Normal Surface Distribution; and Circular Disc of Confidence

Ali Astaneh, Vancouver, B.C.

The main objective of this article is to present a short account of normal surface distributions, as an extension of normal curve distribution theory in one dimensional statistics. One can extend the notions of the mean, the standard deviation, z-scores, the normal curve, the confidence interval, and … , to the case of a set of data in the form of a set of ordered pairs corresponding to a set of points on the coordinate plane.

To begin with, given a set of data $\{(x_1,y_1),(x_2,y_2),\ \ldots,\ (x_n,y_n)\}$ of ordered pairs, one can define the mean μ and the standard deviation σ for this set of data as follows

$\mu = (\ \overline{x}\ ,\ \overline{y}\ )$, where $\overline{x}$ and $\overline{y}$ are the means of the measurements $x_1,x_2,\ \ldots\ ,\ x_n$ and $y_1,y_2,\ \ldots,\ y_n$ respectively. The standard deviation can be defined as $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2+(y_i-\overline{y})^2}$.

Then one can present the following as the defining function defining the proposed ‘Normal Surface’,

$$f(x,y) = \frac{1}{\pi}\ e^{-x^2-y^2}\ ,$$

Observe that, as required for the proposed extension we have

$$\frac{1}{\pi}\iint e^{-(x^2+y^2)}dxdy = 1$$

Then, in statistical terms, one can use the integral

$$\frac{1}{\pi}\iint_{\sqrt{x^2+y^2}\leq Z} e^{-(x^2+y^2)}dxdy$$

to keep track of z-scores over the disc of confidence defined by

$$\sqrt{x^2+y^2} < z$$

as the natural generalization of confidence interval in one dimension.

One can then proceed and build up a theory for this two dimensional version of the statistical theory , say parallel to that of approximating binomial distributions by normal curves, such as circle of confidence , or circular disk of confidence. Interestingly the two dimensional version of Chebychev's inequality in the case of normal surface distribution actually turns out to be an equation. I will first recall Chebychev's inequality in the one dimensional case as follows.

Let X be a random variable defined on a set of single valued data with mean μ and standard deviation σ. Then for each positive number k,

$$P(\mu - k\sigma \le X \le \mu + k\sigma) \ge 1 - \frac{1}{k^2}$$

The following lemma shows that the two dimensional version of Chebychev's inequality for normal surface distributions is indeed an equation. The first part of the lemma can be shown by using a polar coordinate substitution in the appropriate double integral, and the other two parts follow from the first.

Lemma

(i) For a given set of data consisting of ordered pairs of measurements, which are normally distributed on the plane with mean $\mu = (\bar{x}, \bar{y})$ and standard deviation σ as defined above, for any positive number k,

$$P(|X - \mu| < k\sigma) = 1 - \frac{1}{e^{(k\sigma)^2}} \quad .$$

Here, $|X - \mu|$ simply means the distance between $X(x,y)$ and $\mu = (\bar{x}, \bar{y})$ on the coordinate plane.

(ii)For any positive number k, $100\,[1 - \frac{1}{e^{(k\sigma)^2}}]\,\%$ of data are within k standard deviation from the mean $\mu = (\bar{x}, \bar{y})$.

(iii) As a converse statement to (ii), for any given positive number R, $R\%$ of data are within $\frac{1}{\sigma}\sqrt{\ln(1 - R/(100)}$ standard deviation from the mean.

And finally, it should be obvious that the above extension to two dimension can be generalized to the case of set of data represented in n dimension.