The probability distribution of a forecasted extreme

William J.J. REY
Philips Research Laboratories
Prof. Holstlaan 4 (WY81)
5656JX Eindhoven, Netherlands
william.rey@philips.com

Keywords: Bootstrap, Empirical Extreme, Integrated Circuit.

1. Problem statement

An integrated circuit has 1024 sensors and each of these sensors is characterised by a delay, x_i. To be functional, all sensors must present delays that are below a certain limit X.

During the design phase of the circuit, one will estimate the percentage of circuits that fail due to a too large $\max\{x_1, ..., x_{1024}\}$ where each of the x_i-delays can be seen as randomly drawn from a population P. The population P is little known; only a moderate size simulation gives a sample $\{x_1, ..., x_n\}$, $n = 101$; the data set is available on request.

The difficulty is here that the sample is so small that the maximum can be far away from the observed population. Formally, the problem is the estimation of

$$\text{Yield}(X) = Prob\left\{\max\{x_1, ..., x_{1024}\} \leq X \mid \text{Given sample } x_1, ..., x_{101}\right\}.$$

2. A (simple-minded) treatment

A two-step solution is the obvious approach: First, estimate $F_1(X)$, the distribution of the sample $\{x_1, ..., x_{101}\}$ and, second, evaluate $F_{1024}(X) = [\ F_1(X)\ ]^{1024}$, the distribution of the maximum.

Many distributions have tails that look *exponential and the following discussion applies to such a distribution* as can be seen from plots of sample $\{x_1, ..., x_{101}\}$. This quotation, very slightly adapted from Andrews (1973), does not cover the full story. Not only you could think at different sorts of distributions, but you could also wonder how far the distribution model can be formally tested. Also, we observe that the distribution of $\max\{x_1, ..., x_{1024}\}$ only depends on the right tail of $F_1(X)$, the distribution of x_i.

As we do not know $F_1(X)$, we estimate it by

$$\hat{F}_1(X) = \frac{[\#(x_i \leq X)] - h}{101 + 1 - 2\,h}, \qquad h = \frac{1}{2}$$

here written in the usual notation; by simulation on a true negative exponential model, we saw that $h = \frac{1}{2}$ yields the less biased estimate of the tail.

For sake of economy in number of parameters, we adopt the negative exponential model. Thus, we assume that $\hat{F}_1(X)$ is a realisation of $F_1(X)$ with

$$F_1(X) = 1 - e^{-g(X)} \quad \text{and} \quad g(X) = a + b\,X.$$

Seeing the plot of $\hat{g}(X) = -\ln[1 - \hat{F}_1(X)]$, the dashed line of Figure 1, the linear assumption in full line appears to make sense. The fit $\hat{g}(x) = \hat{a} + \hat{b}\;X$ is here least squares on the upper quarter of the observations, i.e. for $\hat{F}_1(X) > \frac{3}{4}$; the estimate of F_{1024} is little sensitive on this cut-off value $\frac{3}{4}$. $\hat{g}(X) \approx -\ln[1 - \hat{F}_1(X)]$ greatly depends on sample $\{x_1, ..., x_{101}\}$ and this leads us to bootstrap the sample; remark that the objections of Deheuvels et al. (1993) do not apply, we are bootstrapping the full sample and not the distribution of the maximum.

Eventually, the evaluation of $\hat{F}_{1024}$ is derived from a bootstrap procedure under a model of negative exponential tail. We generate each drawing of $\hat{F}_{1024}$ by generating a bootstrap sample $\{x_1^\star, ..., x_{101}^\star\}$, estimating $F_1^\star(X)$ as $1 - \exp[-(\hat{a}^\star + \hat{b}^\star\;X)]$ and drawing at random a single variate from $\hat{F}^\star_{1024}(X) = [\hat{F}_1^\star(X)]^{1024}$. Figure 2 let see on the left the original distribution $\hat{F}_1$ (in full line) and, on the right, the forecast of $\hat{F}_{1024}$ in dashed line; this forecast is built up with the help of 999 drawings, each constructed from 101 $x_i^\star$-values.

Fig. 1. $-\ln[1 - \hat{F}_1(X)]$ and $\hat{g}(X)$

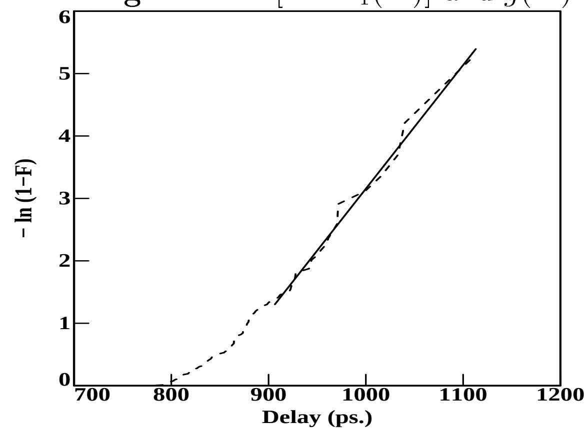
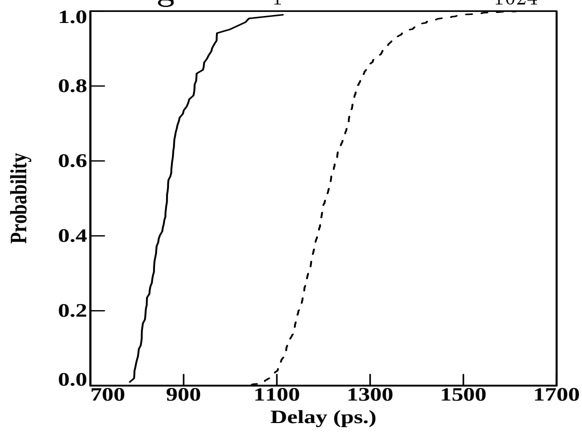


Fig. 2. $\hat{F}_1$ $\hat{F}_{1024}$



REFERENCES

Andrews, D.F. (1973). A general method for the approximation of tail areas. *Annals of Statistics*, **1**, 367-372.

Deheuvels, P., Mason, D.M. and Shorack, G.R. (1993). Some results on the influence of extremes on the bootstrap, *Annales de l'Institut Henri Poincaré - Prob. & Statist.*, **29**, 83-103.

RESUME

Un circuit intégré comporte 1024 capteurs, chacun charactérisé par un retard x_i. En sorte d'être fonctionnel, chacun de ces retards doit être inférieur à un certain seuil X et on se demande quelle est la probabilité d'un tel événement. La difficulté réside dans le peu de connaissance quant à la population des x_i; on ne la connait que par un petit échantillon de taille 101. Formellement, on veut évaluer $Prob\Big\{\max\{x_1, ..., x_{1024}\} \leq X \mid \text{vu l'échantillon } x_1, ..., x_{101}\Big\}$.