The use of conditional probability functions in range data analysis and simulation

DAVID K. LAMBERT AND THOMAS R. HARRIS

Authors are associate professor and professor, respectively, Department of Agricultural Economics, University of Nevada, Reno.

Abstract

Managers and range scientists are interested in the response of such variables as forage production and animal performance to various environmental and management factors. Due to the inability to control many of the factors affecting range systems, production responses should include distributional information in addition to their expected values. Recent developments in the estimation of conditional probability distribution functions provide the range scientist with a practical procedure to more fully characterize variable responses. The conditional probability distribution approach is applied to an analysis of forage production data from the literature. An illustration of the procedure in range decision analysis derives distributional information on animal performance and net return under several different steer stocking levels.

Key Words: data, estimation, forage production, risk, range simulation

A continuing problem in range management is the gap between the data that are necessary versus those which are available for analyzing alternative management strategies. For example, functional relationships between weather and forage production and the response of range vegetation to different treatments and grazing regimes are seldom known with certainty.

Management decisions must incorporate this uncertainty to avoid the very real possibility of economic disaster resulting from unanticipated drought, insect infestations, animal disease, or any of the myriad of environmental factors that can affect the range livestock operation. Many researchers have included stochastic elements in the models. Vantassell et al. (1987) simulated calf weights as a function of alternative management and weather variables, generating deviations from expected weights as a function of assumedly normally distributed residuals from their estimating regression equations. Riechers et al. (1989) ignored even this use of limited distributional information available from least squares techniques, and used expected values of forage standing crop resulting from varying precipitation levels. Rodriguez and Roath (1987) considered different levels of forage production in their dynamic programming analysis of short run grazing management decisions. However, each individual production level was assumed to be equally likely to occur. It is unlikely that production levels actually are uniformly distributed over the assumed range of values.

This paper presents a procedure to directly incorporate stochastic response functions for forage production conditional on annual precipitation. A particular level of precipitation results not in a

given quantity of forage growth, but rather yields a distribution of possible growth responses. For example, 25 cm of precipitation may yield forage growth of less than 125 kg ha^{-1} with a probability of 0.10, less than 150 kg with a probability of 0.25, and so on. An application of the procedure is used to derive conditional forage production estimates from data reported in Sneva and Hyder (1962). The paper concludes with a simple example of the use of the conditional probability density function approach in ranch management.

The Conditional Distribution

The conditional distribution approach derives estimates of the distribution of the dependent variable (e.g., forage production) for a given level of 1 or more conditioning variables (e.g., precipitation, previous utilization, site condition). Several conditional distributions have been explored in the literature (Taylor 1984, 1990; Nelson and Preckel 1989; Moss et al. 1991). Taylor's (1984) univariate approach is adopted here for determining the conditional distribution of range forage production under different precipitation levels.

Taylor (1984) advocates the use of a conditional distribution function based on the hyperbolic tangent function. The hyperbolic tangent and its derivative, the square of the hyperbolic secant, bear close resemblance to a cumulative distribution function (cdf) and its related probability density function (pdf), respectively (Fig. 1).
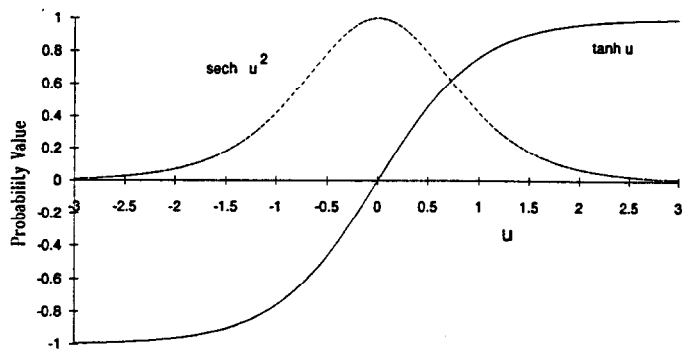

Fig. 1. Values of the hyperbolic tangent and the square of the hyperbolic secant.

The following transformation of the hyperbolic tangent function yields values between 0 and 1, and can be used an an empirical cdf relating foraging yield Y to annual precipitation amounts R:

$$F(Y/R) = 0.5 + 0.5 \tanh [P(Y,R)] \tag{1}$$

where $F(Y/R)$ is the conditional cdf of yield for a given level of precipitation, $P(Y,R)$ is a polynomial function of Y and R, and tanh is the hyperbolic tangent function, $\tanh u = \dfrac{(e^u - e^{-u})}{(e^u + e^{-u})}$

Equation [1] describes a flexible cumulative distribution function on Y for any given value of R.

Taylor (1984) points out that in field studies, as opposed to controlled experiment station trials, a range of values for Y is not observed for each distinct value of R. Consequently, similar values of the conditioning variable(s) must be grouped and the mean of the group can be taken as the single observation R_i for the ith grouping. The r_i observations Y_{ij}, j=1,...,r_i, associated with the values of precipitation included in group i are then sorted in ascending order and assigned cumulative distribution values z_{ij} equal to j(1/r_i), where r_i is the total number of Y values for (pooled) observations R_i.

Given these values of R_i, Y_{ij}, and z_{ij}, the following relationship is estimated using ordinary least squares (OLS) regression

$$z_{ij} = P(Y_{ij}, R_i) + \epsilon_{ij} \tag{2}$$

The degree of the polynomial and the number of interaction terms between Y and R are determined by the researcher to give the most parsimonious relationship among the variables given the size of the data set.

Taylor (1984) also points out that, given experimental data, the OLS estimates of the coefficients in [2] will be biased. This bias results because a nonstochastic variable, z, enters the regression as the dependent variable, and the stochastic variables, P(Y,R), are treated as independent variables. The independent variables are therefore correlated with the error term ϵ_{ij}, giving rise to the biased OLS coefficient estimates. The OLS estimates can, however, be used as starting values in maximum likelihood estimation procedures (Greene 1990) to derive the unbiased coefficients.

The pdf of the conditional hyperbolic tangent distribution results from differentiation of [1],

$$f(Y/R) = 0.5 \ P'(Y_{ij}, R_i) \ sech^2[P(Y_{ij}, R_i)] \tag{3}$$

where $P'(Y_{ij}, R_i)$ is the partial derivative of the polynomial function P with respect to Y, evaluated at the point (Y_{ij}, R_i). The hyperbolic secant of u, or sech u, is defined as the hyperbolic tangent divided by the hyperbolic sine. Using the definition of tanh u from above, sech u = tanh u•[2($e^u - e^{-u}$)].

The likelihood function can thus be formed for the independent and identically distributed observations i and r:

$$L(\beta) = \prod_{i=1} \prod_{j=1} 0.5 \ P'(Y_{ij}, R_i) \ sech^2[P(Y_{ij}, R_i)] \tag{4}$$

For estimation of the coefficient vector β, it is more convenient to rewrite [4] in logarithmic form:

$$\ln L(\beta) = (n \sum_{i=1}^{n} r_i) \ln(0.5) + \sum_{i=1}^{n} \sum_{j=1}^{r_i} \ln[P'(Y_{iij}, R_i)] \tag{5}$$
$$2 \sum_{i=1}^{n} \sum_{j=1}^{r_i} \ln[sech[P(Y_{ij}, Ri)]]$$

Maximum likelihood procedures are available in most statistical software packages for the direct maximization of equation [5].

Sneva and Hyder (1962) report annual precipitation and forage production observations as percentages of normal for various range sites representative of the Intermountain West. Eighty-eight paired observations were used to estimate an empirical distribution of forage and precipitation using the procedures outlined above.

The following polynomial function was determined to be best among the alternatives evaluated. Maximum likelihood parameter estimates and their associated t-statistics (in parentheses) are reported below:

$$u = P(Y,R) = -0.1563 + 0.0590Y - (6.304E-5)Y^2 - 0.0509R \tag{6}$$
$$(-0.400) \quad (12.089) \quad (-6.460) \quad (-9.928)$$

Use of [6] results in pdf's that are conditional upon precipitation levels R (Fig. 2). Conditional forage yields corresponding to different levels of the cdf can be derived by inverting [1] and solving for Y for different values of α, the desired cdf value:

$$\alpha = F(Y,R) = 0.5 + 0.5 \tanh (P(Y,R)) \tag{7}$$
$$= 0.5 + 0.5 \left(\frac{(e^u - e^{-u})}{(e^u + e^{-u})} \right)$$

Solving [7] for u yields:

$$u = 0.5 (\ln(\alpha) - \ln(1-\alpha)) \tag{8}$$

Since u is quadratic, Y, expressed as percentage of average forage production, can be solved for any specified values of α and R:
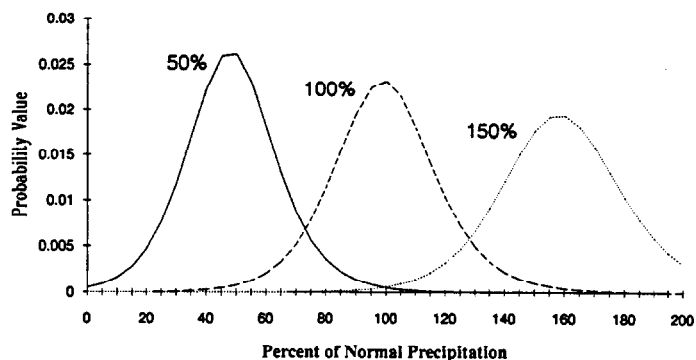
Fig. 2. Values of the probability density function for forage production for precipitation levels equal to 50, 100, and 150% of normal calculated from substituting equation [6] in equation [3].

$$Y = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad (9)$$

where Y will be the 2 roots of the quadratic expression and, from the polynomial in [6], $a = 6.304E-4$, $b=0.0590$, and $c=(-0.1563 - 0.5(\ln(\alpha)-\ln(1-\alpha))-0.0509R)$. No difficulty was encountered in choosing the "more reasonable" of the 2 roots of this quadratic function in the empirical application.

The value of the conditional pdf approach increases when observations corresponding to a given level of the conditioning variable are not normally distributed. The hyperbolic tangent function, or any of the other distributional forms that have been used, preserves nonnormalities that are likely to arise in field research. Direct estimation of the mean and of higher moments of the forage production conditional upon precipitation levels can be obtained by performing the following integrations:

$$E(Y) = \int_{\infty}^{\infty} Y \, f(Y/R) \, dY, \text{ and} \qquad (10)$$

$$\mu_k = \int_{\infty}^{\infty} (Y-E(Y))^k \, f(Y/R) dY \qquad (11)$$

where E(Y) is the expected level of forage production given R and μ_k is the kth central moment of the distribution of Y. Solutions to [10] and [11] can be obtained by use of numerical integration procedures.

A Ranch Decision Making Application

A relatively simple decision problem, and one that is commonly faced by ranchers of the Intermountain West, concerns the choice of purchasing short yearlings in the spring, grazing them through the season, and selling them as feeder cattle in the fall. One source of uncertainty confounding the decision is the amount of forage production that will be available and subsequent animal weight gains. Variation in expected forage production, as well as the shape of the cumulative distribution of forage availability, introduces a considerable amount of uncertainty into the likely net returns resulting from the stocker decision. Additional uncertainty surrounds rates of gain and animal forage intake, but this simple example will abstract from these other sources of risk. The distributional information relating precipitation and forage production obtained from evaluation of Sneva and Hyder's data can provide some of the quantitative information useful in analyzing the stocker decision.

It is assumed in this application of the distributional model that the rancher has 810 ha of improved range available to be grazed. Forage production in an average year is assumed to be 150 kg ha^{-1} of consumable forage. Production varies, however, depending upon precipitation totals.

The rancher is assumed to choose the optimal number of days to

grazing to maximize expected net return from the venture, evaluated over all possible states of nature. A state of nature in this example is defined as a particular outcome of the random variable, annual precipitation. Returns in this example are equal to the ending weight of the animal times the futures price reported in February, 1992 for October delivery, minus the initial purchase price of the steer. Ending weight is dependent on the number of days of grazing available which, in turn, is dependent on state of nature, or annual precipitation. Gain is approximated by a decreasing linear function of time on the range. Daily gain is assumed to be 1.36 kg day^{-1} initially, falling to 0.454 kg day^{-1} at the end of 5 months. Ending weight is thus a concave quadratic function of days of grazing.

The number of days of grazing available under each precipitation state of nature depends upon the total available forage, divided by daily consumption by the animals. Daily forage consumption is assumed to equal 2.5% of animal weight. The final functional constraint limits the number of steers by the total quantity of available forage under the most restrictive precipitation state of nature.

The example problem can be expressed:

$$
\begin{array}{ll}
\text{Maximize} & \sum_{j=1}^{30} \theta_j \; S(p_f W_j - p_o W_o) \qquad (12) \\
\text{subject to} & W_j - g_j D_j \leq W_o \quad \text{for all } j \\
& D_j - \dfrac{afj}{cS} \leq 0 \quad \text{for all } j \\
& \quad\quad\quad\quad\quad\quad \text{for all } j \\
& c_j S \leq afj \quad\quad \text{for all } j
\end{array}
$$

Parameter and variable definitions are presented in Table 1.

Table 1. Parameter and variable definitions for the example problem.

Variables	Parameters
W_j – Ending weight under j	θ_j – probability of precipitation state of nature j (=1/30)
D_j – Number of days of use under j	
g_j – daily gain (kg) of steer on improved range ($g_j = (1.36-0.0059 \frac{D_j}{2})$)	S – Number of steers to purchase
	p_f – Future sale price (October feeder cattle futures price quoted for week of 2/15/92
c_j – daily forage consumption (2.5% of average animal weight over the season, or $c_i = 0.025*(\frac{W_j + W_o}{2})$	p_o – California price of 450 lb (204 kg) steer quoted on 2/15/92
	W_o – Beginning weight (204 kg)
	a – hectares of range (810 ha)
	f_j – conditional forage supply under j

Precipitation levels were derived from annual data at Elko, Nev., for the period January 1930 through December 1984. Analysis of the sample autocorrelation and partial autocorrelation functions found no significant year to year correlation structure in the series. Consequently, precipitation levels could be generated assuming independence between years.

Goodness of fit tests failed to reject the hypothesis that the annual data fit the gamma distribution. Maximum likelihood estimators were calculated for the distribution's 2 parameters ($\alpha = 11.28$ and $\beta = 0.86$). Annual precipitation was generated using the random number generating routine for the gamma distribution (GGAMR) from the International Mathematical and Statistical Libraries' (IMSL) computer collection of statistical subroutines.

Thirty precipitation levels were randomly generated. Days of grazing, ending weights, and enterprise net returns were calculated for stocking levels of 300, 325, and 350 steers. Ten model solutions were obtained for each of the 3 stocking levels corresponding to α values of 0.10 through 1.00, in increasing increments of 0.10. The

lower values of α would represent a more conservative estimate of forage production for a given level of precipitation. For example, forage production corresponding to an α value equal to 0.20 would mean that there was an 80% probability that production would exceed this value.

Table 2. Mean ending weight of individual steers and days of grazing for 300, 325, and 350 steers under increasing values of α.

Proba-bility	Ending weight			Grazing Period		
(α)	300	325	350	300	325	350
	--------(kg)--------			--------(Days)------		
0.1	310	305	300	103	96	90
0.2	316	311	306	111	103	97
0.3	319	314	309	116	108	101
0.4	321	316	311	119	111	104
0.5	323	318	313	122	114	106
0.6	325	319	315	124	116	109
0.7	326	321	316	127	118	111
0.8	327	322	318	129	120	113
0.9	329	324	319	132	123	115
1.0	332	328	323	138	129	120

A similar interpretation can be applied to the results reported in Table 2. For example, using the 0.20 value of α again, there is an 80% probability that ending weights will exceed 316 kg when 300 steers are purchased, 311 kg when 325 steers are purchased, and 306 kg when 350 steers are purchased. Identical quantities of forage are available for the common α value, but the lower stocking level allows a greater number of days of grazing, with the subsequent gain in animal weight.

Net returns are sensitive to α as well (Fig. 3). If the rancher is
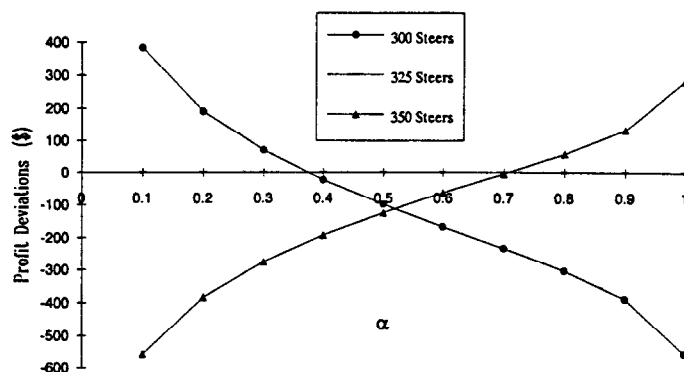

Fig. 3. Deviations in mean profits associated with choosing 300 and 350 steers relative to 325 steers at different values of the α probability parameter determining forage availability.

conservative, the lower stocking level is preferred. Up to an α value of about 0.38, purchasing 300 head will provide the highest mean net return, calculated over the 30 precipitation observations. For those who prefer more risk, the higher stocking rate is preferred since net returns are greater for the 350 head option when α exceeds 0.70. Purchasing 325 head is preferred for values of α between 0.38 and 0.70.

Summary

Many field studies have been used to find associations among variables of interest, such as forage production and animal weight gains, and conditioning environmental and management variables. Ordinary regression techniques are usually applied to determine these relationships (Bransby et al. 1988, Hart et al. 1988a, Hart et al. 1988b). However, concentration upon expected values of the dependent variables ignores much of the distributional information available in the data. Confidence intervals may be reported for the fitted values of the dependent variables, but these confidence

intervals impose normality on the regression error terms, often incorrectly.

Recent work in the simulation of agricultural systems has concentrated on directly estimating cumulative probability distributions for dependent variables, conditional upon 1 or more levels of variables assumed to be important in affecting the shape of the distribution functions. This paper has described the use of one such distribution, the hyperbolic tangent function, in providing distributional information on forage yields conditional on annual precipitation levels. The hyperbolic tangent function is particularly useful for determining expected values of the dependent variables, yielding values similar to regression fitted values. In addition, the function allows numerical integration techniques to be applied for determining second, third, and higher order moments of the distribution.

A simple stocker problem was used to illustrate the usefulness of the procedure in providing distributional information resulting from alternative management decisions. The procedures can, however, be equally valuable for hypothesis testing based on examination of the estimated conditional probability functions of the dependent variables resulting from different levels of the environmental and management conditioning variables.

This application abstracts from the complex interaction of many random variables that affect animal production in an extensive range operation. It is no easy matter to introduce multivariate distributions when 2 or more of the important variables are random and are correlated. *Ad hoc* procedures have been proposed (Condra and Richardson 1978). More recently, Moss et al. (1991) proposed a more rigorous procedure to impose multivariate distributions in agricultural simulation models. However, since the purpose of this paper is to present estimation procedures for conditional probability density functions, some accuracy in the example is foregone to emphasize both the procedures and the potential gains from using conditional distributions in range studies.

Literature Cited

Bransby, D.I., B.E. Conrad, H.M. Dicks, and J.W. Drane. 1988. Justification for grazing intensity experiments: analyzing and interpreting grazing data. J. Range Manage. 41:274–279.

Greene, W.H. 1990. Econometric analysis. Macmillan Publ. Co., N.Y.

Hart, R.H., M.J. Samuel, P.S. Test, and M.A. Smith. 1988a. Cattle, vegetation, and economic responses to grazing systems and grazing pressure. J. Range Manage. 41:282–286.

Hart, R.H., J.W. Waggoner Jr., T.G. Dunn, C.C. Kaltenbach, and L.D. Adams. 1988b. Optimal stocking rate for cow-calf enterprises on native range and complementary improved pastures. J. Range Manage. 41:435–441.

Moss, C.B., O.A. Ramirez, and W.G. Boggess. 1991. Modeling nonnormality in multivariate distributions using an inverse hyperbolic sine transformation to normality. Food and Resource Econ. Dep. Staff Pap. 91–16. Univ. Florida, Gainesville.

Nelson, C.H., and P.V. Preckel. 1989. The conditional beta distribution as a stochastic production function. Amer. J. Agr. Econ. 71:370–378.

Richardson, J.W., and G.D. Condra. 1978. A general procedure for correlating events in simulation models. Working paper, Dep. Agr. Econ., Texas A&M Univ., College Station.

Riechers, R.K., J.R. Conner, and R.K. Heitschmidt. 1989. Economic consequences of alternative stocking rate adjustment tactics: a simulation approach. J. Range Manage. 42:165–171.

Rodriguez, Abelardo, and L. Roy Roath. 1987. A dynamic programming application for short-term grazing management decision. J. Range Manage. 40:294–298.

Sneva, Forrest A., and D.N. Hyder. 1962. Estimated herbage production on semiarid ranges in the intermountain region. J. Range Manage. 15:88–93.

Taylor, C.R. 1984. A flexible method for empirically estimating probability functions. West. J. Agr. Econ. 9:66–76.

Taylor, C.R. 1990. Two practical procedures for estimating multivariate nonnormal probability density functions. Amer. J. Agr. Econ. 72:210–217.

Vantassell, L.W., R.K. Heitschmidt, and J.R. Conner. 1987. Modeling variation in range calf growth under conditions of environmental uncertainty. J. Range Manage. 40:310–314.