартной нормальной случайной величине $\eta\in N(0,1)$, а оценка $\hat{\lambda}_{01}$ сходится по распределению к случайной величине $1/(1+\eta^2)$. Отсюда следует, что закон распределения случайной последовательности

$$\frac{\sqrt{n}(\hat{J}(x;t,\hat{\lambda}_{01}) - J(x;t))}{\hat{\sigma}} = \frac{\sqrt{n}(\hat{J}(x;t) - J(x;t))}{\hat{\sigma}} - \frac{\hat{\lambda}_{01}\sqrt{n}\hat{\Delta}}{\hat{\sigma}}$$

сходится к распределению случайной величины

$$\xi = \eta - \frac{\eta}{1+\eta^2} = \frac{\eta^3}{1+\eta^2}, \eta \in N(0,1),$$

с математическим ожиданием $M\xi = 0$ и дисперсией $D\xi = \sigma^2_\xi$.

2. Вторая оценка λ_o. Желание обеспечить сходимость оценки коэффициента λ_o к единице при $\Delta_F = 0$ приводит ко второй оценке

$$\hat{\lambda}_{02} = \frac{1}{1+n\left|\hat{\Delta}/\hat{\sigma}\right|^\alpha} = \frac{1}{1+n^{1-\frac{\alpha}{2}}\left|\sqrt{n}\hat{\Delta}/\hat{\sigma}\right|^\alpha},\ \alpha > 2.$$

Если $\Delta_F \neq 0$, то при $n\to\infty$

$$\left|\hat{\Delta}/\hat{\sigma}\right|^\alpha \xrightarrow{p} \left|\Delta_F/\sigma_F\right|^\alpha,\ \hat{\lambda}_{02} \xrightarrow{p} 0,\ \sqrt{n}\hat{\lambda}_{02}\hat{\Delta} \xrightarrow{p} 0.$$

Отсюда следует, что асимптотические распределения оценок $\hat{J}(x;t,\hat{\lambda}_{02})$ и $\hat{J}(x;t)$ одинаковы и совпадают с нормальным законом $N(0,\sigma^2_F)$.

Если $\Delta_F = 0$, то при $n\to\infty$ имеют место следующие сходимости по распределению случайных последовательностей: $\left|\sqrt{n}\hat{\Delta}/\hat{\sigma}\right|$ сходится к случайной величине $|\eta|^\alpha$, оценка λ_{o2} – к единице, а $\sqrt{n}(\hat{J}(x;t,\hat{\lambda}_{02}) - J(x;t))/\hat{\sigma} = \sqrt{n}\hat{\Delta}(1-\hat{\lambda}_2)/\hat{\sigma}$ – к нулю.

Возможны и другие виды $\hat{\lambda}_0$ в адаптивных оценках (12), приводящих к новым свойствам.

4. Имитационное моделирование

С целью выяснения качества предложенных в работе адаптивных оценок ВБР при конечных объемах выборок было проведено имитационное компьютерное моделирование при следующих условиях:

1. Выборки генерировались из экспоненциального закона распределения $F(x) = 1 - \exp(-\theta x)$ с параметром $\theta = 1{,}0$. Объемы n выборок изменялись в диапазоне от 5 до 125 с интервалом (шагом) равным 5.

2. Генерирование событий A_t и A_{t+x} проводилось при $t = 0{,}10536$, $x = 0{,}5$. При этом для каждой сформированной выборки соблюдалось условие: имеется хотя бы одно наблюдение с событием A_t, иначе выборка исключается из рассмотрения.

3. Сравнение СКО оценок с разной величиной смещения Δ проводилось при изменении значений параметра θ в диапазоне от 0,4 до 1,8 с интервалом (шагом), равным 0,2 и фиксированным объёмом выборки $n = 15$;

4. Для каждого значения n из заданного диапазона числовые результаты эксперимента рассчитывались по серии выборок объема $K = 1000000$ (с одинаковым объемом наблюдений n в каждой выборке);

5. Квадрат отклонения (ошибки) оценок от истинного значения

$$S_j^2 = (J_{n,j} - J)^2, \quad j = \overline{1,3},$$

где $J_{n,1}$ – параметрическая оценка $\Psi(x;t,\theta)$, $J_{n,2}$ – непараметрическая оценка $\hat{J}(x;t)$, $J_{n,3}$ – комбинированная оценка $\hat{J}(x;t,\hat{\lambda}_{01})$, J – истинное значение $J(x,t,F)$.

6. Критерии оценки СКО, вычисленные по серии K выборок,

$$Q_j = \frac{1}{K}\sum_{i=1}^{K} S_{j,i}^2, \quad j = \overline{1,3},$$

где Q_1 – оценка СКО для $\Psi(x;t,\theta)$, Q_2 – оценка СКО для $\hat{J}(x;t)$, Q_3 – оценка СКО для $\hat{J}(x;t,\hat{\lambda}_{01})$.

5. Результаты моделирования

Результаты моделирования экспериментов представлены на рис. 1 и 2. Дадим краткий комментарий результатов.

Эксперимент 1. Цель эксперимента – оценка и анализ изменения СКО комбинированной оценки $\hat{J}(x;t,\hat{\lambda}_{01})$ по сравнению с непараметрической в условиях конечных объёмов выборки n. Параметрической оценкой была взята величина $\Psi(x;t,\theta)$ с СКО равной нулю. Выигрыш СКО определялся только свойствами комбинированной оценки.

На рис. 1 по горизонтальной оси указан объём выборок n, а по вертикальной оси – величина Q оценок СКО непараметрической и комбинированной оценок. Из рисунка виден почти двукратный выигрыш СКО комбинированной оценки по сравнению с непараметрической уже при $n \geq 5$. Этот выигрыш сохраняется и при увеличении объёмов выборок за пределы диапазона условий моделирования, что подтверждает теоретические результаты.

Эксперимент 2. Цель эксперимента – оценить изменение выигрыша СКО комбинированной оценки $\hat{J}(x;t,\hat{\lambda}_{01})$ по сравнению с оценкой СКО непараметрической оценки при изменении величины θ. Эксперимент проведён при фиксированном объёме выборок $n = 15$.

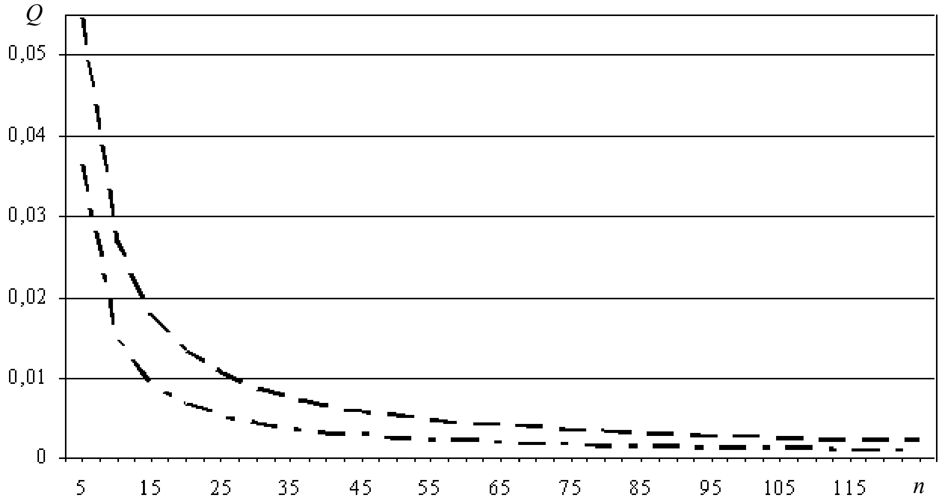

Рис. 1. Зависимости оценок СКО непараметрической (пунктирная линия) и комбинированной (штрих-пунктирная линия) оценок от объёма выборок n

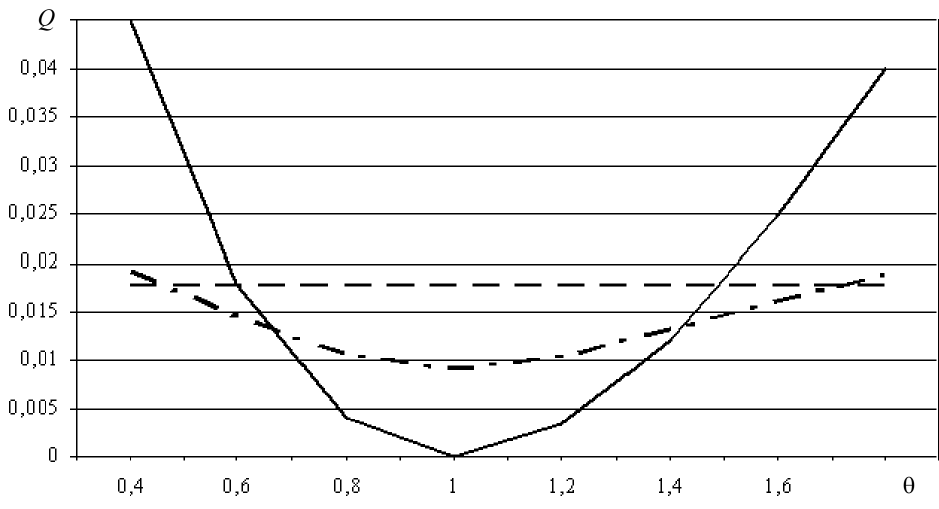

Рис. 2. Зависимости оценок СКО параметрической (сплошная линия), комбинированной (штрих-пунктирная линия) и непараметрической (пунктирная линия) оценок от смещения параметра θ (истинное значение $\theta = 1$) при объёме выборки $n = 15$

На рис. 2 по горизонтальной оси указана величина параметра θ. По вертикальной оси указана величина Q оценок СКО параметрической (сплошная линия), непараметрической (пунктирная линия) и комбинированной (штрих – пунктирная линия) оценок. Все кривые статистически значимо различаются на уровне значимости $< 0{,}05$, за исключением областей пересечения кривых. Рисунок демонстрирует, что при $\theta = 1$ СКО параметрической оценки равно нулю, а оценки СКО непараметрической и комбинированной оценок соответствуют величинам, представленным на рис. 1 при $n = 15$. При смещении параметра θ в меньшую или большую сторону от $\theta = 1$ выигрыш СКО комбинированной оценки плавно уменьшается по сравнению с оценкой СКО непараметрической оценки. При значениях $\theta < 0{,}5$ и $\theta > 1{,}7$ выигрыш отсутствует. Однако уменьшение оценки СКО и выигрыша происходит значительно медленнее, чем у параметрической. Поэтому при значениях $0{,}5 < \theta < 0{,}7$ и $1{,}4 < \theta < 1{,}7$ выигрыш СКО комбинированной оценки имеет место и по сравнению с оценкой СКО параметрической оценки, что подтверждает преимущества комбинированных оценок. Преимущества комбинированных оценок подтверждаются и для $\hat{J}(x;t,\hat{\lambda}_{02})$ и в других экспериментах, не представленных в данном пункте.

Заключение

В работе предложены статистические комбинированные оценки для ВБР объекта, в которых используется знание исследователя о возможных параметрических моделях функций распределения наработки до отказа или самой ВБР. Оценки являются взвешенными суммами непараметрических и параметрических оценок ВБР. Найден оптимальный коэффициент взвешивания и показано, что при любом конечном объеме выборки комбинированные оценки являются более точными по критерию СКО главных частей оценок. Однако применение таких оценок на практике затруднено незнанием оптимального коэффициента взвешивания. В работе приведены две оценки таких коэффициентов и построены адаптивные комбинированные оценки ВБР. Результаты имитационного моделирования показали, что при конечных объемах выборок адаптивные оценки имеют меньшее СКО по сравнению с непараметрической оценкой.

Результаты моделирования, представленные на рисунках, получены с помощью кластера Межрегионального вычислительного центра ТГУ СКИФ Cyberia (skif.tsu.ru). Авторы выражают благодарность сотрудникам Центра за оказанную помощь.

ЛИТЕРАТУРА

1. *Байхельт Ф., Франкен П.* Надежность и техническое обслуживание. Математический подход: пер. с нем. М.: Радио и связь, 1988. 392 с.
2. *Боровков А.А.* Математическая статистика. М.: Наука, 2007. 704 с.

Дмитриев Юрий Глебович
Скрипин Сергей Викторович
Томский государственный университет
E-mail: dmit@mail.tsu.ru, skripinsv@mail.ru
Поступила в редакцию 2 августа 2012 г.

Dmitriev Yury G., Skripin Sergey V. (Tomsk State University). **On a combined estimating of the probability of failure-free operation for the full sample**.

Keywords: the probability of failure-free operation, nonparametric estimation, the combined estimate of the probability, the asymptotic distribution.

There is constructed statistical estimates for the probability of failure-free operation of the system as a weighted sum of nonparametric and parametric estimates of the probability. Asymptotic distributions of combined estimates are found. The accuracy of estimates with finite volume of observations is studied by simulation.