Development of Data-adaptive Probability Plotting and its Applications

Masashi Goto
Department of Informatics and Mathematical Science, Graduate School of Engineering Science, Osaka University.
Machikaneyama-cho 1-3.
Toyonaka, Osaka 560-8531, Japan.
gotoo@sigmath.es.osaka-u.ac.jp

Toshio Shimokawa
Department of Informatics and Mathematical Science, Graduate School of Engineering Science, Osaka University.
Machikaneyama-cho 1-3.
Toyonaka, Osaka 560-8531, Japan.
simokawa@sigmath.es.osaka-u.ac.jp

1. Introduction

In the statistical data analysis process, the data investigation is most basic step and important step. The importance of the data investigation is emphasised and illustrated as initial examination (Cox & snell, 1981; Chat¯eld, 1985,1991; Goto,1986). In the step, the statistical graphics is one of the leading methods. However, ordinary ways of the data investigation have not included only the statistical graphics, but also formal analytical techniques. In the presentation, we try to conduct whole step of the data investigation based only on graphical techniques, and especially focus on graphical diagnosis. Then, as an elementary tool of our graphical diagnosis, we propose the data-adaptive probability plotting(DAPP). This plotting does not provide only distributional information of ordinary plobability plottings, but also shape of empirical distributions. Moreover, we can construct some guardrails on DAPP for providing inferential interpretation of the result of DAPP.

2. Data-adaptive probability plotting

In DAPP, it is assumed that observations have one of data-adaptive distributions. Especially, we propose power-normal Q-Q plotting using the power-normal distribution (Goto et al.,1983). DAPP contains of the transformation curve, which is described based on quantile of the data-adaptive distribution. The shape (transforming) parameters are changed associated with di®erence of shapes of two distribution, until the transformation curve provides the best ¯tting to the data points.

In the presentation, we construct some guardrails on DAPP. The purpose of the guardrails

Figure 1. Example for DAPP.

to support to diagnose in°uence of sample variation and adequacy model (distribution).
Example:The data is the blood suger determinations measured by two methods, glucose oxidase method (g.o.m.) and autoanalyzer method (a.m.) (Kashiwagi, 1979). Figure 1 shows the power-normal Q-Q plotting for this data, with g.o.m. on the vertical axis, and a.m. on the horizontal axis. Let $\tilde{\lambda}_x$ and $\tilde{\lambda}_y$ be the shape parameters on horizontal and vertical axises, respectively. For instance, we put the initial shape parameter $\tilde{\lambda}_x^{(0)} = \tilde{\lambda}_y^{(0)} = 1{:}00$. The transformation curve does not ¯t the data points. Taking notice of di®erences of shapes, we dynamically changed $\tilde{\lambda}_x$ and $\tilde{\lambda}_y$. In case of $\tilde{\lambda}_y = 0{:}79$ and $\tilde{\lambda}_y = ¡\,1{:}70$, the transformation curve provided the best ¯tting to the data points. As a result, the power-normal Q-Q plotting showed that both empirical distributions skewed to left and g.o.m. had more skewer distribution than a.m.

3. Conclusion

In the presentation, we proposed DAPP. Then, we have constructed the guardrail on DAPP for inferential interpretation of the result of DAPP. Some useful aspects of DAPP have been assessed in several practical examples (Shimokawa et al., 2000). As a result, we can say DAPP is °exible and interpretable tool than ordinary probability plottings.

REFERENCES

Cox, D.R. & Snell, E.J. (1981). Applied Statistics, Chapman and Hall.

Chat¯eld, C. (1985). The initial examination of data. J. Roy. Statist. Soc., A143, 214-253.

Chat¯eld, C. (1991). Avoiding statistical pitfalls(with discussion). Statistical Science, 6, 240-268.

Goto, M., Matsubara, Y. and Tsuchiya, Y. (1983). Power-normal distribution and its applications. Rep. Stat. Appl. Res., 30(3), 8-28.

Goto, M. (1986). The process of statistical data analysis. The Japanease Journal of Behaviormetrics(in Japanese), 13(2), 48-63.

RESUME

Nous avons exloitµa un plot de la probabilitµa adaptatif des donnµees ("data-adaptive probability plotting") pour e®ectuer l'analyse des donnµas statistiques avec le graphique statistique seulement. De plus, nous avons constituµa un rail de sµacuritµa sur ce plot de la probabilitµa adaptatif des donnµees pour interpreter le point des donnµees de faxon objective.