Appl. Entomol. Zool. 33 (3): 339-347 (1998)

Probit analysis of preference data

Masayuki Sakuma

Laboratory of Insect Physiology, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Japan

(Received 21 October 1996; Accepted 5 March 1998)

Abstract

A probit analysis method is presented for the analysis of preference data from a behavioural assay of animals, where responses were observed as animals' choice between the treatment and control sides or areas. Standard probit models designed for *all or nothing* phenomena are not applicable to *preference* assays, because a log dose metameter only indirectly correlates with the choice. The present method enabled analysis by regressing the probit transformation of the proportion of net responders, expressed in a preference model equation, on a log dose metameter. The computer program includes the maximization routine of the log-likelihood function, a test of homogeneity and a calculation of potency with fiducial limits. After the explanation of the basic model, as well as generalized regression models including the parameters of natural preference and immunity, a computer output demonstrated the analysis of an olfactometer assay on two cockroach attractants by reporting statistics. By comparing differently generalized models for fitness to the data, a preferable experimental design was formulated.

Key words: Attractant, bioassay, ED_{50}, preference, probit analysis

INTRODUCTION

Recent findings in the study of animal behaviour have invoked various types of bioassays which extract and quantify the target behavioural responses frequently expressed as quantal data. As for *all or nothing* type phenomena such as the presence or absence of a response, data accumulated in a graded series of stimulus levels can be directly analysed by a standard probit analysis or logistic regression (Bliss, 1934; Finney, 1971, 1978; Dobson, 1990), whereas *preference* data and directional choice still remained to be analysed.

Preference assays may be represented by a T-maze olfactometer test of attractants (e.g., Sakuma and Fukami, 1985), where test animals introduced from a stem tube are forced to choose between odour-containing and control branches according to directional choice. Assuming that non-responders approach both the treatment and control sides equally, the proportion of insects in the treatment side may increase from 0.5 to 1.0 by upgrading the dosage or stimulus level of the attractant. These proportions, though apparently different from those of the *all or nothing*, could be subjected to a standard probit analysis if they were converted to the proportion of net responders extending between 0.0 and 1.0. This basic model may be further generalized by considering the *natural preference* caused by laterally unequal conditions at zero dose and the *natural immunity* of test animals incapable of responding even at very high doses.

These ideas could be generally applicable to the dilution assays of attractants. This paper deals with the theory, application and experimental design for the probit analysis of preference data. After illustrating the regression models incorporating the proportion of net responders, I demonstrate analytical procedures for preference data from olfactometer assays. Then, based on a comparison of the fitness between differently generalized models, an experimental plan for the application of probit analysis to preference data is discussed.

MATERIALS AND METHODS

Probit analysis of preference data.

A basic preference model for regression. Preference assay results are usually obtained as the numbers of test animals or their repetitive choices in both treatment (r) and untreated

control (u) sides. Since the same numbers of non-responders as in the control side may be included in the treatment side, the sample proportion of net responders from experiment (p) can be obtained by

$$p=\frac{r-u}{r+u}=\frac{r-u}{n},$$

where $n=r+u$ is the total number of test animals. This quantity was termed as an 'excess percentage index' (Bentley, 1944), 'aggregation index' (Roth and Cohen, 1973) or 'excess proportion index' (Sakuma and Fukami, 1985). The sample proportion of net responders can be rewritten with the sample proportion of treatment or control side animals (p^* or q^*) as

$$p=\frac{2r}{n}-1=2p^*-1=1-2q^*,$$

where $p^*=r/n$ and $q^*=1-p^*=u/n$. Therefore, the population proportions of choices in treatment (P^*) and control sides (Q^*) can be expressed in terms of the population proportion of net responders (P) and non-responders ($Q=1-P$) as

$$P^*=\frac{1+P}{2}=P+\frac{Q}{2},\quad Q^*=\frac{1-P}{2}=\frac{Q}{2}. \qquad (1)$$

Since n test animals react independently, the probability that r individuals exactly choose treatment side is stated by the binomial distribution with the probability function

$$\Pr(r)=\frac{n!}{r!(n-r)!}P^{*r}Q^{*(n-r)}.$$

For an experiment comprising multiple stimulus levels, in which r_i individuals of n_i test animals choose the treatment side in the i-th stimulus level, the joint probability of these results is the product of their probability functions which is called the likelihood function (l) (Fisher, 1922):

$$l=\prod_i \Pr(r_i)$$

Therefore, the log-likelihood function (L) is

$$L=\ln(l)=\sum_i \ln \Pr(r_i)$$

$$=\text{constant}+\sum_i r_i \ln P_i^*+\sum_i (n_i-r_i)\ln Q_i^*, \qquad (2)$$

and the maximum likelihood estimate of any parameter of the distribution of individual preference (θ) is the solution of the equation $\partial L/\partial\theta=0$, that is,

$$\sum_i \frac{r_i}{P_i^*}\frac{\partial P_i^*}{\partial\theta}+\sum_i \frac{(n_i-r_i)}{Q_i^*}\frac{\partial Q_i^*}{\partial\theta}$$

$$=\sum_i \frac{n_i(p_i^*-P_i^*)}{P_i^*Q_i^*}\frac{\partial P_i^*}{\partial\theta}=0 \qquad (3)$$

where $p_i^*=r_i/n_i$ and a suffix i of each parameter corresponds to the i-th simulus level. When preference assays are conducted on a graded series of directional stimuli, the probit analysis would be applicable to the net response. If the distribution of net responders P is expressed by the normal distribution, a linear relation between the logarithmic transform of stimulus intensity (x) and the probit transform of $P(Y)$ can be obtained by a linear regression

$$Y=\alpha+\beta x, \qquad (4)$$

where α and β are parameters representing the intercept and the regression coefficient, respectively, and the ordinate to the normal curve at point Y is

$$Z=\frac{\partial P}{\partial Y}=\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}Y^2}. \qquad (5)$$

To obtain the maximum likelihood estimators of parameters α and β a set of Eq. (3) must be solved simultaneously by numerical iteration (Finney, 1971), which is conducted with a maximization routine until the increments of α and β, become negligible (Appendix). The approximate variances and covariances of a and b, i.e. the estimators of respective parameters α and β, can be obtained as the elements of the inverse Hessian matrix ((A2) in Appendix) evaluated at convergence.

Generalization. The preference assay discussed above has assumed that non-responders (their proportion: $Q=1-P$) choose both treatment and control sides equally (Fig. 1a); however, the choice of non-responders could be biased by conditional unbalance other than the target directional cue. If the natural preference, i.e. the proportion of animals choosing the treatment side at zero dose, is C (Fig. 1b), the population proportion of choices in treatment side (P^*) and control side (Q^*) can be expressed by the proportion of responders (P) and C as

$$P^*=CQ+P=C+(1-C)P,$$
$$Q^*=(1-C)Q=1-C-(1-C)P. \qquad (6)$$

If C is already known experimentally or systematically, the parameter (C) may be replaced with a constant (C_0). Thus, in an idealistic 2-

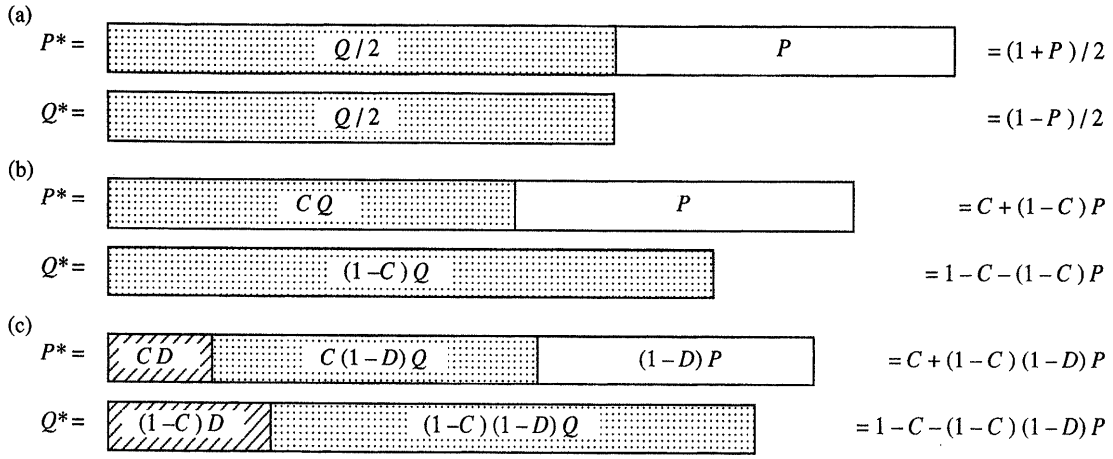

Fig. 1. Models for preference data: a basic model (a), general model for the biased choice of non-responders (b) and general model for both biased choice and natural immunity (c). The population proportions of animals in the treatment (P^*) and control sides ($Q^*=1-P^*$) are expressed by those of responders (P) and non-responders ($Q=1-P$), natural preference to the treatment side (C) and natural immunity (D).

choice system where C_0 is supposed to be 1/2, Eq. (6) is reduced to a basic model (1).

Sometimes a population includes animals that do not respond even at a dose sufficient to induce responses. If a proportion D of animals is considered as natural immunity or non-responders at very high doses as in Abbot's formula (Finney, 1971, 1978), P^* and Q^* may be illustrated as in Fig. 1c, and then extended from (6) to

$$\begin{aligned} P^* &= C\{D+(1-D)Q\}+(1-D)P \\ &= C+(1-C)(1-D)P, \\ Q^* &= (1-C)\{D+(1-D)Q\} \\ &= 1-C-(1-C)(1-D)P. \end{aligned} \quad (7)$$

Here, both population proportions of net responders (P) and non-responders (Q) bear a relation not to the total, as they did before, but to the natural potential ($1-D$), while the natural immunity (D) may also be divided between treatment and control sides according to the natural preference (C). Indeed the natural immunity (D) can be replaced with a constant (D_0), though it may be rarely obtainable before the analysis except in the case of a complete response ($D_0=0$). In contrast, a natural preference (C) can be replaced with a predictable constant (C_0), as in Eq. (6).

If D is high and/or C clearly deviates from the expected value, an estimation of C and/or D is required together with α and β. The same optimization routine as described above is applicable, except that P^* and Q^* in Eq. (3) are replaced by those given by (7) to obtain the estimators of C ($\hat{C}$) and D ($\hat{D}$), as well as a and b. Either of these may be replaced with a constant C_0 or D_0, if it is already known. The matrix of variances and covariances for Eq. (7), in the fully generalized condition, is presented as (A4) in the Appendix.

Computer program. The computer program employed in this paper incorporated a maximization algorithm: either the *variable metric method* (*quasi-Newton method*) (Fletcher and Powell, 1963) or the *downhill simplex method* (Nelder and Mead, 1965) which are listed by Press et al. (1992). Statistical functions appearing in Kurose et al. (1986) were also used. Since these maximization methods work directly on the log-likelihood function (2), the exact values of the variances and covariances have to be calculated by solving the inverse Hessian matrix ((A2), (A4) in Appendix) with respect to the final approximates by means of an LU decomposition and back-substitution routine; therefore, the weighting coefficient and other auxiliary parameters ((A1), (A3) in Appendix) are used in the calculation of the matrix elements.

The program begins by determining provisional values for the parameters by unweighted linear regression. Then the maximization routine optimizes parameters from the provisional values, and the asymptotic variances and covariances are calculated with respect to the final approximates. The latter part

of the program includes the validity check and the potency estimation with its fiducial limits after Finney (1978).

The total system of the program was designed to deal with not only *preference*, but also *all or nothing* type data. One can choose any combination of models from the differently generalized formulae and statistical distributions of normal, logistic and extreme value (complementary-log-log) models (Dobson, 1990). The program is written in the ANSI version of C language, and the present analysis was performed by an object compiled by a Think C compiler (Symantec, U.S.A.) running on a Macintosh personal computer (Apple Computer, U.S.A.).

The fitness of models. The fitness of models to the assay results was compared by evaluating Akaike's information criteria (AIC) expressed by

$$\mathrm{AIC} = -2\hat{L} + 2m, \qquad (8)$$

where $\hat{L}$ and m are the maximum log-likelihood and the number of parameters, respectively (Akaike, 1976).

RESULTS

Olfactometer assay of a cockroach attractant pheromone

The preference data to be analysed were obtained by an olfactometer assay specially designed for an aggregation attractant pheromone of the German cockroach (Sakuma and Fukami, 1985, 1990a, b). In the olfactometer, test animals climbed up the stem of a T-maze wire, sustained in a T-tubing, and chose the direction of either the odour-containing or odourless side at the T-junction of the wire. The insects were eventually caught in a trap attached to each end of the wire, and then the attractiveness or *chemotaxis* response was recorded as the excess proportion of animals in the treatment side to the control side.

Constitution of data sets

Each dose group consists of a log dose (x), the number of animals (n) and responses (r). Here a dose was measured as the amount of chemicals in μmol/min continuously applied onto a dispenser. The total number (n) is the sum of both trap catches, whereas the number of responses (r) is not that of net responders, but that of the trap catch in the treatment side. The data set included the zero dose data of the control experiments, where n and r are also accumulated into L in Eq. (2) by assuming no net responders at zero dose: $P=0$ and $Q=1$.

Parallel line assay of 1-DMA-2-M-2-P and 2-DMA-2-M-1-P

An example presented here is a parallel line assay on the attractiveness of the pheromone component, 1-dimethylamino-2-methyl-2-propanol (1-DMA-2-M-2-P) and its non-component isomer, 2-dimethylamino-2-methyl-1-propanol (2-DMA-2-M-1-P). The probit analysis of the assay was conducted on a fully generalized preference model (7) in order to cover all the details.

Table 1 shows the history of analysis. The suffix of each estimator represents the corresponding preparation: 1 and 2 denote 1-DMA-2-M-2-P and 2-DMA-2-M-1-P respectively. After the model was specified, the program began by regressing data sets simultaneously with independent intercepts (a_1, a_2) and regression coefficients (b_1, b_2), as well as a common natural immunity ($\hat{D}$) and preference to the treatment side ($\hat{C}$). The maximization routine, here variable metric method, optimized the estimates by initiating from provisional values obtained by unweighted regression. After 16 iterations, the maximum log-likelihood was optimized at $\hat{L}=-4.153592$, which was completely identical to that presented by the downhill simplex method after 661 iterations. Equations corresponding to (4) are taken from the estimates of parameters,

$$Y_1 = 11.0315 + 2.0281x,$$
$$Y_2 = 5.2728 + 2.0187x.$$

$\hat{C}$ and $\hat{D}$ were directly optimized as $\hat{C}=0.4735\pm0.0189$ (S.E.) and $\hat{D}=0.1603\pm0.0202$ (S.E.), that is, 47.36% of the animals naturally chose the treatment side and 16.03% of them were not attracted even at very high doses.

Then a parallel line regression on the same data sets, with the common regression coefficient (b), was used to estimate $\hat{L}=-4.153812$ and the parameters after 12 iterations. Variances and covariances of the

Table 1. Data output from the computer program applied to a 2-choice olfactometer assay on two cockroach attractants

Type of Model (Activated Parameters)
1: Normal 8 : Preference (a, b, C: Natural Preference, D: Immunity)
Configuration
2 Preparations
15 Dose Groups, of which: 7 refer to 1-DMAM-2-M-2-P
8 refer to 2-DMAM-2-M-1-P

Optimization Algorithm: Davidon-Fletcher-Powell (variable metric method)

The Maximum Log-Likelihood: L = -4.153592 iterations : 16

Parameters Estimates from Independent Regression

#	Parameter	Std. Error	Description
a1	11.031517	2.132521	1-DMAM-2-M-2-P
b1	2.028141	0.398387	Regression Coefficient
a2	5.272789	1.017850	2-DMAM-2-M-1-P
b2	2.018697	0.401564	Regression Coefficient
C	0.473543	0.018859	Natural Preference
D	0.160322	0.020160	Immunity

Observed and Expected Frequencies
1. 1-DMAM-2-M-2-P

#	Dose	Log(Dose)	n	Obs.r	Exp.r	Dev.	Obs.*p*	Exp.*p*
1	0.000e+00	-	204	105	96.60	8.40	0.5147	0.4735
2	1.000e-06	-6.0000	206	104	109.18	-5.18	0.5049	0.5300
3	3.162e-06	-5.5000	211	147	141.98	5.02	0.6967	0.6729
4	1.000e-05	-5.0000	219	180	182.46	-2.46	0.8219	0.8331
5	3.162e-05	-4.5000	203	183	183.32	-0.32	0.9015	0.9030
6	1.000e-04	-4.0000	214	200	195.77	4.23	0.9346	0.9148
7	3.162e-04	-3.5000	205	183	187.69	-4.69	0.8927	0.9156

2. 2-DMAM-2-M-1-P

#	Dose	Log(Dose)	n	Obs.r	Exp.r	Dev.	Obs.*p*	Exp.*p*
1	0.000e+00	-	210	98	99.44	-1.44	0.4667	0.4735
2	1.000e-04	-4.0000	202	95	95.88	-0.88	0.4703	0.4747
3	3.162e-04	-3.5000	202	92	98.92	-6.92	0.4554	0.4897
4	1.000e-03	-3.0000	208	123	118.42	4.58	0.5913	0.5693
5	3.162e-03	-2.5000	206	150	151.22	-1.22	0.7282	0.7341
6	1.000e-02	-2.0000	207	178	179.61	-1.61	0.8599	0.8677
7	3.162e-02	-1.5000	220	205	200.23	4.77	0.9318	0.9101
8	1.000e-01	-1.0000	211	191	193.14	-2.14	0.9052	0.9153

The Maximum Log-Likelihood: L = -4.153812 iterations : 12

Parameters Estimates from Parallel line Regression

#	Parameter	Std. Error	Description
a1	11.003897	1.664522	1-DMA-2-M-2-P
a2	5.283530	0.793836	2-DMA-2-M-1-P
b	2.023012	0.311144	Common Regression Coefficient
C	0.473562	0.018807	Natural Preference
D	0.160306	0.020154	Immunity

Matrix of Variances, Covariances, and (Correlations)

	a1	a2	b	C	D
a1	2.770635	(0.982017)	(0.996005)	(0.405891)	(0.548421)
a2	1.297595	0.630175	(0.981474)	(0.358782)	(0.548423)
b	0.515836	0.242421	0.096810	(0.439921)	(0.533737)
C	0.012707	0.005357	0.002574	0.000354	(0.376384)
D	0.018397	0.008934	0.003347	0.000143	0.000406

Observed and Expected Frequencies
1. 1-DMAM-2-M-2-P

#	Dose	Log(Dose)	n	Obs.r	Exp.r	Dev.	Obs.*p*	Exp.*p*
1	0.000e+00	-	204	105	96.61	8.39	0.5147	0.4736
2	1.000e-06	-6.0000	206	104	109.24	-5.24	0.5049	0.5303
3	3.162e-06	-5.5000	211	147	142.00	5.00	0.6967	0.6730
4	1.000e-05	-5.0000	219	180	182.41	-2.41	0.8219	0.8329
5	3.162e-05	-4.5000	203	183	183.29	-0.29	0.9015	0.9029
6	1.000e-04	-4.0000	214	200	195.77	4.23	0.9346	0.9148
7	3.162e-04	-3.5000	205	183	187.70	-4.70	0.8927	0.9156

2. 2-DMAM-2-M-1-P

#	Dose	Log(Dose)	n	Obs.r	Exp.r	Dev.	Obs.*p*	Exp.*p*
1	0.000e+00	-	210	98	99.45	-1.45	0.4667	0.4736
2	1.000e-04	-4.0000	202	95	95.88	-0.88	0.4703	0.4747
3	3.162e-04	-3.5000	202	92	98.89	-6.89	0.4554	0.4895
4	1.000e-03	-3.0000	208	123	118.37	4.63	0.5913	0.5691
5	3.162e-03	-2.5000	206	150	151.23	-1.23	0.7282	0.7341
6	1.000e-02	-2.0000	207	178	179.65	-1.65	0.8599	0.8679
7	3.162e-02	-1.5000	220	205	200.24	4.76	0.9318	0.9102
8	1.000e-01	-1.0000	211	191	193.14	-2.14	0.9052	0.9154

Chi-Square Test

	DF	SS	MS	*p*
Parallelism	1	0.000439	0.000439	0.983279
Heterogeneity	9	8.307185	0.923021	0.503510
Total	10	8.307624	0.830762	0.598816

Since heterogeneity is small (p> 0.050), a percentage point of normal distribution (u) is used.

Normal or t-Deviates (Both Tails) and Index of Regression Significance, g
u (0.050)=1.959960 g (0.050)=0.090870

Asymptotic Variance for Log(ED50)

#	Preparation	Log(ED50)	Variance	Std.Error
1	1-DMA-2-M-2-P	-5.439363	0.004248	0.065180
2	2-DMA-2-M-1-P	-2.611715	0.004309	0.065646

Median Effective Dose Estimates

#	Preparation	ED50 (95% Fiducial Limits)	Log(ED50) (95% Fiducial Limits)
1	1-DMAM-2-P	3.6361e-06 (2.6102e-06, 4.8454e-06)	-5.43936 (-5.58333, -5.31467)
2	2-DMAM-1-P	2.4450e-03 (1.7558e-03, 3.2727e-03)	-2.61171 (-2.75552, -2.48510)

Asymptotic Variance for log(R)

#	Preparation	Log(R)	Variance	Std.Error
1	1-DMA-2-M-2-P	0.000000	0.000000	0.000000
2	2-DMA-2-M-1-P	-2.827649	0.008166	0.090364

Relative Potency (R) Estimates

#	Preparation	R (95% Fiducial Limits)	Log(R) (95% Fiducial Limits)
1	1-DMA-2-M-2-P	1.0000e+00 (1.0000e+00, 1.0000e+00)	0.00000 (0.00000, 0.00000)
2	2-DMA-2-M-1-P	1.4871e-03 (9.6729e-04, 2.2754e-03)	-2.82765 (-3.01444, -2.64294)

Natural Preference and Immunity Estimates

#	Description	Proportion (95% Fiducial Limits)
C	Natural Preference	0.473562 (0.436701, 0.510424)
C/(1-C)	Natural Bias	0.899560 (1.042583, 0.775255)
D	Natural Immunity	0.160306 (0.120806, 0.199806)

parameters are tabulated in a matrix. The equations for the parallel line assay are

$$Y_1 = 11.0039 + 2.0230x,$$
$$Y_2 = 5.2835 + 2.0230x,$$

and $\hat{C}$, $\hat{D}$ were almost identical to the above results as $\hat{C} = 0.4736 \pm 0.0188$ (S.E.) and $\hat{D} = 0.1603 \pm 0.0202$ (S.E.).

The model of P^* in Eq. (7) was suggested to be pertinent, because the deviations of the observed frequencies from the expected appeared to be randomly distributed. The total residuals of the regression, containing heterogeneity and 'parallelism' (meaning deviations from 'parallelism'), may be obtainable as a likelihood ratio, $-2\hat{L} = 8.307624$ equivalent to $\chi^2_{[10]}$. As the heterogeneity component, obtained from the regression with independent regression coefficients, was $\chi^2_{[9]} = 8.307185$, the remaining parallelism was $\chi^2_{[1]} = 0.000439$. The values were small enough to imply no worries. Indeed, this likelihood ratio test of homogeneity and parallelism can be replaced by a conventional χ^2 test on sum of squares of the deviations (Finney, 1971) giving practically the same results in heterogeneity and parallelism at $\chi^2_{[9]} = 8.225024$ and $\chi^2_{[1]} = 0.001492$, respectively.

Then the potencies and their fiducial limits were estimated. Since heterogeneity was small, a normal deviate ($t = 1.960$) was used in Fieller's theorem to calculate the 95% fiducial limits (Finney, 1971, 1978). The regression equations gave means of the net responder distributions (P) at $m_1 = -5.4394$ and $m_2 = -2.6117$, thus, the median effective doses (ED_{50}) for 1-DMA-2-M-2-P and 2-DMA-2-M-1-P were 3.636 (2.610, 4.845) pmol/min and 2.445 (1.756, 3.273) μmol/min respectively: 95% limits are in parentheses. The relative potency of 2-DMA-2-M-1-P to 1-DMA-2-M-2-P was estimated to be 0.001487 (0.0009673, 0.002275) suggesting an essential difference in activity between the two isomers.

Fitness of the models to the assay results

The assay results on 1-DMA-2-M-2-P and 2-DMA-2-M-1-P were analysed again with the general model (7) by fixing the natural preference (C) at $C_0 = 0.5$ and/or the natural immunity (D) at $D_0 = 0$. The fitness of the variably restricted models to the assay results can be expressed as Akaike's information criteria (AIC). Table 2 shows the estimates of parameters a_1, a_2, b, $\hat{C}$, $\hat{D}$ and/or constants C_0, D_0 as well as the number of parameters (m) and AIC. The best fitness represented by the least AIC was recorded by the fully generalized model with parameters C and D, and essentially the same AIC was observed in a model with $C_0 = 0.5$ and D. However, the other models including the basic one gave large AIC values.

DISCUSSION

The method described here enables the quantitative analysis of the dose-response relation

Table 2. Fitness of preference models[a] to the assay on two cockroach attractants[b]

Activated parameters and (constants)	Values of estimates					m[c]	AIC[d]
	a_1[e]	a_2[f]	b[g]	$\hat{C}$[h] (C_0[i])	$\hat{D}$[j] (D_0[k])		
a_1, a_2, β, ($C_0 = 0.5$, $D_0 = 0$)	4.6334	2.1045	0.9144	(0.5000)	(0.0000)	3	72.8395
a_1, a_2, β, C, ($D_0 = 0$)	4.4771	2.1038	0.8572	0.4471	(0.0000)	4	67.8522
a_1, a_2, β, ($C_0 = 0.5$), D	11.6850	5.5503	2.1664	(0.5000)	0.1698	4	18.5032
a_1, a_2, β, C, D	11.0039	5.2835	2.0230	0.4736	0.1603	5	18.3076

[a] The examined preference models were the general model (7) in the text as well as those derived from it by fixing a parameter(s) C and/or D as a constant(s).
[b] The same preparations as in Table 1.
[c] The number of activated parameters.
[d] Akaike's Information Criterion defined by Eq. (8).
[e, f] Intercepts of the regression formula corresponding to the preparations for 1-DMA-2-M-2-P and 2-DMA-2-M-1-P, respectively.
[g] Common regression coefficient in a parallel-line assay.
[h, i] Natural preference or its known constant in parentheses.
[j, k] Natural immunity or its known constant in parentheses.

in chemical *attractants* or even physical cues, which have been analysed qualitatively or semi-quantitatively. Although the idea of the basic model (1) has been briefly explained (Sakuma and Fukami, 1985), the generalization of the model and a theoretical description have not as yet been conducted. The basic model was already applied to the analysis of olfactometer assay results (Sakuma and Fukami, 1985, 1990a, b), where a symmetrical environment to non-responders and almost complete responses at higher doses were realized. Here the model was generalized with two more parameters, natural preference (C) and immunity (D), which extends its applicable area to the preference assays in general.

As shown in Table 2, when an appropriate model was employed, the fitness of the model to the experimental results was significantly improved. The models with immunity (D) exhibited good fitness, whereas the fitness of those assuming no immunity was unsatisfactory. The low regression coefficient (b) in the latter models may be derived from an underestimation of net responses. In contrast to the significance of the immunity, the employment of natural preference (C) did not improve the fitness dramatically, probably because the assay was conducted in a sufficiently balanced 2-choice condition.

This generalization of the model, however, may require additional effort in accumulating low and/or high dose data to cope with extra degrees of freedom. The investigator should, therefore, optimize the assay system so that a simpler model may become applicable. In an ordinal preference test, a balanced system can easily be realized, but the immunity is barely controllable. The immunity originates from the non-responsiveness of test insects to a directional cue, or the incapability of an assay design to discriminate the components of orientation behaviour. Both may be reduced to some extent, but are still inevitable. Consequently, the use of a model with an extra parameter D and a predictable constant C_0 should be adequate in many cases.

Since D, as well as C, contributes nothing to the estimation of potency, it should be rationalized. Sometimes, the dose-response relation was completely inverted at excessive dosages. Actually two attractants examined here induced an avoidance response to the cockroaches when they were supplied at doses over 1.0 mmol/min. In such compounded responses, dose groups should be carefully selected for the purpose of the assay. For the evaluation of the attractant activity, dose groups analysed here were exclusively contained within a sigmoidal increase and plateau of the dose-response curve. In this context, the estimates of immunity (16.03%) could include the animals suffering from repellent activity.

The present examples were performed on a large sample, which provides sufficient numbers of dose groups and animals to estimate C and D. When the number of test animals is limited and $C \approx 0.5$ and a small value for D is confirmed, the use of a 2-choice basic model is preferable. The probit analysis of a preference assay based on net response, however, still demands more samples than in the *all or nothing* type analysis. In order to obtain an ED_{50} value with a sufficiently small fiducial interval, the basic model requires at least 3 dose groups of 30 (preferably 50) animals each, and then the dose groups should be allocated to where the responses are between 40 and 60, around 75 and more than 90% of the animals.

Even if the sample size is too small to evaluate meaningful potencies in assays on single preparations, it often becomes possible in the parallel line assay involving several preparations: a hypothetical common regression coefficient lessens the mean variance and, thus, the fiducial intervals. Therefore, the comparison between the potencies of purified fractions (Sakuma and Fukami, 1990b; Sakuma et al., 1997a), and attractant analogues (Sakuma et al., 1997b), etc. may be possible in a relatively small sample size.

However, parallelism or equal variances between the preparations is prerequisite for a parallel line assay, as mentioned in the results section. When the homogeneity of an assay is not confirmed, the analysis of variances between the mean square of parallelism and that of heterogeneity (denoted as the *heterogeneity factor*) substitutes for the usual χ^2-analysis, and then the index of regression significance (g) and

the term of variance in Fieller's theorem are multiplied by the heterogeneity factor if this is thought appropriate (Finney, 1971). Not often in plain analyses using a basic model, but sometimes in precise one using a general model, problems may arise in parallelism. These are mainly arise from the difference in C and D between the preparations by independent analyses, which directly links to the different regression coefficients (b). If the difference can not be rationalized by the use of the same assay apparatus and animals in the same group, etc., the test attractants may be categorized in the separate groups of response mechanism.

The methods described here open a path to the quantitative analysis of preference data. Any strategies, including those discussed here, could be applicable according to the purpose of an experiment, and the computer program enables simple and precise data analysis providing all statistics necessary for complete reporting. The program, written in the ANSI version of C language, can be obtained from the author.

ACKNOWLEDGEMENTS

I express my sincere gratitude to Dr. S. Miyai of Forestry and Fisheries Research Council Secretariat, Ministry of Agriculture, Forestry and Fisheries, Japan for invaluable comments and critical readings of the manuscript.

REFERENCES

Akaike, H. (1976) What is Akaike's information criterion—its meaning and perspective. *Mathematical Sci.* (*Suri kagaku*) 153: 5-11 (in Japanese).

Bentley, E. W. (1944) The biology and behaviour of *Ptinus tectus* Boie (Coleoptera, Ptinidae), a pest of stored products. V. Humidity reactions. *J. Exp. Biol.* 20: 152-159.

Bliss, C. I. (1934) The method of probits. *Science* 79: 38-39.

Dobson, A. J. (1990) *An Introduction to Generalized Linear Models*. Chapman and Hall, London. 174 pp.

Finney, D. J. (1971) *Probit Analysis*, 3rd ed. Cambridge University Press, Cambridge. 333 pp.

Finney, D. J. (1978) *Statistical Method in Biological Assay*, 3rd ed. Charles Griffin & Company Ltd., London. 508 pp.

Fisher, R. A. (1922) On the mathematical foundation of theoretical statistics. *Philos. Trans. A* 222: 309-368.

Fletcher, R. and M. J. D. Powell (1963) A rapidly convergent descent method for minimization. *Computer J.* 6: 163-168.

Kurose, Y., I. Matsushima and T. Matsuo (1986) *Scientific Subroutine Library in C*. Keigaku Shuppan Co., Tokyo. 278 pp. (in Japanese).

Nelder, J. A. and R. Mead (1965) A simplex method for function minimization. *Computer J.* 7: 308-313.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (1992) *Numerical Recipes in C*, 2nd ed. Cambridge University Press, Cambridge. 994 pp.

Roth, L. M. and S. Cohen (1973) Aggregation in Blattaria. *Ann. Entomol. Soc. Am.* 66: 1315-1323.

Sakuma, M. and H. Fukami (1985) The linear track olfactometer: An assay device for taxes of the German cockroach, *Blattella germanica* (L.) (Dictyoptera: Blattellidae) toward their aggregation pheromone. *Appl. Entomol. Zool.* 20: 387-402.

Sakuma, M. and H. Fukami (1990a) Dose/response relation in taxes of nymphs of the German cockroach, *Blattella germanica* (L.) (Dictyoptera: Blattellidae) to their aggregation pheromone. *Appl. Entomol. Zool.* 25: 9-16.

Sakuma, M. and H. Fukami (1990b) The aggregation pheromone of the German cockroach, *Blattella germanica* (L.) (Dictyoptera: Blattellidae): Isolation and identification of the attractant components of the pheromone. *Appl. Entomol. Zool.* 25: 355-368.

Sakuma, M., H. Fukami and Y. Kuwahara (1997a) Aggregation pheromone of the German cockroach, *Blattella germanica* (L.) (Dictyoptera: Blattellidae): controlled release of attractant amines by salt formation. *Appl. Entomol. Zool.* 32: 143-152.

Sakuma, M., H. Fukami and Y. Kuwahara (1997b) Attractiveness of alkylamines and aminoalcohols related to the aggregation attractant pheromone of the German cockroach, *Blattella germanica* (L.) (Dictyoptera: Blattellidae). *Appl. Entomol. Zool.* 32: 197-205.

APPENDIX

Variances and covariances of the parameters in the basic model

If not specifically mentioned, the notation of each parameter is the same as that used in the text. If a weighting coefficient (w) is defined by

$$w_i=\frac{Z_i^2}{4P_i^*Q_i^*}=\frac{Z_i^2}{(1+P_i)(1-P_i)}, \tag{A1}$$

a quadratic Taylor expansion of Eq. (3) with respect to the parameters α and β can be expressed as

$$\delta a \sum_i n_i w_i+\delta b \sum_i n_i w_i x_i=\sum_i n_i w_i\left(\frac{p_i-P_i}{Z_i}\right),$$

$$\delta a \sum_i n_i w_i x_i+\delta b \sum_i n_i w_i x_i^2=\sum_i n_i w_i x_i\left(\frac{p_i-P_i}{Z_i}\right).$$

where, δa and δb are the increments of the parameters improving each initial guess. At convergence, the approximate variances (v_{aa}, v_{bb}) and covariances (v_{ab}, v_{ba}) of a and b, the estimators for parameters α and β, can be obtained as the elements of inverse matrix of the second differential coefficients:

$$\begin{pmatrix} v_{aa} & v_{ab} \\ v_{ba} & v_{bb} \end{pmatrix}=\begin{pmatrix} \sum_i n_i w_i & \sum_i n_i w_i x_i \\ \sum_i n_i w_i x_i & \sum_i n_i w_i x_i^2 \end{pmatrix}^{-1} \tag{A2}$$

Variances and covariances of the parameters in the general model

If a weighting coefficient (w_i) and auxiliary variates (x_i', x_i'') are defined by the following equations,

$$w_i=\frac{Z_i^2}{\left[\frac{C}{(1-C)(1-D)}+P_i\right]\left[\frac{1}{1-D}-P_i\right]},$$

$$x_i'=\frac{1-(1-D)P_i}{Z_i},\quad x_i''=\frac{-(1-C)P_i}{Z_i}, \tag{A3}$$

the inverse Hessian matrix including additional two parameters C and D in the Eq. (7) becomes

$$\begin{pmatrix} v_{aa} & v_{ab} & v_{aC} & v_{aD} \\ v_{ba} & v_{bb} & v_{bC} & v_{bD} \\ v_{Ca} & v_{Cb} & v_{CC} & v_{CD} \\ v_{Da} & v_{Db} & v_{DC} & v_{DD} \end{pmatrix} = \begin{pmatrix} \sum_i n_i w_i, & \sum_i n_i w_i x_i, & \sum_i \frac{n_i w_i x_i'}{(1-C)(1-D)}, & \sum_i \frac{n_i w_i x_i''}{(1-C)(1-D)} \\ \sum_i n_i w_i x_i, & \sum_i n_i w_i x_i^2, & \sum_i \frac{n_i w_i x_i x_i'}{(1-C)(1-D)}, & \sum_i \frac{n_i w_i x_i x_i''}{(1-C)(1-D)} \\ \sum_i \frac{n_i w_i x_i'}{(1-C)(1-D)}, & \sum_i \frac{n_i w_i x_i x_i'}{(1-C)(1-D)}, & \sum_i \frac{n_i w_i x_i'^2}{(1-C)^2(1-D)^2}, & \sum_i \frac{n_i w_i x_i' x_i''}{(1-C)^2(1-D)^2} \\ \sum_i \frac{n_i w_i x_i''}{(1-C)(1-D)}, & \sum_i \frac{n_i w_i x_i x_i''}{(1-C)(1-D)}, & \sum_i \frac{n_i w_i x_i' x_i''}{(1-C)^2(1-D)^2}, & \sum_i \frac{n_i w_i x_i''^2}{(1-C)^2(1-D)^2} \end{pmatrix}^{-1} \tag{A4}$$