Kansas State University Libraries [New Prairie Press](https://newprairiepress.org/) 

[Conference on Applied Statistics in Agriculture](https://newprairiepress.org/agstatconference) 1992 - 4th Annual Conference Proceedings

# A MARKOV CHAIN MODEL TO ASSESS RESISTANCE OF CATTLE TO HORN FLIES

Edward Gbur

C. Dayton Steelman

Follow this and additional works at: [https://newprairiepress.org/agstatconference](https://newprairiepress.org/agstatconference?utm_source=newprairiepress.org%2Fagstatconference%2F1992%2Fproceedings%2F8&utm_medium=PDF&utm_campaign=PDFCoverPages)

 $\bullet$  Part of the [Agriculture Commons](http://network.bepress.com/hgg/discipline/1076?utm_source=newprairiepress.org%2Fagstatconference%2F1992%2Fproceedings%2F8&utm_medium=PDF&utm_campaign=PDFCoverPages), and the Applied Statistics Commons



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.](https://creativecommons.org/licenses/by-nc-nd/4.0/)

## Recommended Citation

Gbur, Edward and Steelman, C. Dayton (1992). "A MARKOV CHAIN MODEL TO ASSESS RESISTANCE OF CATTLE TO HORN FLIES," Conference on Applied Statistics in Agriculture. [https://doi.org/10.4148/](https://doi.org/10.4148/2475-7772.1396) [2475-7772.1396](https://doi.org/10.4148/2475-7772.1396) 

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

## **A MARKOV CHAIN MODEL TO ASSESS RESISTANCE OF CATTLE TO HORN FLIES**

Edward Gbur and **C.** Dayton Steelman University of Arkansas

#### **Abstract**

The horn fly is an economically important external permanent parasite of cattle. As part of a research project focused on alternatives to chemical control of the horn fly, a study was conducted to determine the degree of innate resistance of individual cattle to the horn fly. A fly resistant cow was defined as one whose horn fly counts were in the lower quartile of the weekly fly count distributions for a herd more often than would be expected by chance. A Markov chain model was formulated and a small sample test for fly resistance was developed. The model and procedure are illustrated using data collected on a herd of Charolais cows. Tests of the Markov chain and stationarity assumptions are discussed and applied to the data.

**Keywords:** Bernoulli process, Independence, stationarity.

## **1. Introduction**

The horn fly, Haematobia irritans (L.), is an economically important external permanent parasite of cattle. Drummond (1987) estimated losses to producers attributable to the horn fly in the U.s. at \$150 million annually. In the past, control of the horn fly has been through the use of insecticides in the form of ear tags or dusts and sprays. Ear tags and sprays are commonly used on all animals for the entire fly season. As a result, the horn fly has developed resistance to the insecticides, thereby requiring higher and higher dosages or increasingly potent chemicals to attain control. This paper arose from collaborative work on an ongoing entomology research project at the University of Arkansas which is focused on alternatives to chemical control measures.

The entomological literature contains numerous studies of the behavior of horn fly populations over time, e.g., Morgan and Thomas (1974) and Palmer and Bay(1981). Several papers have explicitly noted variation among cows in a herd at fixed points in time (e.g., Campbell, 1976), but none have tracked individual animals over time. One of the purposes of the University of Arkansas project was to determine if there were consistent patterns in the horn fly counts over time among cows in a herd relative to other cows in the same herd. The existence of cows with consistently high or low fly counts relative to other cows in the herd would have implications for non-insecticide based management strategies to deal with the horn fly problem. If a producer could recognize cows in his herd with fly counts below any injury or economic threshold, then these animals would not need to be treated for horn fly control. On the other hand, animals with fly counts which, if untreated, would rise above the threshold can be treated selectively and eventually culled from the herd.

The objective of this paper is to develop a statistical procedure which, under the assumption of differences among animals in the herd, would enable the entomologist to identify animals with consistently low horn fly counts relative to other animals in the same herd.

#### 2. Development **of a** Model

Weekly horn fly counts on individually identified cows in herds of several different breeds collected during a preliminary study showed that some animals had horn fly counts which were consistently smaller than most of the other animals in their herd. This led to the notion that a fly resistant animal is one which has consistently (for "many" weeks of the fly season) "small" horn fly counts relative to other animals in the herd. Since the horn fly population density naturally changes over the course of the season, comparison of the magnitude of fly counts must be for each fixed time point rather than over time. Moreover, the shape of the weekly distributions change over time, making parametric distributional assumptions unrealistic. Thus, we defined a fly resistant cow to be one whose horn fly counts are in the lower quartile of the weekly fly distributions more often than would be expected by chance. Using this definition, the changing shape of the weekly fly count distributions does not affect the determination of the animal's status. It does, however, require that the same set of animals be sampled each week of the season.

To develop a statistical model based on the above notion of a fly resistant cow, for  $t = 1, ..., T$ , define the random variable  $X_i$  by,

- $X<sub>1</sub> = 1$  if the cow's fly count fell in the lower quartile of the fly distribution for the  $t<sup>th</sup>$  week of the fly season
- $X_i = 0$  otherwise.

Let  $p = P[X_1 = 1]$ . Note that p is assumed to independent of time. Then intuitively,

 $H_0: p \le 0.25$ 

 $H_1: p > 0.25$ 

can be tested using the statistic  $X = \Sigma_1 X_1$ , with "large" values of the statistic leading to rejection of  $H_0$ . If  $H_0$  is rejected for a particular cow, then the cow will be said to be fly resistant.

Under the above model, the data for a particular cow in a particular fly season has been reduced to a sequence of Bernoulli random variables  $X_1$ , ...,  $X_T$ , each with the same p. Since the  $X_i$ 's are determined from the horn fly counts on the same cow on consecutive weeks relative to the same set of animals in the herd, we need to account for potential week to week dependence. Initially we shall assume a simple first order autocorrelation structure for this dependence. Let  $\rho = \rho(X_t, X_{t+1})$  be the correlation between consecutive  $X_i$ 's, independent of time.

The above dependent Bernoulli trial model can be recast as a two state stationary Markov chain. Let the first state correspond to the cow's horn fly count falling in the lower quartile of the fly distribution for the week. Let

 $(1)$ 

$$
\varrho = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}
$$

represent the transition probability matrix of the chain, where

 $p_{11} = P[X_t = 1 | X_{t-1} = 1],$  $p_{22} = P[X_1 = 0 | X_{11} = 0 ]$ .

Under this formulation, the statistic  $X = \Sigma_t X_t$  represents the occupation time for the first state of the chain.

Using the definition of correlation and the fact that  $X_t$  has a Bernoulli distribution, a straightforward calculation yields

$$
\rho = (p_{11} - p)/(1 - p) \t\t(2)
$$

and hence,

$$
p_{11} = p + q\rho \t{1}
$$

where  $q = 1 - p$ . In addition, it can be shown that

$$
p_{22} = q + p\rho \tag{4}
$$

Thus, the distribution of X can be expressed in terms of the parameters T, p, and *p.* 

The correlation  $\rho$  must satisfy the restriction

 $\max\{-p/q, -q/p\} \le \rho \le 1$ .

For  $p = 0.25$ , this reduces to  $-0.33 \le p \le 1$ . In addition,  $p = 0$ corresponds to independent trials (simple binomial model).

An implicit expresssion for the pdf of X is given in Pedler (1971; 1980, equation 2.1). Although the expression is not directly useful for calculating individual probabilities, the probabilities can be obtained from a recursion formula for the cdf. Following Pedler (1980), let

 $F_T(x) = P[X \le x ; T, p, \rho]$ 

denote the cdf. Then

$$
F_T(x) = p_{22}F_{T-1}(x) + p_{11}F_{T-1}(x-1) - \rho F_{T-2}(x-1)
$$
 (5)

subject to the boundary conditions



and where we define  $F_0(0) = 1$ . The distributions for T = 21, p = 0.25, and several values of  $\rho$  are shown in Figure 1. A listing of the SAS IML code used to calculate the distribution is given in Appendix 1. The effect of  $\rho$  on the upper tail probabilities, and hence, on the test of  $(1)$ , is evident from the figure.

### 3. Inference in the Model

Under the Markov chain formulation of the problem, the critical region for the hypothesis of fly resistance (1) is an upper tailed region based on the test statistic X. If the correlation  $\rho$  were known, then the p-value

$$
p_{obs} = P [ X > x_{obs} ] = 1 - FT(x_{obs})
$$

can be calculated using the recursion formula (5) with  $p = 0.25$ and a known T (number of weeks in the fly season).

To obtain an estimate of  $\rho$ , let  $n_{ii}$  be the number of one step transitions from state i to state j; i.e., the number of times in the Bernoulli sequence that  $X_t = j$  follows  $X_{t-1} = i$ . Let  $n_{i+} = \sum_j n_{ij}$ . Then, assuming  $X_1$  is non-stochastic and that all  $n_{i+} > 0$ , the maximum likelihood estimators of the  $p_{ii}$  are given by

$$
\hat{\mathbf{p}}_{ii} = \mathbf{n}_{ii}/\mathbf{n}_{i+} \tag{6}
$$

Substituting into (2) for  $p_{11}$  from (6) and the sample proportion of ones in the sequence for p, we obtain an estimator

$$
\hat{\rho} = (\hat{p}_{11} - \hat{p})/(1 - \hat{p}) \tag{7}
$$

for  $\rho$ .

The estimator (7) was used in the recursion formula (5) for the cdf to calculate p-values. The effect of using the estimated  $\rho$  in place of the true value on the calculation of the p-value has not been investigated.

The estimation procedure described above was based on the assumption that all  $n_{i+} > 0$ ; that is, the Bernoulli sequence contains at least one 0 and at least one 1. Lehoczky (1990) discusses two alternative approaches when some  $n_{i+}$  are zero. Rather than use one of these more complicated approaches, we chose a simple non-statistical solution for cows for which an  $n_{i+}$  was zero. If  $\hat{p} = 1$  (the cow's fly count was in the lower quartile every week), we declared the cow to be fly resistant without a formal test of (1). In the much more common case when  $\hat{p} = 0$  (the cow's fly count was never in the lower quartile), we automatically declared the cow not to be fly resistant.

#### 4. Numerical Example of Cow Classification

As an example, we consider data collected in 1989 on a herd of 17 Charolais cows located at the University of Arkansas Beef Cattle Research Station. Horn fly counts were recorded on each animal for 21 consecutive weeks. Further details of the experiment can be found in Steelman et al. (1991). A plot of the weekly fly counts for a cow with relatively low fly counts and one with relatively high fly counts is shown in Figure 2. The mean weekly fly count for the entire herd is also shown in the figure. The relative difference between resistant and susceptible animals is clear from the graph.

To illustrate the classification of an individual cow, consider the data for cow #375 in 1989. The actual horn fly counts were plotted in Figure 2. The Bernoulli sequence obtained from these counts was

1 1 0 0 1 1 1 1 0 1 0 1 0 1 1 0 1 1 1 1 O.

Thus, we and  $\hat{\mathbf{p}}_{22}$  = close to the lower bound of -0.33 for  $\rho$  under the hypothesis  $p$ 0.25. have  $T = 21$ , 1/6 = 0.1667.  $X = 14$ ,  $\hat{p} = 0.6667$ ,  $\hat{p}_{11} = 8/14 = 0.5714$ , From (7), we obtain  $\widehat{\rho}$  = -0.2859 which is

The pdf and cdf of  $X = \Sigma_t X_t$  using the recursion formula (5) with the estimated parameters from cow #375 are given in Table 1. The binomial distribution ( $\rho = 0$ ) is included for comparison. From Table 1,

$$
p_{obs} = P[X > 14]
$$
  
= 1 - F<sub>21</sub>(14)  
< 0.00001

and we conclude that cow  $#375$  can be classified as fly resistant for 1989.

## **5. Tests of Model Assumptions**

In a seminal paper, Anderson and Goodman (1957) developed easily applied inference procedures for several fundamental questions related to Markov chain models. In this section we describe tests for (i) the hypothesis that the  $X_t$  are independent against the alternative of a first order Markov chain and (ii) the stationarity of a first order Markov chain.

**Independence:** If the X, are independent, then the conditional probabilities  $p_{ii} = P[X_i = j | X_{i-1} = i]$  are the same as the corresponding unconditional probabilities  $p_i$ . In the special case of a two state Markov chain this reduces to



 $H_1$ : the  $p_{ii}$  depend on i  $\overline{\phantom{a}}$ 

Anderson and Goodman (1957) formally treat the transition probability matrix of the chain as the following  $2 \times 2$  contingency table whose row sums are 1.



Then the independence hypothesis (8) is formally equivalent to the hypothesis of row homogeneity in the above contingency table. Asymptotically, the usual  $\chi^2$  test for row homogeneity is equivalent to the likelihood ratio test for independence in the context of a first order Markov chain. The asymptotics are valid if either

(1) the length T of a single sample from the chain approaches infinity

or

(2) the number of samples from chain approaches infinity where each sample is of fixed length T.

In the latter case, there are additional conditions on the behavior of the distribution of the number of individuals in each state at the initial time.

Wang and Scott (1989) provide a small sample alternative to the asymptotic  $\chi^2$  test using a bootstrap procedure. Their approach is based on the largest non-unit eigenvalue of the transition probability matrix. For a two state Markov chain, this eigenvalue is given by  $\lambda = p_{11} - p_{21}$  so that the independence hypothesis (8) can be restated as  $H_0$ :  $\lambda = 0$ .

stationarity: A Markov chain is said to be stationary if the transition probabilities  $p_{ij}$  are independent of time. Let  $p_{ij}(t)$ denote time dependent transition probabilities. Then, for a two state first order Markov chain, the hypothesis of stationarity can be stated as

 $H_0$ :  $p_{11} (t) = p_{11}$  $p_{22}(t) = p_{22}$ , for  $t = 1, ..., T$  (9)

 $H_i$ :  $p_{ii}(t)$  depend on t.

Anderson and Goodman (1957) reformulate the stationarity hypothesis (9) as a hypothesis of row homogeneity in each of a pair of  $(T - 1) \times 2$  contingency tables defined by  $X_{i-1} = i$  as shown below.

X,



The  $\chi^2$  test for homogeneity is applied to each table separately and the sum of the two statistics is used to test (9). In contrast to the test for independence, the asymptotics for the validity of the  $\chi^2$  test for stationarity can only be based on increasing the number of samples from the chain and not on increasing the length of a single sample.

## 6. Numerical Example of the Tests of Model Assumptions

Intuitively, consistency in the modeling process dictates that the Markov dependence and stationarity assumptions should apply to the entire population of animals in the herd rather than separately to individual animals. Thus, it appears that the data should be combined over cows and the validity of the  $\chi^2$  tests

should be based on a large number of samples (cows) of fixed length from the same chain. The difficulty is that, although each cow's data may fit a first order stationary Markov chain, the transition probabilities  $p_{ij}$  and hence, the unconditional probability p, clearly vary across animals. By definition, not all cows can be fly resistant  $(p < 0.25)$ . Thus, the results obtained from testing (8) and (9) using the combined data must be interpreted cautiously in an exploratory rather than confirmatory mode. We will interpret the results as providing information about an underlying common process which generates the data for all animals but will not attempt to extract information about the values of the parameters, which are cow specific.

The work of Gardner (1990) may provide some insight into the problem. He takes a latent variable approach to modeling the variability in the transition probabilities among samples under the assumption that each sample follows a first order stationary Markov chain. The problem here is analogous to the problem of a common form for a regression model for all items in a population but with item specific regression coefficients.

With the above caveat firmly in mind, we proceed to apply the Anderson-Goodman tests to the 1989 data on the Charolais herd. A Bernoulli sequence based on 21 weeks of fly counts was constructed for each of the 17 cows in the herd and the necessary contingency tables were tabulated.

**Independence:** The  $2 \times 2$  contingency table needed to test the independence hypothesis (8) is given below.

> x, 1 0 1 38 47 85  $X_{t-1}$  $0 \mid 47 \mid 208 \mid 255$

From the data in the table,  $\chi^2$  = 23.472, which, when compared to a  $\chi^2$  distribution with 1 df, yielded a p-value of approximately 0.000001. Thus, it appears that the independence assumption can be rejected in favor of the assumption of a first order Markov chain.

From the table,  $\hat{p}_{11}$  = 38/85 = 0.4471,  $\hat{p}$  = 85/340 = 0.25, and from  $(7)$ ,  $\hat{p} = 0.2628$ . Contrasting these estimates with those calculated previously for cow #375 provides additional evidence that parameter estimates based on the combined data probably have little meaning.

**Stationarity:** The pair of 20  $\times$  2 tables required to test the stationarity hypothesis (9) are given Table 2. From the table based on  $X_{l-1} = 1$  we obtain  $\chi^2 = 12.116$  and from the table based on  $X_{l-1} = 0$ ,  $\chi^2 = 9.694$ . Combining the results gives  $\chi^2 = 21.810$ , which, when compared to a  $\chi^2$  distribution with 38 df, yielded a pvalue of 0.9837. Hence, it appears that the underlying first order Markov chain is stationary.

#### 7. Conclusion

We have developed a procedure which can be used to identify cows in a herd which exhibit some degree of innate resistance to horn flies relative to their herdmates. The Markov chain formulation accounts for the dependence structure in the data arising from repeatedly sampling the same set of animals. At the same time, it overcomes the problem of a changing shape for the fly count distribution. The procedure has been applied to data collected on six breeds of beef cattle in a three year study. The cows identified as fly resistant within each herd remained consistent over years provided the herd composition did not change greatly (steelman et al., 1993).

#### 8. References

- Anderson, T. W. and L. A. Goodman (1957). Statistical inference about Markov chains. Annals of Math. Stat. 28, 89-110.
- Campbell, J. B. (1985). Highlights of veterinary entomology in the United States. <u>Nigeria J. Entomology</u> 6, 111-119.
- Drummond, R. o. (1987). Economic aspects of ectoparasites of cattle in North America. In The Economic Impact of Parasitism in Cattle, W. H. D. Learning and J. Guerrero, ed. Proceedings of the MSD AqVet. Symposium. XXIII World Veterinary Congress, Montreal, 9-24.
- Gardner, W. (1990). Analyzing sequential categorical data: Individual variation in Markov chains. Psychometrika 55, 263- 275.
- Lehoczky, J. (1990). Statistical methods. In Stochastic models, D. P. Heyman and M. J. Sobel, ed. Amsterdam: North Holland, 255-293.
- Morgan, C. E. and G. D. Thomas (1974). Annotated bibliography of the horn fly Haematobia irritans (L.), including references on the buffalo fly, H. exiqua (de Meijere), and other species belonging to the genus Haematobia. U.S. Dept. Agric. Misc. Publ. 1278.
- Palmer, W. A. and D. C. Bay (1981). A review of the economic importance of the horn fly, Haematobia irritans (L.). Protective Ecology 3, 237-244.
- Pedler, P. J. (1971). Occupation times for two-state Markov chains. J. Applied Prob. 8, 381-390.
- Pedler, P. J. (1980). Effect of dependence on the occupation time in a two-state stationary Markov chain. JASA 75, 739-746.
- Steelman, C. D., A. H. Brown, E. E. Gbur, and G. Tolley (1991). Interactive response of the horn fly (Diptera: Muscidae) and selected breeds of beef cattle. J. Economic Entomology 84, 1275-1282.
- Steelman, C. D., E. E. Gbur, G. Tolley, and A. H. Brown (1993). Individual variation within breeds of beef cattle in resistance to horn fly (Diptera: Muscidae). J. Medical Entomology 30, 414-420.

Wang, D. Q. and D. J. Scott (1989). Testing a Markov chain for independence. Comm. in Stat., Theory and Methods 18, 4085-4103.

## Appendix 1

The following SAS program calculates the cdf of the occupation time for the first state of the Markov chain,  $X = \Sigma_i X_i$ , using the recursion formula (5) from Pedler (1980). The program was written by Kevin Thompson.

```
%Macro Markov( /*COF of 
Occup. time of state 1 of Markov chain */ 
/* Parameter Default 
Description */ 
   data = \begin{array}{cc} \n= & \text{last} \\ \n= & \text{first} \\ \n= & \text{constant} \end{array} , \begin{array}{cc} \n= & \text{first} \\ \n= & \text{first} \\id 
   t 
                          /*Input Var(s): ID-type variables */ 
                           , /*Input Var: No. of trials */ 
   P 
   rho 
   out 
                  =t 
                  =p 
                  =rho 
                  = data 
                            , /*Input \text{Var: } P(X = 1) */
                            , /*Input Var: Corr(Xt, Xt-1) */ 
                            , /*Output SAS Data Set */ 
   x 
   cdf 
   ) ; 
                  = x 
                  = cdf 
                            , /*Output Var: Occupat. time of state 1 */ 
                              /*Output Var: CDF of X */ 
proc iml; start; \overline{\phantom{a}} /* Begin IML step */
  reset nocenter nolog name; <br>use &DATA var{ &ID &T &P &RHO }; /* Input SAS DS and vars */
  use &DATA var{ &ID &T &P &RHO };
  &X=. ; &COF=. ; /* Initialize output vars*/ 
  create &OUT var{ &IO &T &P &RHO /* Output SAS DS and vars*/ 
                       \&X\&CDF };
  setin &DATA point 1; read current; /* Read first param. set */
  do data; /* Do until end of data */ 
       * Create temporary vars and 
initialize recursion conditions; 
      p11 = \&P + (1 - \&P) * \Ρp22 = (1 - \&P) + \&P* \Ρ/* P[ Xt=1 | Xt-1=1 ] */
      matrix = j( & \& T+1, & \& T+1, 1);
                                              /* P[ Xt=0 | Xt-1=0 ] */matrix[2,1] = (1-\&P);* Recursion formula (5); 
      do it = 3 to (6T+1);
          matrix [\n    it, 1] = matrix [\n    it-1, 1] * p22;do ix = 2 to &x+1;
             _matrix [it, i x] = p22*_matrix [ it -1, i x ]p11*_matrix[ _it-1, i-x-1 ] -
                                      \overline{\text{R}R}HO* matrix[ \overline{\text{it}}-2, \overline{\text{it}} ix-1 ];
          end; 
      end; 
      * Print to LISTING ; 
      cdf = matrix[ <math>&T+1, ]</math>;print &IO &T[format=best5.) 
&P[format=best5.) 
              &RHO[format=best5.) 
cdf[format=best5.) 
      reset noname; 
      * Write to output SAS Data Set; 
      do \&X = 0 to \&T;\&CDF = matrix[ \&T+1, \&X+1 ];
```
append; end; read next; end; **/\* Read next param set /\* End DO DATA loop**  finish; run; reset log; quit; %MEND Markov; data corr; length figure \$9; input year figureS t p numerat denomin; **/\* End IML step /\* End Macro**  rho=numerat/denomin; **/\* Calcuate rho to best precision \*/**  cards; 1989 Histogram 1989 Histogram 21 0.25 **-1** 4 1989 Histogram 21 0.25 1989 Histogram 21 0.25 1989 Histogram 21 0.25 1989 Histogram 21 0.25 19 20 1989 Charolais 21 0.25 -0.2859 1989 Charolais 1989 Charolais 21 0.25 run; 21 0.25 **-1** 3 21 0.25 0 1 1 4 1 2 0.0000 0.2628 1 1 1 %Markov(data=corr, id=year figure, t=t, p=p, rho=rho,  $out=cdf, x=x, cdf=cdf);$ proc print data=cdf; quit; **\*/ \*/ \*/ \*/** 



Table 1. Pdf and cdf of X for Charolais cow #375 in 1989 ( $\hat{\rho}$  =  $-0.2859$ ) and for the binomial model ( $\rho = 0$ )

Table 2. Summary of the data used to test the assumption of stationarity of the first order Markov chain for the Charolais herd in 1989.





Figure 1. Distribution of the occupation time X =  $\Sigma_{\text{i}}\text{X}_{\text{i}}$  for T = 21,  $p = 0.25$ , and selected values of  $p$ .

 $\frac{8}{6}$ 



Figure 2. Weekly horn fly counts for selected Charolais cows and mean fly count for the herd in 1989.