Modelling Probability Distributions from Data and its Influence on Simulation

Hörmann, Wolfgang; Bayar, Onur

Modelling Probability Distributions from Data and its Influence on Simulation

Wolfgang Hörmann,Onur Bayar

Department of Applied Statistics and Data Processing
Wirtschaftsuniversität Wien

Preprint Series

Preprint 29
January 2000

http://statmath.wu-wien.ac.at/

MODELING PROBABILITY DISTRIBUTIONS FROM DATA AND ITS INFLUENCE ON SIMULATION

Wolfgang Hörmann and Onur Bayar
Bogazici University Istanbul
80815 Bebek-Istanbul, Turkey

This work was supported by the Austrian Academy of Science, APART scholarship.

Abstract. Generating random variates as generalisation of a given sample is an important task for stochastic simulations. The three main methods suggested in the literature are: fitting a standard distribution, constructing an empirical distribution that approximates the cumulative distribution function and generating variates from the kernel density estimate of the data. The last method is practically unknown in the simulation literature although it is as simple as the other two methods. The comparison of the theoretical performance of the methods and the results of three small simulation studies show that a variance corrected version of kernel density estimation performs best and should be used for generating variates directly from a sample.

Introduction.

It is well known that the choice of the input distribution is a crucial task for building a stochastic simulation model. If the inputs of the real system we are interested in are observable, it is possible to collect data. In this case the choice of the input distribution for the stochastic simulation model is a statistical problem, which can be called the modelling of probability distributions from data. The problem can be solved in a parametric approach by estimating the parameters of a suitable standard distribution or in a non-parametric approach by estimating the unknown distribution. We are convinced that due to its greater flexibility the non-parametric approach should be used unless there are profound a priori reasons (eg. of physical nature) favouring a certain standard distribution.

In stochastic simulation we are interested not only in estimating the input distribution but also in generating random variates from that distribution. This task is called ”generating variates from empirical distributions” or ”generalising a sample” in the simulation literature (see eg. [1] and [5]). As these names indicate, the problem of estimating (or modelling) the input distribution is often hidden behind a procedure to generate random variates from data. Perhaps that is the reason that no comparison of the quality of the estimation of the different methods was done till now, allthough there is a developed statistical theory discussing the optimal estimation of densities. Especially kernel density estimation is well suited for modelling input distributions, as variate generation from these estimates is very simple. This was already observed in the monographs [4], [2] and [6] but seems to be widely unknown in the simulation literature.

Therefore this paper compares the theoretical properties of these different methods of generating random variates from data and will demonstrate with simple examples that the choice of the method can have an influence on simulation results.

Sampling from Empirical Distributions

We are given a random sample of size n, denoted by $X_1, X_2, \ldots, X_n$. s will denote the sample standard deviation. Of course the simplest method of sampling from the empirical distribution is **naive resampling**. We just take randomly numbers of the sample. If the sample is based on a continuous random variable this method has the obvious drawback, that only a small number of different values can be generated.

To overcome these problems two well known simulation text-books ([1] and [5]) suggest to use a linear interpolation of the empirical cumulative distribution function (CDF) for generating random variates. The algorithm suggested in [5] (we shall call it **ELK** in the sequel) is only generating points between the minimum and maximum of the sample, whereas the algorithm suggested in [1] (called **EBFS**) in this paper) uses an exponential tail on the right hand side of the sample. Both algorithms are simple to implement.

There is another simple adaptation of naive resampling called smoothed bootstrap in the statistic literature. Do not only resample but add to any of the resampled numbers some noise, ie. a continuous random variable with 0 expectation and small variance. It is not difficult to see, that smoothed bootstrap is the same as generating random variates from a density-estimate by using the kernel method, but it is not even necessary to compute the estimated density.

Algorithm **KDE**: (Kernel Density estimation)

```
(0) Set-up: Choose the smoothing parameter $b$ (see below for the formula).
(1) Generate a random integer $I$ uniformly distributed on $(1, 2, \ldots, n)$
(2) Generate a random variate $W$ from the noise distribution
(3) Return $Y = X_I + bW$
```

The density of the random noise distribution W is called kernel and will be denoted by $k(x)$. Clearly $k(x)$ must be a density function and should be symmetric around the origin. As we want to change the variance of the random noise we introduce the scale parameter b (called bandwidth or smoothing parameter in density estimation); the random variable bW has the density $k(x/b)/b$. The random variate Y generated by Algorithm KDE is the equiprobable mixture of n noise distributions, each centered around one of the sample points. This implies that the density of Y (denoted f_Y) is the sum of n translated versions of $k(x)$ multiplied with $1/n$. f_Y is the kernel density estimate of the unknown distribution and is called $\hat{f}$ in the literature.

$$f_Y(x) = \frac{1}{nb} \sum_{i=1}^{n} k\left(\frac{x - X_i}{b}\right)$$

Of course there remains the question of the choice of the bandwidth b and the kernel function $k(x)$. Here we can use the results of the theory of density estimation as presented eg. in [6] or [7]. To minimise the mean integrated squared error we use a very simple and robust variant of estimating the optimal bandwidth b as given in [6]. $b = \alpha(k)\, 1.364\, \min(s, R/1.34)\, n^{-1/5}$,
where the constant $\alpha(k)$ is 0.776 for the Gaussian and 1.351 for the rectangular kernel respectively. s denotes the standard deviation and R the interquartile range of the sample. There are lots of much more complicated ways to determine b published in literature. For an overview see [3], where the L_1-error (ie. the mean integrated absolute error) of many different bandwidth selection procedures is compared. The method we use is a mixture of the methods called "reference: L_2, quartile" and "reference: L_2, std. dev" in [3]. The results of the simulation study show that with the exception of some very strangely shaped multimodal distributions the performance of this very simple choice of b is not bad. And we are not interested in an optimal estimation of the density here but in constructing an empirical distribution that is "as close as possible" to the theoretic distribution in all aspects.

The last question that has to be solved before we can use Algorithm KDE is the choice of the kernel. Asymptotic theory shows that the MISE is minimal for the Epanechnikov kernel $f(x) = (1 - x^2)3/4$ but some other kernels have allmost the same efficiency. Therefore we can choose the kernel by also considering other properties, eg. the speed and simplicity of our generation algorithm. In that respect the rectangular kernel (ie. uniformly distributed noise) is of course the best choice, but it has the theoretical draw-back that the estimated density is not continuous. Due to the nice statistical interpretation we prefer Gaussian noise and will use it in the sequel.

Algorithm KDE guarantees that the density function of the empirical distribution approximates the density of the unknown true distribution as good as possible with respect to the mean integrated squared error. On the other hand we clearly see, that for algorithm KDE the variance of the empirical distribution is always larger than the variance of the observed sample. This can be a disadvantage in simulations that are sensitive against changes of the variance of the input distributions. To overcome this problem it is possible to force the empirical distribution to have the same variance as the sample in the following way (suggested in [6]).

Algorithm **KDEVC**:

```
(0) Set-up: Compute the mean $\bar{x}$, the standard deviation $s$ and the interquartile range $R$ of the sample.
    Compute $b = \alpha(k)\, 1.364\, \min(s, R/1.34)\, n^{-1/5}$
(1) Generate a random integer $I$ uniformly distributed on $(1, 2, \ldots, n)$
(2) Generate a random variate $W$ from the noise distribution
(3) Return $Y = \bar{x} + (X_I - \bar{x} + bW)/(1 + b^2\sigma_k^2/s^2)^{1/2}$ ($\sigma_k^2$ denotes the variance of the kernel)
```

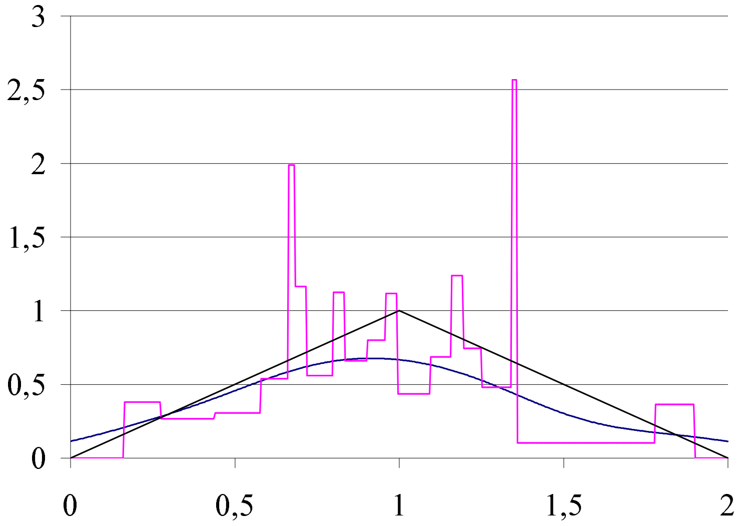

Figure 1: A triangular density with the empirical densities of ELK (step function) and KDE

Remark: Positive random variables are interesting for many applications. Method KDE can cause problems for such applications as it will also generate negative variates. The easiest way out is the so-called mirroring principle. Instead of a negative number Y simply return $-Y$. Unfortunately the mirroring principle disturbs the variance correction. They can be used together but the resulting empirical distribution has a smaller variance than the sample. This can only be a practical problem if the sample of a positive distribution has many values close to zero.

Comparison of Methods

Expectation and Variance

An important concern for many simulations is the expectation and the variance of the input distribution. The best we can hope to reach is that the expectation and the variance of the empirical distribution are equal to the sample mean and sample variance of the observed sample as these are the best available estimates for the unknown values. All methods described above produce random variates that have as expectation the sample mean. Only for ELK the result is slightly different.

Concerning the variance the situation is more complicated: For kernel density estimation we know that the variance of the empirical distribution is larger than the sample variance. A simple calculation shows that $V(KDE) = s^2((n-1)/n + b^2\sigma_k^2)$ which is $s^2(1 - 1/n + 1.058\,n^{-2/5})$ for the Gaussian kernel and our choice of b. For eg. $n = 100$ the variance factor $V(\text{emp. distr})/s^2 = 1.164)$ which shows that it may be wise to consider the variance corrected version of the algorithm KDEVC which has by design a factor of one for any sample size. A second possibility to reduce the variance factor of Algorithm KDE is the use of a smaller bandwidth b. For the limiting case $b \to 0$ Algorithm KDE coincides with naive resampling and thus has a variance factor of $(n-1)/n$.

For the two methods based on linear approximation of the CDF (ELK and EBFS) there is no simple formula for the variance of the empirical distribution as the variance depends on the sample. So we computed the variance of the empirical distribution in a small simulation study. Our results show that for ELK the variance of the empirical distribution is always smaller than the sample variance, for EBFS the variance is due to the added tail always bigger. The factor $V(\text{emp. distr})/s^2$ is strongly influenced by the shape of the theoretic distribution. For samples of size 100 we observed factors up to 1.12 for EBFS and down to 0.91 for ELK.

The fine structure of the empirical distribution

Thanks to L. Devroye there exists a theoretical result about the quality of the local approximation of the unknown density by the methods ELK and EBFS. He showed (see [1] p. 132) that for n towards infinity the density of the empirical distribution does not even converge against the correct distribution. In contrast to this poor behaviour we know that for method KDE the estimated density converges and we even have (approximately) minimised the mean integrated squared error. For method KDEVC the optimal approximation of the density is slightly disturbed by the correction of the variance. Nevertheless we know that asymptotically KDEVC has very nice approximation properties because it coincides with KDE. As this theoretical argument seems to be unimpressive for simulation practitioners we try to illustrate the consequence of this theoretical result in Figure 1. It compares for a "well behaved" sample of size 20 (from a triangular distribution) the empirical density of method ELK (which is practically identical with EBFS) and of method KDE. Looking at Figure 1 we can also understand that the high peaks in the ELK-density occur when two sample points are comparatively close together and this happens in practically all samples.

The distance between the theoretical and the empirical distributions

There are different distance measures suggested in the literature to compute the distance between two distributions. In density estimation the L_2-difference (integrated squared difference) and the L_1- difference (integrated absolute difference) are of major importance. Another possibility is to use the L_1-difference or the L_2-difference between the two CDFs. As method KDE is based on estimating the density whereas ELK and EBFS are based on an approximation of the CDF we thought that these four measures would favour automatically one group of the algorithms discussed here. Therefore we decided to use a third class of distance measures for our comparison: The test-statistics of three well known goodness-of-fit tests, the Chi-square test, the Kolmogorov-Smirnov test and the Anderson-Darling test. For the Chi-square test we took the number of equiprobable classes as $[\sqrt{m}]$. Then the test-statistic divided through m converges against $\int(\hat{f}(x)-f(x))^2/f(x)dx$ a weighted squared difference of empirical and theoretical density. ($\hat{f}$ denotes the density of the empirical distribution of the different methods.) These considerations clearly show that the chi-square test is more sensitive to deviations in the tails than to deviations in the centre of the distribution. The Kolmogorov-Smirnov test measures the maximal absolute difference between the empirical CDF and the theoretical CDF. Its test statistic is $D_m = \sup(|F_m(x) - F(x)|)$ (where F_m denotes the emprical CDF of the sample). The power of the KS-test for deviations in the tails is low. The Anderson-Darling test on the other hand was designed to detect discrepancies between the tails of the distributions. Like the KS-test it compares the theoretical and the empirical CDF but it uses a weighted L_2-difference to compare the CDFs. The test statistic is $A_m^2 = m\int(F_m(x)-F(x))^2 f(x)/(F(x)(1-F(x)))dx$

Then for each random sample of size $n = 100$ of the different theoretical distributions and for each of the five different methods we generated 40 different samples of the empirical distributions with size $m = 3000$. The computed the average test statistics for all these experiments and repeated them for 40 different samples of the theoretical distribution.

Table 1 gives the final average of the different test staistics. These results can be seen as a (stochastic) distance measure between the theoretical distribution and the empirical distribution. They are (like eg. the mean intagrated squared error) a measure for the deviation between empirical and theoretical distribution averaged over different samples from the theoretical distribution. We can see that all averages are in the critical region of the respective tests. This is not surprising. As the empirical distribution is based on a sample of size 100 only, a much larger sample ($m = 3000$) from the empirical distribution cannot have exactly the same properties as a sample from the correct distribution. In the chi-square and in the Kolmogorov-Smirnov tests methods KDE and KDEVC perform considerably better than EBFS, ELK and naive resampling. For the Anderson-Darling tests the differences are small but even there KDEVC performs best. The results do not only show that kernel density estimation performs better than the other methods, they also show that the variance corrected method performs in almost all cases better than the original version. We were astonished that for these three very different distance measures and for four quite different distributions the same method for constructing the empirical distribution is best or close to best in all cases. We think that this result is a strong argument in favour of method KDEVC.

We added the last column of Table 1 to compare the discussed methods with the method of fitting a standard distribution (FSD). Of course FSD performs best if we fit the correct distribution but Table 1 shows that KDEVC is in most cases not far away which means that we do not loose much in using KDEVC instead of a standard distribution. The two mixture distributions were chosen such that their shape is not far away from a standard distribution. If we assume – as we do for FSD – that the data come from a standard normal distribution with unknown parameters, we would estimate the parameters μ and σ and then conduct a chi-square test. The power of the test (the probability to reject the hypothesised normal distribution) is 0.5 if the unknown true distribution of the sample is our normal equiprobable mixture of N(0,1) & N(3,1) and the sample size is 100. Thus the poor results for fitting a normal distribution to the normal mixture are not artificial numbers. They have practical relevance as the power of the goodness-of-fit tests is often too low to show the deviation from the hypothesised standard distribution. For the gamma mixture we used a distribution, which is even closer to a gamma distribution. Only for 15 % of all samples of size 100 the chi-square test rejects the hypothesised gamma distribution. The distance measures show that the quality of the approximation of FSD and KDEVC is about the same. Of course it is no problem to find examples where FSD performs arbitrarily poor. Just take a theroetic distribution with a shape far away from any standard distribution.

Table 1: Average Test Statistics and standard errors (in brackets).

	Critical value 5%	EBFS	ELK	Naive resampling	KDE	KDEVC	Standard distribution
Theoretical distribution: Gamma(2)							
χ^2-mean(SE)	71.0	1218 (24)	1270 (25)	1642 (31)	319 (10)	241 (7)	110 (6)
KS-mean*1000	24.8	79 (2)	79 (2)	84 (2)	56 (2)	54 (2)	44 (2)
AD-mean	2.5	27 (2)	27 (2)	27 (2)	31 (2)	24 (2)	17 (2)
Theoretical distribution: Gamma mixture: G(2)&G(6)							
χ^2-mean(SE)	71.0	1196 (23)	1256 (24)	1651 (28)	268 (6)	230 (6)	220 (5)
KS-mean*1000	24.8	84 (3)	84 (3)	88 (3)	57 (2)	63 (2)	68 (2)
AD-mean	2.5	32 (2)	32 (2)	32 (2)	30 (2)	28 (2)	628 (2)
Theoretical distribution: Normal(0,1)							
χ^2-mean(SE)	71.0	1290 (26)	1240 (23)	1633 (30)	220 (12)	155 (7)	113 (7)
KS-mean*1000	24.8	82 (2)	83 (2)	87 (2)	59 (2)	54 (2)	46 (2)
AD-mean	2.5	30 (2)	30 (2)	31 (2)	30 (2)	22 (2)	19 (2)
Theoretical distribution: Normal Mixture: N(0,1)&N(3,1)							
χ^2-mean(SE)	71.0	1261 (29)	1229 (26)	1601 (33)	350 (18)	209 (9)	513 (7)
KS-mean*1000	24.8	80 (3)	80 (3)	80 (3)	62 (2)	65 (3)	92 (2)
AD-mean	2.5	32 (3)	31 (3)	31 (2)	35 (3)	26 (2)	44 (2)

Influence on Simulation results

Changing the method of modelling the empirical distribution is not more than changing the fine structure and perhaps slightly the variance of the input distribution of a simulation model. It is to be expected that many simulations, which have as ouput averages of a large number of input random variables, are not very sensitive to small changes in the fine structure of the input distribution. For example it is known that the average waiting time in the M/G/1 queue is only influenced by the expectation and the variance of the service time distribution and not by its shape. And it is even better known that the distribution of the sample mean of a large sample is always very close to normal. The parameters of that normal distribution are again only influenced by the expectation and the variance of the underlying distribution and not by its shape. These are arguments why the choice of the method will not have a big influence on many simulation results. Nevertheless we try to get some insight into this question by looking at three examples. The first simulation model we tried is the M/G/1 queue. The inter-arrival times are taken exponential with expectation 1, the service times are modelled from samples of different gamma distributions, using the different empirical methods described above. Then we simulated the model starting with an empty system and observed the average waiting time (AVW) and the maximal number in queue (MAXNIQ). We repeated this experiment for several different samples of the theoretical service-time distribution to get an average over different samples. We assumed in advance that this model is probably very stable with respect to small changes of the fine structure of the distribution but we tried it because of its importance and because we thought that the tail-modelling of the empirical distribution could have some influence on the results. The results given in Table 2 mainly show that there is little to choose between the different methods to fit an empirical distribution, all methods have about the same performance and rarely differ more than one standard error. The second interesting result is that the size of the error when using an empirical instead of the correct distribution strongly depends on ρ, the utilisation factor of the system. The results for $\rho = 0.4$ and $n = 100$ are better than those for $\rho = 0.9$ and $n = 500$.

Table 2: M/G/1-queue: Average Error and its Standard Error (in brackets)

	ρ	a of Γ distr.	n	KDE	KDEVC	EBFS	Naive Resampling	ELK
AVW*100	0.9	10	100	137(21)	137(21)	138(21)	138(21)	132(19)
MAXNIQ*100	0.9	10	100	340(41)	341(40)	341(42)	343(40)	332(37)
AVW*100	0.9	10	500	53(4)	52 (4)	52(4)	52(4)	53(4)
MAXNIQ*100	0.9	10	500	142(11)	137(10)	138(11)	140(11)	143(10)
AVW*100	0.4	2	100	3(0.3)	3(0.3)	3(0.3)	3(0.2)	3(0.2)
MAXNIQ*100	0.4	2	100	37(3)	37(3)	47(4)	36(3)	38(3)
AVW*100	0.4	2	500	1.5(0.1)	1.5(0.1)	1.5(0.1)	1.5(0.1)	1.5(0.1)
MAXNIQ*100	0.4	2	500	18(1)	19(1)	21(2)	17(1)	18(1)
AVW*100	0.4	10	100	1.0(0.1)	1.0(0.1)	1.0(0.1)	1.0(0.1)	1.0(0.1)
MAXNIQ*100	0.4	10	100	12(1)	11(1)	13(1)	11(1)	12(1)
AVW*100	0.4	10	500	0.5(0.03)	0.5(0.03)	0.5(0.03)	0.5(0.03)	0.5(0.03)
MAXNIQ*100	0.4	10	500	5(0.4)	6(0.4)	6(0.4)	5(0.4)	5(0.4)

Due to the very small differences between the methods for the M/G/1-queue we looked for simulation examples that are influenced by the fine structure of the distribution. So we tried the following: We take a sample of size 50 of a gamma distribution and compute the maximal and the minimal distance between two neighbouring points. What happens in that experiment if the gamma distribution is replaced by an empirical distribution constructed from a sample of size $n = 100$ or 500 of the correct gamma distribution? We repeated each experiment 10000 times and arrived at the results given in Table 3.

Table 3: Average minimal and maximal distances, (standard errors in brackets)

	n	EBFS	ELK	Naive resampling	KDE	KDEVC	Correct distrib.
Theoretical distribution: Gamma(2)							
min*105 (SE)	100	55(0.7)	54(0.7)	0 (0)	172(2)	168(2)	163(2)
min*105 (SE)	500	76(1)	77 (1)	12 (0.6)	173(2)	167(2)	163(2)
max *100 (SE)	100	196 (2)	125 (1)	147 (1)	138(1)	129(1)	150(1)
max *100 (SE)	500	172 (2)	139 (1)	148 (1)	146(1)	141(1)	150(1)
Theoretical distribution: Gamma(20)							
min*105 (SE)	100	209 (3)	199(3)	0 (0)	679(7)	624(6)	628(6)
min*105 (SE)	500	306 (4)	304 (4)	53 (3)	662(7)	628(6)	628(6)
max *100 (SE)	100	706 (5)	291 (1)	342 (2)	339(2)	310(2)	323(2)
max *100 (SE)	500	482 (4)	314 (2)	331 (2)	338(2)	324(2)	323(2)

The interpretation of Table 3 with respect to the minimal distance is simple. Naive resampling is useless if the fine structure of the distribution is of any importance, even though the sample generated from the empirical distribution had only size $m = 50$ whereas $n = 100$ or even $n = 500$ data points were available. The second observation is that the fine structure of EBFS and ELK are only slightly better whereas those of KDE and KDEVC are much better with results close to the results using the correct distribution. Interesting is the fact that the results of the variance corrected method are better than those of the standard method. If we look at the results for the maximal distance we see that naive resampling works better than expected. The results of EBFS are worse than expected, although EBFS assumes exponential tails, which should be an advantage. KDE and KDEVC again show good results.

Our last example can be interpreted as part of a computer-system simulation. Two processes work with the same file. They start at the same time and the time between two file-accesses follows the same distribution (gamma(10, 0.1)). Now we want to estimate the probability that the two processes try to access the file at "almost the same time", ie. that the time difference is smaller than a given tolerance. What happens in this example with the simulation results if again the gamma distribution is replaced by an empirical distribution which is constructed from a sample from the correct distribution? Our results are given in Table 4. As we have observed in Table 3 the empirical distributions constructed by KDE and KDEVC have about the same behaviour as the correct distribution. EBFS and ELK are considerably worse whereas naive resampling is totally useless for this example.

Table 4: Estimated probability of file access at the same time, (SE of estimate in brackets)

	n	tol	EBFS	ELK	Naive resampling	KDE	KDEVC	Correct distrib.
Prob*10^6(SE)	100	10^{-5}	306 (25)	282 (24)	10328 (143)	202(20)	240 (22)	167 (10)
Prob*10^6(SE)	500	10^{-5}	242 (22)	248 (22)	2156 (66)	186(19)	214 (21)	167 (10)
Prob*10^6(SE)	100	10^{-4}	2592 (72)	2628 (72)	12142 (155)	1960(63)	1978(63)	1898 (30)
Prob*10^6(SE)	500	10^{-4}	2278 (67)	2234 (67)	3956 (89)	2044(64)	2026(64)	1898 (30)

Future Work

It is also possible to use kernel functions that have heavier tails than the normal distribution, for example the density of the t-distribution or of the logistic distribution. Allthough not used in density estimation they could be interesting for our purpose as they allow to generate distributions with the same behaviour as the given sample but different tail behaviour. They could be used in simulation studies to test the influence of the tails of the input distribution on the final results. We tried as kernels the Gaussian, the uniform, the logistic and the t-distribution (with 3 degrees of freedom). There were clear differences between different used kernels in the results of the maximal distance in Table 3, which is obviously sensitive to the tail behaviour of the input distribution. As the results for all other tables were practically the same for all different kernels we have only reported the results of the Gaussian kernel. Nevertheless we think that the use of different (heavier tailed) kernels in simulation studies would deserve future discussion. An additional advantage of the kernel method is the possibility to generalize it to higher dimensions. This is important as with the exception of the normal distribution few standard distributions are commonly used to model multivariate data. We will present the details in a subsequent paper.

Conclusions

The first of the final conclusions from the above investigations is in our opinion that methods that generate random variates directly from data are important and useful tools in simulation studies. They are easy to use and more flexible than fitting standard distributions to data. They should be used whenever there are no a priori reasons for using a certain standard distribution. The second conclusion is even more obvious. Use kernel density estimates to construct the empirical distribution function. Allthough the question which variant should be taken is not fully solved here, we think that the results presented in this paper clearly favour the variance corrected version (KDEVC) allthough one could find applications where the original version KDE performs better.

Sampling from kernel density estimates is a simple task. There is mathematical theory that shows the good theoretical behaviour of these estimates, and the empirical results of this paper confirm that these good theoretical properties can lead to more accurate results in simulation studies. Thus it is an important tool for modelling input distributions in simulation studies.

References

[1] P. Bratley, B. L. Fox, and E. L. Schrage. A Guide to Simulation. Springer-Verlag, New York, 2 edition, 1987.

[2] L. Devroye. Non-Uniform Random Variate Generation. Springer-Verlag, New-York, 1986.

[3] L. Devroye. Universal smoothing factor selection in density estimation, theory and practice. Test, 6 (1997), 223–320.

[4] L. Devroye and L. Györfi. Nonparametric Density Estimation: The L_1 View. John Wiley, New-York, 1985.

[5] A. Law and D. Kelton. Simulation Modeling and Analysis. Mc-Graw-Hill, New-York, 1991.

[6] B. Silverman. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London, 1986.

[7] M. Wand and M. Jones. Kernel Smoothing. Chapman and Hall, London, 1995.