# Ecological risk assessment: bootstrap estimates of the extinction risk based on a stochastic model

Hiroshi Hakoyama[1] and Yoh Iwasa[2,3]

[1] Stock Assessment Section, Subarctic Fisheries Resources Division, Hokkaido National Fisheries Research Institute, Katsurakoi 116, Kushiro, Hokkaido 085-0802, JAPAN

[2] Department of Biology, Faculty of Science, Kyushu University, CREST, Japan Science and Technology Corporation (JST), Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-81, JAPAN

[3] Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812-81, JAPAN

**Key Words:** ecological risk assessment, stochastic model, extinction time, confidence interval, bootstrap, harvest model, model aggregation

**Abstract**

We have studied a method to evaluate the ecological risk in terms of the decrease of mean extinction time $T$ (or log$T$) of populations. The method bases on a stochastic differential equation model that we call canonical model (logistic growth with environmental and demographic stochasticites) and maximum likelihood estimate of the parameters (growth rate $r$, carrying capacity $K$, and environmental noise $\sigma_e^2$) from a time series data of population size. In this paper, we develop a method to evaluate the approximate confidence interval of the estimated parameters using bootstrap computer simulation. In cases of short data, the parameters (especially $r$) tend to be missestimated because the maximum likelihood estimates are not unbiased estimates. Therefore, we develop a new simple method based on bootstrap to improve these biased estimates and to obtain better estimates. This adjusted method reduces bias in estimates eminently. We also study the effect of model simplification (aggregation) of three harvest models (constant, proportional and threshold harvest) to canonical model. Proportional harvest model shows complete aggregation. log$T$ of constant harvest model significantly correlated with log$T$ estimated using by canonical model. However, mean extinction time of threshold harvest model is difficult to estimate by canonical model.

## 1. Introduction

The mean extinction time of natural populations provides a risk criterion for ecological risk assessment (Nakanishi 1995; Iwasa 1998; Hakoyama and Iwasa 1998). We studied a method to estimate mean extinction time from time series data of population size based on a stochastic model (canonical model; Iwasa 1998).

The estimate method was based on a maximum likelihood estimator with some approximations, then the estimate is not always correct, especially in short time-series data. Time-series data of natural populations are usually shorter than 50 generations, therefore the estimate bias in short data set was a problem of the method. As the other problem, because we do not know the distribution of the estimate of mean extinction time, we could only calculate a point estimation but could not calculate confidence intervals of the estimate.

In this paper, we attempt to solve these problems. We develop a new method to obtain a better estimate than the simple maximum likelihood method, based on bootstrap computer simulation. Bootstrapping method also provides approximate confidence intervals of estimates. We also examine the robustness of the canonical model to three harvest models (complex structured models), and examine the error in estimates when the canonical model is applied to the structured populations.

## 2. Canonical Model

This section is a brief review of our previous study on a stochastic single population model (canonical model; see Iwasa 1998 and Hakoyama and Iwasa 1998). The dynamics of population size $X$ at time $t$ is expressed in terms of stochastic differential equation (canonical model):

$$\frac{dX}{dt} = rX\left(1-\frac{X}{K}\right) + \sigma_e \xi_e(t) \circ X + \xi_d(t) \bullet \sqrt{X}, \qquad (r, K, \text{and } \sigma_e > 0), \qquad (1)$$

where $r$ is a growth rate, and $K$ is a carrying capacity, $\xi_e(t)$ is the white noise, and $\sigma_e$ is the intensity of the environmental fluctuation. The mean extinction time $T_K$, for a population following Eq. (1) is:

$$T_K(r, K, \sigma_e) = \frac{2}{{\sigma_e}^2} \int_0^K \int_0^\infty e^{-R(y-x)} \left(\frac{y+D}{x+D}\right)^{R(K+D)+1} \frac{1}{(y+D)y} dy dx, \qquad (2)$$

where $R \equiv \dfrac{2r}{{\sigma_e}^2 K}$ and $D \equiv \dfrac{1}{{\sigma_e}^2}$.

To apply the model to natural populations, we need three parameters ($r$, $K$, and $\sigma_e$). These can be estimated from a time-series data of population size using by a maximum likelihood method. Suppose that we sample population size $X(t)$ at $n$+1 points with regular intervals $\{X(t_0)=x_0, X(t_0+\tau)=x_1, X(t_0+2\tau)=x_2, \ldots, X(t_0+n\tau)=x_n\}$. If the demographic stochasticity is neglected (if population size moderately large), carrying capacity $K$ is equal to the average population size:

$$K = E[X(t)]. \tag{3}$$

A demographic stochasticity causes a small bias: $K > E[X(t)]$. Assuming small fluctuations around the population average ($\sigma_e^2 << r$), the likelihood function $L$ is:

$$\ln L = -\frac{1}{2}\ln(2\pi\alpha) - \frac{n}{2}\ln\left(2\pi\alpha(1-\beta^2)\right) - \frac{x_0^2}{2\alpha} - \frac{1}{2\alpha(1-\beta^2)}\sum_{i=0}^{n-1}\left(x_{i+1}-\beta x_i\right)^2 \tag{4}$$

in which $\alpha = \frac{1}{2r}\left(\sigma_e^2 K^2 + K\right)$, $\beta = \mathrm{e}^{-r\tau}$.

## 3. Confidence Intervals

### 3.1 *Bootstrap Confidence Intervals*

Approximate confidence intervals of parameters of canonical model, Eq. (1) can be found by bootstrapping computer simulation (see Dennis and Taper 1994). First, we calculate maximum likelihood estimates of parameters ($\hat{r}$, $\hat{K}$ and $\hat{\sigma}_e$) from time series data of population size, maximizing ln$L$, Eq. (4). Second, using the parameters estimated and a computer simulation model that parallels to canonical model, we generate repeatedly (for example, $n$ = 5000) time series of the same length as the original data. We start each simulation from $K$ and use the population size at the 30th generation time (as moderately long time) for the initial population size of computer-generate data. For each computer-generate data, we calculate maximum likelihood estimates of parameters, denoted $\hat{r}^*$, $\hat{K}^*$ and $\hat{\sigma}_e^*$. Third, we take the 2.5th and the 97.5th sample percentiles of the 5000 $\hat{r}^*$ values for 95% confidence intervals of $r$, $K$, and $\sigma_e$.

Figure 1 shows the distribution of $\hat{r}^*$, $\hat{K}^*$, $\hat{\sigma}_e^*$ and $\log\hat{T}^*$. Here we do not use real data to estimate $\hat{r}$, $\hat{K}$ and $\hat{\sigma}_e$, but we set these parameters arbitrarily. Clearly, the estimate of $\hat{r}$ is under large bias, and $\hat{r}$ tend to be overestimated ($E[\hat{r}^*]$ is larger than $\hat{r}$). On the other hand, the estimate bias of $\hat{K}^*$ and $\hat{\sigma}_e^*$ is relatively small ($E[\hat{K}^*]$ and $E[\hat{\sigma}_e^*]$ are close to $\hat{K}$ and $\hat{\sigma}_e$, respectively). $E[\log\hat{T}^*]$ is also larger than $\log\hat{T}$, and this bias of $\log\hat{T}(\hat{r},\hat{K},\hat{\sigma}_e)$ attributes to the estimate bias of $\hat{r}$ mostly. Clearly, in the case of short data, the maximum likelihood estimate based on Eq. 3 and Eq. 4 is quite misleading.

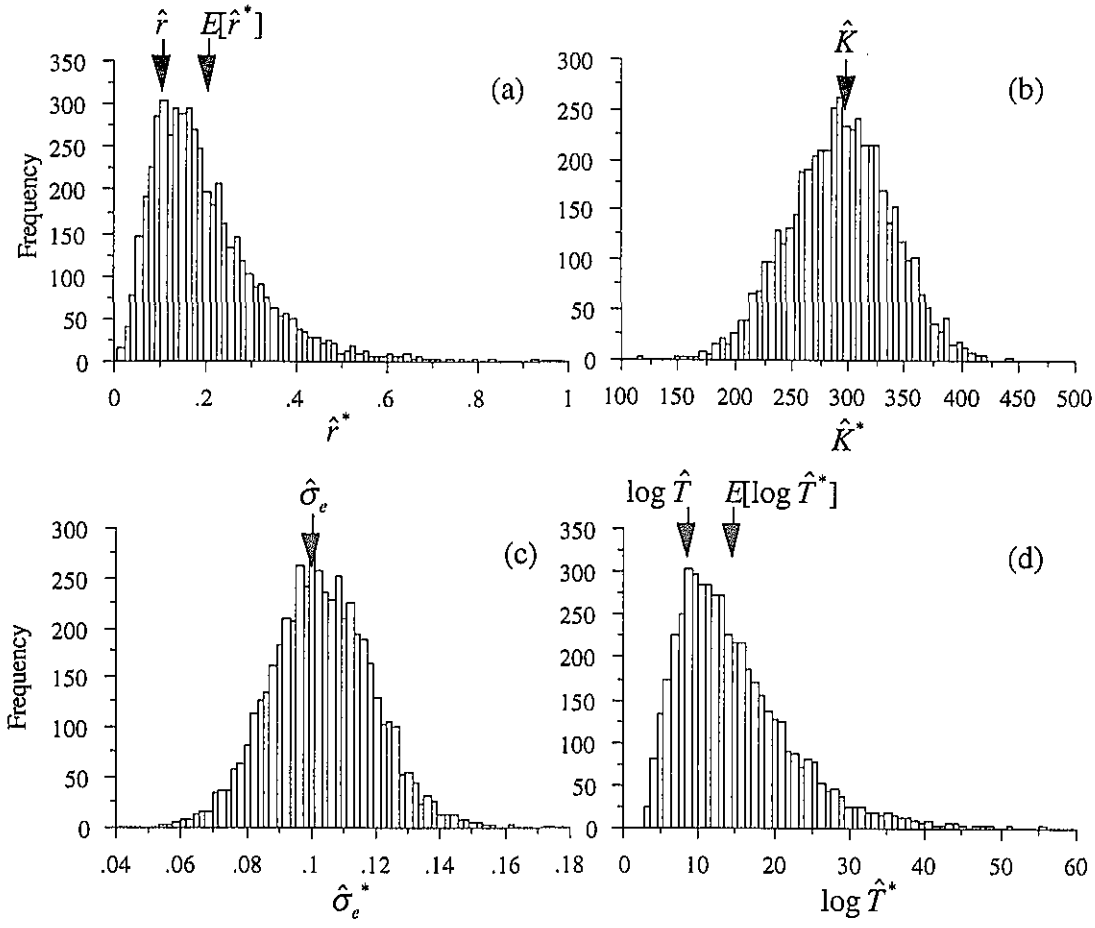


Fig. 1 The distribution of (a) $\hat{r}^*$, (b) $\hat{K}^*$, (c) $\hat{\sigma}_e^*$ and (d) $\log \hat{T}^*$. We set $\hat{r}$ = 0.1, $\hat{K}$ = 300 and $\hat{\sigma}_e$ = 0.1. $n_{\text{time series}}$ = 50. $n_{\text{computer-generate}}$ = 5000. We calculate $\log \hat{T}^*$ using by an approximate regression formula (Eq. A1, Appendix).

### 3.2 *Data length and confidence intervals*

As shown in Fig. 2, the bootstrap confidence intervals of parameters become small with long data. Because of the nature that a maximum likelihood estimate approaches to an unbiased estimate with long data, the bootstrap average of $r$ and $\sigma_e$ approaches to the true value (Fig. 2a, c).

When bootstrap averages deviate from true values, the confidence intervals are also not correct (for example, in $r$ with data length = 10; Fig. 2). Because the bootstrap average of $r$ and $\sigma_e$ is overestimate (Fig. 2a, c), the true confidence intervals must be smaller than the bootstrap confidence intervals.

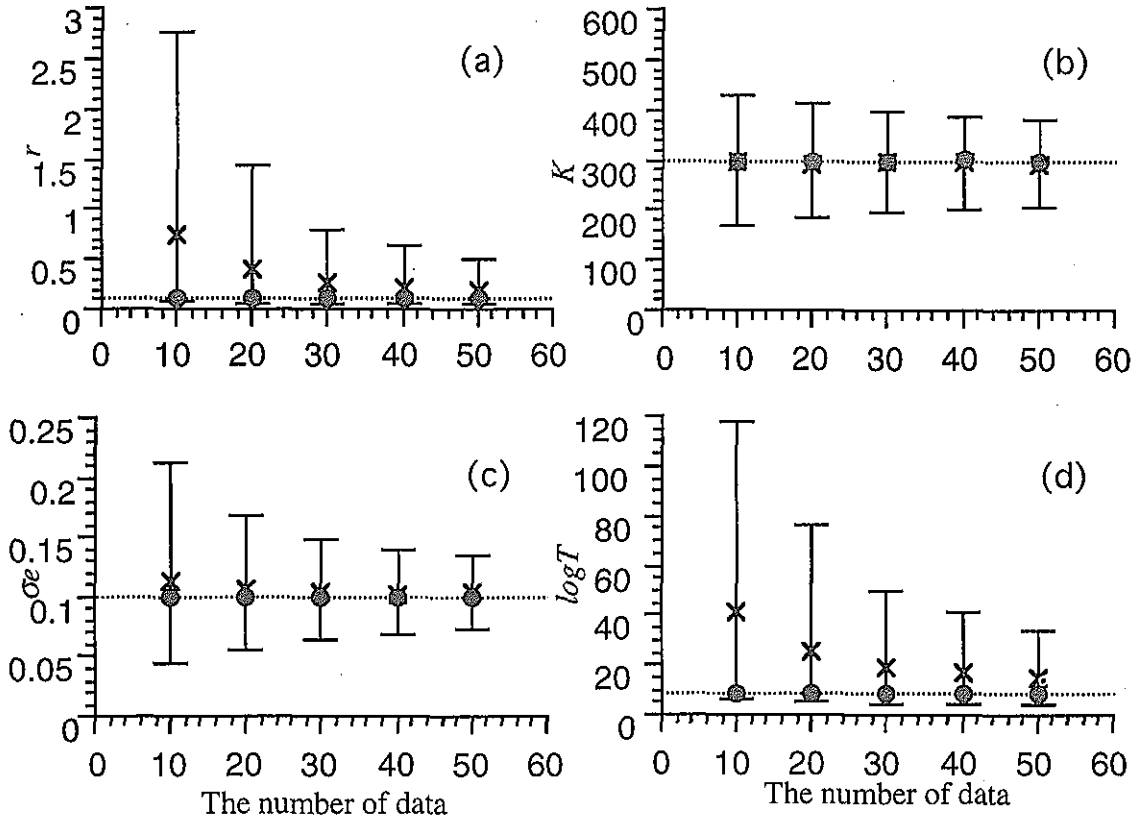


Fig. 2 The relationship between the number of data length ($n_{\text{time series}}$), and the bootstrap average (×) and 95% confidence intervals (bars) of (a) $r^*$, (b) $K^*$, (c) $\sigma_e^*$ and (d) $\log T^*$. We set $r = 0.1$, $K = 300$ and $\sigma_e = 0.1$. Dot lines indicate the true values. $n_{\text{computer-generate}} = 5000$. We calculate $\log T^*$ using by an approximate regression formula (Appendix). Black circles (•) represent the adjusted expectation of parameters calculated from the bootstrap averages (×). See text for detail.

### 3.3 *Adjusted estimation based on bootstrapping*

As shown in 3.1 and 3.2 sections, the maximum likelihood estimate of three parameters is clearly misleading in short time series data. Since available time series data of natural population sizes must be short, we need a better estimate method to improve the maximum likelihood estimate.

Here we develop a simple method based on bootstrap computer simulation. First, we suppose that there exists monotone-increasing transformations between a maximum likelihood estimate and a bootstrap average (see Efron 1987):

$$\hat{r} = g_r\left(E[\hat{r}^*]\right), \hat{K} = g_K\left(E[\hat{K}^*]\right) \text{ and } \hat{\sigma}_e = g_{\sigma_e}\left(E[\hat{\sigma}_e^*]\right), \tag{5}$$

where $g_r$, $g_K$ and $g_{\sigma_e}$ are some monotone transformations. In this case, the unbiased estimation of parameters is:

$$r = g_r(\hat{r}),\ K = g_K\left(\hat{K}\right) \text{ and } \sigma_e = g_{\sigma_e}\left(\hat{\sigma}_e\right). \tag{6}$$

If $E\left[\hat{r}^*\right]$ is larger than $\hat{r}$, the unbiased estimation of $r$ is smaller than $\hat{r}$, and if $E\left[\hat{r}^*\right]$ is smaller than $\hat{r}$, the unbiased estimation of $r$ is larger than $\hat{r}$. Parameters $K$ and $\sigma_e$ also have similar nature. Therefore, the approximate bias-corrected estimation of parameters ($\hat{r}_i$, $\hat{K}_i$ and $\hat{\sigma}_{e,i}$) can be obtained by repeat calculations of the bootstrap average of three parameters:

$$\hat{r}_i = \frac{\hat{r}_1\hat{r}_{i-1}}{E\left[\hat{r}_{i-1}{}^*\right]},\ \hat{K}_i = \frac{\hat{K}_1\hat{K}_{i-1}}{E\left[\hat{K}_{i-1}{}^*\right]} \text{ and } \hat{\sigma}_{e,i} = \frac{\hat{\sigma}_{e,1}\hat{\sigma}_{e,i-1}}{E\left[\hat{\sigma}_{e,i-1}{}^*\right]}, \qquad (i = 2, 3, 4,\ldots), \tag{7}$$

where $\hat{r}_1 = \hat{r}$, $\hat{K}_1 = \hat{K}$ and $\hat{\sigma}_{e,1} = \hat{\sigma}_e$. We do not need to know the monotone transformations, $g$, but the repeat calculations provide the better estimation automatically. For each computer-generated $\hat{r}^*$, $\hat{K}^*$ and $\hat{\sigma}_e{}^*$, we can also calculate the approximate bias-corrected estimation, then we can obtain better bootstrap confidence intervals. In Fig. 2, we show the approximate bias-corrected estimations for the bootstrap average of three parameters; $E\left[\hat{r}^*\right]$, $E\left[\hat{K}^*\right]$ and $E\left[\hat{\sigma}_e{}^*\right]$. The estimations (black circles, •) are close to true values (Fig. 2). The number of repeat calculation $i$ is 19. Table 1 shows an example of a convergence from $\hat{r}_1$, $\hat{K}_1$ and $\hat{\sigma}_{e,1}$ (= $E\left[\hat{r}^*\right]$, $E\left[\hat{K}^*\right]$ and $E\left[\hat{\sigma}_e{}^*\right]$, respectively) to $\hat{r}_{19}$, $\hat{K}_{19}$ and $\hat{\sigma}_{e,19}$.

Note that not only the estimation of $r$ and $\sigma_e$, but also that of $K$ is improved by the bias-correct method (Table 1). The estimation of $K$ have systematic bias (underestimate) because of the approximation in maximum likelihood functions (Eq. 3 and 4). Namely, we may improve the systematic biases from approximations in maximum likelihood functions, and exclude the constraints of the assumptions.

Table 1. A convergence from $\hat{r}_1$, $\hat{K}_1$ and $\hat{\sigma}_{e,1}$ to $\hat{r}_{19}$, $\hat{K}_{19}$ and $\hat{\sigma}_{e,19}$.

$n_{\text{computer-generate}}$ is 2000, but the time series with $\sum x_{j+1}x_j \leq 0$ or went extinct are not used for calculations. $r = 0.1$, $K = 300$ and $\sigma_e = 0.1$. $n_{\text{time series}} = 10$.

| $i$ | $\hat{r}_i$ | $\hat{K}_i$ | $\hat{\sigma}_{e,i}$ | $E[\hat{r}_i^*]$ | $E[\hat{K}_i^*]$ | $E[\hat{\sigma}_{e,i}^*]$ | $n$ |
|---|---|---|---|---|---|---|---|
| 1 | .739846 | 296.173 | .112679 | 1.305251 | 295.423 | .117219 | 1481 |
| 2 | .419361 | 296.925 | .108314 | 1.027982 | 294.817 | .116671 | 1664 |
| 3 | .301817 | 298.291 | .104607 | .952832 | 296.505 | .116346 | 1726 |
| 4 | .234352 | 297.957 | .101309 | .889183 | 295.367 | .111481 | 1749 |
| 5 | .194993 | 298.770 | .102398 | .805724 | 299.214 | .114864 | 1776 |
| 6 | .179050 | 295.733 | .100450 | .835005 | 293.990 | .112219 | 1763 |
| 7 | .158645 | 297.928 | .100861 | .824804 | 296.422 | .113491 | 1752 |
| 8 | .142304 | 297.678 | .100140 | .780895 | 294.918 | .112408 | 1758 |
| 9 | .134823 | 298.945 | .100381 | .773165 | 297.003 | .110975 | 1750 |
| 10 | .129013 | 298.109 | .101922 | .777683 | 296.152 | .113411 | 1773 |
| 11 | .122736 | 298.130 | .101264 | .781341 | 296.952 | .113916 | 1775 |
| 12 | .116218 | 297.348 | .100164 | .741014 | 292.382 | .111519 | 1777 |
| 13 | .116035 | 301.203 | .101205 | .748730 | 295.564 | .113768 | 1810 |
| 14 | .114658 | 301.823 | .100236 | .779885 | 297.591 | .113748 | 1763 |
| 15 | .108771 | 300.385 | .099294 | .744181 | 299.008 | .111352 | 1783 |
| 16 | .108138 | 297.537 | .100477 | .782080 | 294.154 | .114339 | 1784 |
| 17 | .102298 | 299.579 | .099018 | .746287 | 296.023 | .111815 | 1805 |
| 18 | .101415 | 299.731 | .099783 | .731621 | 297.676 | .112360 | 1762 |
| 19 | .102555 | 298.217 | .100066 | .729842 | 297.649 | .112248 | 1808 |

## 4. Model Aggregation

### 4.1 *Harvest models*

We can use the estimate method from time series data of population size to aggregate a complex model to the canonical model. As complex models, we examine three harvest models:

$$\frac{dx}{dt} = rx\left(1-\frac{x}{K}\right) + \sigma_e\xi_e(t)\circ x + \xi_d(t)\bullet\sqrt{x} - y(x) \tag{8}$$

where $y(x) = a$(constant harvest model), $-bx$ (proportional harvest model) and $\begin{cases} 0 \text{ for } x \leq c \\ \infty \text{ for } x > c \end{cases}$ (threshold harvest model) (see Lande et al. 1995).

First, we generate a sample path of population size using a simulation model that is parallel to each harvest model ($n_{\text{time series}}$ = 1000). Second, we estimated three parameters, $\hat{r}$, $\hat{K}$ and $\hat{\sigma}_e$ by fitting to the sampled population fluctuations to the canonical model, and calculated the mean extinction time log$T_k$ using Eq. (2). Finally, we compared it with true log$T_k$ of harvest models obtained from Eq. 9, 10 and 11.

### 4.2 *Constant harvest model*

The mean extinction time of constant harvest model is:

$$T_{K,a} = \frac{2}{\sigma_e^2}\int_0^K\int_0^\infty e^{-R(y-x)}\left(\frac{y+D}{x+D}\right)^{R(K+D)+2a+1}\left(\frac{x}{y}\right)^{2a}\frac{1}{(y+D)y}dydx, \qquad (9)$$

The estimate method based on the canonical model overestimates the true log$T_k$ of constant harvest model (Fig. 3), but there is a significant positive correlation between the estimate based on the canonical model and the true log$T_k$ ($r^2$ = 0.915, $n$ = 11, $p$ < 0.0001).

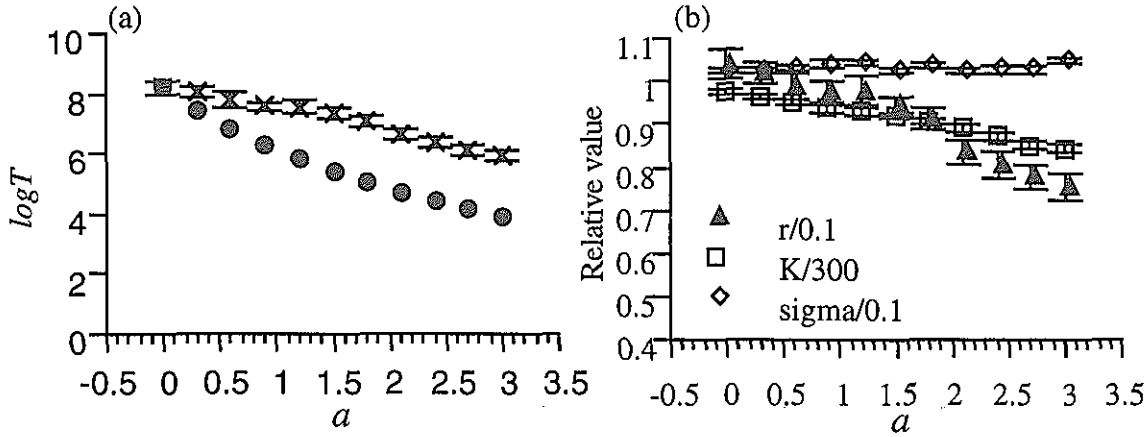


Fig. 3 Model aggregation of constant harvest model. (a) The relationship between harvest rate $a$ and log$T$. (b) The relationship between harvest rate $a$ and three parameters estimated. Black circles show the true value of log$T$ calculated by Eq (9). × and bars are bootstrap average ± SE ($n$ = 30) (log$T$ estimated). $n_{\text{time series}}$ = 1000. $r$ = 0.1 $K$ = 300 and $\sigma_e$ = 0.1.

### 4.3 *Proportional harvest model*

Proportional harvest model is similar to a model when a population is exposed to toxic chemical substances in the environment, the effect causes a constant decrease in the survival rate per generation (Hakoyama and Iwasa 1998). Because the mean extinction time of proportional harvest model is:

$$T_k\left(r-b, K-\frac{bK}{r}, {\sigma_e}^2\right), \tag{10}$$

where $T_K$ is the formula obtained for the canonical model (Eq. 2). Therefore, model aggregation is perfect (Fig. 4).

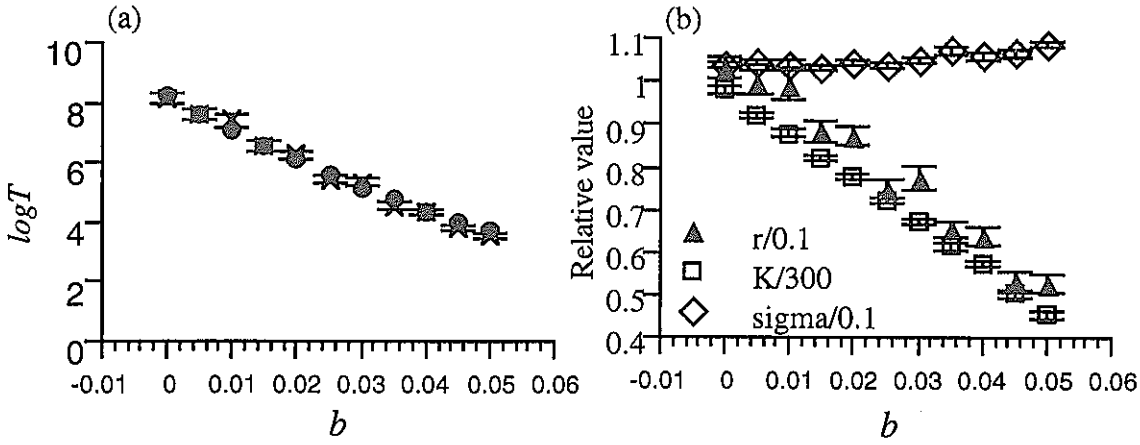


Fig. 4 Model aggregation of proportional harvest model. (a) The relationship between harvest rate $b$ and log$T$. (b) The relationship between harvest rate $b$ and three parameters estimated. Black circles show the true value of log$T$. × and bars are bootstrap average ± SE ($n$ = 30) (log$T$ estimated). $n_{\text{time series}}$ = 1000. $r$ = 0.1 $K$ = 300 and $\sigma_e$ = 0.1.

### 4.4 *Threshold harvest model*

The mean extinction time of threshold harvest model is:

$$T_K = \frac{2}{{\sigma_e}^2}\int_0^K\int_0^c e^{-R(y-x)}\left(\frac{y+D}{x+D}\right)^{R(K+D)+1}\frac{1}{(y+D)y}dydx, \tag{11}$$

The estimate method based on the canonical model quite overestimates the true log$T_k$ of threshold harvest model (Fig. 3), and there is no significant positive correlation between the estimate based on the canonical model and the true log$T_k$. The estimated value of log$T_k$ have local maximum around $c$ = 200 (Fig. 5a). Note that the estimated value of $r$ have local maximum around $c$ = 150 ($K$ / 2) (Fig. 5b).

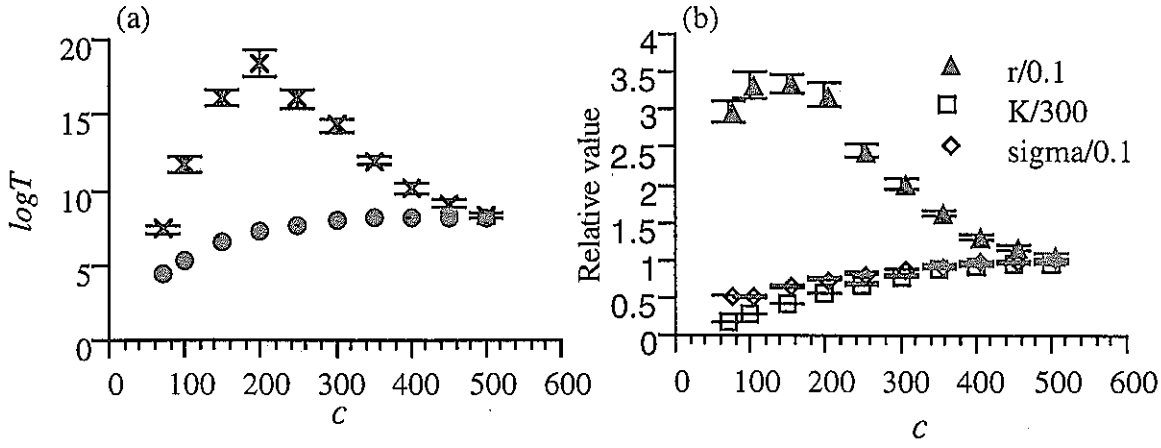


Fig. 5 Model aggregation of threshold harvest model. (a) The relationship between threshold $c$ and log$T$. (b) The relationship between threshold $c$ and three parameters estimated. Black circles show the true value of log$T$ calculated by Eq (11). × and bar are bootstrap average ± SE ($n$ = 30) (log$T$ estimated). $n_{\text{time series}}$ = 1000. $r$ = 0.1 $K$ = 300 and $\sigma_e$ = 0.1.

## 5. Appendix

### 5.1 *Regression formula*

It takes a long time to calculate log$T$ using by the integral formula (Eq. 2), Then, we made a regression formula to calculate approximate log$T$. We calculate some log$T$ for some parameter sets (combination of following parameters; $r$ = 0.1, $K$ = {1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3}, $\sigma_e^2$ = {.00001, .0001, .0005} and {.001, .002, ..., .098, .099, .1}), and using a non-linear regression method and a parameter scaling rule, derived an empirical equation:

$$\ln T_{regression} = -\ln\left(\frac{r}{0.1}\right) + \left(1.12073\left(\frac{0.1\sigma_e^2}{r}\right)^{0.318121} - 0.0267559\right)\ln\left(\frac{rK}{0.1}\right)^{\left(-9.70471\left(\frac{0.1\sigma_e^2}{r}\right)^{0.176337}+8.07769\right)} + \left(-1.93776\left(\frac{0.1\sigma_e^2}{r}\right)^{0.113793} + 2.56977\right). \quad \text{(A1)}$$

This formula is a good approximation of Eq. 2, when parameters range in the above parameter sets that was used Eq. A1 or the corresponding scaled parameter range (see Hakoyama and Iwasa 1999 for detail).

## 6. Acknowledgment

This work has been supported by CREST (Core Research for Evolutional Science and Technology) of Japan Science and Technology Corporation (JST). Hokkaido National Fisheries Research Institute have also supported this work partially.

## 7. References

Dennis, B. and Taper, M.L. 1994. Density dependence in time series observations of natural populations: estimation and testing. Ecological Monographs, 64: 205-224.

Efron, B. 1987. Better bootstrap confidence intervals. Journal of American Statistical Association, Vol. 82, No. 397:171-185.

Hakoyama, H. and Iwasa, Y. 1999. Extinction risk of density-dependence populations, estimated from fluctuating population size. (preparing)

Hakoyama, H. and Iwasa, Y. 1998. Ecological risk assessment: a new method of extinction risk assessment and its application to a freshwater fish (*Carassius auratus* subsp.) Proceeding of the International Workshop on Ecological Risk, (org. J. Nakanishi) Yokohama, Japan: 93-110.

Iwasa, Y. 1998. Ecological risk assessment by the use of the probability of species extinction. Proceeding of the International Workshop on Ecological Risk, (org. J. Nakanishi) Yokohama, Japan: 42-49.

Lande, R., Engen, S. and Saether, B-E. 1995. Optimal harvesting of fluctuating populations with a risk of extinction. the American Naturalist, 145:728-745.

Nakanishi, J. 1995. Environmental risk theory. Iwanami Publ. Com., Tokyo (In Japanese)