Colorado State University

SEMINAR IN STATISTICS

Thursday, July 9, 2009
10:00 a.m.
Statistics 008

Confidence Regions for Level Curves and A Limit Theorem for the Maxima of Gaussian Random Fields

Joshua French
Ph.D. Candidate, Dissertation Defense
Department of Statistics
Colorado State University

Abstract

One of the most common display tools used to represent spatial data is the contour plot. Informally, a contour plot is created by taking a "slice" of a three-dimensional surface at a certain level of the response variable and projecting the slice onto the two-dimensional coordinate-plane. The "slice" at each level is known as a level curve.

Consider a Gaussian random field $\{Z(s) : s \in D\}$, where s is the location in the continuous two-dimensional region of interest $D \subset \Re^2$. The level curve for the process Z at level u is defined to be $I_u = \{s : Z(s) = u\}$. From the observed data $z(s_1), \dots, z(s_N)$, one can predict $\hat{Z}(s)$ for any location $s \in D$ using kriging or some other procedure, and then construct the estimated level curve $\hat{I}_u = \{s : \hat{Z}(s) = u\}$ as an approximation of I_u. We present two methods for constructing confidence regions for the level curves of a contour plot.

The first method proposed is an extension of Lindgren (1995) and Wameling (2003) based on level crossings. A series of rectangular confidence regions are constructed along $\hat{I}_u$ which should individually intersect the true level curve with high confidence. The boxes extend in directions perpendicular to the estimated level curve and the widths of the boxes are chosen so that the edge of each box touches the neighboring box and there are no gaps between the boxes along the estimated level curve. The heights of the boxes are chosen by simulating realizations of $Z(s) \mid z(s_1), \dots, z(s_N)$, and then taking the appropriate quantiles of the set of nearest level crossings for the realizations. The heights of the boxes give insight into the approximate distance between the estimated level curve and the true level curve.

The second method constructs a confidence region for I_u through hypothesis testing, adjusting the critical value to control the simultaneous Type I error rate. Our goal is to construct a confidence region S for the true level curve such that $P(I_u \subseteq S) = 1 - \alpha$. Instead of finding S

directly, we adopt a different approach and find a set *R* which does not intersect I_u with high confidence, so that $P(\{R \cap I_u\} = \emptyset) \geq 1 - \alpha$. Consequently, the set $S = R^c$ will satisfy our goal since $P(S \subseteq R^c) \geq 1 - \alpha$. The region *R* is constructed by testing $H_0 : Z(s) = u$ versus $H_a : Z(s) \neq u$, and taking *R* to be the union of all *s* for which we conclude that $Z(s) \neq u$. Using kriging, we construct a test statistic which has a standard normal distribution. The critical value is adjusted to control the simultaneous Type I error rate through empirical simulation of the test statistic.

We conclude by introducing a limit theorem for the distribution of the maxima of a triangular sequence of stationary Gaussian random fields. The result is an extension of the work presented by Hsing (1996) to two-dimensions. The result was motivated by the desire to control the simultaneous Type I error rate of hypothesis tests at locations on an $n \times n$ lattice where the test statistics are Gaussian and correlated. Under certain dependence and limiting conditions we show that the maximum of the random fields exhibits extremal clustering in the limit. Consideration is then given to the use of this result in approximating $P(\max_{1 \leq i,j \leq n} Z_{i,j} \leq u)$, where $\{Z_{i,j}\}$ is a stationary Gaussian random field on a square $n \times n$ lattice of equally spaced locations.

<u>Advisory Committee</u>

Dr. Richard A. Davis (Advisor)
Dr. F. Jay Breidt (Committee Member)
Dr. Daniel S. Cooley (Committee Member)
Dr. Robin M. Reich (Forest, Rangeland and Watershed Stewardship, Outside Member)