CONFIDENCE REGIONS FOR
LINEAR REGRESSION

PAUL G. HOEL
UNIVERSITY OF CALIFORNIA, LOS ANGELES

1. Introduction

It is well known that Student's t-distribution yields a confidence interval for the ordinate, y, of a regression line corresponding to any fixed valued of x, under the assumption that the sample x's are fixed variates and the corresponding sample ordinates are independently normally distributed about the regression line with a common variance. Less well known is the result of Hotelling and Working [1] in which a confidence band is obtained for the entire regression line, although with the additional assumption that the common variance of the sample ordinates is known.

Confidence bands are a particularly useful tool in those sampling problems that produce an estimate of a curve, such as a growth curve. Very often such curves can be treated as special cases of linear regression in several variables. It is not difficult to extend the methods employed in [1] to linear regression in several variables and thus obtain confidence bands for such curves.

In attempting to obtain a confidence band, it is desirable to seek for one that is as narrow as possible, in some sense, over the range of interest. The confidence band obtained in [1] was derived with mathematical convenience in mind, rather than with optimum properties dominant. The purpose of this paper is to derive confidence bands from an optimum point of view and to study the extent to which the confidence band of [1] is optimum. For simplicity of explanation, the discussion will be limited mostly to the regression line; however a generalization to linear regression in several variables is straightforward.

2. General confidence bands

This section will be concerned with deriving the equations that define a fairly general confidence band for a regression line. Consider a fixed set of x's: $x_1, x_2, \ldots, x_n$. Let y_i corresponding to x_i be normally distributed with mean $a + \beta(x_i - \bar{x})$ and variance σ^2, and let the y_i be independently distributed. It will be assumed that σ^2 is known; however in a later section this restriction will be removed. Let

$$
(1) \qquad a^* = \bar{y}, \qquad \beta^* = \frac{\sum (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2},
$$

$$
u = \frac{\sqrt{n}\,(a - a^*)}{\sigma}, \qquad v = \frac{\sqrt{n}\,s_1\,(\beta - \beta^*)}{\sigma},
$$