Link between Gaussian anamorphosis and the rank-based inverse normal transform

August 2, 2016

1 Definitions

- $p(x)$ probability density function (PDF),
- $g(\chi)$ special case of a PDF (Gaussian, mean 0, variance 1):

$$g(\chi) = \frac{1}{\sqrt{2\pi}} \exp -\frac{\chi^2}{2}, \tag{1}$$

- $P(x)$ cumulative density function (CDF) of $p(x)$:

$$P(x) = \int_{x'=-\infty}^{x} p(x')dx', \tag{2}$$

- $G(\chi)$ CDF of $g(\chi)$:

$$G(\chi) = \int_{\chi'=-\infty}^{\chi} g(\chi')d\chi'. \tag{3}$$

2 Gaussian anamorphosis

A Gaussian anamorphosis transforms a random variable, x, which is distributed according to an arbitrary 1D PDF, $p(x)$ into a new variable, χ, which is distributed according to a Gaussian (here $g(\chi)$). The requirement is that there is a mapping between x and χ such that:

$$p(x)dx = g(\chi)d\chi, \tag{4}$$

subject to the boundary conditions that $p(-\infty) = 0$ and $g(-\infty) = 0$. Integrating each side of (4) from this boundary gives $\int_{x'=-\infty}^{x} p(x')dx' = \int_{\chi'=-\infty}^{\chi} g(\chi')d\chi'$, i.e. that:

$$P(x) = G(\chi). \tag{5}$$

This can be made into the Gaussian anamorphosis transform:

$$\chi = G^{-1}\left(P(x)\right), \tag{6}$$

where G^{-1} is the inverse Gaussian CDF function.

3 Rank-based inverse normal transform

Consider a sample of ordered variables represented by the set X:

$$\begin{aligned} X &= \{x_1, x_2, \ldots, x_N\}, \\ \text{where } x_{i+1} &\geq x_i. \end{aligned} \tag{7}$$

Suppose that members of X represent possible draws from $p(x)$, then $p(x)$ may be represented by the following sum of Dirac delta-functions:

$$p(x) \sim \frac{1}{N}\sum_{i=1}^{N}\delta(x - x_i), \tag{8}$$

(i.e. the closer the delta-functions, the denser the PDF). The CDF may therefore be represented by:

$$\begin{aligned} P(x) &\approx \int_{x'=-\infty}^{x}\frac{1}{N}\sum_{i=1}^{N}\delta(x'-x_i)dx', \\ &= \frac{1}{N}\sum_{i=1}^{N}\int_{x'=-\infty}^{x}\delta(x'-x_i)dx', \\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(x_i \leq x). \end{aligned} \tag{9}$$

Here $\mathbf{1}(x_i \leq x)$ (the indicator function) is defined by:

$$\begin{aligned} \mathbf{1}(x_i \leq x) &= \int_{x'=-\infty}^{x}\delta(x'-x_i)dx', \\ &= \tilde{r}_i = \begin{cases} 1 & \text{if } x_i \leq x \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{10}$$

For an arbitrary x, the indicator function returns a quantity called the relative rank, $\tilde{r}_x$. This is an integer between 0 and N essentially informing where amongst the sample x lies, e.g. if x lies between x_i and x_{i+1} (strictly $x_i \leq x < x_{i+1}$) then $\tilde{r}_x = i$. Using (10) (and calling it $\tilde{r}_x$) with (9) and (6) leads to the following special 'rank' form of the anamorphosis:

$$\chi = G^{-1}\left(\frac{\tilde{r}_x}{N}\right). \tag{11}$$

Compare this to the following previously documented rank-based inverse normal transform - see [Chipilski (2016)] and references therein:

$$\chi = G^{-1}\left(\frac{r_x - c}{N - 2c + 1}\right), \tag{12}$$

where c is an empirical constant, $c \in [0, 0.5]$ and r_i is a slightly different definition of the relative rank used in (11). In [Chipilski (2016)], r_i is defined as $r_i = i + 1$ when $x_i < x \leq x_{i+1}$ (assuming here for simplicity no repeated values in the sample), which is different from the definition of $\tilde{r}_i = i$ when $x_i \leq x < x_{i+1}$.

Consider the following limits:

- When N is large, i.e. $N \gg -2c + 1$, the denominator of (12) is approximated by N.
- When $r_x \gg c$ the numerator is approximated by r_x (when N is large, occurrences where this condition is not met are very unlikely).
- When N is large, $x_{i+1} \approx x_i$, and so $\tilde{r}_i \approx r_i$.

Under these conditions (i.e. most practical situations) (11) and (12) are practically equivalent.

References

[Chipilski (2016)] Chipilski H.G., The search for Gaussian moisture variables at the convective scale, MMet dissertation, Univ. of Reading, 2016.