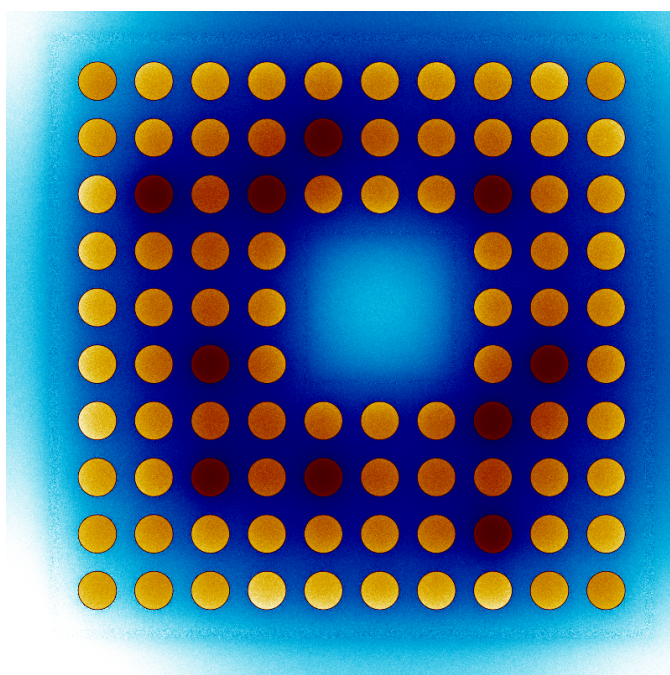




RESEARCH REPORT

VTT-R-00371-14



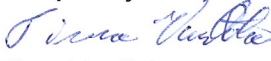


Statistical Tests and the Underestimation of Variance in Serpent 2

Authors: Toni Kaltiaisenaho

Confidentiality: Public



Report's title Statistical Tests and the Underestimation of Variance in Serpent 2	
Customer, contact person, address	Order reference
Project name KÄÄRME 2013/Safir2014	Project number/Short name 77464
Author(s) Toni Kaltiaisenaho	Pages 67
Keywords Serpent, Monte Carlo, statistical tests, underestimation of variance	Report identification code VTT-R-00371-14
Summary <p>In this report, we present three statistical tests implemented in Serpent 2 Monte Carlo Code. Two of the tests are normality tests which are used for detecting non-normally distributed Monte Carlo estimates. The third test is a so-called "drift-in-mean" test which is used for detecting changing distributions of estimates. The statistical tests are studied using four test problems: a BWR lattice case, a two-dimensional slab case and two three-dimensional pin-cell array cases, one with a homogeneous and the other with a heterogeneous fuel profile. The underestimation of variance is studied in the BWR lattice case and in the pin-cell array cases.</p> <p>The results show that undersampling and strongly non-converged fission source distribution can cause non-normal distributions of estimates, which can be detected by the normality tests. The drift-in-mean test is shown to be capable of detecting also smaller changes in the distributions. If the system is well-behaved, e.g. the system is adequately sampled and the source distribution has converged, the tests are likely to pass with a probability related to the used significance level.</p> <p>The underestimation of variance was no major problem in the BWR case. In the heterogeneous fuel pin case, however, the true standard deviation was 15 times higher than the predicted one, which was due to strong inter-cycle correlations. The underestimation was reduced using multiple cycles per batch. In the heterogeneous fuel pin case the underestimation of variance was caused by undersampling. Increasing the number of neutron histories per cycle decreased the underestimation in this case.</p>	
Confidentiality	Public
Espoo 17.2.2014	
Written by  Toni Kaltiaisenaho, Research Trainee	Reviewed by  Petri Kotiluoto, Research Team Leader
Accepted by  Timo Vanttola, Head of Research Area	
VTT's contact address	
Distribution (customer and VTT)	
<i>The use of the name of the VTT Technical Research Centre of Finland (VTT) in advertising or publication in part of this report is only permissible with written authorisation from the VTT Technical Research Centre of Finland.</i>	

Contents

1	Introduction	4
2	Monte Carlo neutron transport	5
2.1	Neutron tracking in Serpent	5
2.2	The k -eigenvalue method	7
2.3	Result estimation	7
2.3.1	Estimates for k_{eff}	8
2.4	Statistics of the estimates	9
3	Concerns in Monte Carlo statistics	11
3.1	Convergence of the source distribution	11
3.2	Number of neutron histories	11
3.3	Inter-cycle correlations	12
3.4	Motivation for using statistical tests	13
4	Normality tests	14
4.1	Shapiro–Wilk test	15
4.1.1	Royston’s extension	16
4.2	Moment tests	18
4.2.1	Skewness test	20
4.2.2	Kurtosis test	20
4.2.3	D’Agostino–Pearson test	21
5	Drift-in-mean test	22
6	Implementation and usage of the statistical tests	24
6.1	Implementation of the tests	24
6.2	Input and output	24
6.3	Important notes	25
7	BWR lattice test problem	26
7.1	Results of the statistical tests	26
7.2	Underestimation of variance	28
8	Slab test problem	30
9	Pin-cell array test problem	35
9.1	Homogeneous fuel pin	35
9.1.1	Results of the statistical tests	36
9.1.2	Underestimation of variance	39
9.2	Heterogeneous fuel pin	39
9.2.1	Results of the statistical tests	40
9.2.2	Underestimation of variance	42
10	Conclusions	44
	References	46

Appendices

A BWR	49
B Homogeneous fuel pin	52
C Heterogeneous fuel pin	64

1 Introduction

Monte Carlo methods have been widely used for neutron transport calculations since the 1950s [1]. In criticality calculations Monte Carlo methods are applied for solving the k -eigenvalue problem, which results in estimates of the multiplication factor and different reaction rates. An important topic in k -eigenvalue calculations is the accuracy of the results and the reliability of the confidence intervals. The fission source convergence, correct sampling of neutron histories and inter-cycle correlations need to be taken into account in order to obtain accurate results [2]. The inter-cycle correlations, which result in the underestimation of variance, have been regarded as one of the major challenges on full-core reactor analysis [3].

One purpose of this work is to implement statistical tests in Serpent Monte Carlo neutron transport code. Three tests are implemented, which can be used to detect abnormalities in Monte Carlo statistics. Two normality tests, namely the Shapiro–Wilk test and the D’Agostino–Pearson test, are used for detecting non-normalities in the statistics. The third test is a “drift-in-mean” test which can identify changes in the distributions of the estimates. The behaviour of the statistical tests are studied using four test problems: a BWR lattice case, a two-dimensional slab case and two three-dimensional pin-cell array cases, one with a homogeneous and the other with a heterogeneous fuel profile. The other subject of this work is to study the underestimation of variance in Serpent. The BWR lattice and the pin-cell array cases are used for this study.

This work is organized as follows: In Sec. 2 the basics of the Monte Carlo neutron transport are presented. The concerns in Monte Carlo statistics are then discussed in Sec. 3. The normality tests are presented in Sec. 4 and the drift-in-mean test in Sec. 5. The implementation and the usage of the tests are discussed in Sec. 6. The tests and the underestimation of variance are studied using the BWR lattice, slab and pin-cell array cases in Sec. 7, 8 and 9, respectively. Finally, we conclude in Sec. 10.

2 Monte Carlo neutron transport

Monte Carlo is a stochastic method which is used for simulating systems in many fields of physics, engineering, finance and others. The basic idea of the method is to simulate some known process using randomly generated values of input variables. The desired output variables are calculated and collected during the simulation. Due to the stochastic nature of the method, the results are random variables. Therefore, the simulation is often repeated several times to obtain estimates of the means and variances of the results.

There are many advantages in using the Monte Carlo method in neutron transport calculations. The method is closely linked to the actual transport process of neutrons. The linearity of the process allows each neutron to be simulated separately, which enables the use of parallel computation. The method can be used for simulating complex geometries, which can be a difficult task for deterministic methods. Another advantage of using the Monte Carlo method is the capability to use evaluated nuclear data in a continuous-energy form which is based on experimental measurements and theoretical nuclear models. In deterministic codes, the continuous-energy nuclear data is discretized into groups, which introduces inaccuracies in the results. A major disadvantage of the Monte Carlo neutron transport is the high computational cost when solving large-scale problems.

In this section, the basics of the Monte Carlo neutron tracking in Serpent and the k -eigenvalue method are described first. Then, the reaction rate estimators are presented. Finally, we discuss the statistics of the estimates.

2.1 Neutron tracking in Serpent

The basic idea of the Monte Carlo neutron transport is to follow neutrons and record their interactions with media in a user-defined geometry. The simulation proceeds in cycles, or generations, in which neutrons are followed one by one. All the events related to a neutron, such as scatterings, capture and fission, form the history of that neutron. In another words, a cycle consists of neutron histories.

The neutron path in a medium consists of multiple straight-line segments, which join the interaction points together. In the first cycle the start location of the path is usually randomly selected from a uniform distribution. In the following cycles the start location is determined by a fission site from the previous cycle. The end location of the path is either an absorption site or the location on the boundary of the geometry where the neutron escapes from the system. The path between two interaction points, called the free path length, is sampled from an exponential distribution and is given by [4]

$$l = -\frac{\ln \xi}{\Sigma_{t,m}}, \quad (2.1)$$

where ξ is a uniformly distributed random variable on the interval $[0, 1]$ and $\Sigma_{t,m}$ is the macroscopic total cross section of the material.

On the interaction site, the target nucleus k is selected with a probability

$$P_k = \frac{\Sigma_{t,k}}{\Sigma_{t,m}}, \quad (2.2)$$

where $\Sigma_{t,k}$ is the macroscopic total cross section of the nucleus k . The reaction type with the target nucleus is selected with a probability

$$P_i = \frac{\Sigma_{i,k}}{\Sigma_{t,k}}, \quad (2.3)$$

where $\Sigma_{i,k}$ is the macroscopic cross section of reaction type i with the nucleus k . The possible reaction types are capture, fission and scattering. All reactions emitting non-neutron secondary particles belong to capture type, such as (n,γ) -, (n,p) - and (n,α) -reactions. In Serpent, a capture reaction terminates the neutron history. A fission reaction also ends the neutron history and generates new source neutrons for the next cycle. Elastic and inelastic scattering reactions change the direction and energy of the neutron.

The user-defined geometry consists of homogeneous material regions called cells, which each have their own macroscopic total cross sections. Therefore, the free path length distribution is different in each cell. When a neutron moves from one cell to another, the path length sampled in the first cell isn't statistically valid in the next one. A common way to solve this problem in Monte Carlo neutron transport codes is to stop the neutron at the boundary surface of the two cells and then resample the free path length. In practice, this requires a comparison between the sampled path length and the distance to the nearest surface, which is the general idea of the surface-to-surface ray-tracing method.

Another possible method for a neutron tracking procedure is the delta-tracking method, in which the macroscopic total cross sections are homogenized by introducing virtual collisions. Virtual collisions are defined as collisions which preserve the energy and direction of a neutron, hence having no effect on the simulation statistics. The idea of the delta-tracking method is to add an adequate virtual cross section to every material cross section, so that the resulting cross section has the same value in the entire geometry. This resulting cross-section is called a majorant cross section $\Sigma_{\text{maj}}(E)$ which is defined to be the maximum of the total cross sections at each energy. The free path length is sampled using $\Sigma_{\text{maj}}(E)$ in every cell, which eliminates the need for surface tracking. Virtual collisions are handled by rejection sampling, which means that a real collision is accepted with a probability

$$P_m(E) = \frac{\Sigma_{t,m}(E)}{\Sigma_{\text{maj}}(E)} \quad (2.4)$$

and a virtual one with a probability $1 - P_m(E)$. If a virtual collision occurs, a new free path length is sampled.

The delta-tracking method has some major advantages over the surface-tracking method. As the calculation of the surface distances is not needed, the tracking is often faster in complex geometries. Another advantage is the straightforward handling of complicated surfaces and objects. The delta-tracking method has some drawbacks,

though. In a highly heterogeneous geometry, e.g. in a geometry containing localized heavy absorbers such as burnable absorber pins or control rods, the majorant cross section is dominated by the cross section of the heavy absorber. However, the volume fraction of the heavy absorber is small, and the $P_m(E)$ becomes low outside the heavy absorber. This results in a waste of computing time as the virtual collision frequency increases. Another disadvantage of the delta-tracking is that the common track-length flux estimator can't be used. Instead, the less efficient collision flux estimator is usually used with the delta-tracking method. The flux estimators are discussed in more detail in Sec. 2.3.

To overcome the efficiency problems introduced by localized heavy absorbers, Serpent uses a combination of surface-tracking and delta-tracking methods [4]. The default delta-tracking method is switched to surface tracking when $P_m(E)$ is low.

2.2 The k -eigenvalue method

In Serpent, the default calculation method is the k -eigenvalue method which is used in stationary, self-sustaining fission chain reaction simulations. The method is used for calculating an estimate of the effective multiplication factor k_{eff} , which determines the criticality of the system under study. When k_{eff} is below unity, the system is subcritical which means that the number of fissions decreases from generation to generation. When k_{eff} is above unity, the number of fissions increases, and the system is said to be supercritical. If k_{eff} is equal to unity, the system is critical, meaning that the number of fissions is a constant. The k -eigenvalue method is also used for calculating estimates of different reaction rates.

A k -eigenvalue simulation proceeds in cycles consisting of neutron histories. As already stated, the neutron source distribution is provided by the fission distribution of the previous cycle. The source size usually changes from cycle to cycle due to randomness or the nature of the simulation. To prevent the source size from increasing or decreasing, neutrons are either removed from or added to the source after each cycle.

At the beginning of the simulation the source may be far from the stationary converged distribution. Therefore, an adequate number of cycles must be skipped before the gathering of statistics can be started. The skipped cycles are called inactive cycles, whereas the cycles used for recording statistics are called active cycles. The active cycles are continued until a desired level of accuracy is obtained. The final estimates are calculated from the cycle-wise estimates as presented in Sec. 2.4.

2.3 Result estimation

In neutron transport calculations, the Monte Carlo method is usually used for estimating reaction rates of type

$$R = \frac{1}{V} \int_V \int_E f(\mathbf{r}, E) \phi(\mathbf{r}, E) d^3r dE, \quad (2.5)$$

where V is some volume element, E is some energy domain, $f(\mathbf{r}, E)$ is a response function representing some physical parameter and $\phi(\mathbf{r}, E)$ is a scalar neutron flux. The response function can be for example a macroscopic reaction cross section. In the Monte Carlo method, reaction rates are obtained by forming estimates (or estimators) from scores which are recorded during the simulation in the given volume and energy interval. The estimators can be categorized into analog and implicit estimators.

The basic idea of analog estimators is to count the number of reactions occurring in the given volume and energy interval. A simple example of an analog estimate is the total reaction rate estimate obtained by calculating the total number of reactions in a cycle. An implicit estimate, on the other hand, is calculated by multiplying an appropriate flux estimate by a desired response function. Given a flux estimate ϕ and a response function f , the implicit reaction rate estimate is given as

$$R_{\text{imp}} = \sum_{i=1}^I f_i \phi_i, \quad (2.6)$$

where the summation is over the scored values in the given volume and energy interval. The variance of the implicit estimator is usually smaller than the variance of the analog estimator because a score is recorded even without an actual reaction taking place. The physical reaction rate corresponding to Eq. (2.5) is obtained by normalizing the analog or implicit estimate.

The most commonly used flux estimator is the track length estimator which is based on recording track lengths. The reaction rate of the estimate is given by

$$R_{\text{tle}} = \sum_{i=1}^I f_i l_i, \quad (2.7)$$

where l_i is the track length. A score is recorded each time a neutron creates a track, including when a neutron passes the boundary of a cell. Another common estimator is the collision flux estimator

$$R_{\text{cfe}} = \sum_{i=1}^I \frac{f_i}{\Sigma_{t,i}}, \quad (2.8)$$

where $\Sigma_{t,i}$ is the macroscopic total cross-section in the scored material. As the name suggests, a score is recorded each time a collision occurs. The resulting estimate is the expected rate of the interactions corresponding to the response function. The collision estimator is less efficient than the track-length estimator which doesn't require an interaction taking place in the cell for the score to be recorded. The efficiency difference is seen especially in small volumes.

2.3.1 Estimates for k_{eff}

Serpent provides three estimates for k_{eff} . The analog estimate for k_{eff} (also known as the generation estimate) is defined as the ratio of the source sizes of two subsequent

cycles

$$k_{\text{eff,ana}} = \frac{M_{j+1}}{M_j}. \quad (2.9)$$

The implicit estimate for k_{eff} is the absorption estimate which is calculated as the ratio of the production rate and the loss rate of neutrons

$$k_{\text{eff,imp}} = \frac{\bar{\nu}R_f}{R_f + R_c - R_{(n,xn)} + R_L}, \quad (2.10)$$

where $\bar{\nu}$ is the average number of neutrons emitted in a fission, R_f is the total fission rate, R_c is the total capture rate, $R_{(n,xn)}$ is the total inelastic scattering production rate of neutrons and R_L is the total neutron leakage rate.

The third estimate for k_{eff} is the collision estimate which is calculated as the ratio of the collision estimate of born fission neutrons and the source size

$$k_{\text{eff,col}} = \frac{\bar{\nu}R_f}{M_1}, \quad (2.11)$$

where M_1 is the initial source size.

2.4 Statistics of the estimates

In the Monte Carlo method, the final estimate of the studied parameter is usually given as the sample mean of single estimates obtained by repeating the simulation. In k -eigenvalue calculations, single estimates are calculated from the scores recorded during active cycles. In Serpent, the scores are collected into batches to form batch-wise estimates using, for example, the collision estimator. If the scores collected to a batch are from a single cycle, then a batch-wise estimate is equal to a cycle-wise estimate.

In Serpent, the final estimate is the sample mean calculated as

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n, \quad (2.12)$$

where N is the number of batches and x_n is a batch-wise estimate. The accuracy of the result is reported as the standard deviation of \bar{x} , which is also known as the standard error of the mean, and is denoted by

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{N(N-1)} \sum_{n=1}^N (x_n - \bar{x})^2}. \quad (2.13)$$

The standard deviation of \bar{x} is a measure of variation in the results of multiple identical and independent simulations. A more convenient form for the accuracy is the relative standard error (or relative statistical error) given by

$$RSE = \frac{\sigma_{\bar{x}}}{\bar{x}}. \quad (2.14)$$

The standard error of the mean can be interpreted as a confidence interval, in which the sample mean falls with the corresponding probability. In order to determine the confidence intervals, the distribution of the mean must be known. According to the central limit theorem, the distribution of \bar{x} approaches to a normal distribution if the variables x_n are independent and identically distributed (iid) and have a finite mean μ and a variance σ^2 [5], i.e.

$$\lim_{N \rightarrow \infty} P \left(a < \frac{\bar{x} - \mu}{\sigma/\sqrt{N}} < b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx. \quad (2.15)$$

If N is sufficiently large, \bar{x} is approximately normally distributed with a mean μ and a variance $\sigma_{\bar{x}}^2$, i.e. $\bar{x} \sim \mathcal{N}(\mu, \sigma_{\bar{x}}^2)$. Now, the standard deviation $\sigma_{\bar{x}}$ is easy to interpret as a confidence interval, e.g. , the probability that \bar{x} lies within one $\sigma_{\bar{x}}$ of the true mean is around 68%.

The batch estimate x_n is also approximately normally distributed if x_n is formed by summing random variables together as in Eq. (2.6) and the assumptions of the central limit theorem hold. If x_n is calculated as a product or ratio of random variables, then it is not normally distributed in general. However, the product of two independent normally distributed random variables $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ is approximately normally distributed if the ratios μ_1/σ_1 and μ_2/σ_2 are large enough [6]. Also, the distribution of the ratio of two independent normally distributed random variables can be approximated by a normal distribution under certain conditions [7]. The estimates for k_{eff} , which are products and ratios of random variables, are usually considered normally distributed [8].

3 Concerns in Monte Carlo statistics

The main concerns related to the accuracy of the results of a Monte Carlo neutron transport simulation are the convergence of the source distribution, the number of neutron histories and the inter-cycle correlations [2]. Here, we discuss the nature and the effects of these issues, and describe how they can possibly be detected by using simple statistical tests.

3.1 Convergence of the source distribution

As described in Sec. 2.2, the k -eigenvalue method is used for calculating estimates for k_{eff} and reaction rates in a system having a stationary source distribution. However, if the source distribution hasn't converged when the collecting of the statistics is started, the estimates and their uncertainties will not be accurate. The sufficient number of inactive cycles can be found by studying the convergence of both k_{eff} and the source distribution. The convergence of merely k_{eff} is not a sufficient criterion to begin the active cycles because the source distribution can converge more slowly than k_{eff} [9].

A common way to measure the convergence of the source is to use the Shannon entropy [10]. This method involves creating a three-dimensional mesh over the fissionable regions, and counting the fraction of fission sites in each mesh bin. The Shannon entropy is then calculated as

$$H_{src} = - \sum_{J=1}^{N_s} P_J \log_2(P_J), \quad (3.1)$$

where N_s is the number of mesh bins and P_J is the number of source neutrons in mesh bin J divided by the total number of source neutrons. To determine when the source has converged, it is advised to study the behaviour of both k_{eff} versus cycle and the Shannon entropy versus cycle [11].

Although the Shannon entropy is a useful tool for determining the convergence of the source distribution, it should be used with caution in problems in which the source convergence is difficult to obtain. An example of such a problem is a fuel pin with a symmetric flux profile having a low minimum at the centre. In this particular case, the source can converge into a stable, strongly asymmetric distribution which is physically incorrect [12]. Therefore, the possible peculiarities of the studied problem should be taken into account when conducting simulations.

3.2 Number of neutron histories

The number of neutron histories used per cycle affects the number of recorded scores. In general, the greater the number of source neutrons is, the better the obtained statistics are. The number of neutron histories must be large enough so that the problem space is adequately sampled. If some region of the geometry is poorly

sampled, the reaction rate estimates and their errors at the region will be incorrect. This also results in the underestimation of variance as the obtained reaction rate distribution is likely to be incomplete.

As described in Sec. 2.2, the source size is adjusted to be constant after each cycle. This normalization introduces bias in k_{eff} and local reaction rates [2]. However, this bias is inversely proportional to the number of neutron histories, and it can be reduced or eliminated by using an adequate number of source neutrons. It is recommended that at least 10,000 source neutrons should be used per cycle to prevent the bias [2].

3.3 Inter-cycle correlations

An important issue in Monte Carlo criticality calculations is the underestimation of variance caused by inter-cycle correlations. The standard deviation calculated with Eq. (2.13) is unbiased when there are no correlations between the batch-wise estimates. However, in k -eigenvalue calculations, the source distribution in each cycle is provided by the fission distribution of the previous cycle. If the concentration of fission neutrons is high at some location of the geometry, the concentration is likely to be high in the next cycle as well. This means that there exists a positive correlation between the source distributions, which leads to the underestimation of variance [11].

The underprediction of uncertainties can be studied by comparing the real and the apparent standard deviation. The apparent standard deviation is the mean of the standard deviations of S independent simulations, and is given by

$$\bar{\sigma}_a = \frac{1}{S} \sum_{k=1}^S \sigma_{\bar{x}_k}, \quad (3.2)$$

where $\sigma_{\bar{x}_k}$ is calculated as in Eq. (2.13). The real standard deviation is the sample standard deviation of the estimates for S independent simulations, and is denoted by

$$\hat{\sigma}_r = \sqrt{\frac{1}{S-1} \sum_{k=1}^S (\bar{x}_k - \bar{\bar{x}})^2}, \quad (3.3)$$

where \bar{x}_k is the estimate for each simulation and $\bar{\bar{x}}$ is the mean of the estimates. The measure of the uncertainty underestimation is the ratio of $\hat{\sigma}_r$ and $\bar{\sigma}_a$ which is denoted by

$$f_\sigma = \frac{\hat{\sigma}_r}{\bar{\sigma}_a}. \quad (3.4)$$

In 2D full core simulations, some reported values of f_σ have been between 3 and 5 [2, 13]. With more challenging configurations, the value of f_σ can be above 40 [12].

One simple way to obtain more accurate estimate of variance is to use a batch means method, in which multiple cycle-wise estimates are included in each batch. It can be shown that the batch means are less correlated than the individual cycles [14].

The batch means method has been shown to reduce the underestimation of variance in three-dimensional PWR full-core models [15, 16]. This method is also available in Serpent.

3.4 Motivation for using statistical tests

Simple statistical tests can be used for detecting a non-converged source distribution and poor sampling of neutron histories. The tests are aimed for identifying abnormalities in the distributions of the batch-wise estimates of k_{eff} and reaction rates. As described in Sec. 2.4, the batch-wise estimates are normally distributed if the assumptions of the central limit theorem hold, or the estimates are otherwise approximately normally distributed. Deviations from the normal distribution indicate that the scored values are not identically distributed, or the number of scores is too small. These causes of non-normality are related to the non-converged source distribution and poor sampling.

If the source distribution hasn't converged when the collecting of the results is started, the distributions of the estimates of k_{eff} and reaction rates will change as the active cycles proceed. If the change is significant, the distributions will be non-normal, which can be detected by normality tests. However, the change can also be so small that normality tests can't identify any deviations from a normal distribution. Therefore, a so-called "drift-in-mean" test is needed to detect changes in the distributions.

The non-normalities of the estimate distributions may also be due to insufficient number of recorded scores. This issue arises when some region of the geometry is poorly sampled, i.e. the number of neutron histories is too low. This is a problem especially when the volume in which the scores are recorded is small and the interaction probability is low.

To conclude this section, normality tests can be used for detecting poor sampling of neutron histories and possibly a non-converged source distribution. The selected normality tests for Serpent are presented in Sec. 4. A test suited especially for identifying a non-converged source is the drift-in-mean test which is presented in Sec. 5.

4 Normality tests

Numerous different methods exist for testing the normality of a given data set. The simplest ones are graphical methods which can be useful for quick checking of the normality. However, they don't provide any conclusive evidence of the normality and their usage can be cumbersome when there are numerous data sets to be studied. Therefore, statistical tests have been created to provide mathematical evidence of the non-normality of the data.

The most commonly used normality tests, which belong to frequentist statistical tests, are tests against the null hypothesis H_0 that the data is normally distributed. The first step in the testing procedure is to compute a test statistic, which follows some known distribution. The test statistic is then converted into a p -value which is the probability of getting a test statistic equal to or smaller than the one observed under the null hypothesis. The null hypothesis is rejected if the p -value is less than a predetermined significance level. If the p -value is greater than the significance level, one fails to reject the null hypothesis. For example, if the significance level is chosen to be 0.05 and the obtained p -value is less than this, one can conclude that with 95% probability the data isn't normally distributed.

Although normality tests can be useful for detecting non-normalities, they should be used with caution. One problem is in the testing against the assumption of normality. By rejecting the null hypothesis, one can conclude with a certain probability that the underlying data set is not drawn from a normal distributed random variable. But if one fails to reject the null hypothesis, one can't conclude that the random variable really is normally distributed. Another limitation is the effect of the sample size on normality tests. When the size of the data set is small, normality tests may not find the deviations from the normal distribution. On the contrary, when the sample size is large, even small deviations from the normality are detected, and thus the null hypothesis is rejected even though the distribution would otherwise look like normally distributed. It is therefore suggested that normality tests should be used together with graphical methods [17].

There are contradicting results regarding which normality test is the best or optimal one [18]. When choosing a normality test, one should take into account the sample size and prior information of the distribution. The recommended tests in the literature include the Shapiro–Wilk test and the D'Agostino–Pearson test [17–20]. These two tests have very different approaches to the normality testing. The Shapiro–Wilk test belongs to the regression-type tests, whereas the D'Agostino–Pearson test is based on the sample skewness and kurtosis. In some comparisons the Shapiro–Wilk test has been powerful for testing asymmetric distributions, but it performs well with symmetric distributions as well [18, 21]. The D'Agostino–Pearson test has a good power for testing symmetric short-tailed distributions [18].

Due to their different approaches and the recommendations in the literature, the Shapiro–Wilk test and the D'Agostino–Pearson test were chosen to be implemented in Serpent. The main ideas of these tests are presented in this section.

4.1 Shapiro–Wilk test

The Shapiro–Wilk test and its variations are one of the most popular and recommended normality tests. The idea of the test is to compare two estimates of the variance: the best linear unbiased estimator and the sample estimate. Here, the Shapiro–Wilk test statistic is derived and discussed briefly.

Let x_1, \dots, x_n be an ordered iid random sample from a standard normally distributed random variable X , i.e. $X \sim \mathcal{N}(0, 1)$. Furthermore, let $\mathbf{m} = [m_1, \dots, m_n]^T$ denote the vector of the expected values of x_1, \dots, x_n and $\mathbf{V} = (v_{ij})$ the corresponding covariance matrix, that is

$$\begin{aligned} E(x_i) &= m_i, & i &= 1, \dots, n, \\ \text{cov}(x_i, x_j) &= v_{ij}, & i, j &= 1, \dots, n. \end{aligned} \quad (4.1)$$

Also, let y_1, \dots, y_n be an ordered iid random sample from a random variable Y . The objective is to test whether this sample is from a normal distribution with a mean μ and variance σ^2 , that is, may y_i be expressed as

$$y_i = \mu + \sigma x_i, \quad i = 1, \dots, n. \quad (4.2)$$

Using generalized least-squares theorem, μ and σ can be estimated by minimizing $(\mathbf{y} - \mu \mathbf{1} - \sigma \mathbf{m})^T \mathbf{V}^{-1} (\mathbf{y} - \mu \mathbf{1} - \sigma \mathbf{m})$, where $\mathbf{y} = [y_1, \dots, y_n]^T$ and $\mathbf{1} = [1, \dots, 1]^T$ [22]. According to the Gauss-Markov theorem, least-squares estimators have the smallest variance of any unbiased linear estimators. Therefore, these estimates are called the best linear unbiased estimators (BLUEs). The BLUEs of μ and σ are [23]

$$\begin{aligned} \hat{\mu} &= \frac{\mathbf{m}^T \mathbf{V}^{-1} (\mathbf{m} \mathbf{1}^T - \mathbf{1} \mathbf{m}^T) \mathbf{V}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1} \mathbf{m}^T \mathbf{V}^{-1} \mathbf{m} - (\mathbf{1}^T \mathbf{V}^{-1} \mathbf{m})^2}, \\ \hat{\sigma} &= \frac{\mathbf{1}^T \mathbf{V}^{-1} (\mathbf{1} \mathbf{m}^T - \mathbf{m} \mathbf{1}^T) \mathbf{V}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1} \mathbf{m}^T \mathbf{V}^{-1} \mathbf{m} - (\mathbf{1}^T \mathbf{V}^{-1} \mathbf{m})^2}. \end{aligned} \quad (4.3)$$

If the sample y_1, \dots, y_n is normally distributed, then $\mathbf{1}^T \mathbf{V}^{-1} \mathbf{m} = \mathbf{m}^T \mathbf{V}^{-1} \mathbf{1} = 0$ and $\mathbf{1}^T \mathbf{V}^{-1} = \mathbf{1}^T$, and $\hat{\mu}$ and $\hat{\sigma}$ can be expressed as

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}, \\ \hat{\sigma} &= \frac{\mathbf{m}^T \mathbf{V}^{-1} \mathbf{y}}{\mathbf{m}^T \mathbf{V}^{-1} \mathbf{m}}. \end{aligned} \quad (4.4)$$

Furthermore, let

$$S_n^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (4.5)$$

be the unbiased estimate of $(n-1)\sigma^2$. We can now define the Shapiro–Wilk test statistic as

$$W = \frac{R^4 \hat{\sigma}^2}{C^2 S_n^2} = \frac{(\mathbf{a}^T \mathbf{y})^2}{S_n^2} = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4.6)$$

where

$$\begin{aligned}
 R^2 &= \mathbf{m}^T \mathbf{V}^{-1} \mathbf{m}, \\
 C^2 &= \mathbf{m}^T \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m}, \\
 \mathbf{a}^T &= [a_1, \dots, a_n] = \frac{\mathbf{m}^T \mathbf{V}^{-1}}{(\mathbf{m}^T \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m})^{1/2}}.
 \end{aligned} \tag{4.7}$$

The constants R and C are introduced to Eq. (4.6) in order to obtain a normalized coefficient vector \mathbf{a} , that is

$$\mathbf{a}^T \mathbf{a} = \sum_{i=1}^n a_i^2 = 1. \tag{4.8}$$

Another important property of \mathbf{a} is that [23]

$$-a_i = a_{n-i+1}, \quad i = 1, \dots, n. \tag{4.9}$$

For odd n , i.e. $n = 2l + 1, l \in \mathbb{N}$, the middle coefficient a_{l+1} is zero.

From the definition (4.6) we can see that the W statistic is, up to a constant, the ratio of the BLUE of σ^2 and the unbiased estimate of σ^2 . Only the numerator of W depends on the normality assumption (4.2). If the vector \mathbf{y} really is normally distributed, then the numerator and denominator of W are both estimating, up to a constant, the variance. Therefore, a value of W close to unity indicates that the sample \mathbf{y} is normally distributed.

A major drawback of the W statistic is that its distribution can't be given in an explicit form. The distribution is needed for converting the test statistic into a p -value. One option is to conduct empirical simulations for obtaining test statistics corresponding to desired significance levels. The other way is to approximate the distribution of W or its transformations. Another challenge in W statistic is that the values of \mathbf{m} and \mathbf{V}^{-1} are unknown for large sample sizes. This problem can be solved by either finding approximations for the elements of \mathbf{m} and \mathbf{V} or approximating \mathbf{a} directly. The direct approach avoids the computation of matrix \mathbf{V}^{-1} which can be computationally demanding for large sample sizes.

4.1.1 Royston's extension

Probably the most common way to compute the Shapiro–Wilk test is to use Royston's extension [24, 25]. This method has been implemented by many numerical and statistical software packages, such as Matlab, R and SAS. The idea of the Royston's extension is to approximate \mathbf{a} directly with a vector $\tilde{\mathbf{a}}$ which is then used for computing a test statistic W_R . Using a normalizing transformation, W_R is transformed into an approximately normally distributed test statistic z_{W_R} . The Shapiro–Wilk test with Royston's extension is a one-tailed test in a sense that only large positive values of z_{W_R} indicate non-normality.

As a basis for approximating \mathbf{a} , Royston used vector \mathbf{c} given as

$$\mathbf{c} = \frac{\tilde{\mathbf{m}}}{(\tilde{\mathbf{m}}^T \tilde{\mathbf{m}})^{1/2}}, \tag{4.10}$$

where $\tilde{\mathbf{m}}$ is an approximation of \mathbf{m} . The elements of $\tilde{\mathbf{m}}$, which are also known as Blom scores [26], are denoted by

$$\tilde{m}_i = \Phi^{-1} \left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}} \right), \quad i = 1, \dots, n, \quad (4.11)$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function. Vector \mathbf{c} is used in the Weisberg-Bingham test statistic [27]

$$WB = \frac{(\mathbf{c}^T \mathbf{y})^2}{S^2} = \frac{(\sum_{i=1}^n c_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (4.12)$$

which is asymptotically equal to the W statistic. Up to a constant, \mathbf{c} differs from \mathbf{a} mainly in the first two and last two components. Using polynomial regression analysis of $a_n - c_n$ and $a_{n-1} - c_{n-1}$ on $u = n^{-\frac{1}{2}}$, Royston approximated a_n and a_{n-1} as

$$\tilde{a}_n = c_n + 0.221157u - 0.147981u^2 - 2.07119u^3 + 4.434685u^4 - 2.706056u^5, \quad (4.13)$$

$$\tilde{a}_{n-1} = c_{n-1} + 0.042981u - 0.293762u^2 - 1.752461u^3 + 5.682633u^4 - 3.582633u^5. \quad (4.14)$$

The rest of the components of \mathbf{a} can be approximated by normalizing each \tilde{m}_i as

$$\tilde{a}_i = \phi^{-1/2} \tilde{m}_i \quad \text{for} \quad \begin{cases} i = 2, \dots, n-1 & , \quad n \leq 5 \\ i = 3, \dots, n-2 & , \quad n > 5 \end{cases}, \quad (4.15)$$

where

$$\phi = \begin{cases} \frac{\tilde{\mathbf{m}}^T \tilde{\mathbf{m}} - 2\tilde{m}_n^2}{1 - 2\tilde{a}_n^2} & , \quad n \leq 5 \\ \frac{\tilde{\mathbf{m}}^T \tilde{\mathbf{m}} - 2\tilde{m}_n^2 - 2\tilde{m}_{n-1}^2}{1 - 2\tilde{a}_n^2 - 2\tilde{a}_{n-1}^2} & , \quad n > 5 \end{cases}. \quad (4.16)$$

Approximating \mathbf{a} as $\tilde{\mathbf{a}} = [\tilde{a}_1, \dots, \tilde{a}_n]^T$, the test statistic W_R can be defined as

$$W_R = \frac{(\tilde{\mathbf{a}}^T \mathbf{x})^2}{S_n^2} = \frac{(\sum_{i=1}^n \tilde{a}_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (4.17)$$

Using a proper transformation, the W_R test statistic can be transformed into an approximately standard normally distributed test statistic. For sample sizes $4 \leq n \leq 11$, Royston fitted the empirical distribution of $\ln(1 - W_R)$ with a three-parameter lognormal distribution

$$f_{\ln(1-W_R)}(t) = \frac{1}{(t - \gamma_R)\sigma_R\sqrt{2\pi}} \exp \left\{ -\frac{[\ln(t - \gamma_R) - \mu_R]^2}{2\sigma_R^2} \right\}, \quad (4.18)$$

and for $12 \leq n \leq 2,000$, he fitted the empirical distribution of $1 - W_R$ with a two-parameter lognormal distribution

$$f_{1-W_R}(t) = \frac{1}{t\sigma_R\sqrt{2\pi}} \exp \left\{ -\frac{[\ln t - \mu_R]^2}{2\sigma_R^2} \right\}. \quad (4.19)$$

The parameters γ_R , μ_R and σ_R were then estimated by smoothing them as functions of n . As a result, Royston proposed an approximately standard normally distributed test statistic

$$z_{W_R} = \begin{cases} \frac{-\ln[\gamma_R - \ln(1 - W_R)] - \mu_R}{\sigma_R} & , \quad 4 \leq n \leq 11 \\ \frac{\ln(1 - W_R) - \mu_R}{\sigma_R} & , \quad 12 \leq n \leq 2,000 \end{cases} \quad (4.20)$$

where for sample sizes $4 \leq n \leq 11$

$$\begin{aligned} \gamma_R &= -2.273 + 0.459n \\ \mu_R &= 0.544 - 0.39978n + 0.025054n^2 - 6.714 \cdot 10^{-4}n^3 \\ \sigma_R &= \exp(1.3822 - 0.77857n + 0.062767n^2 - 0.0020322n^3) \end{aligned} \quad (4.21)$$

and for $12 \leq n \leq 2,000$

$$\begin{aligned} \mu_R &= -1.5861 - 0.31082 \ln n - 0.083751(\ln n)^2 + 0.0038915(\ln n)^3 \\ \sigma_R &= \exp[-0.4803 - 0.082676 \ln n + 0.0030302(\ln n)^2]. \end{aligned} \quad (4.22)$$

Although the maximum sample size used here is 2,000, Royston claimed that the test is valid up to $n = 5,000$ [24]. Both of these limits are used in numerical and statistical softwares, e.g. SAS uses the former whereas Matlab and R use the latter.

4.2 Moment tests

Moment tests are based on the use of sample moments of a distribution. Here, we present the third and the fourth moment and use them in the skewness and kurtosis tests for normality. These test are then combined in the D'Agostino–Pearson test.

The third and fourth standardized moments of the distribution of a random variable X are called skewness and kurtosis and are defined as, respectively,

$$\sqrt{\beta_1} = \frac{E[(X - \mu)^3]}{(E[(X - \mu)^2])^{3/2}}, \quad (4.23)$$

$$\beta_2 = \frac{E[(X - \mu)^4]}{(E[(X - \mu)^2])^2}, \quad (4.24)$$

where E is the expected value operator and μ is the mean of the distribution. Skewness measures how much the probability distribution leans to one side of the mean. Positive $\sqrt{\beta_1}$ corresponds to skewness to the right and negative to the left. As the normal distribution is symmetric about its mean, its skewness is zero.

Kurtosis can be interpreted as a measure of tail thickness. The normal distribution has kurtosis equal to 3. Distributions with $\beta_2 > 3$ have heavier tails than a normal distribution and distributions with $\beta_2 < 3$ have lighter tails. Examples of distributions with different values of skewness and kurtosis are illustrated in Fig. 4.1.

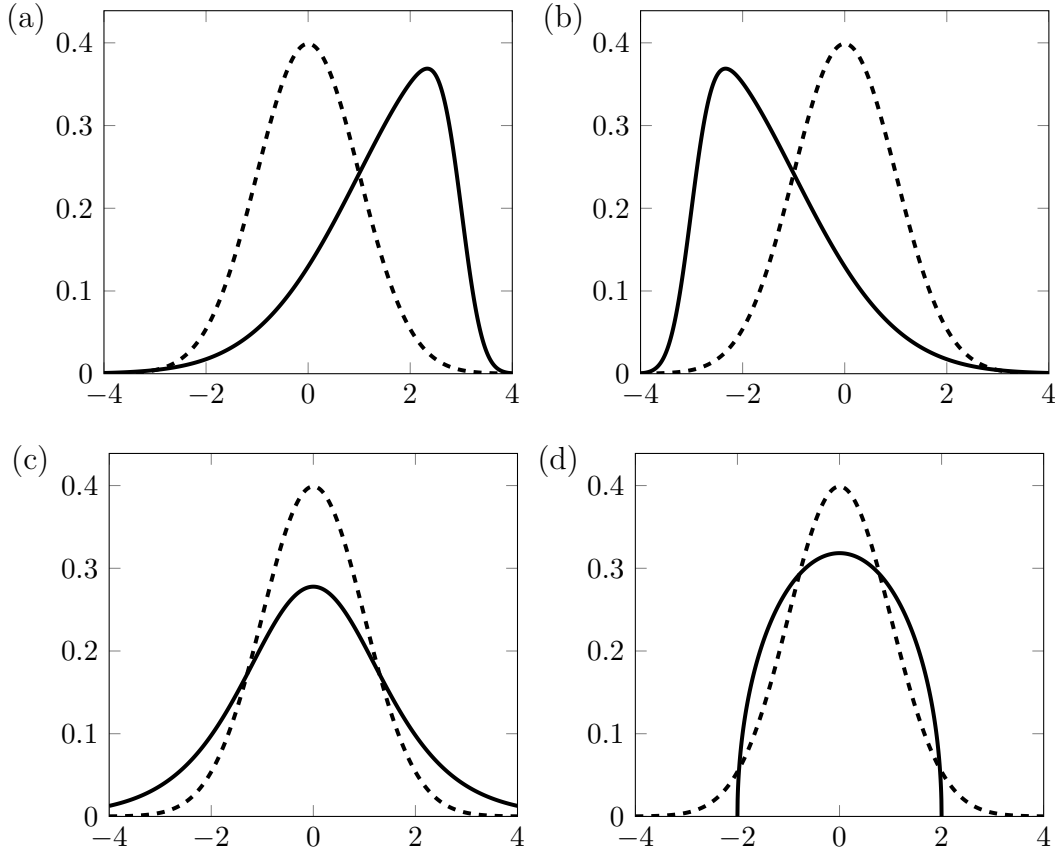


Figure 4.1: Visualization of skewness and kurtosis. The solid line in (a) is a left-skewed distribution ($\sqrt{\beta_1} < 0$), in (b) a right-skewed distribution ($\sqrt{\beta_1} > 0$), in (c) a heavy-tailed distribution ($\beta_2 > 3$) and in (d) a light-tailed distribution ($\beta_2 < 3$). The dashed line in each case presents a normal distribution ($\sqrt{\beta_1} = 0$, $\beta_2 = 3$).

The empirical counterparts of skewness and kurtosis are the sample skewness defined as

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{3/2}} \quad (4.25)$$

and the sample kurtosis defined as

$$b_2 = \frac{m_4}{m_2^2} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^2}, \quad (4.26)$$

where m_k is the k -th central sample moment and \bar{x} is the sample mean. The sample skewness and kurtosis are estimators for theoretical skewness and kurtosis. The value of $\sqrt{b_1}$ is expected to be close to 0 and the value of b_2 close to 3 if the sample is from a normal distribution. Deviations from these values indicate that the given data set doesn't follow a normal distribution.

4.2.1 Skewness test

It can be shown that $\sqrt{n}\sqrt{b_1}$ is asymptotically normally distributed with a mean 0 and variance 6 [21], that is

$$\sqrt{n}\sqrt{b_1} \xrightarrow{D} N(0, 6) \quad \text{as } n \rightarrow \infty. \quad (4.27)$$

The problem in using $\sqrt{n}\sqrt{b_1}$ as a test statistic is that its convergence is relatively slow, and hence it can't be used for small samples ($n < 150$). To overcome this problem, D'Agostino proposed a transformation to approximate the distribution of $\sqrt{b_1}$ as [28]

$$Z(\sqrt{b_1}) = \delta \ln \left(\frac{Y}{\nu} + \sqrt{\frac{Y^2}{\nu^2} + 1} \right), \quad (4.28)$$

where

$$Y = \sqrt{b_1} \left(\frac{(n+1)(n+3)}{6(n-2)} \right)^{1/2}, \quad (4.29)$$

$$\tau = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \quad (4.30)$$

$$\omega^2 = \sqrt{2(\tau - 1)} - 1, \quad (4.31)$$

$$\delta = 1/\sqrt{\ln \omega}, \quad (4.32)$$

$$\nu = \sqrt{2/(\omega^2 - 1)}. \quad (4.33)$$

$Z(\sqrt{b_1})$ is approximately standard normally distributed under the null hypothesis that the data is normally distributed. The approximation works satisfyingly for sample sizes of $n > 8$.

4.2.2 Kurtosis test

It can be shown that $\sqrt{nb_2}$ is asymptotically normally distributed with a mean $\mu = 3$ and variance $\sigma^2 = 24$ [21], i.e.

$$\sqrt{n}(b_2 - 3) \xrightarrow{D} N(0, 24) \quad \text{as } n \rightarrow \infty. \quad (4.34)$$

The convergence of $\sqrt{nb_2}$ to its asymptotic value is very slow. This problem can be solved with an appropriate transformation to the standard normal distribution. The most used transformation is the one given by Anscombe and Glynn [29], which is

$$Z(b_2) = \sqrt{\frac{9A}{2}} \left[1 - \frac{2}{9A} - \left(\frac{1 - 2/A}{1 + \xi \sqrt{2/(A-4)}} \right)^{1/3} \right], \quad (4.35)$$

where

$$A = 6 + \frac{8}{\sqrt{\beta_1(b_2)}} \left[\frac{2}{\sqrt{\beta_1(b_2)}} + \sqrt{1 + \frac{4}{\beta_1(b_2)}} \right], \quad (4.36)$$

$$\xi = \frac{b_2 - E[b_2]}{\sqrt{\text{Var}(b_2)}}. \quad (4.37)$$

In Eq. (4.36) $\sqrt{\beta_1(b_2)}$ is the third moment of b_2 denoted by

$$\sqrt{\beta_1(b_2)} = \frac{6(n^2 - 5n + 2)}{(n + 7)(n + 9)} \sqrt{\frac{6(n + 3)(n + 5)}{n(n - 2)(n - 3)}}. \quad (4.38)$$

Eq. (4.37) is the normal approximation of b_2 where $E[b_2]$ is the expectation value and $\text{Var}(b_2)$ is the variance of b_2 . Under the normality assumption they are denoted by

$$E[b_2] = \frac{3(n - 1)}{(n + 1)}, \quad (4.39)$$

$$\text{Var}(b_2) = \frac{24n(n - 2)(n - 3)}{(n + 1)^2(n + 3)(n + 5)}. \quad (4.40)$$

The variable $Z(b_2)$ is approximately standard normally distributed under the normality assumption for sample sizes of $n \geq 20$.

4.2.3 D'Agostino–Pearson test

The D'Agostino–Pearson test (also known as the D'Agostino–Pearson K^2 test) is an omnibus test which combines the skewness and kurtosis tests described in Sec. 4.2.1 and 4.2.2 [30]. Here, omnibus means that the test is capable of detecting departures from normality due to either skewness or kurtosis. The test statistic is defined as

$$K^2 = Z^2(\sqrt{b_1}) + Z^2(b_2), \quad (4.41)$$

where $Z(\sqrt{b_1})$ is given by Eq. (4.28) and $Z(b_2)$ by Eq. (4.35). The test statistic K^2 is χ_2^2 -distributed under the null hypothesis and assuming that $Z(\sqrt{b_1})$ and $Z(\sqrt{b_1})$ are independent.

5 Drift-in-mean test

The drift-in-mean test can be used for testing whether the mean of an estimate is drifting during active cycles. An appropriate statistical test for this purpose is an unequal variance t -test (also known as Welch's t -test), which tests whether the means of two groups differ from each other. The test statistic of the test is defined as

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{s_1}{n_1} + \frac{s_2}{n_2}}}, \quad (5.1)$$

where \bar{x}_i is the mean, s_i is the unbiased estimator of the variance and n_i is the sample size of the group $i = \{1, 2\}$. The test statistic follows approximately Student's t -distribution with degrees of freedom of [31]

$$\nu \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}. \quad (5.2)$$

In order to calculate the p -value from the test statistic t , the cumulative distribution function of the t -distribution is needed. For $t > 0$, it can be calculated as

$$F_\nu(t) = 1 - \frac{1}{2} I_{u(t)}\left(\frac{\nu}{2}, \frac{1}{2}\right), \quad (5.3)$$

where $I_{u(t)}$ is the regularized incomplete beta function and

$$u(t) = \frac{\nu}{t^2 + \nu}. \quad (5.4)$$

For negative values of t , $F_\nu(t)$ is obtained by symmetry. The regularized incomplete beta function is defined as

$$I_u(a, b) = \frac{B_u(a, b)}{B(a, b)}, \quad (5.5)$$

where $B(a, b)$ and $B_u(a, b)$ are the ordinary and incomplete beta functions, respectively, which are defined as

$$B(a, b) = \int_0^1 v^{a-1}(1-v)^{b-1} dv, \quad (5.6)$$

$$B_u(a, b) = \int_0^u v^{a-1}(1-v)^{b-1} dv. \quad (5.7)$$

The ordinary beta function can be calculated using its other form

$$B(a, b) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \quad (5.8)$$

Rather than calculating Eq. (5.8) directly, it is usually calculated by taking an exponential of the logarithms of the gamma functions. The natural logarithm of the gamma function is provided, e.g., by the C standard library. The incomplete beta function can be calculated using its series form

$$B_u(a, b) = \frac{x^a}{a} \left(1 + a \sum_{j=1}^{\infty} \frac{(1-b)(2-b)\cdots(j-b)}{j!(a+j)} x^j \right), \quad (5.9)$$

which can be truncated after it has converged properly. The accuracy of the computation of $I_u(a, b)$ can be increased by taking an advantage of the property

$$I_u(a, b) = 1 - I_{1-u}(b, a). \quad (5.10)$$

A good rule of thumb for using this form is when $u > a/(a + b + 2)$. This holds only for calculating the cumulative distribution function, not in general.

A good approximation for calculating the p -value for large values of ν is to use normal distribution as [32]

$$F_\nu(t) \approx \Phi(z(t, \nu)), \quad (5.11)$$

where Φ is the cumulative distribution function of the standard normal distribution and

$$z(t, \nu) = \sqrt{\frac{\ln(1 + t^2/\nu)(\nu - 1)^2}{\nu - 1.5}}. \quad (5.12)$$

For example, for $\nu = 400$, the relative error for calculating the p -value 0.01 is around 3.5×10^{-6} , which is small enough for this purpose. The relative error of this approximation decreases as ν increases.

6 Implementation and usage of the statistical tests

6.1 Implementation of the tests

The Shapiro–Wilk test with Royston’s extension and the D’Agostino–Pearson test were implemented as presented in Sec. 4. The algorithm for the inverse of the standard normal cumulative distribution function used in Eq. (4.11) was taken from Ref. [33]. The skewness and kurtosis tests presented in Sec. 4.2.1 and 4.2.2, respectively, were implemented separately, although their results are not printed in the output file but only used in the D’Agostino–Pearson test.

The drift-in-mean test was implemented as presented in Sec. 5. The cumulative distribution function of the Student’s t -distribution is calculated using the regularized incomplete beta function when $\nu < 400$. When $\nu \geq 400$, the normal approximation given by Eq. (5.11) is used. However, in the simulations presented in Sec. 7, 8 and 9, $F_\nu(t)$ was approximated with the standard normal distribution as

$$F_\nu(t) \approx \Phi(t). \quad (6.1)$$

It was tested that this approximation didn’t produce any significant error to the results as the number of active cycles was large in all the tested cases.

The implementation of the drift-in-mean test enables to divide the input data into two or more parts, and the test is then performed on the adjacent parts. The default, hard-coded option is to divide the data in half. This means that the default test studies whether the distribution of the first half of the active cycles differs from the distribution of the second half. The direction of the drift can be determined from the sign of the test statistic. A positive test statistic indicates that the mean is increasing, whereas a negative test statistic means that the mean is decreasing.

The chosen significance levels for this work were 0.05 for the normality tests and 0.01 for the drift-in-mean test. However, due to an error in the p -value calculation, an incorrect 0.1 significance level was used for the Shapiro-Wilk test. The error is explained in further detail in Sec. 7.1. As a side note, MCNP code uses both 0.05 and 0.01 significance levels for normality testing and produces a warning message if the p -value is smaller than the latter [8]. Similar feature could be well implemented in Serpent, too.

6.2 Input and output

The statistical tests are performed automatically for the analog, implicit and collision estimates of k_{eff} , mean population size and weight, and transport runtime and CPU usage. For detectors, the statistical tests can be activated by adding “`dhis = 1`” to the detector definition. The tests are performed only on the active cycles.

The output of the statistical tests is printed in Matlab m-file format to a file named “<input>_stat.m”, where “<input>” is the name of the Serpent input file. The variables in the file are named as “STATTESTS_<parameter>”, where “<parameter>” is the name of the tested parameter defined in Serpent. The output file contains three columns for each statistical test. The first column presents the test statistic, the second column the p -value and the third column the pass value of the test. The pass value is 1 if the test passed and 0 if it failed. The test statistic, p -value and the pass value can also be NaN (not a number). All the tests produce a NaN output if every element in the tested data is the same. This happens, for example, when a detector bin remains empty in all cycles. Another cause of a NaN output is that the number of samples in the data set is too small. The Shapiro–Wilk and the D’Agostino–Pearson tests produce a NaN output if the number of samples is less than 4 and 20, respectively.

6.3 Important notes

The p -value of a test is the probability of obtaining the sample by chance assuming that the null hypothesis of the test is true. Therefore, it’s possible that the random variable under study is normally distributed even if the test fails. This behaviour can be adjusted by changing the significance level.

As noted in Sec. 2.4, batch-wise estimates are not normally distributed in general if they are formed by products or ratios of random variables. Therefore, the way the estimates are calculated must be taken into account before making any conclusions about the results of the normality tests.

The Shapiro–Wilk test is assumed to be accurate up to sample sizes of 5,000. However, the test is also performed for sample sizes above 5,000, but the results may not be reliable. Therefore, the D’Agostino–Pearson test should be considered more trustworthy if the sample size exceeds 5,000.

The normality tests should be used with some caution. If the sample size is too small, the tests might not find the deviations from the normal distribution. With large sample sizes, even small deviations from the normal distribution may result in rejection of normality. Therefore, it is recommended that the tests should be used together with graphical methods.

7 BWR lattice test problem

In this section, the implemented statistical tests and the underestimation of variance are studied using a BWR lattice test case, which demonstrates the calculation of fast neutron flux in a boiling water reactor (BWR) lattice. This test problem should be well-behaved in a sense that the source is expected to converge quickly, and the underestimation of variance should be small. The uncertainty underprediction is studied for the estimate of fast neutron flux, and the behaviour of the statistical tests are investigated with the cycle-wise estimates of fast neutron flux and k_{eff} . The effect of the number of neutron histories per cycle and the number of inactive cycles are studied both with the underestimation of the variance and the statistical tests.

The BWR lattice contains seven types of fuel pins in a 10×10 assembly and an asymmetrically positioned moderator channel as presented in Fig. 7.1. The fast neutron flux ($E > 1$ MeV) is calculated in the cladding and in the inner and outer moderator channel walls using four detectors. Detector 1 and 2 are used for calculating the average and pin-wise fast flux in the cladding, respectively, whereas detector 3 and 4 are used for calculating the fast flux in the inner and outer moderator channel walls, respectively. The complete definition of the BWR case can be found in the Serpent user's manual [34].

Three sets of cases with 2,000, 10,000 and 50,000 source neutrons per cycle were studied. In each set, the numbers of inactive cycles used were 10, 100, 200, 400, and 800. The number of active cycles in each test configuration was 800. 50 independent simulations were run for each case.

7.1 Results of the statistical tests

The statistical tests are investigated here by comparing their pass rates in the tested configurations. A pass rate is calculated by taking the mean of the pass values of the simulation runs. The used significance levels were 0.05 for the D'Agostino–Pearson test and 0.01 for the drift-in-mean test. Due to a small error in the p -value calculation, a 0.05 significance level was used for both tails of the normal distribution of the Shapiro–Wilk test statistic. This resulted in incorrect 0.1 significance level because only a large positive Shapiro–Wilk test statistic indicates non-normal data. Therefore, the pass rates of the Shapiro–Wilk test are lower than the ones of the D'Agostino–Pearson test in this section, and also in Sec. 8 and 9.

The pass rates of the statistical tests performed on the estimates of fast neutron flux are presented in Fig. A.1–A.4 in Appendix A. A common feature in all the cases is that the number of inactive cycles doesn't have any consistent effect on the pass rates. Also, the pass rates of the drift-in-mean test are close to unity regardless of the number of source neutrons per cycle or the number of inactive cycles. These results imply that the distributions of the flux estimates didn't change much during the active cycles.

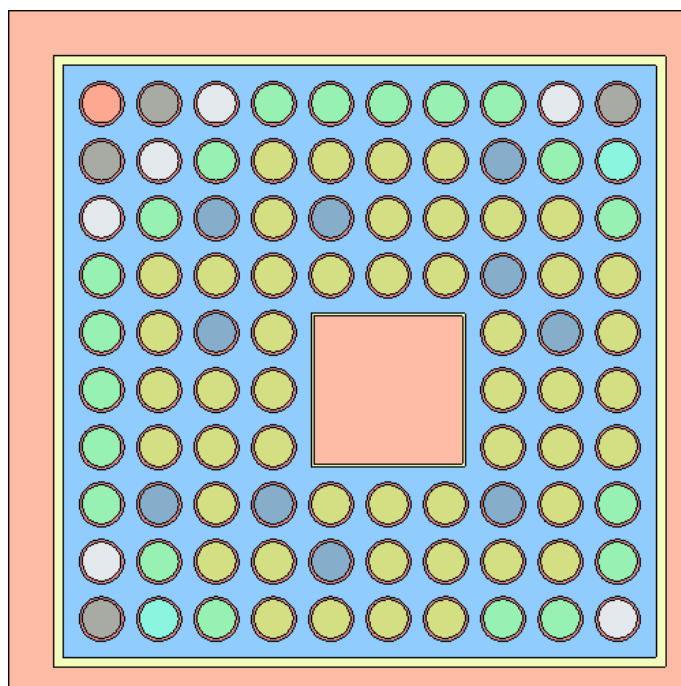


Figure 7.1: The BWR lattice geometry.

The pass rates of the normality tests vary significantly between the estimates. With the average fast flux in the cladding (Fig. A.1), the pass rates of the normality tests are between 0.8 and 1, and they increase slightly with the number of neutron histories per cycle. The cycle-wise distribution is very normal-like even with just 2,000 neutron histories per cycle. However, the pin-wise fast flux is clearly non-normal when too few source neutrons per cycle are used. This can be seen in Fig. A.2 presenting the average pass rate which is calculated as an average of the means of the pin-wise pass values. With 2,000 source neutrons per cycle, the average pass rates of the normality tests are zero, but with 10,000 and 50,000 source neutrons per cycle, they increase to around 0.5 and 0.8, respectively. The standard deviations of the average pass rates are reasonably small, which means that the means of the pin-wise pass values don't change much from run to run. The differences between the normality tests' pass rates of the average and pin-wise fluxes are caused by the different detector volumes. When the number of neutron histories per cycle is small, the number of recorded scores per pin in the cladding is not enough to form a normal distribution. The total number of recorded scores in the cladding is sufficient, however.

The effect of the detector volume can also be seen with the estimates of fast flux in the inner and outer moderator channel walls. When the number of neutron histories per cycle is small, the normality tests' pass rates of the fast flux in the smaller volume inner channel walls (Fig. A.3) are lower than in the larger volume outer channel walls (Fig. A.4). In the inner channel walls, the pass rates increase from around 0.4 to 0.9 when the number of histories per cycle is increased from 2,000 to 50,000. In the outer channel walls the increase is smaller, from around 0.8 to 0.9.

The pass rates of the implicit, collision and analog estimates of k_{eff} are presented in

Fig. A.5, A.6 and A.7 in Appendix A, respectively. There are no major differences in the pass rates between the estimates, and the number of inactive cycles and source neutrons per cycle don't have much of an effect on the pass rates. The pass rates of the D'Agostino–Pearson test are between 0.9 and 1 in all cases, while the pass rates of the Shapiro–Wilk test are slightly lower.

The pass rates of the analog estimates of the prompt and delayed neutron k_{eff} are illustrated in Fig. A.8 and A.9, respectively. The number of neutron histories per cycle has a clear impact on the normality pass rates of the delayed neutron k_{eff} . With 2,000 source neutrons per cycle, the Shapiro–Wilk test's pass rate is close to zero and D'Agostino–Pearson test's is around 0.1. With 10,000 source neutrons per cycle, the normality tests' pass rates are increased to between 0.7 and 0.8, and with 50,000 source neutrons per cycle, to above 0.8. The drift-in-mean pass rate is close to unity regardless of the number of neutron histories per cycle. The pass rates of the prompt neutron k_{eff} behave like the pass rates of the total k_{eff} .

To summarize the results of the statistical tests, the number of neutron histories per cycle must be above 50,000 to produce adequate statistics for the calculated fast flux estimates and for the analog estimate of prompt neutron k_{eff} . The number of inactive cycles didn't have any major effect on the pass rates. The pass rates of the Shapiro–Wilk test were lower than the ones of the D'Agostino–Pearson test in most cases, which was expected due to the error in the p -value calculation. The pass rates of the drift-in-mean test were close to unity in all cases, meaning that the distributions of the tested variables were stable.

7.2 Underestimation of variance

The underestimation of variance was studied for the pin-wise estimates of the fast neutron flux in cladding. The ratio of the real and the apparent standard deviation f_{σ} was calculated as presented in Sec. 3.3 using the standard deviations and flux estimates of the 50 simulations.

In Fig. 7.2, the average of the pin-wise values of f_{σ} is plotted as a function of the number of inactive cycles for the tested number of neutron histories per cycle. The average value of f_{σ} is close to unity in all cases. Neither the number of inactive cycles nor the number of neutron histories per cycle have any significant effect on f_{σ} . The standard deviation of f_{σ} is quite stable, which means that the variation between the pin-wise values of f_{σ} doesn't change much from case to case. The spatial distribution of the pin-wise f_{σ} was also random-like in all the tested configurations. In conclusion, the apparent variance of the pin-wise fast flux is very close to the true variance in this test problem.

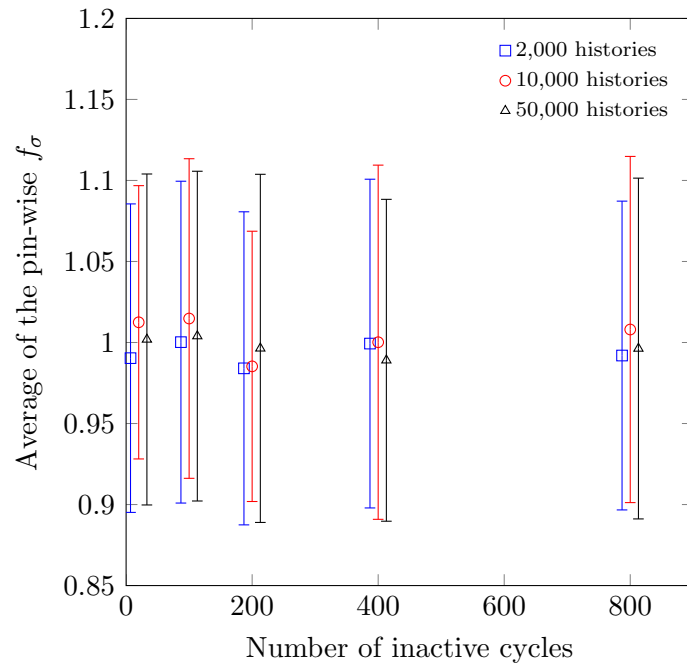


Figure 7.2: Average of the pin-wise f_σ versus the number of inactive cycles. The points are adjusted on the horizontal axis in order to clarify the error bars.

8 Slab test problem

In this section, the statistical tests are studied using a slab geometry with a strongly non-stationary source distribution. The motif for this case is to see how well the tests perform when k_{eff} is changing during the simulation. The chosen geometry for this case was a 300×300 cm square consisting of five fuel regions. Two of the regions consisted of natural UO_2 fuel and the rest had UO_2 fuel enriched to 3.35, 3.5 and 3.65 wt-% ^{235}U . A reflective boundary condition was used in both dimensions, which means that neutrons were diverted back to the geometry from the boundaries. The initial source distribution was set by a point source in the centre of the fuel region 1. The geometry is presented in Fig. 8.1.

The non-stationary source distribution combined with the fuel regions of different enrichment ratios should result in increasing k_{eff} as the simulation proceeds. This behaviour should lead to a non-normal k_{eff} distribution if the number of inactive cycles is not too large. After a sufficient number of inactive cycles, the source distribution should be converged and hence k_{eff} should be normally distributed. In order to see how the statistical tests perform, the slab case was run using 1, 10, 25, 50, 100, 200, 400, 800, 1,600, 3,200, 8,000 and 14,000 inactive cycles. The number of active cycles was 3,000 and the number of source neutrons per cycle 10,000 in each case. 50 independent simulations were run for each case, and the pass rates of the statistical tests were calculated for the analog, implicit and collision estimates of total k_{eff} and for the analog estimates of prompt and delayed neutron k_{eff} .

To illustrate the evolution of the source distribution, a simulation with 1 inactive and 14,000 active cycles was run in which the relative fission rate distribution was plotted after the selected numbers of cycles. The plots are presented in Fig. 8.2. During the first steps of the simulation, the distribution is concentrated close to the initial point source, as can be seen from Fig. 8.2a and 8.2b. After 25-200 cycles (Fig. 8.2c-8.2f), the distribution begins to smooth out in the fuel region 1. Some of the neutrons born in the fuel regions 1 and 2 travel to the region 3 which starts to produce neutrons (Fig. 8.2g). Similarly, the neutrons born mainly in the regions 3 and 4 will reach the region 5, which also begins to generate neutrons (Fig. 8.2h-8.2i). Because the region 5 contains the highest enriched fuel, it begins to dominate the fission rate distribution when the source starts to convergence (Fig. 8.2j-8.2l). It should be noted that Fig. 8.2 illustrates the typical observed source convergence. In some simulation runs, however, the source distribution resembled Fig. 8.2j even after 14,000 cycles.

The pass rates of the statistical tests performed on the estimates of k_{eff} are plotted in Fig. 8.3. The pass rates of the implicit (Fig. 8.3a) and the collision estimates (Fig. 8.3b) are very similar. Between 1 and 200 inactive cycles, the normality tests give a pass rate of around 0.1 and the drift-in-mean test gives a pass rate of zero. These results are expected, as it's clear from Fig. 8.2a-8.2f that the source has not converged when the collecting of the statistics is started. When the number of inactive cycles is increased from 400 to 1,600, the pass rates increase from below 0.1 to around 0.6. At this interval, the source distribution begins to shift to the left side of the geometry (Fig. 8.2g-8.2i), which means that more and more runs are

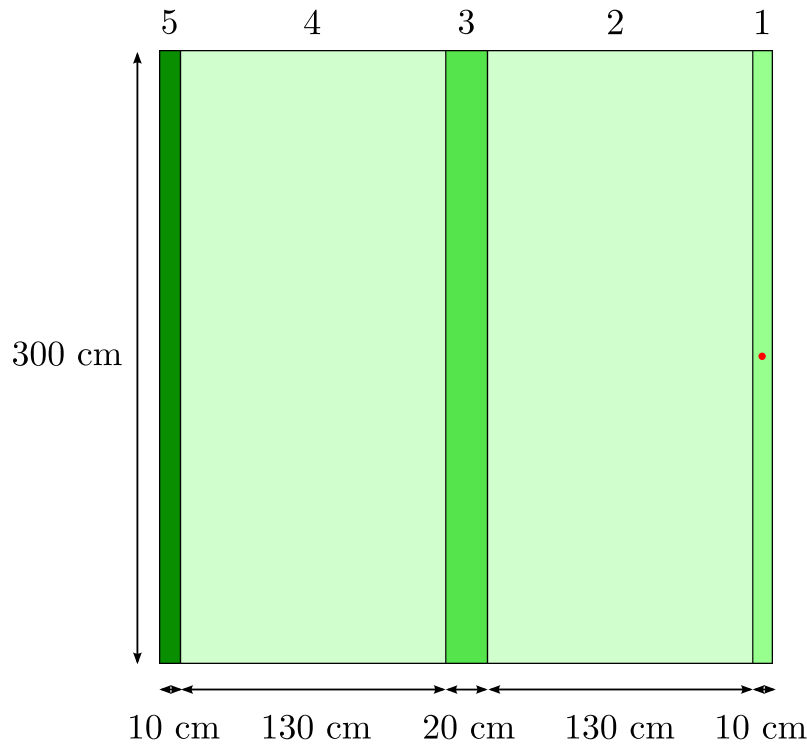


Figure 8.1: The slab geometry. The UO_2 fuel regions are 1: 3.35 wt-% ^{235}U , 2: natural uranium, 3: 3.5 wt-% ^{235}U , 4: natural uranium and 5: 3.65 wt-% ^{235}U . The initial point source is marked as a red circle in the fuel region 1.

likely to pass the tests. As already noted, in some of the simulation runs the source distribution convergence was either faster or slower than in the case of Fig. 8.2, and therefore some runs pass the tests and others don't at this interval. When the number of inactive cycles is 3,200 and above, the pass rate of the drift-in-mean test is close to unity and the pass rates of the normality tests are around 0.9.

The pass rates of the normality tests in the analog estimate (Fig. 8.3c) differ significantly from the implicit and collision estimates. Between 10 and 200 inactive cycles, the pass rates of the normality tests are already around 0.6. Above 400 cycles, the pass rates are over 0.8 and approach to around 0.9 as the source converges. The drift-in-mean test, however, doesn't show this kind of behaviour, and it's very alike in all three estimates. The drift-in-mean test was studied in more detail by plotting histograms of the first and second halves of the cycle-wise values of the analog estimate of k_{eff} . The histograms showed that the k_{eff} distribution was really drifting when the test failed, which confirms that the test gave reliable results in the studied case. An example of a passed and a failed histogram is given in Fig. 8.4.

The pass rates of the analog estimate of prompt neutron k_{eff} (Fig. 8.5a) are quite similar to the ones of the analog estimate of total k_{eff} . The pass rates of the analog estimate of delayed neutron k_{eff} (Fig. 8.5b) are very different, however. The normality tests give a zero pass rate in all points, whereas the pass rate of the drift-in-mean test increases from around 0.6 to close to unity as the source converges. Histogram plots of the delayed neutron k_{eff} showed that the distributions were right-skewed and also quite stable during the active cycles, which confirms that the outputs of

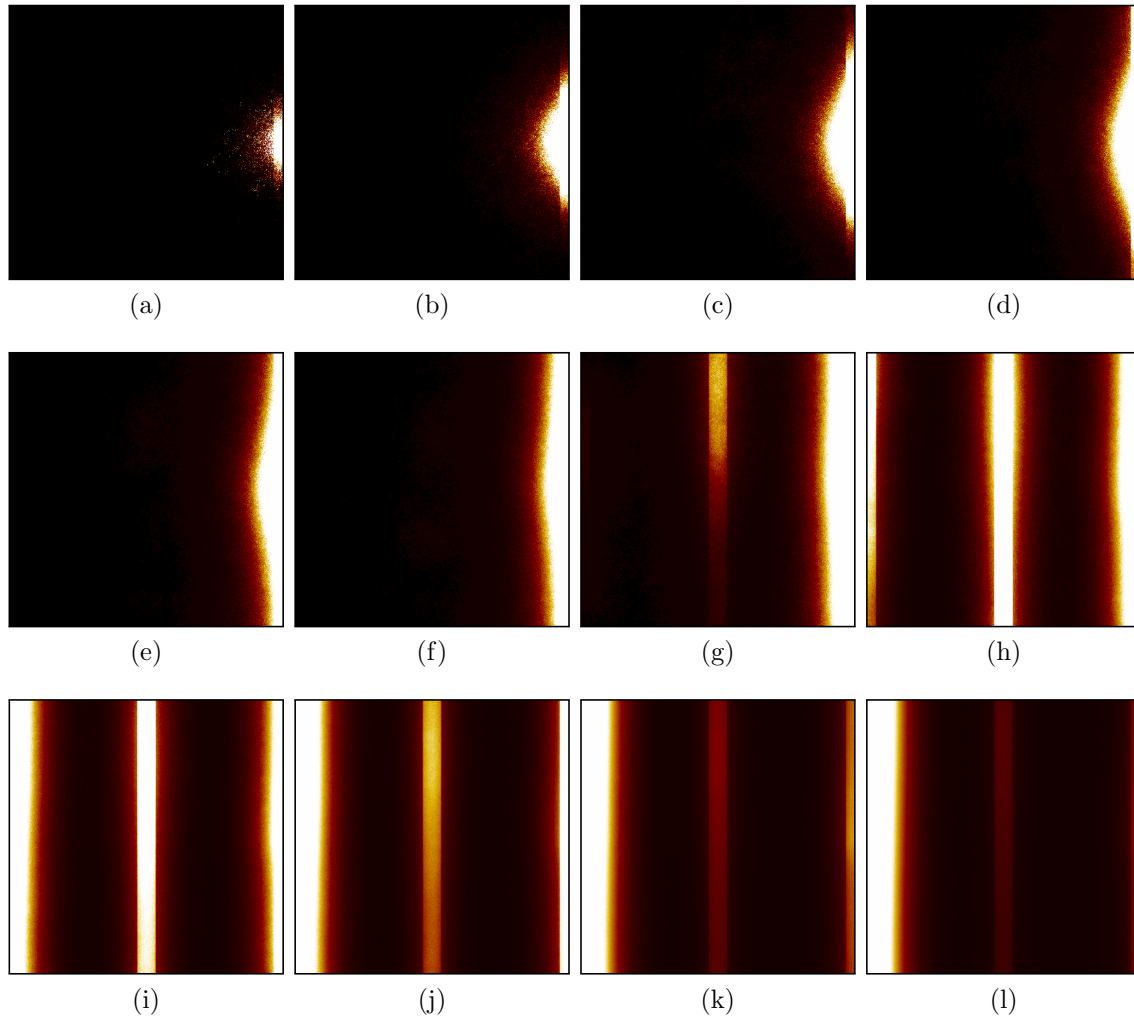


Figure 8.2: The relative fission power in the slab case after (a) 1, (b) 10, (c) 25, (d) 50, (e) 100, (f) 200, (g) 400, (h) 800, (i) 1,600, (j) 3,200, (k) 8,000 and (l) 14,000 cycles. The figures were plotted in a single simulation with 10,000 source neutrons per cycle.

the statistical tests were correct.

In summary, the Shapiro–Wilk and D’Agostino–Pearson tests gave similar pass rates in all the cases. The pass rates of the Shapiro–Wilk test were slightly lower on average, which was most likely due to the error in the p -value calculation discussed in Sec. 7.1. The drift-in-mean test was evidently stricter than the normality tests when the source hadn’t converged. The analog estimate of k_{eff} was clearly more normally distributed than the implicit and collision estimates when the source hadn’t converged. The reason for the difference is somehow related to the way the estimates are calculated, but the exact analysis of this phenomenon is left outside of this work.

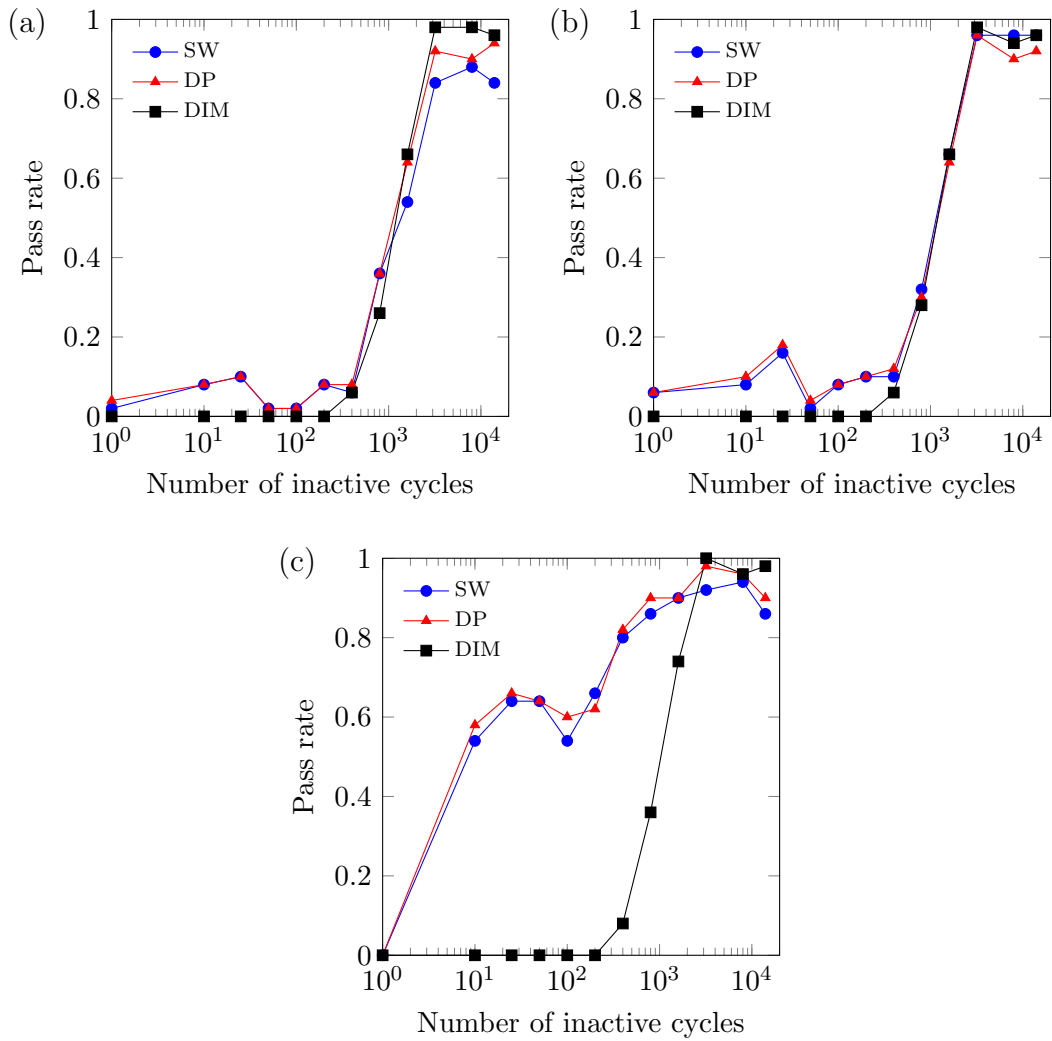


Figure 8.3: Pass rates of the statistical tests for the (a) implicit, (b) collision and (c) analog estimates of k_{eff} as a function of number of inactive cycles.

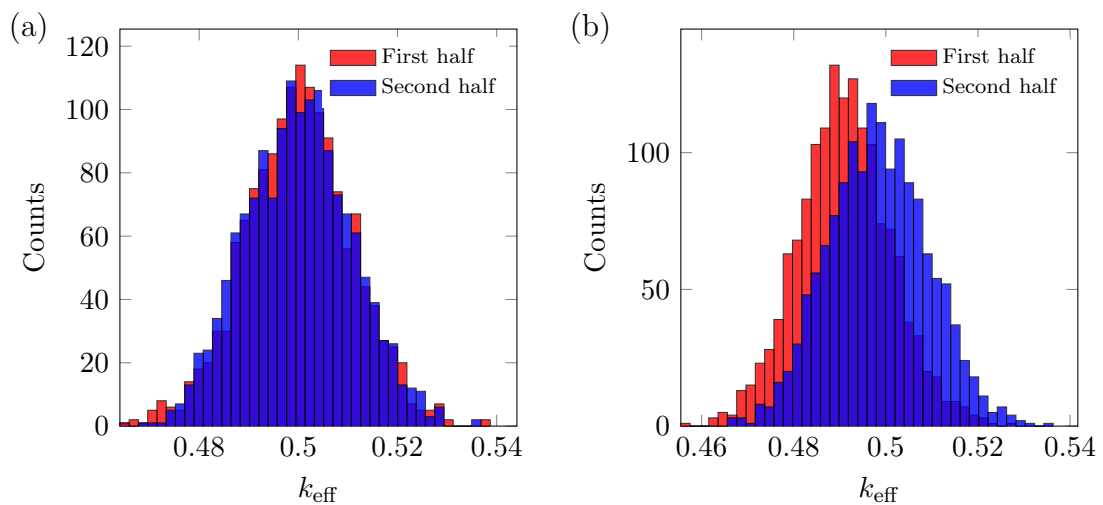


Figure 8.4: Example histograms of the first and second halves of the analog estimates of k_{eff} . The drift-in-mean test passes in (a) and fails in (b).

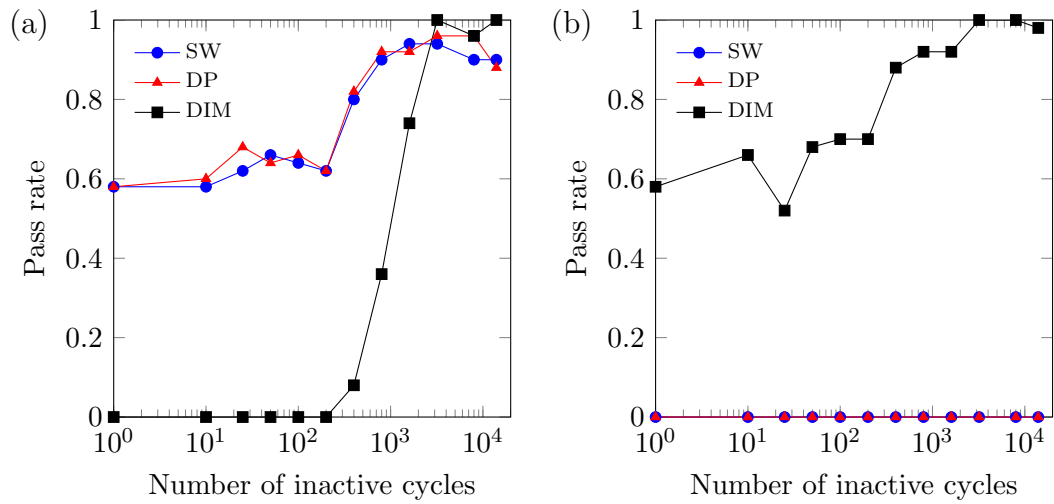


Figure 8.5: Pass rates of the statistical tests for analog estimates of (a) prompt neutron and (b) delayed neutron k_{eff} as a function of number of inactive cycles.

9 Pin-cell array test problem

In this test problem, the underestimation of variance and the statistical tests are studied in an infinite array of irradiated light water reactor (LWR) fuel pins. The problem is defined as “Pin-cell array with irradiated fuel” in Ref. [35] and it contains three cases, each having a different fuel profile. The case selected for this test (case 2.2. in Ref. [35]) represents an axially asymmetric fuel composition, for which it should be difficult to determine when the source has converged into a stationary distribution. This case has also been used to test the underestimation of variance in MCNP5 and KENO Monte Carlo transport codes [12]. In addition to the asymmetric fuel composition, a homogeneous fuel configuration is also studied.

The horizontal and vertical cross sections of the test geometry are presented in Fig. 9.1. Reflective boundary conditions were used in the horizontal directions and a vacuum boundary condition in the vertical direction. The temperature used for the fuel, moderator and cladding was 300 K. The thermal scattering data used for the moderator was evaluated at 293.6 K.

The underestimation of variance was studied using the estimate of the neutron flux, which was tallied in 50 equal-sized volumes spanning the axial length of the fuel on an eight-group energy grid given in table 9.1. 50 independent simulations were run for the tested configurations, and the ratios of the real and the apparent standard deviations were calculated for the flux tallies. The statistical tests were studied by calculating their pass rates for the flux tallies and for the estimates of k_{eff} .

9.1 Homogeneous fuel pin

In the homogeneous fuel pin configuration, all the fuel regions shown in Fig. 9.1b contained the same fuel, namely UO_2 enriched to 1.0 wt-% ^{235}U . The enrichment ratio was selected so that the estimate of k_{eff} would be close to unity. The minimum and maximum estimates of k_{eff} from all the runs of different test configurations were around 1.0390 and 1.0397.

Three sets of cases with 50,000, 100,000 and 250,000 source neutrons per cycle were studied. In each set, the tested numbers of inactive cycles were 10, 25, 50, 100, 400 and 1,600. The batch means method discussed in Sec. 3.3 was studied separately by using 2, 4 and 8 cycles per batch with 400 inactive cycles and 100,000 neutron histories per cycle. The number of active cycles in all these cases was 1,000.

Fig. 9.2a illustrates the mean flux profiles of the energy groups taken over the fluxes of 50 simulation runs. The energy groups 6, 7 and 8 have the highest fluxes, while the groups 3, 4 and 5 have the lowest. The mean relative statistical errors of the fluxes are presented in Fig. 9.2b which shows that the maximum errors are obtained at the ends of the pin and the minimum errors at the centre. The mean flux and error profiles are symmetrical in all the energy groups. However, the flux and error profiles of individual runs were not all symmetrical, e.g. some flux profiles were right- or left-skewed. It was observed that increasing the number of inactive cycles up to

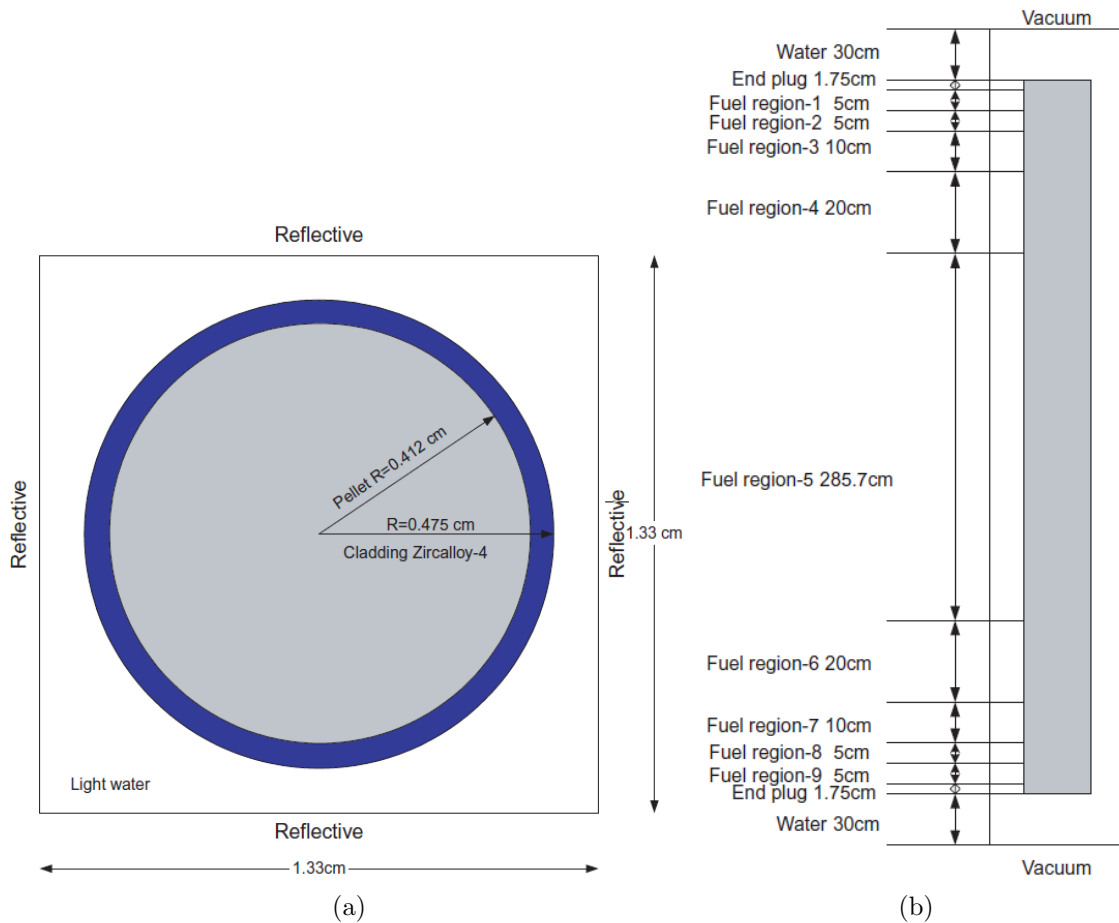


Figure 9.1: Geometry of the fuel pin used in the pin-cell array test problem. Figure (a) presents the horizontal cross section and (b) the vertical cross section [35].

100 decreased the mean relative error.

9.1.1 Results of the statistical tests

The pass rates of the D'Agostino–Pearson test performed on the flux estimates are given in Fig. B.1, B.2 and B.3 in Appendix B as a function of the axial position of the pin for the cases with 50,000, 100,000 and 250,000 source neutrons, respectively. With 50,000 source neutrons per cycle, the pass rates of all the energy groups are peaked when the number of inactive cycles is 10. The pass rates of the energy groups 3, 4, and 5 are non-zero at the centre, unlike the pass rates of the other groups. When the number of inactive cycles is increased to 25, the pass rates are still close to zero near the ends, but non-zero at the centre in all the energy groups. When the number of inactive cycles is 50 and above, the peaks have disappeared, and the maximum pass rate is obtained at the centre and the minima at the ends in all the energy groups. The maximum pass rate is around 0.8 in all the groups and the minimum pass rate is between 0 and 0.2 depending on the group.

When the number of neutron histories per cycle is increased to 100,000, the peaks

Table 9.1: Energy grid used in the pin-cell array test problem.

Energy group	Lower boundary (MeV)	Upper boundary (MeV)
1	1.0×10^{-11}	5.0×10^{-8}
2	5.0×10^{-8}	1.5×10^{-7}
3	1.5×10^{-7}	2.75×10^{-7}
4	2.75×10^{-7}	6.25×10^{-7}
5	6.25×10^{-7}	3.0×10^{-6}
6	3.0×10^{-6}	1.7×10^{-2}
7	1.7×10^{-2}	8.2×10^{-1}
8	8.2×10^{-1}	20

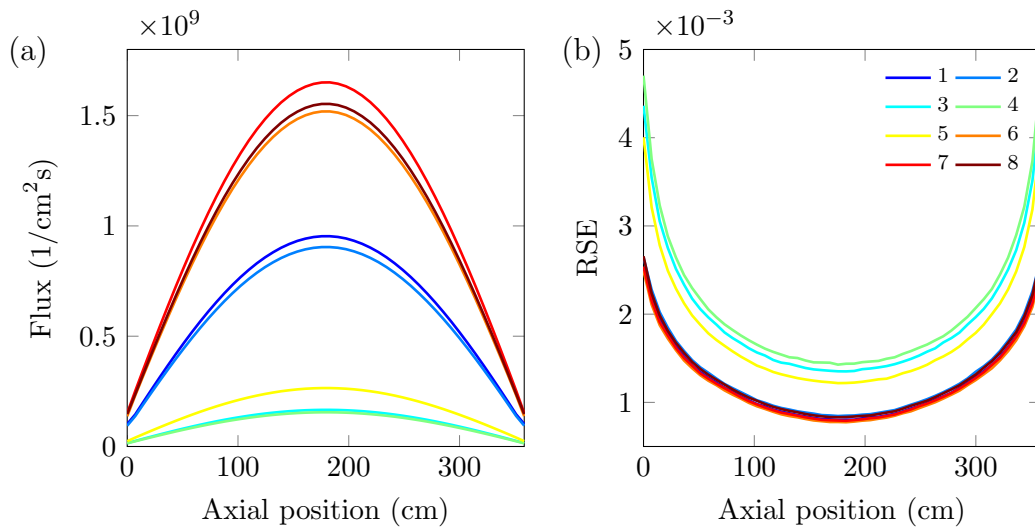


Figure 9.2: (a) The flux profiles of the eight energy groups and (b) their relative statistical errors in the homogeneous fuel pin case. The plots are averages of 50 independent simulations which used 1,600 inactive and 1,000 active cycles with 250,000 neutron histories per cycle.

of the D’Agostino–Pearson test’s pass rates become narrower, and it takes more inactive cycles for the peaks to disappear. For example, when the number of inactive cycles is 50, the pass rates of the energy groups 1, 2, 6, 7 and 8 are clearly peaked (Fig. B.2c), unlike in the case with 50,000 source neutrons (Fig. B.1c). The energy groups 3, 4 and 5 have slightly higher pass rates than the other groups, except near the ends of the pin where their pass rates are close to zero. The pass rate profiles seem to be converged after 100 inactive cycles. The maximum pass rates are around 0.9 which are obtained close to the centre of the pin.

The peaks of the pass rates become even narrower when the number of neutron histories per cycle is increased to 250,000, and they retain their shapes longer than with 100,000 source neutrons per cycle. Again, the pass rate profiles converge after 100 inactive cycles. The pass rates are increased most in the energy groups 3, 4 and 5, which have maximum pass rates of above 0.9. At the ends, the pass-rates of the groups 3, 4 and 5 are clearly larger than with 50,000 and 100,000 source neutrons per cycle, with a value of about 0.4. The energy groups 1, 2, 6, 7 and 8 have their

maximum pass rates of around 0.8 at the centre, while their minimum pass rates are obtained at the axial positions of around 70 cm and 300 cm.

A possible explanation to the peaks observed in the D'Agostino–Pearson test's pass rates is that the initial flat source distribution coincides reasonably well with the final distribution at the locations of the peaks. Therefore, the flux distribution doesn't vary much at these locations, and the distribution is more normal than in the other parts of the pin. When the number of neutron histories per cycle is increased, the resolution of the source distribution increases, which means that the coinciding parts are narrowed and thus, the peaks become narrower.

The small pass rates of the D'Agostino–Pearson test close to the ends of the pin in the energy groups 3, 4 and 5 are most likely caused by poor sampling. This is because these groups have the lowest fluxes which minimum values are obtained at the ends. Also, increasing the number of neutron histories per cycle increased the pass rates most at the ends of the pin in these groups.

The pass-rates of the drift-in-mean test performed on the flux estimates are presented in Fig. B.4–B.6 in Appendix B. The pass rates are very low in every case, regardless of the number of neutron histories per cycle. The pass rates increase only slightly with the number of inactive cycles up to 100 cycles. Above 100 inactive cycles, the maximum pass rates are obtained at the ends and close to the centre of the pin. Fig. B.7, which illustrates the average pass rates taken over the total pass rates of the simulation runs, shows that the standard deviation of the average drift-in-mean pass rate is very large in most cases. This tells that some runs gave high drift-in-mean pass rates while others gave very low. For example, of the 50 runs of the case with 250,000 source neutrons per cycle and 400 inactive cycles, the maximum drift-in-mean pass rate of all the flux tallies was 0.87 in one run and the minimum was 0.03 in another run. Clearly, the converged source is difficult to obtain in some runs, which was already observed by investigating the flux profiles. Because the pass rates of the independent runs differ so much from each other, the cycles were very likely strongly correlated. Another observation from Fig. B.7 is that there is some variation in the pass rates of the normality tests as well, though it is smaller than the variation in the drift-in-mean test. Fig. B.7 also shows the already noted observation that the pass rates of the normality tests increase with the number of neutron histories per cycle.

The statistical tests performed on the estimates of k_{eff} behave quite well in most cases (Fig. B.8–B.10 in Appendix B). In general, the pass rates of the Shapiro–Wilk test are in most cases between 0.85 and 0.95 and the pass rates of the D'Agostino–Pearson test are above 0.9. The differences between the pass rates of the normality tests are most likely due to the error in the p -value calculation discussed in Sec. 7.1. The pass rates of the drift-in-mean test are in most cases above 0.95. The most notable effect of the number of neutron histories per cycle and the number of inactive cycles can be seen with the implicit estimate of k_{eff} (Fig. B.8). When the number of inactive cycles is small, the pass rates of all the tests decrease as the number of neutron histories per cycle is increased. Using 10 inactive cycles, the pass rates of the normality tests and the drift-in-mean test drop from 0.8 to zero and 0.2, respectively, when the number of histories per cycle is increased from 50,000 to

250,000. The pass rates increase with the number of inactive cycles and converge to 0.9 and above after 100 inactive cycles. Similar but more modest behaviour can be seen with the drift-in-mean test in the collision and analog estimates of k_{eff} , and also in the analog estimate of prompt neutron k_{eff} (Fig. B.11). The pass rates of the normality tests performed on the analog estimate of delayed neutron k_{eff} (Fig. B.12) seem to increase slightly as the number of source neutrons is increased.

9.1.2 Underestimation of variance

Fig. B.13, B.14 and B.15 in Appendix B present f_{σ} of the flux tallies as a function of the axial position of the pin for the cases with 50,000, 100,000 and 250,000 source neutrons per cycle, respectively. The shared feature in all the cases is that f_{σ} obtains its minimum values at the centre and at the ends of the fuel pin. Between the ends and the centre, f_{σ} has peaks which shape depends on the number of inactive cycles. When the number of inactive cycles is small, the peaks are higher and narrower, especially when the number of source neutrons per cycle is 100,000 and 250,000. The distribution of f_{σ} is not symmetrical in general, for example in Fig. B.15e the left-side peak is strongly right-skewed. The energy groups 1, 2, 6, 7 and 8 produce the highest values of f_{σ} , which vary from around 3 to 14, depending on the case and the axial location. The energy groups 3, 4 and 5 produce the lowest values, which vary from around 2 to 10. The number of source neutrons per cycle doesn't seem to have a significant effect on the magnitude of f_{σ} .

Increasing the number of cycles used per batch decreases the underestimation of the variance, which can be seen in Fig. 9.3. As the number of cycles used for each batch is increased from 1 to 8, the maximum value of f_{σ} drops from around 13 to 5. Therefore, the batch means method can be used for obtaining a better estimate of the variance.

9.2 Heterogeneous fuel pin

In this case, the pin contains asymmetrically positioned fuel regions of different burn-ups. The ends of the fuel pin consist of short, low-burnup regions which are separated by a long, high-burnup region. The fuel profile is given in table 9.2 and the fuel compositions are presented in table C.1 in Appendix C. Four cases with different number of neutron histories per cycle were studied. In three of them, which had 10,000, 50,000 and 250,000 histories per cycle, the number of inactive cycles was 300 and the number of active cycles was 1,000. The fourth case had 1,600 inactive and 2,000 active cycles with 500,000 source neutrons per cycle. The statistical tests were studied only in the first three cases due to an error in the input file of the case with 500,000 source neutrons.

The mean flux profiles and the mean relative statistical errors taken over 50 runs are plotted in Fig. 9.4 for the case with 500,000 neutron histories. The flux is very low at the bottom of the fuel pin, and the maximum flux is obtained close to the top of the pin. The relative error obtains its maximum at the bottom of the pin.

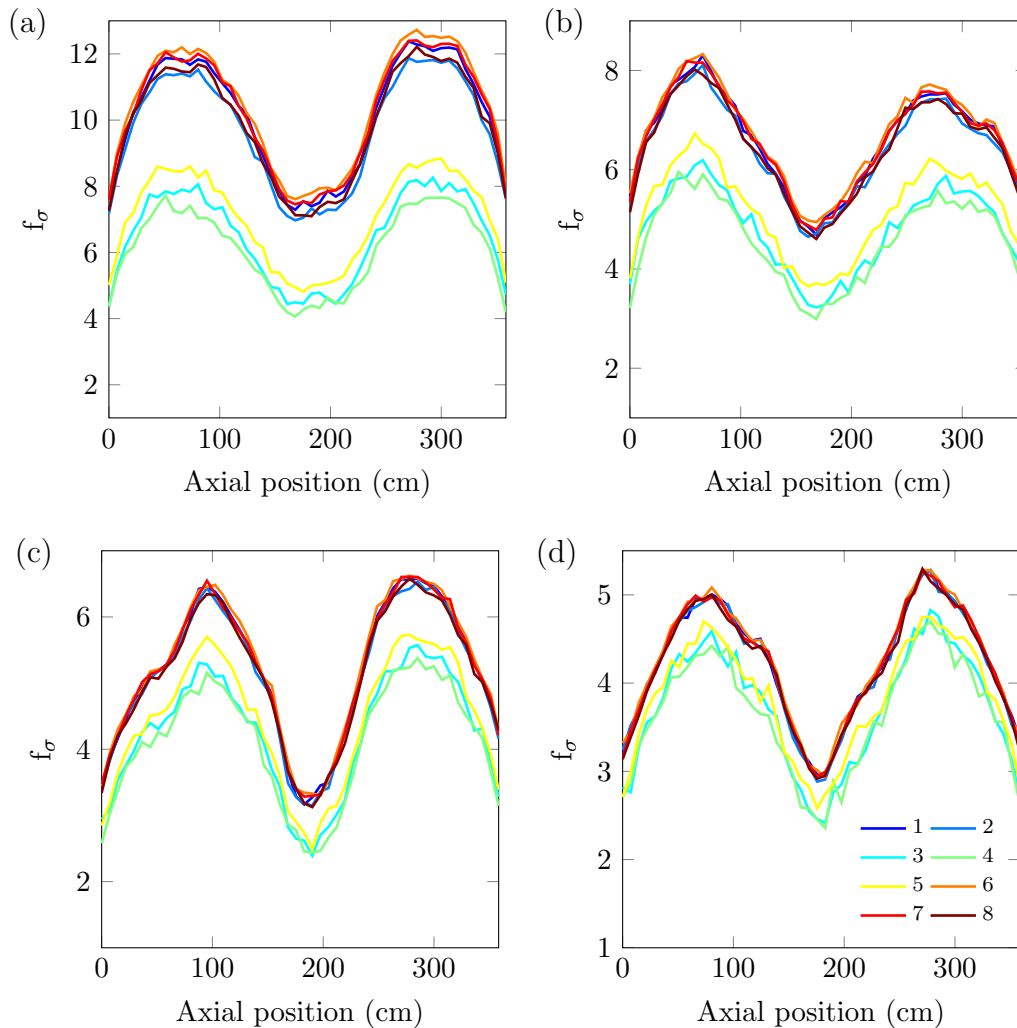


Figure 9.3: The ratio of the real and the apparent standard deviation of the total flux for the eight energy groups in the homogeneous fuel pin case using (a) 1, (b) 2, (c) 4 and (d) 8 cycles for each batch.

The energy groups 6, 7 and 8 have the highest fluxes whereas the groups 3 and 4 have the lowest.

9.2.1 Results of the statistical tests

The very low flux at the bottom of the pin results in small number of recorded scores at the region. This can be seen from Fig. C.1 in Appendix C, which illustrates the pass rates of the D’Agostino–Pearson test. The pass rates are non-zero only in the high-flux region. In some runs, no scores were recorded at the bottom of the pin, and therefore the pass rates couldn’t be calculated at the region. With 50,000 source neutrons per cycle, the pass rates of the energy groups 3, 4 and 5 are between 0.6 and 0.8 in the high-flux region, while the other groups have their maximum values between around 0.8 and 1. Increasing the number of neutron histories per cycle broadens the non-zero region and also increases the pass rates of the energy groups

Table 9.2: Fuel pin composition in the heterogeneous fuel pin case [35].

Fuel region	Fuel material
1	B21G
2	B24G
3	B30G
4	B40G
5	B55G
6	B55G
7	B40G
8	B30G
9	B24G

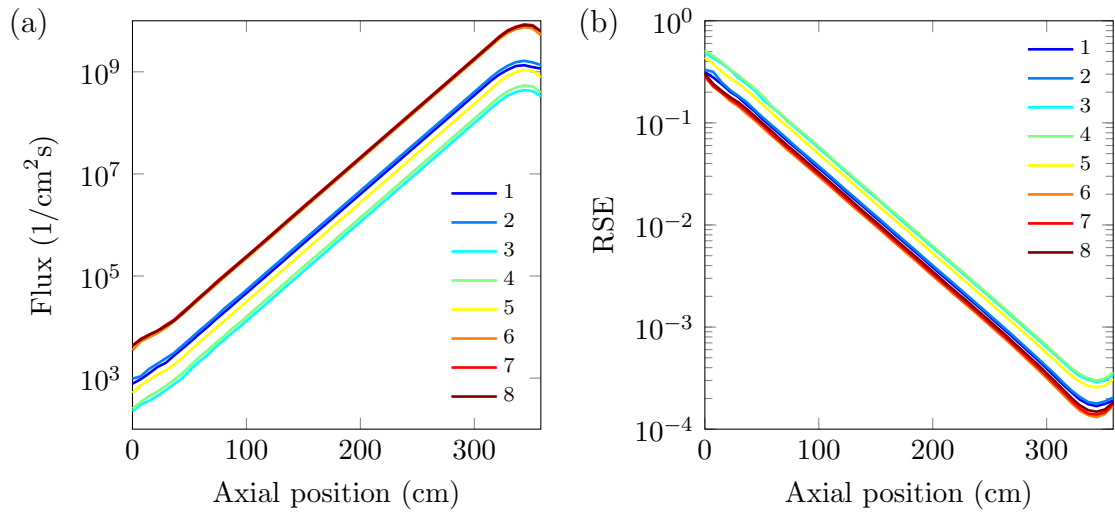


Figure 9.4: (a) The flux profiles of the eight energy groups and (b) their relative statistical errors in the heterogeneous fuel pin case. The plots are averages of 50 independent simulations which used 1,600 inactive and 2,000 active cycles with 500,000 neutron histories per cycle.

3, 4 and 5 to the same level with the other groups. The energy groups 3, 4 and 5 have slightly narrower non-zero region in each case. Similar results were also obtained with the Shapiro–Wilk test.

The drift-in-mean test (Fig. C.2) behaves very differently from the D’Agostino–Pearson test. The pass rates are non-zero in the whole region in which all the runs produced statistics. The number of source neutrons per cycle doesn’t have any significant effect on the pass rates, with the exception that it broadens the region producing statistics. All the energy groups obtain their maximum pass rates of around unity in the high-flux region. The energy groups 3, 4 and 5 have higher overall pass rates than the other groups in every case.

The pass rates of the statistical tests performed on the estimates of k_{eff} are presented in Fig. C.3. The pass rates of the normality tests are relatively stable with all estimates, and most of the pass rates are above 0.9. The pass rates of the drift-in-

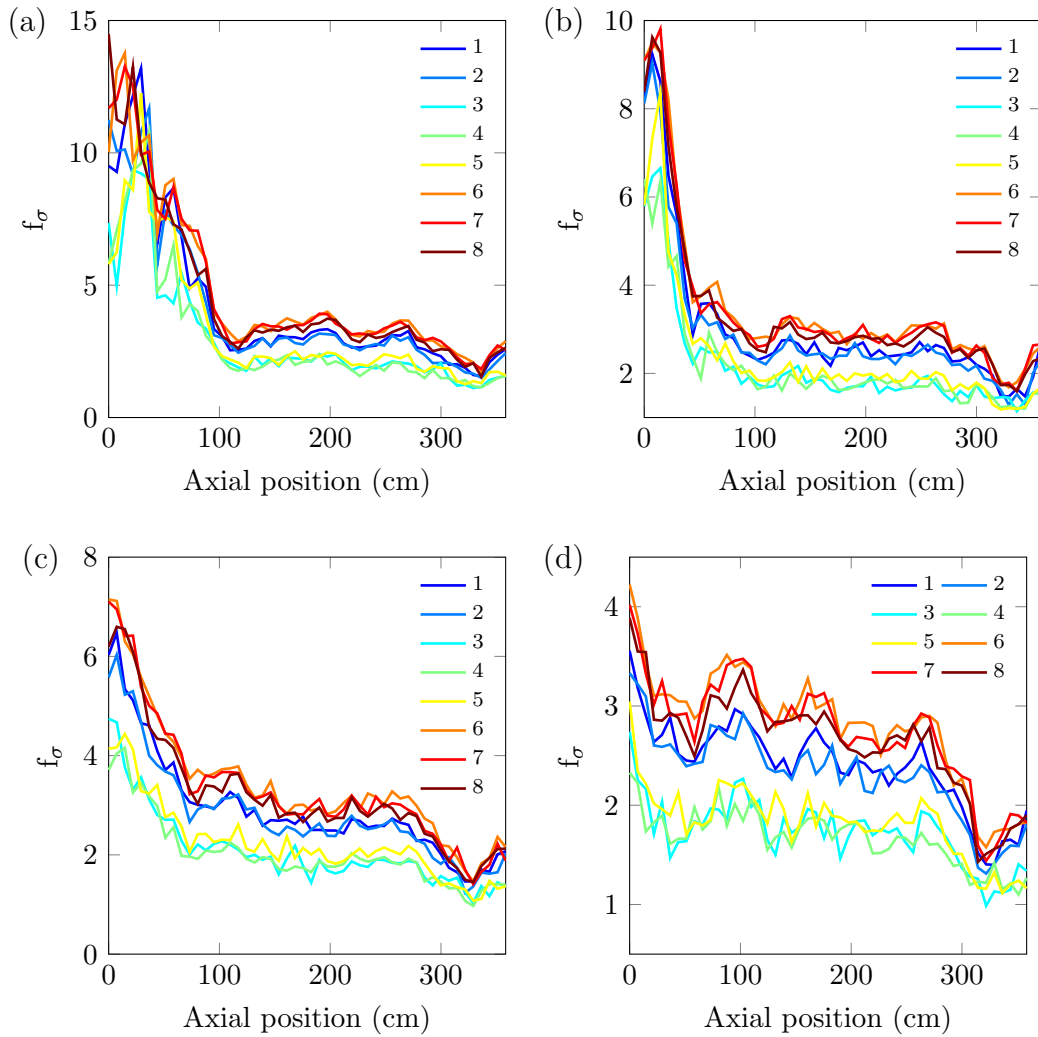


Figure 9.5: The ratio of the real and the apparent standard deviation of the total flux for the eight energy groups in the heterogeneous fuel pin case using (a) 10,000, (b) 50,000, (c) 250,000 and (d) 500,000 neutron histories per cycle.

mean test are close to unity with the collision and analog estimates, while with the implicit estimate the pass rates are between 0.9 and 0.95. The effect of the number of neutron histories per cycle is seen in the analog estimate of delayed neutron k_{eff} (Fig. C.4b), for which the pass rates of the normality tests are around 0.3 when the number of histories per cycle is 10,000. The pass rates increase to around 0.8 when the number of histories per cycle is increased to 50,000 and above.

9.2.2 Underestimation of variance

The ratio of the real and the apparent standard deviation f_σ is plotted in Fig. 9.5 for the tested numbers of source neutrons per cycle. In all the cases, f_σ obtains its maximum close to the bottom and minimum near the top of the pin. The energy groups 3, 4 and 5 have the lowest values in all the cases. The increase in the number of neutron histories per cycle decreases the value of f_σ near the bottom of

the pin, but above 100 cm f_σ is quite stable. With 10,000 source neutrons per cycle, the maximum value of f_σ is around 15, whereas with 500,000 neutrons, it drops to around 4. The minimum value of f_σ is close to unity in all the cases. The 50,000 and 250,000 source neutrons cases were repeated three times in order to see how much f_σ varies. With 50,000 source neutrons per cycle, the obtained maximum values of f_σ were around 7, 10 and 15, whereas with 250,000 source neutrons per cycle, the maximum values were all around 7. This implies that there are larger variations in f_σ when the number of source neutrons per cycle is small.

The results imply that the underestimation of variance in this case is due to poor sampling in the low-flux region at the bottom of the pin. As the number of neutron histories per cycle is increased, more neutrons are sampled in the low-flux region and thus, more statistics are collected. However, the results differ from the ones presented in Ref. [12], in which the underestimation of variance was studied with MCNP5 and KENO using the same case and parameters. With 10,000 histories per cycle, the results are comparable, but with 50,000 and 250,000 source neutrons per cycle, the maximum value of f_σ is much lower with MCNP (below 2.5 and 0.6, respectively) and with KENO (below 1.5 and 0.4, respectively). The differences in the results are possibly due to differences in the standard deviation estimators. For example, KENO uses an iterative approach and lag covariance data to estimate the variance [36]. Nevertheless, the authors of Ref. [12] also concluded that the underestimation of the variance in this case is due to poor sampling in the low-flux region.

10 Conclusions

In this work the statistical tests, namely the Shapiro–Wilk and D’Agostino–Pearson normality tests and the drift-in-mean test, were implemented in Serpent. The tests can be used for detecting a non-converged neutron source distribution and poor sampling which can result in inaccurate estimates of k_{eff} and reaction rates. The statistical tests were studied using four test problems: the BWR lattice problem, the slab case and the two pin-cell array cases. In addition to the statistical tests, the underestimation of variance was studied using the BWR and the pin-cell array cases.

The Shapiro–Wilk and the D’Agostino–Pearson tests behaved similarly in most of the tested cases. When the problem was well-behaved, i.e. the source distribution had converged and the number of neutron histories per cycle was large enough, the pass rates of the D’Agostino–Pearson test were usually between 0.9 and 1 with a significance level of 0.05. Due to an error in the p -value calculation, the pass rates of the Shapiro–Wilk test were lower in general.

In the tested cases, low pass rates of the normality tests were caused by poor sampling of source neutrons and non-converged source distributions. In the BWR lattice case, the pass rates of the normality tests were low in small-volume tally regions when the number of neutron histories per cycle was small. In the heterogeneous fuel pin case, the normality tests failed in the bottom of the fuel pin, where the flux was very low. In both of these cases, the pass rates of the normality tests increased with the number of source neutrons per cycle.

In the slab test problem, the statistical tests were studied using a strongly non-stationary source distribution. The pass rates of the normality tests were close to zero with the implicit and collision estimates of k_{eff} when the source hadn’t been converged, and increased close to unity when the source started to converge. The normality tests’ pass rates increased also in the homogeneous fuel pin case with the number of inactive cycles up to a certain level. The drift-in-mean test, however, proved to be a better tool for detecting a non-converged source distribution. In the slab case the drift-in-mean test was stricter than the normality tests when the source hadn’t been converged. In the homogeneous fuel pin case, the pass rates of the drift-in-mean test were much lower than the ones of the normality tests, and they also varied a lot from run to run. These results imply that the source convergence was difficult to obtain in some runs of the homogeneous fuel pin case, and that the cycles were strongly correlated.

The underestimation of variance was studied in the BWR lattice case using the estimates of pin-wise fast flux. The ratio of the real and the apparent standard deviation f_{σ} was close to unity even with a small number of neutron histories per cycle, so the uncertainty underestimation was no major problem in this case.

The underestimation of variance caused by inter-cycle correlations proved to be a major issue in the homogeneous fuel pin case. The maximum observed value of f_{σ} was close to 15 with 250,000 neutron histories per cycle and 1,600 inactive cycles. Neither the number of neutron histories per cycle nor the number of inactive cycles

had a consistent effect on the maximum value of f_σ . However, increasing the number of cycles used per batch from one to eight decreased the maximum value of f_σ down to 5. Therefore, the batch means method can be used for obtaining a better estimate of the variance. Increasing the number of cycles used per batch above eight should decrease the underestimation even further. In other studies hundreds of cycles per batch have been used [15, 16].

In the heterogeneous fuel pin case the underestimation of variance was caused by poor sampling in the low-flux region at the bottom of the fuel pin. By increasing the number of neutron histories per cycle from 10,000 to 500,000, the maximum value of f_σ obtained at the bottom of the bin decreased from 15 to 4. The minimum value of f_σ , which was obtained in the high-flux region at the top of the bin, was close to unity regardless of the number of neutron histories per cycle.

References

- [1] E. Cashwell and C. Everett, *A Practical Manual on the Monte Carlo Method for Random Walk Problems*, Los Alamos Scientific Laboratory of the University of California (1957).
- [2] F. B. Brown, ““K-effective of the World” and Other Concerns for Monte Carlo Eigenvalue Calculations,” *Progress in Nuclear Science and Technology*, **2**, 738–742 (2011).
- [3] W. R. Martin, “Challenges and Prospects for Whole-Core Monte Carlo Analysis,” *Nuclear Engineering and Technology*, **44**, 2, 151–160 (2012).
- [4] J. Leppänen, “Performance of Woodcock Delta-Tracking in Lattice Physics Applications Using the Serpent Monte Carlo Reactor Physics Burnup Calculation Code,” *Annals of Nuclear Energy*, **37**, 5, 715–722 (2010).
- [5] O. Kallenberg, *Foundations of Modern Probability*, Applied probability, Springer (2002).
- [6] R. Ware and F. Lad, “Approximating the Distribution for Sums of Products of Normal Variables,” Technical Report UCDMS 2003/15, University of Canterbury, England (2003).
- [7] E. Díaz-Francés and F. J. Rubio, “On the Existence of a Normal Approximation to the Distribution of the Ratio of Two Independent Normal Random Variables,” *Statistical Papers*, **54**, 2, 309–323 (2013).
- [8] X-5 Monte Carlo Team, *MCNP - A General Monte Carlo N-Particle Transport Code, Version 5*, LA-UR-03-1987, Los Alamos National Laboratory, 2003.
- [9] F. B. Brown, “On the Use of Shannon Entropy of the Fission Distribution for Assessing Convergence of Monte Carlo Criticality Calculations,” *Proc. PHYSOR 2006*, Vancouver, British Columbia, Canada, Sept. 10–14, 2006.
- [10] T. Ueki, “Information Theory and Undersampling Diagnostics for Monte Carlo Simulation of Nuclear Criticality,” *Nuclear Science and Engineering*, **151**, 3, 283–292 (2005).
- [11] F. Brown, “A Review of Best Practices for Monte Carlo Criticality Calculations,” *Proc. ANS NCSD-2009*, Richland, WA, Sept. 13-17, 2009.
- [12] B. T. Mervin, S. W. Mosher, J. C. Wagner, and G. I. Maldonado, “Uncertainty Underprediction in Monte Carlo Eigenvalue Calculations,” *Nuclear Science and Engineering*, **173**, 3, 276–292 (2013).
- [13] M.-J. Lee, H. G. Joo, D. Lee, and K. Smith, “Multigroup Monte Carlo Reactor Calculation with Coarse Mesh Finite Difference Formulation for Real Variance Reduction,” *Proc. Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2010 (SNA + MC2010)*, Tokyo, Japan, October 17–21, 2010.

- [14] G. S. Fishman, “Grouping Observations in Digital Simulation,” *Management Science*, **24**, 5, 510–521 (1978).
- [15] T. Ueki, “Batch Estimation of Statistical Errors in the Monte Carlo Calculation of Local Powers,” *Annals of Nuclear Energy*, **38**, 11, 2462–2469 (2011).
- [16] D. J. Kelly, T. M. Sutton, and S. C. Wilson, “MC21 analysis of the nuclear energy agency Monte Carlo performance benchmark problem,” *Proc. PHYSOR 2012*, Knoxville, TN, Apr. 15–20, 2012.
- [17] R. B. D’Agostino, A. Belanger, and R. B. J. D’Agostino, “A Suggestion for Using Powerful and Informative Tests of Normality,” *The American Statistician*, **44**, 4, 316–321 (1990).
- [18] B. W. Yap and C. H. Sim, “Comparisons of Various Types Of Normality Tests,” *Journal of Statistical Computation and Simulation*, **81**, 12, 2141–2155 (2011).
- [19] T. Thadewald and H. Büning, “Jarque–Bera Test and its Competitors for Testing Normality – A Power Comparison,” *Journal of Applied Statistics*, **34**, 1, 87–105 (2007).
- [20] J. Zhang and Y. Wu, “Likelihood-Ratio Tests for Normality,” *Computational Statistics & Data Analysis*, **49**, 3, 709–721 (2005).
- [21] J. Hain, “Comparison of Common Tests for Normality,” Master’s thesis, Julius-Maximilians-Universität Würzburg, 2010.
- [22] E. Lloyd, “Least-Squares Estimation of Location and Scale Parameters Using Order Statistics,” *Biometrika* (1952).
- [23] S. S. Shapiro and M. B. Wilk, “An Analysis of Variance Test for Normality (Complete Samples),” *Biometrika*, **52**, 3/4, 591–611 (1965).
- [24] P. Royston, “Approximating the Shapiro-Wilk W-test for Non-Normality,” *Statistics and Computing*, **2**, 3, 117–119 (1992).
- [25] P. Royston, “A Toolkit for Testing for Non-Normality in Complete and Censored Samples,” *Journal of the Royal Statistical Society, Series D - The Statistician*, **42**, 1, 37–44 (1993).
- [26] G. Blom, *Statistical Estimates and Transformed Beta-Variables*, Wiley (1958).
- [27] S. Weisberg and C. Bingham, “An Approximate Analysis of Variance Test for Non-Normality Suitable for Machine Calculation,” *Technometrics*, **17**, 1, 133–134 (1975).
- [28] R. B. D’Agostino, “Transformation to Normality of the Null Distribution of g_1 ,” *Biometrika*, **57**, 3, 679–681 (1970).
- [29] F. J. Anscombe and W. J. Glynn, “Distribution of the Kurtosis Statistic b_2 for Normal Samples,” *Biometrika*, **70**, 1, 227–234 (1983).

- [30] R. D'Agostino and E. S. Pearson, "Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$," *Biometrika*, **60**, 3, 613–622 (1973).
- [31] G. D. Ruxton, "The Unequal Variance t-Test is an Underused Alternative to Student's t-test and the Mann–Whitney U Test," *Behavioral Ecology*, **17**, 4, 688–690 (2006).
- [32] J. R. Gleason, "A Note on a Proposed Student t Approximation," *Computational Statistics Data Analysis*, **34**, 1, 63 - 66 (2000).
- [33] M. Wichura, "Algorithm AS 241: The Percentage Points of the Normal Distribution," *Applied Statistics*, **37**, 3, 477–484 (1988).
- [34] J. Leppänen, *Serpent – a Continuous-energy Monte Carlo Reactor Physics Burnup Calculation Code*, VTT Technical Research Centre of Finland, 2013.
- [35] R. Blomquist, M. Armishaw, D. Hanlon, and O. N. E. Agency, *Source Convergence in Criticality Safety Analyses, Phase I: Results for Four Test Problems*, Number 5431 in Nuclear Science, Nuclear Energy Agency, Organisation for Economic Cooperation and Development (2006).
- [36] S. Goluoglu, L. M. J. Petrie, M. E. Dunn, D. F. Hollenbach, and B. T. Rearden, "Monte Carlo Criticality Methods and Analysis Capabilities in SCALE," *Nuclear Technology*, **174**, 2, 214–235 (2011).

A BWR

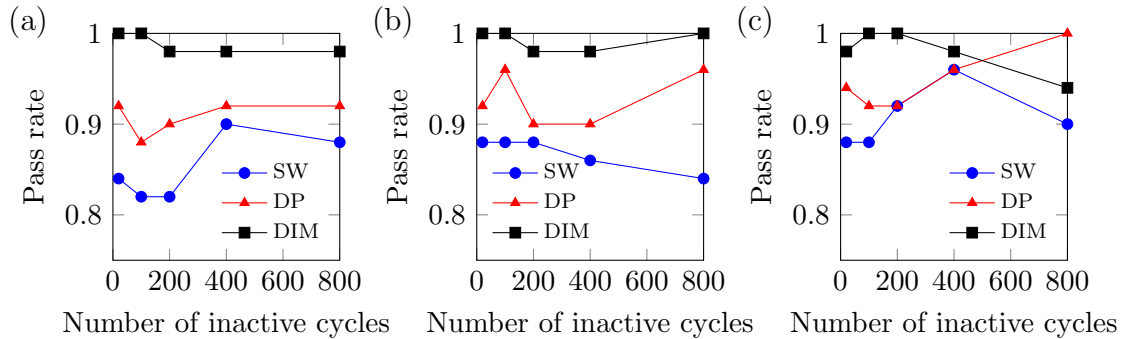


Figure A.1: Pass rates of the statistical tests performed on detector 1 values (average fast flux in cladding) using (a) 2,000, (b) 10,000 and (c) 50,000 neutron histories per cycle.

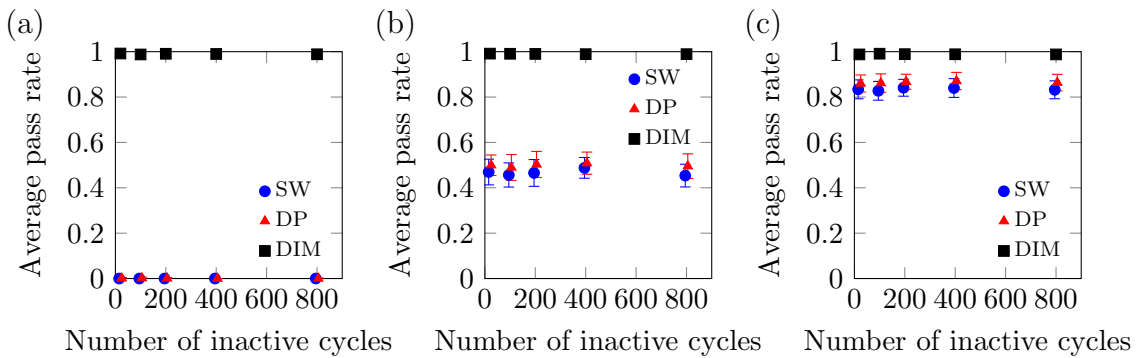


Figure A.2: Average pass rates of the statistical tests performed on detector 2 values (pin-wise fast flux in cladding) using (a) 2,000, (b) 10,000 and (c) 50,000 neutron histories per cycle.

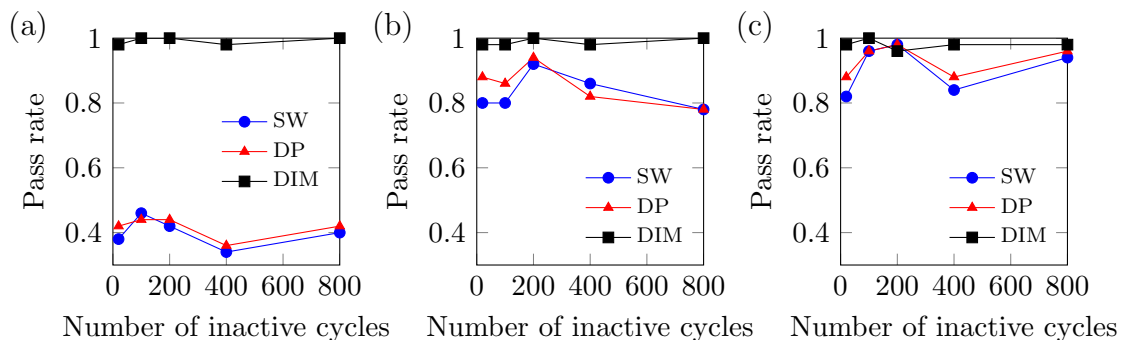


Figure A.3: Pass rates of the statistical tests performed on detector 3 values (fast flux in inner moderator channel wall) using (a) 2,000, (b) 10,000 and (c) 50,000 neutron histories per cycle.

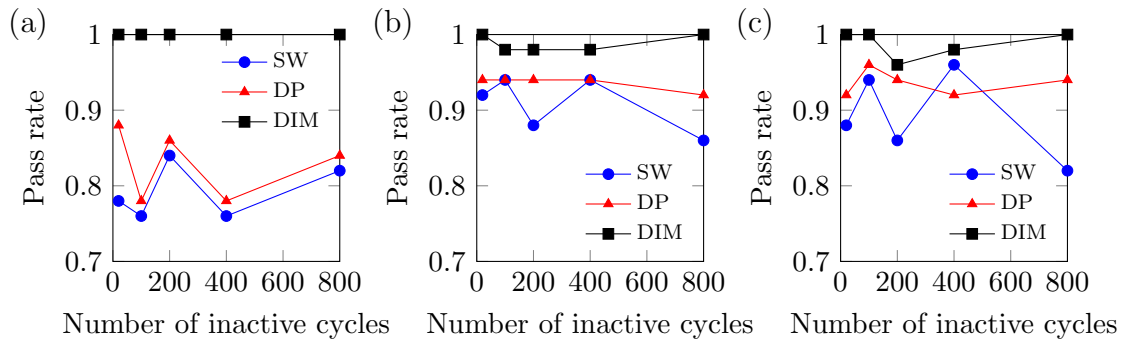


Figure A.4: Pass rates of the statistical tests performed on detector 4 values (fast flux in outer channel wall) using (a) 2,000, (b) 10,000 and (c) 50,000 neutron histories per cycle.

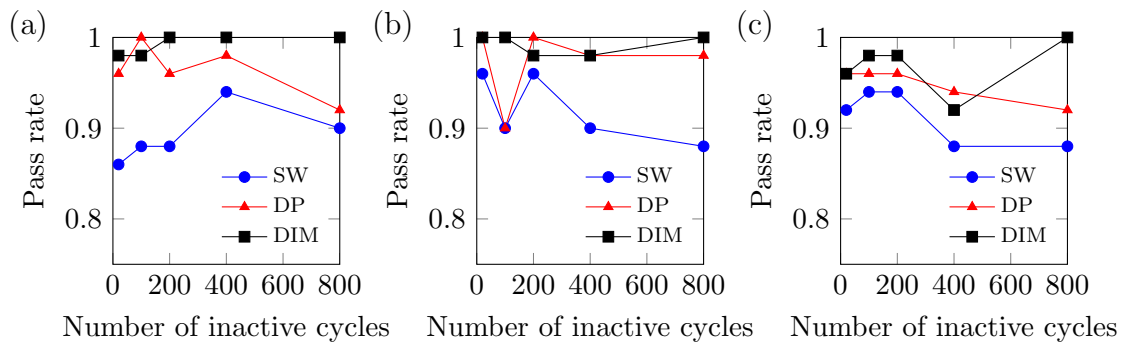


Figure A.5: Pass rates of the statistical tests performed on the implicit estimate of k_{eff} using (a) 2,000, (b) 10,000 and (c) 50,000 neutron histories per cycle.

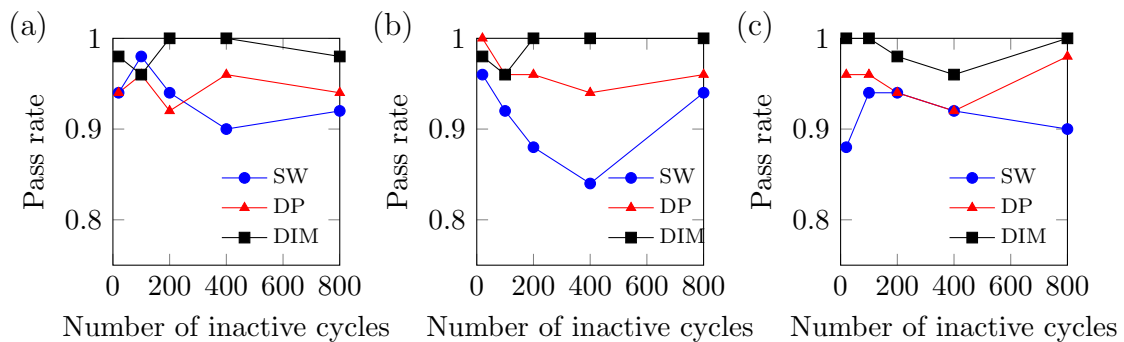


Figure A.6: Pass rates of the statistical tests performed on the collision estimate of k_{eff} using (a) 2,000, (b) 10,000 and (c) 50,000 neutron histories per cycle.

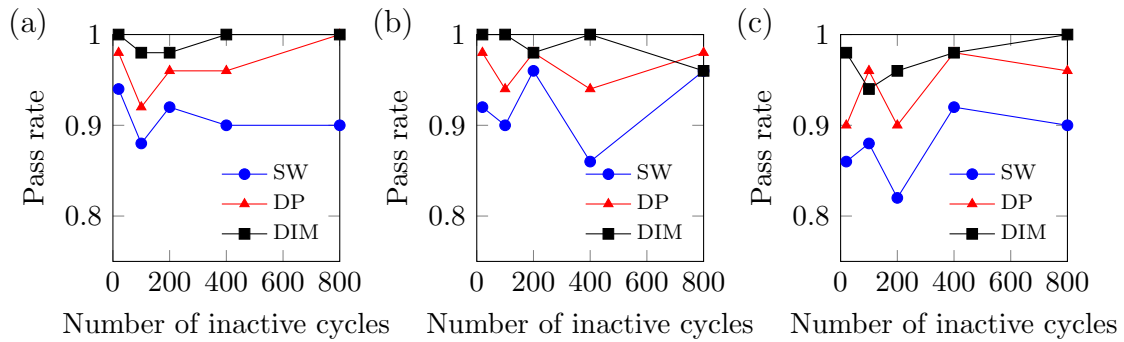


Figure A.7: Pass rates of the statistical tests performed on the analog estimate of k_{eff} using (a) 2,000, (b) 10,000 and (c) 50,000 neutron histories per cycle.

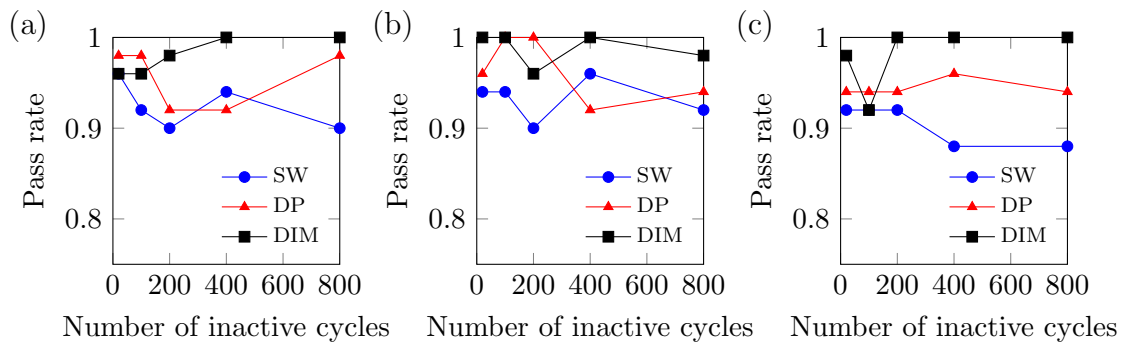


Figure A.8: Pass rates of the statistical tests performed on the analog estimate of prompt neutron k_{eff} using (a) 2,000, (b) 10,000 and (c) 50,000 neutron histories per cycle.

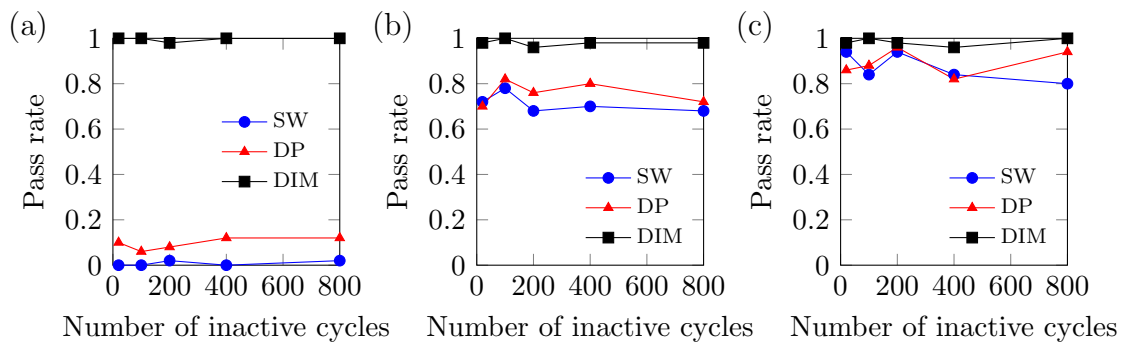


Figure A.9: Pass rates of the statistical tests performed on the analog estimate of delayed neutron k_{eff} using (a) 2,000, (b) 10,000 and (c) 50,000 neutron histories per cycle.

B Homogeneous fuel pin

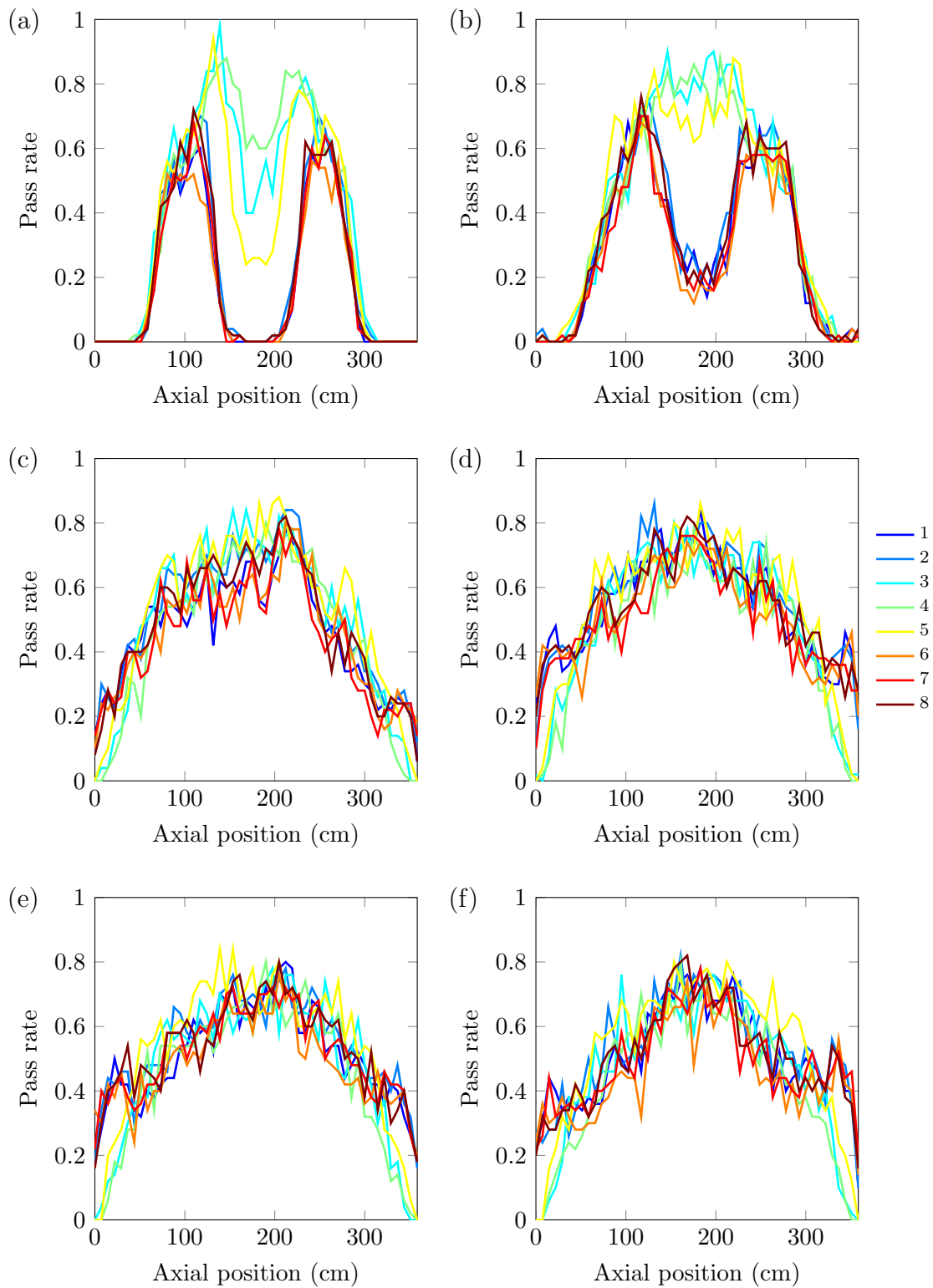


Figure B.1: Pass rates of the D'Agostino–Pearson test for the flux tallies as a function of the axial position of the pin using 50,000 neutron histories per cycle. The numbers of inactive cycles used were (a) 10, (b) 25, (c) 50, (d) 100, (e) 400 and (f) 1,600.

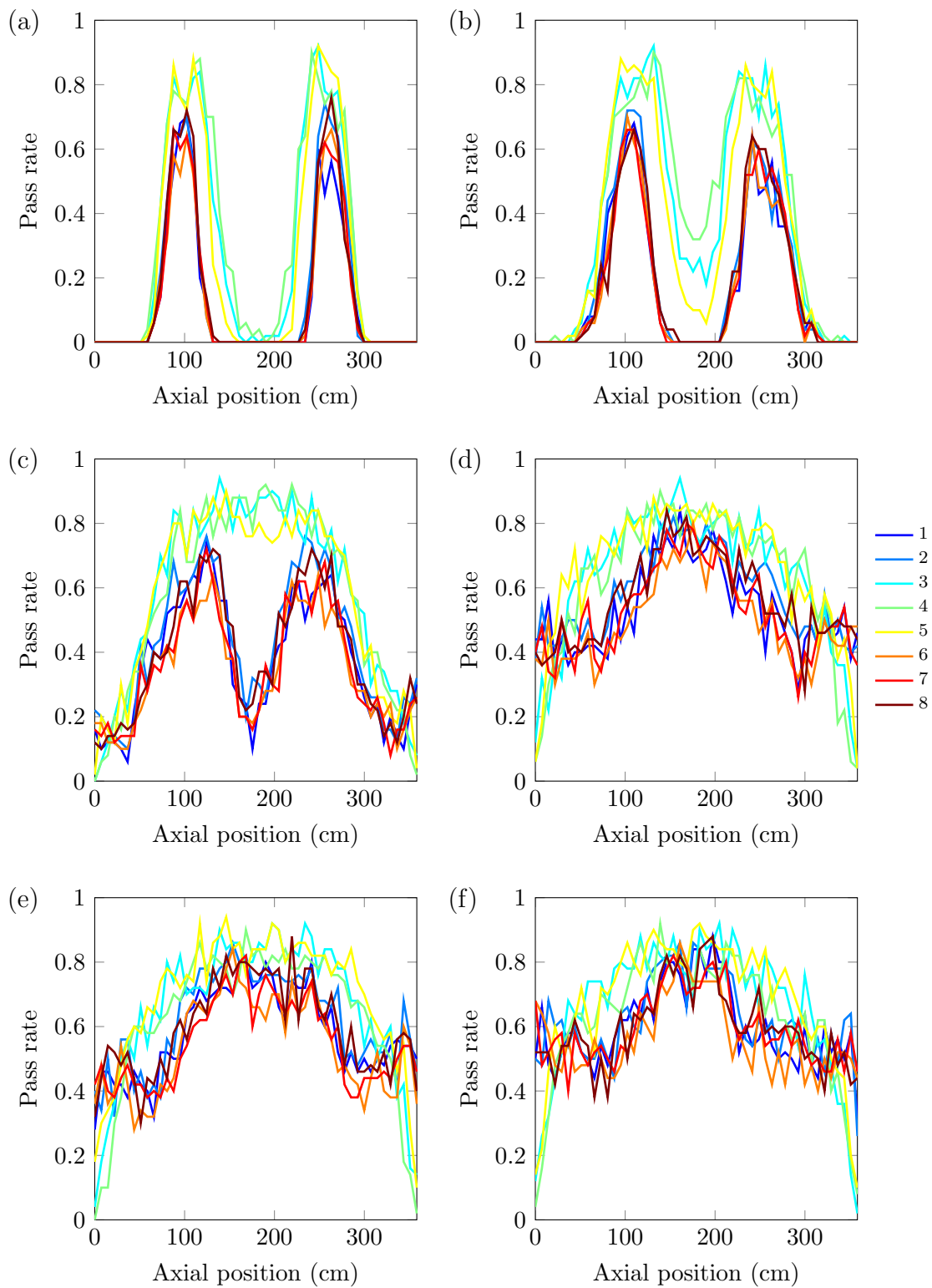


Figure B.2: Pass rates of the D'Agostino–Pearson test for the flux tallies as a function of the axial position of the pin using 100,000 neutron histories per cycle. The numbers of inactive cycles used were (a) 10, (b) 25, (c) 50, (d) 100, (e) 400 and (f) 1,600.

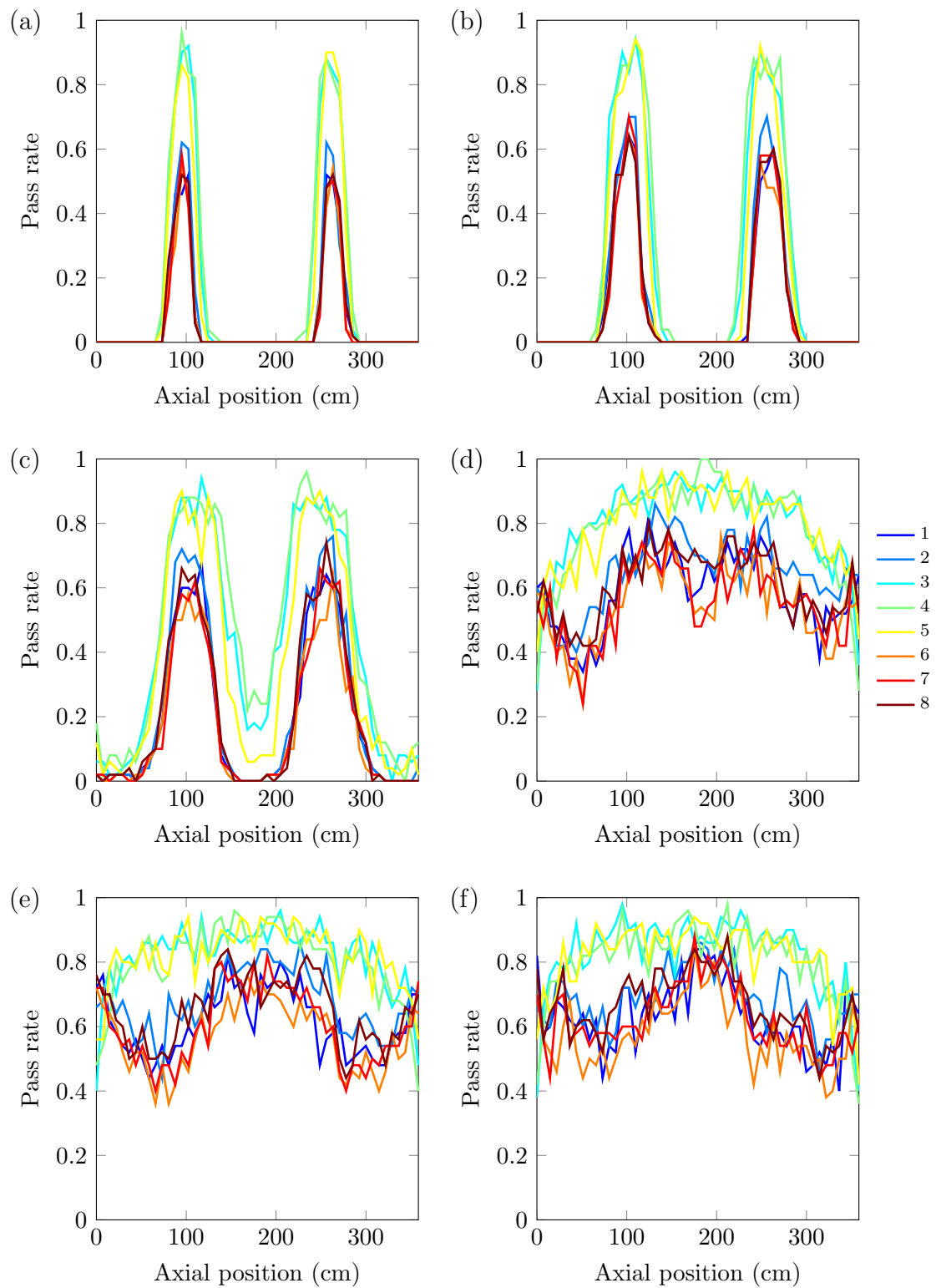


Figure B.3: Pass rates of the D'Agostino–Pearson test for the flux tallies as a function of the axial position of the pin using 250,000 neutron histories per cycle. The numbers of inactive cycles used were (a) 10, (b) 25, (c) 50, (d) 100, (e) 400 and (f) 1,600.

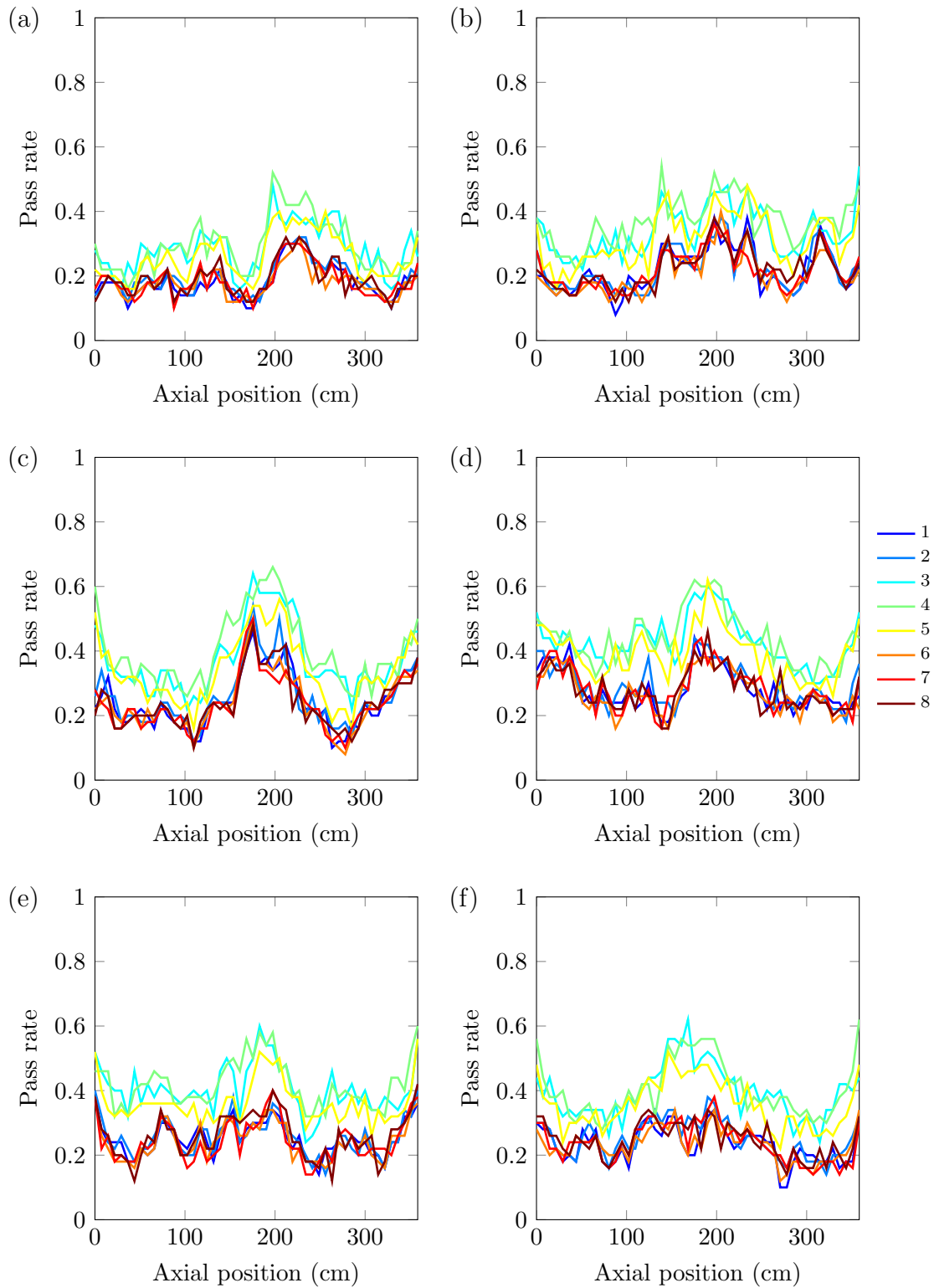


Figure B.4: Pass rates of the drift-in-mean test for the flux tallies as a function of the axial position of the pin using 50,000 neutron histories per cycle. The numbers of inactive cycles used were (a) 10, (b) 25, (c) 50, (d) 100, (e) 400 and (f) 1,600.

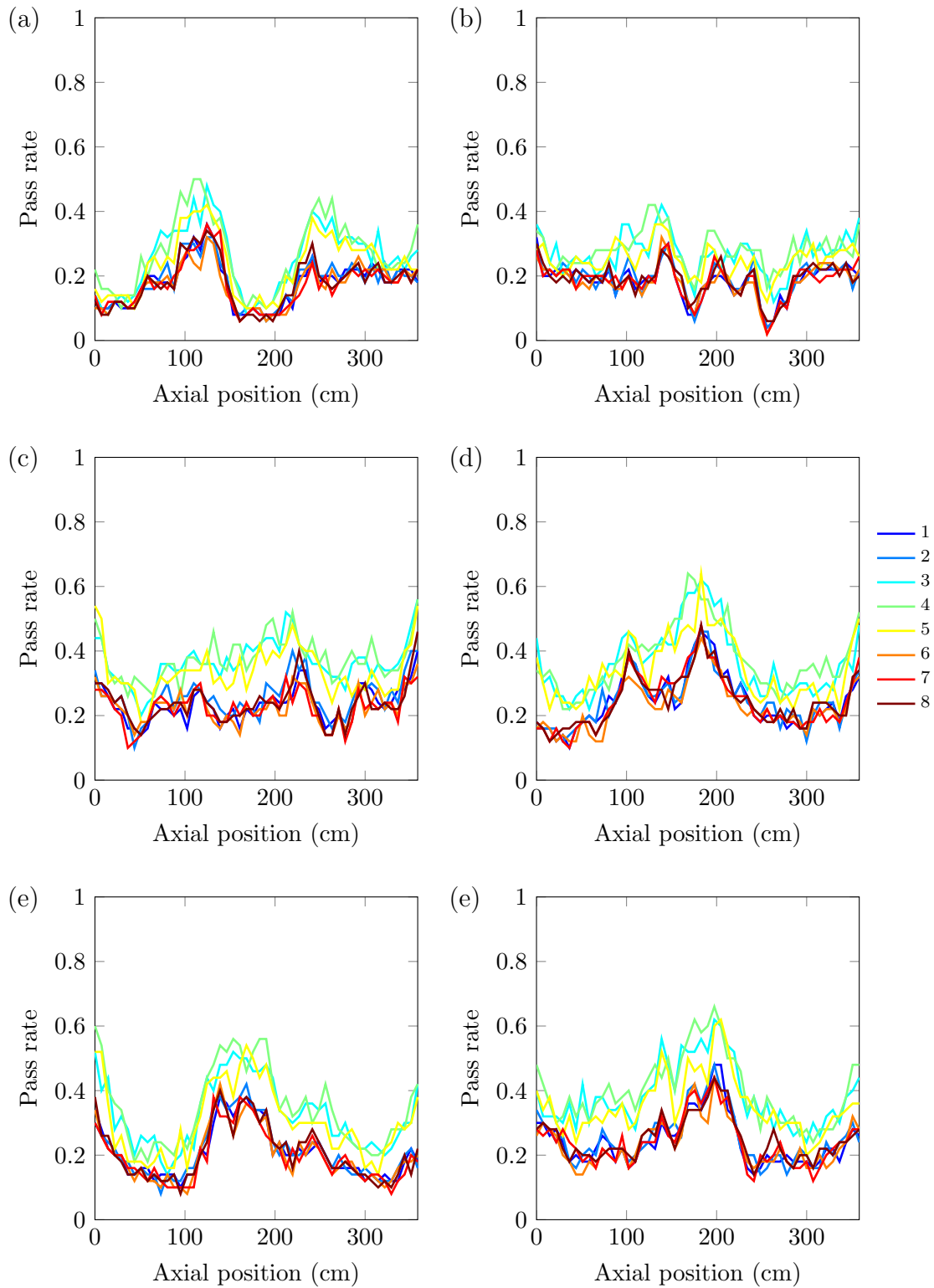


Figure B.5: Pass rates of the drift-in-mean test for the flux tallies as a function of the axial position of the pin using 100,000 neutron histories per cycle. The numbers of inactive cycles used were (a) 10, (b) 25, (c) 50, (d) 100, (e) 400 and (f) 1,600.

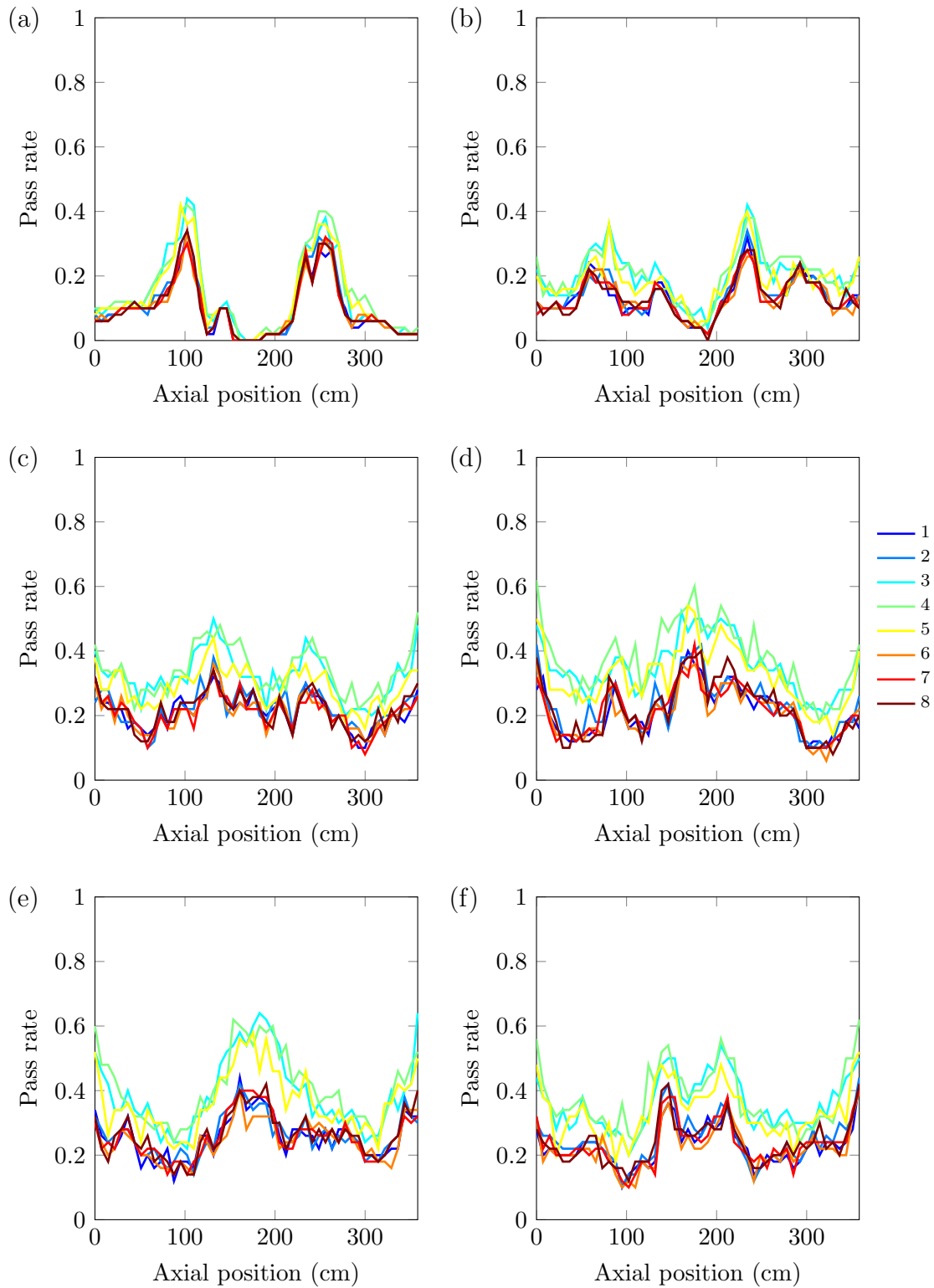


Figure B.6: Pass rates of the drift-in-mean test for the flux tallies as a function of the axial position of the pin using 250,000 neutron histories per cycle. The numbers of inactive cycles used were (a) 10, (b) 25, (c) 50, (d) 100, (e) 400 and (f) 1,600.

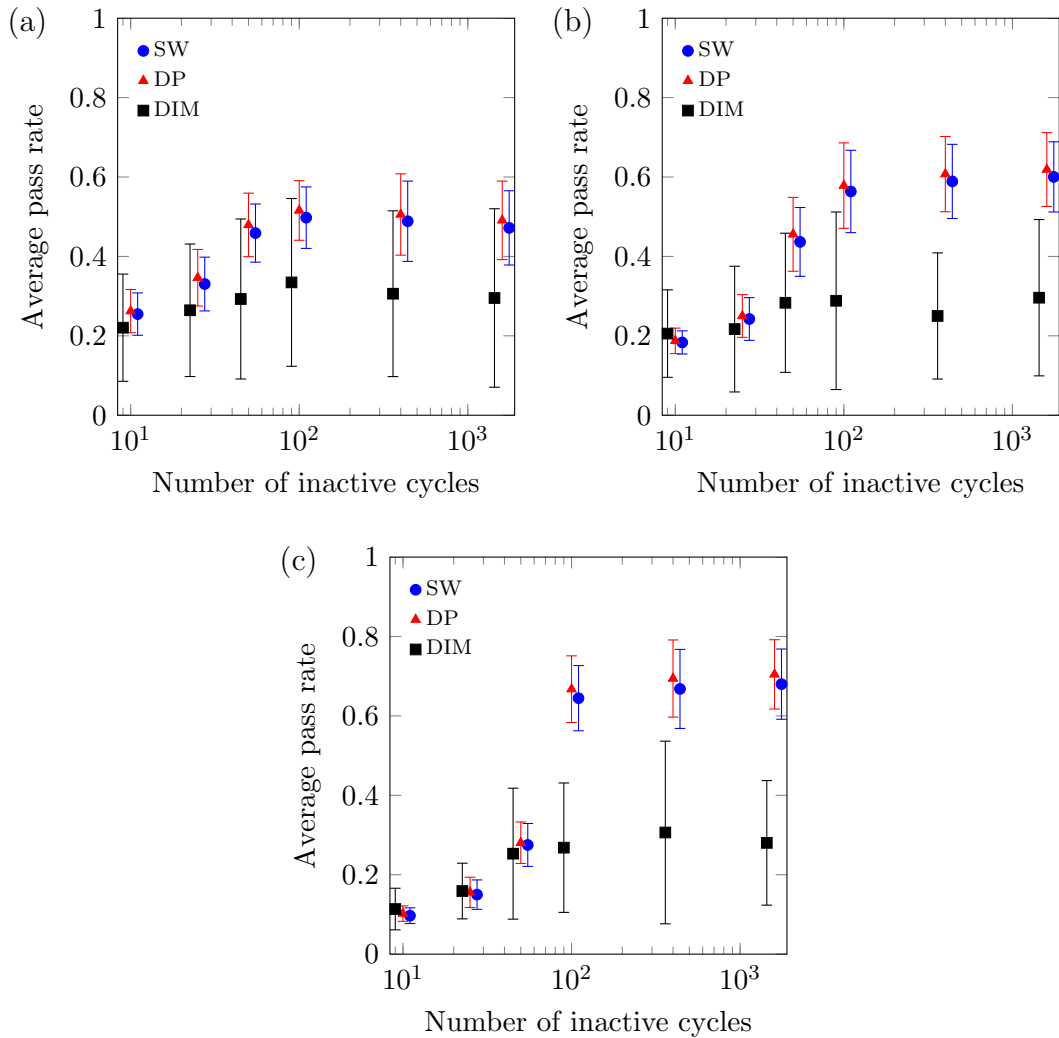


Figure B.7: Average pass rates for the flux tallies using (a) 50,000, (b) 100,000 and (c) 250,000 neutron histories per cycle. The points are adjusted on the horizontal axis in order to clarify the error bars.

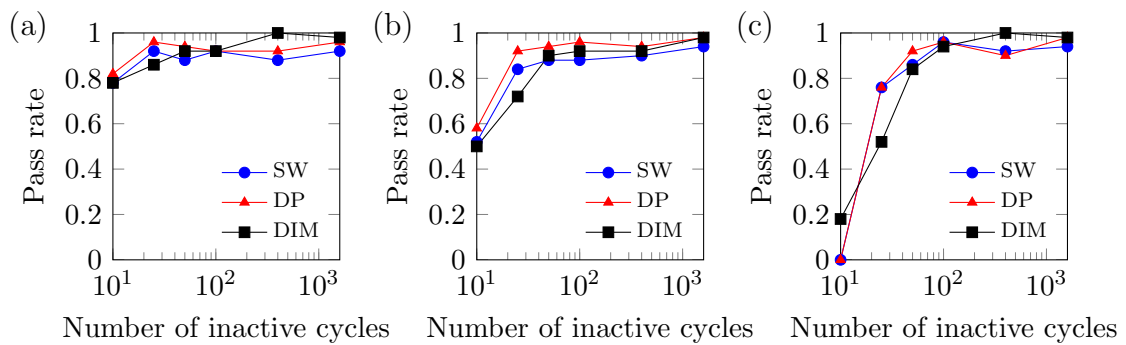


Figure B.8: Pass rates of the statistical tests performed on the implicit estimate of k_{eff} using (a) 50,000, (b) 100,000 and (c) 250,000 neutron histories per cycle.

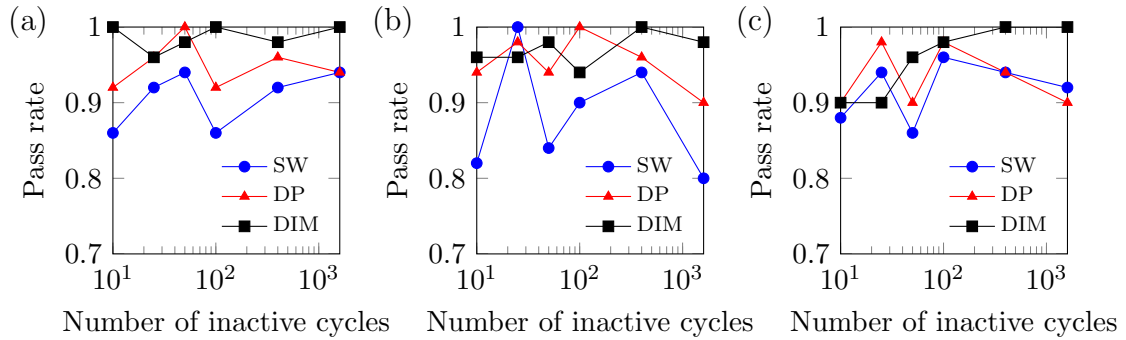


Figure B.9: Pass rates of the statistical tests performed on the collision estimate of k_{eff} using (a) 50,000, (b) 100,000 and (c) 250,000 neutron histories per cycle.

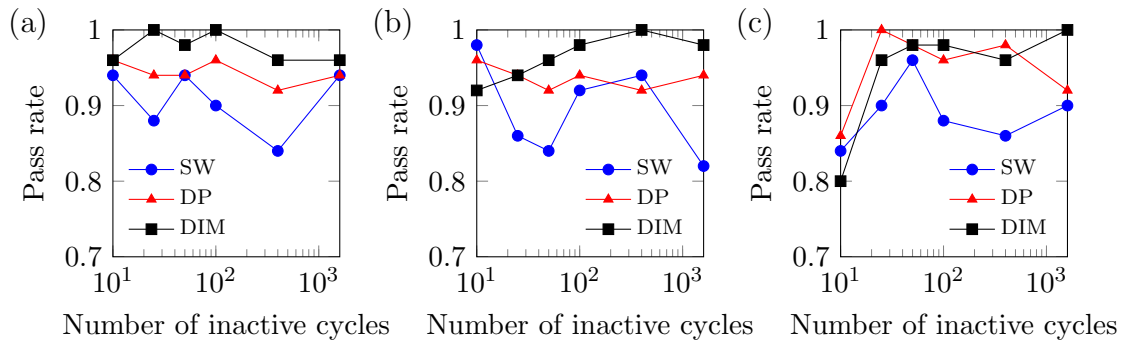


Figure B.10: Pass rates of the statistical tests performed on the analog estimate of k_{eff} using (a) 50,000, (b) 100,000 and (c) 250,000 neutron histories per cycle.

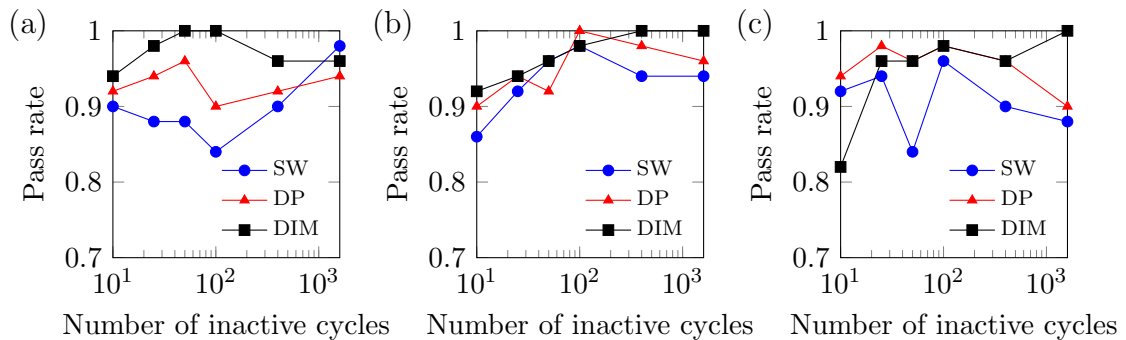


Figure B.11: Pass rates of the statistical tests performed on the analog estimate of prompt neutron k_{eff} using (a) 50,000, (b) 100,000 and (c) 250,000 neutron histories per cycle.

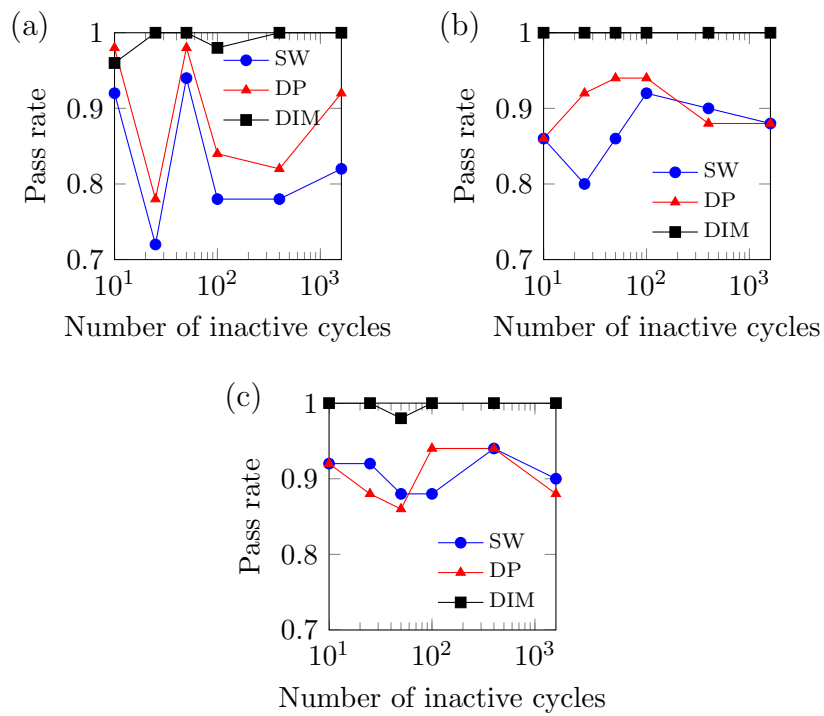


Figure B.12: Pass rates of the statistical tests performed on the analog estimate of delayed neutron k_{eff} using (a) 50,000, (b) 100,000 and (c) 250,000 neutron histories per cycle.

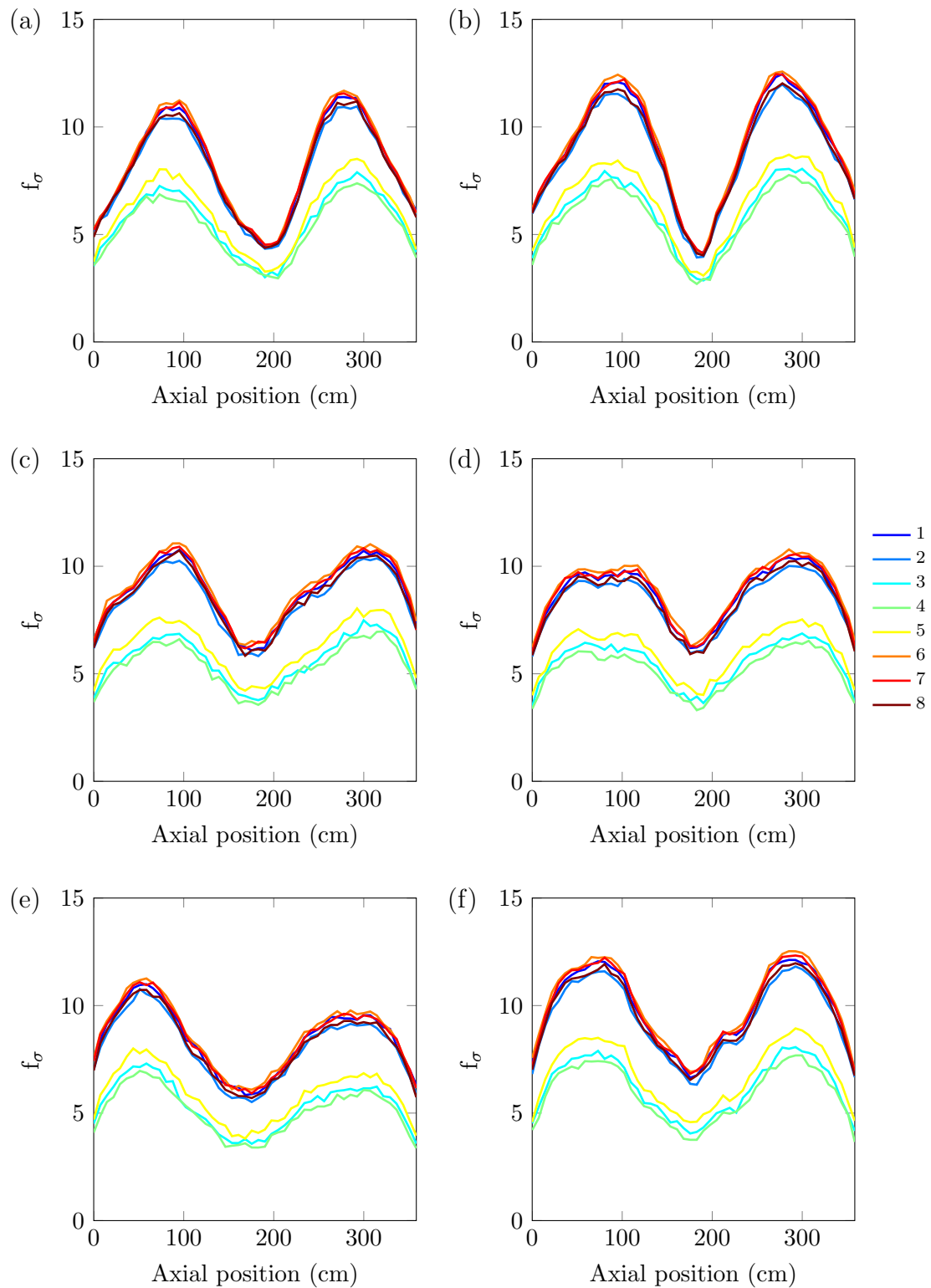


Figure B.13: The ratio of the real and the apparent standard deviation of the total flux as a function of the axial position of the pin for the eight energy groups using 50,000 neutron histories per cycle. The numbers of inactive cycles used were (a) 10, (b) 25, (c) 50, (d) 100, (e) 400 and (f) 1,600.

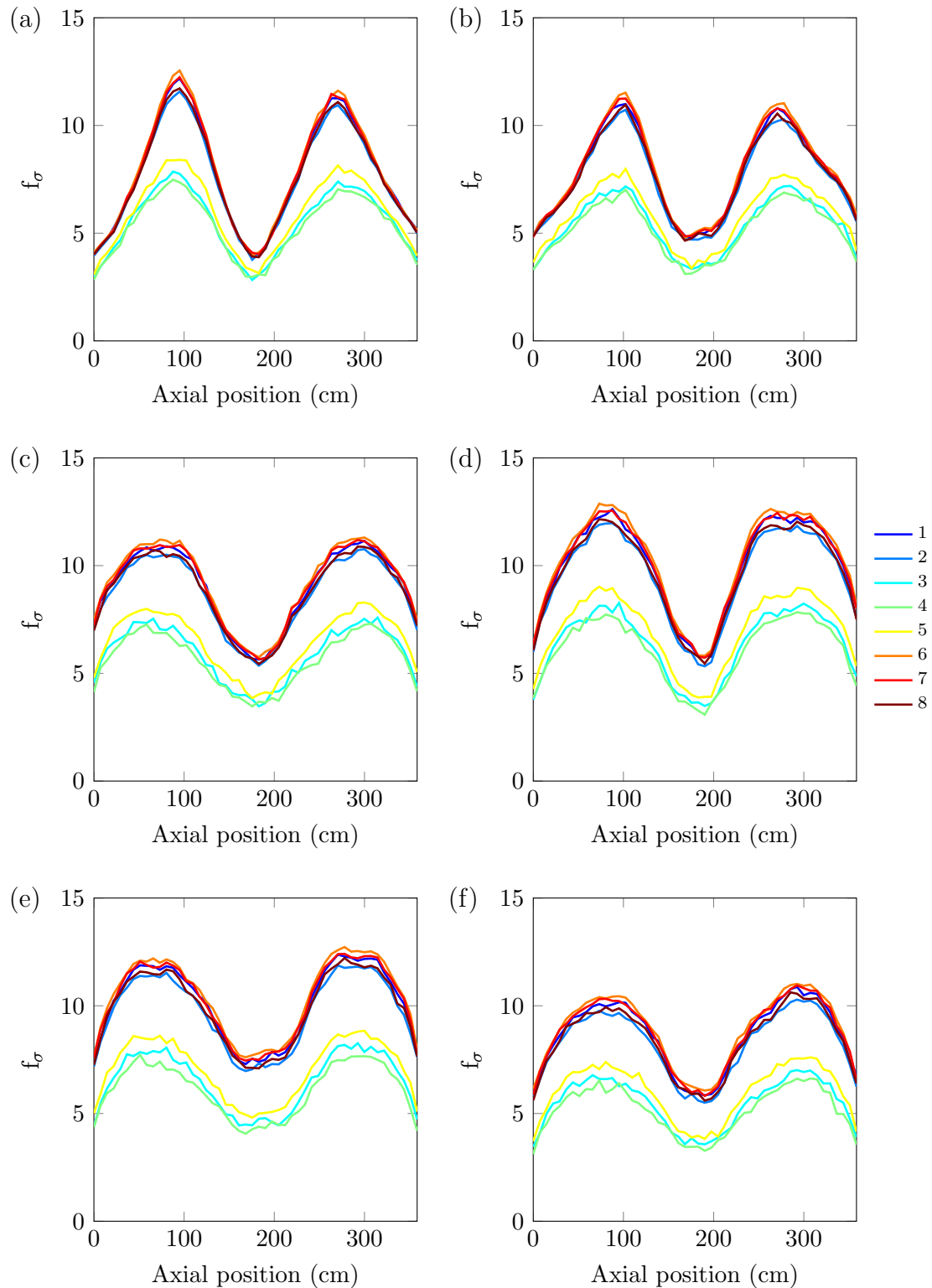


Figure B.14: The ratio of the real and the apparent standard deviation of the total flux as a function of the axial position of the pin for the eight energy groups using 100,000 neutron histories per cycle. The numbers of inactive cycles used were (a) 10, (b) 25, (c) 50, (d) 100, (e) 400 and (f) 1,600.

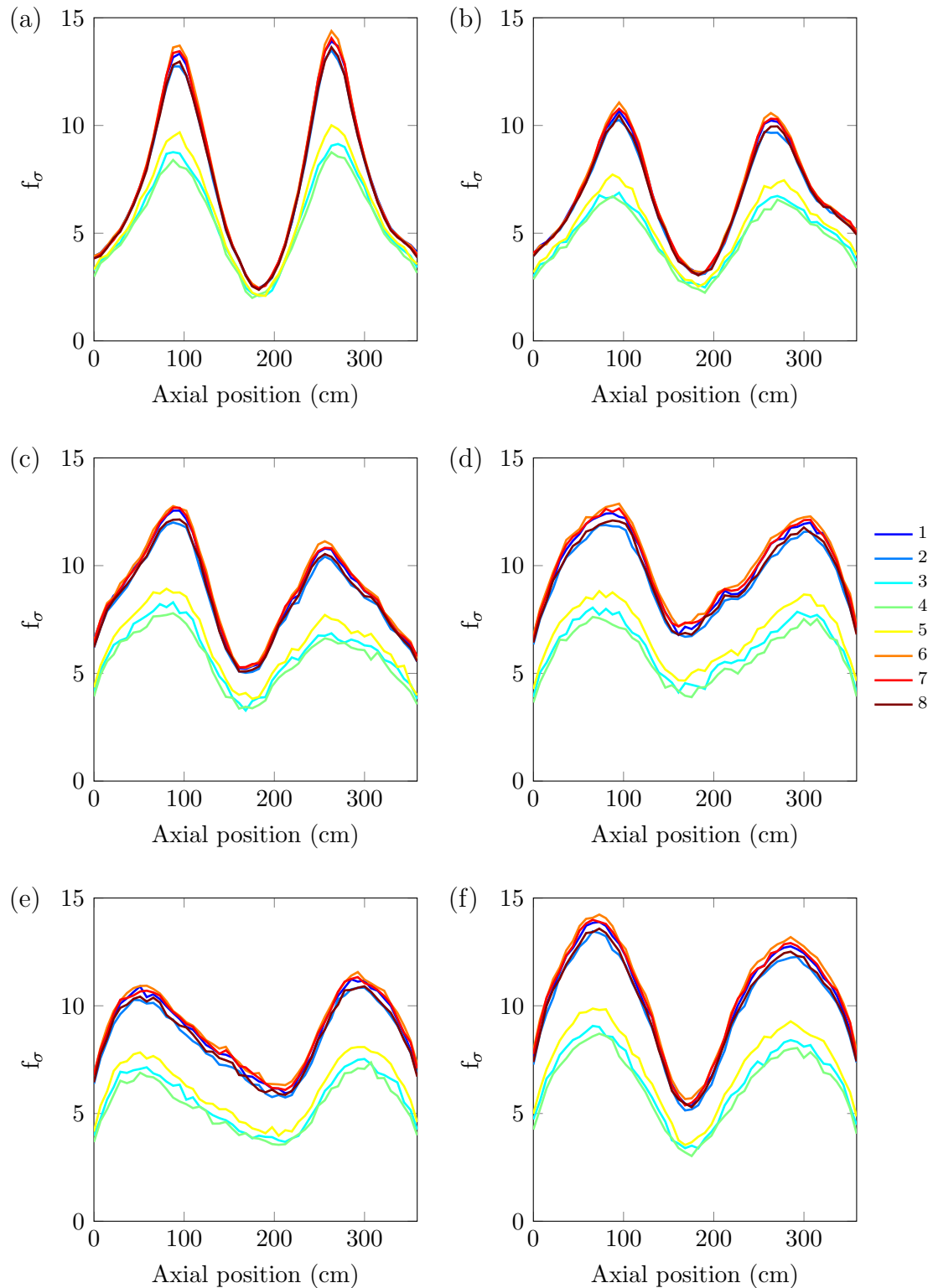


Figure B.15: The ratio of the real and the apparent standard deviation of the total flux as a function of the axial position of the pin for the eight energy groups using 250,000 neutron histories per cycle. The numbers of inactive cycles used were (a) 10, (b) 25, (c) 50, (d) 100, (e) 400 and (f) 1,600.

C Heterogeneous fuel pin

Table C.1: Atom densities of the fuel types used in the heterogeneous fuel pin case [35].

Isotope	B21G (21.57 GWD/MTU)	B24G (24.023 GWD/MTU)	B30G (30.580 GWD/MTU)	B40G (40.424 GWD/MTU)	B55G (54.605 GWD/MTU)
²³⁴ U	6.4862×10^{-6}	6.2881×10^{-6}	5.7868×10^{-6}	5.1152×10^{-6}	4.3280×10^{-6}
²³⁵ U	5.6757×10^{-4}	5.2562×10^{-4}	4.2455×10^{-4}	3.0002×10^{-4}	1.7108×10^{-4}
²³⁶ U	9.4556×10^{-5}	1.0155×10^{-4}	1.1763×10^{-4}	1.3534×10^{-4}	1.4897×10^{-4}
²³⁸ U	2.1732×10^{-2}	2.1693×10^{-2}	2.1585×10^{-2}	2.1413×10^{-2}	2.1142×10^{-2}
²³⁸ Pu	1.1322×10^{-6}	1.4876×10^{-6}	2.7262×10^{-6}	5.3656×10^{-6}	1.0471×10^{-5}
²³⁹ Pu	1.2591×10^{-4}	1.3161×10^{-4}	1.4249×10^{-4}	1.5064×10^{-4}	1.5300×10^{-4}
²⁴⁰ Pu	2.6395×10^{-5}	2.9965×10^{-5}	3.9140×10^{-5}	5.1484×10^{-5}	6.5540×10^{-5}
²⁴¹ Pu	1.3614×10^{-5}	1.5953×10^{-5}	2.1863×10^{-5}	2.9175×10^{-5}	3.6032×10^{-5}
²⁴² Pu	2.2623×10^{-6}	3.0335×10^{-6}	5.6472×10^{-6}	1.0893×10^{-5}	2.0303×10^{-5}
²⁴¹ Am	4.0266×10^{-6}	4.7603×10^{-6}	6.6579×10^{-6}	9.0866×10^{-6}	1.1405×10^{-5}
²⁴³ Am	2.6222×10^{-7}	3.9809×10^{-7}	9.7353×10^{-7}	2.5108×10^{-6}	6.1550×10^{-6}
²³⁷ Np	6.8089×10^{-6}	7.9041×10^{-6}	1.0948×10^{-5}	1.5602×10^{-5}	2.1806×10^{-5}
⁹⁵ Mo	3.1530×10^{-5}	3.4781×10^{-5}	4.3130×10^{-5}	5.4806×10^{-5}	6.9892×10^{-5}
⁹⁹ Tc	3.0319×10^{-5}	3.3465×10^{-5}	4.1531×10^{-5}	5.2746×10^{-5}	6.6975×10^{-5}
¹⁰¹ Ru	2.7452×10^{-5}	3.0548×10^{-5}	3.8751×10^{-5}	5.0881×10^{-5}	6.7875×10^{-5}
¹⁰³ Rh	1.7260×10^{-5}	1.9008×10^{-5}	2.3371×10^{-5}	2.9057×10^{-5}	3.5406×10^{-5}
¹⁰⁹ Ag	1.8879×10^{-6}	2.2230×10^{-6}	3.1774×10^{-6}	4.7134×10^{-6}	6.9836×10^{-6}
¹³³ Cs	3.3140×10^{-5}	3.6549×10^{-5}	4.5257×10^{-5}	5.7263×10^{-5}	7.2263×10^{-5}
¹⁴⁷ Sm	6.4488×10^{-6}	6.9339×10^{-6}	8.0315×10^{-6}	9.2075×10^{-6}	1.0106×10^{-5}
¹⁴⁹ Sm	1.9390×10^{-7}	1.9611×10^{-7}	1.9935×10^{-7}	1.9993×10^{-7}	1.9259×10^{-7}
¹⁵⁰ Sm	7.1174×10^{-6}	8.0472×10^{-6}	1.0496×10^{-5}	1.3997×10^{-5}	1.8627×10^{-5}
¹⁵¹ Sm	5.7531×10^{-7}	6.0064×10^{-7}	6.6160×10^{-7}	7.3765×10^{-7}	8.1779×10^{-7}
¹⁵² Sm	3.0178×10^{-6}	3.3473×10^{-6}	4.1844×10^{-6}	5.3316×10^{-6}	6.7637×10^{-6}
¹⁴³ Nd	2.5793×10^{-5}	2.8067×10^{-5}	3.3502×10^{-5}	3.9984×10^{-5}	4.5992×10^{-5}
¹⁴⁵ Nd	1.8633×10^{-5}	2.0504×10^{-5}	2.5263×10^{-5}	3.1783×10^{-5}	3.9892×10^{-5}
¹⁵³ Eu	2.0026×10^{-6}	2.3329×10^{-6}	3.2546×10^{-6}	4.6705×10^{-6}	6.5858×10^{-6}
¹⁵⁵ Gd	1.1986×10^{-7}	1.4386×10^{-7}	2.2155×10^{-7}	3.6871×10^{-7}	6.0862×10^{-7}
O (natural)	4.6219×10^{-2}	4.6219×10^{-2}	4.6219×10^{-2}	4.6219×10^{-2}	4.6219×10^{-2}

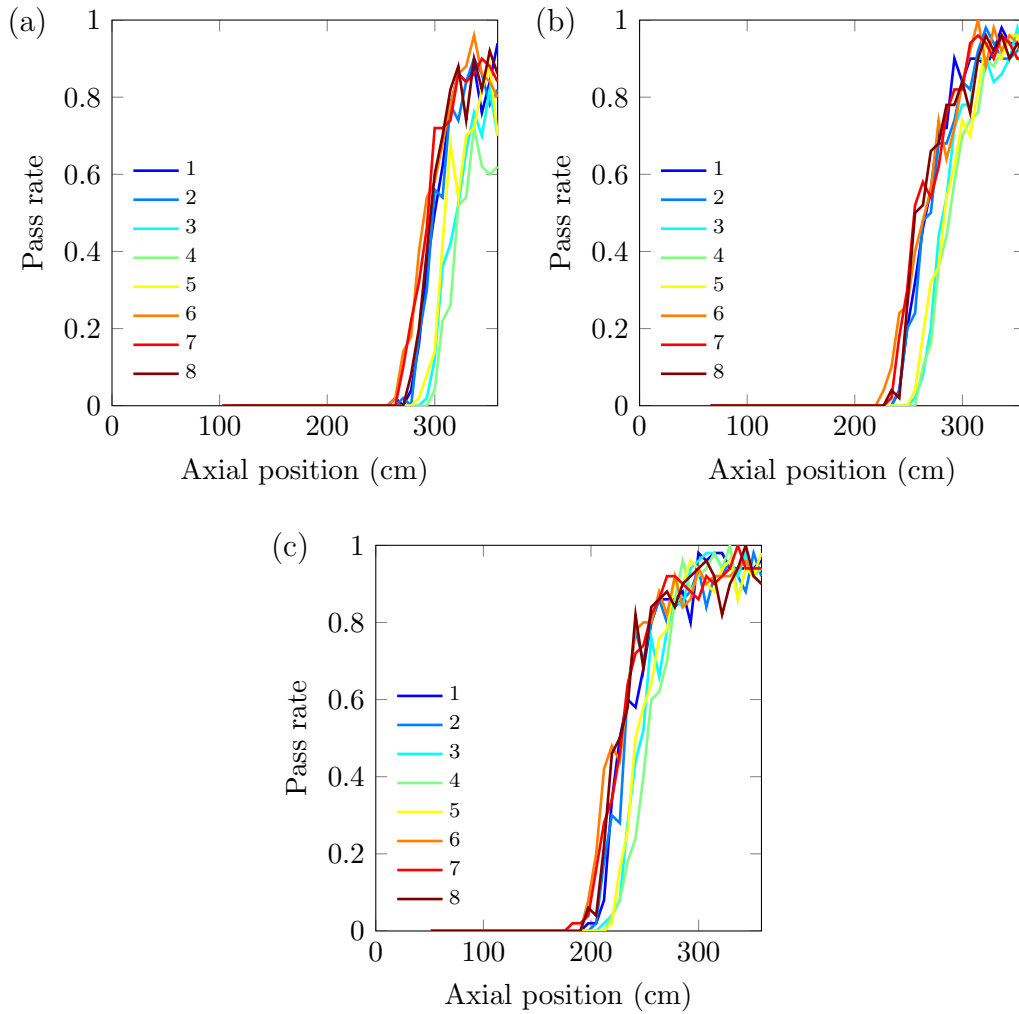


Figure C.1: Pass rates of the D'Agostino–Pearson test for the flux tallies as a function of the axial position of the pin using (a) 10,000, (b) 50,000 and (c) 250,000 neutron histories per cycle. Some simulation runs didn't produce any statistics at the bottom of the fuel pin and therefore these locations have been left blank.

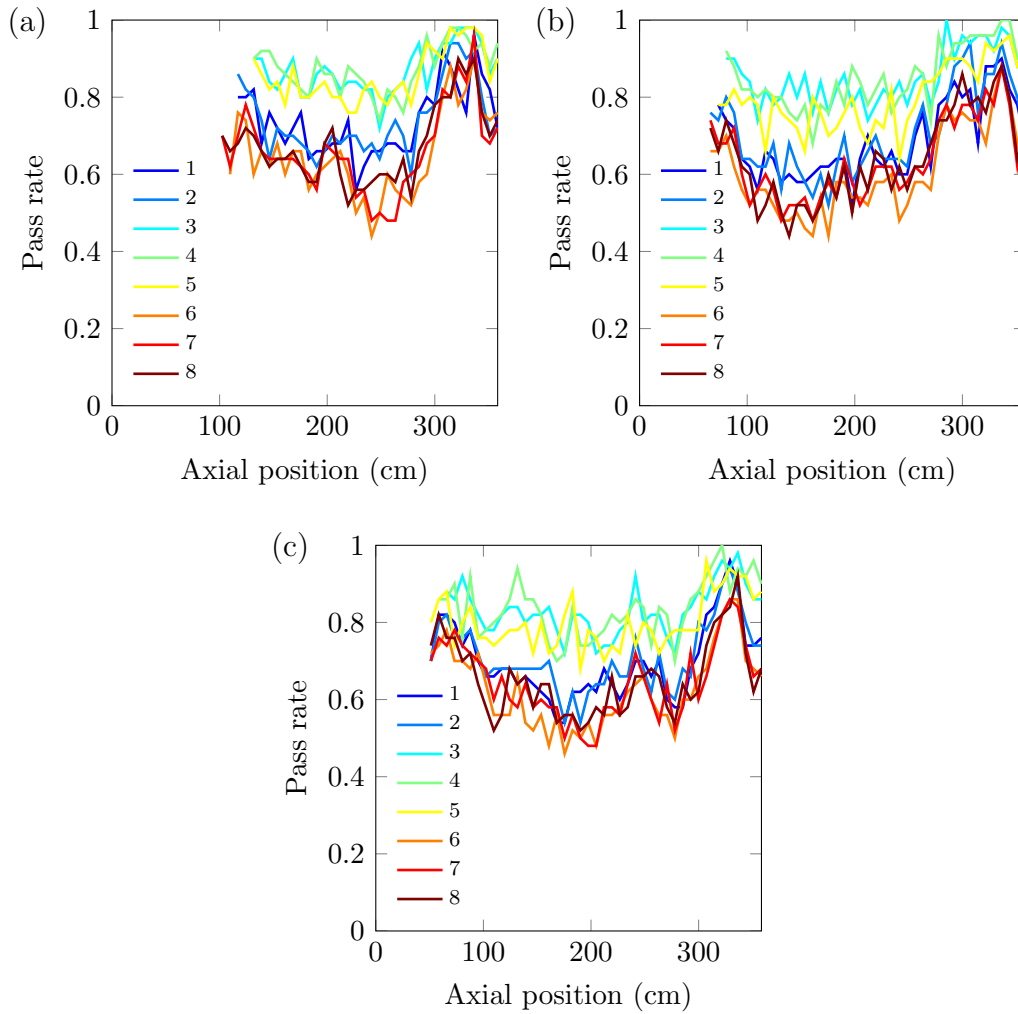


Figure C.2: Pass rates of the drift-in-mean test for the flux tallies as a function of the axial position of the pin using (a) 10,000, (b) 50,000 and (c) 250,000 neutron histories per cycle. Some simulation runs didn't produce any statistics at the bottom of the fuel pin, and therefore these locations have been left blank.

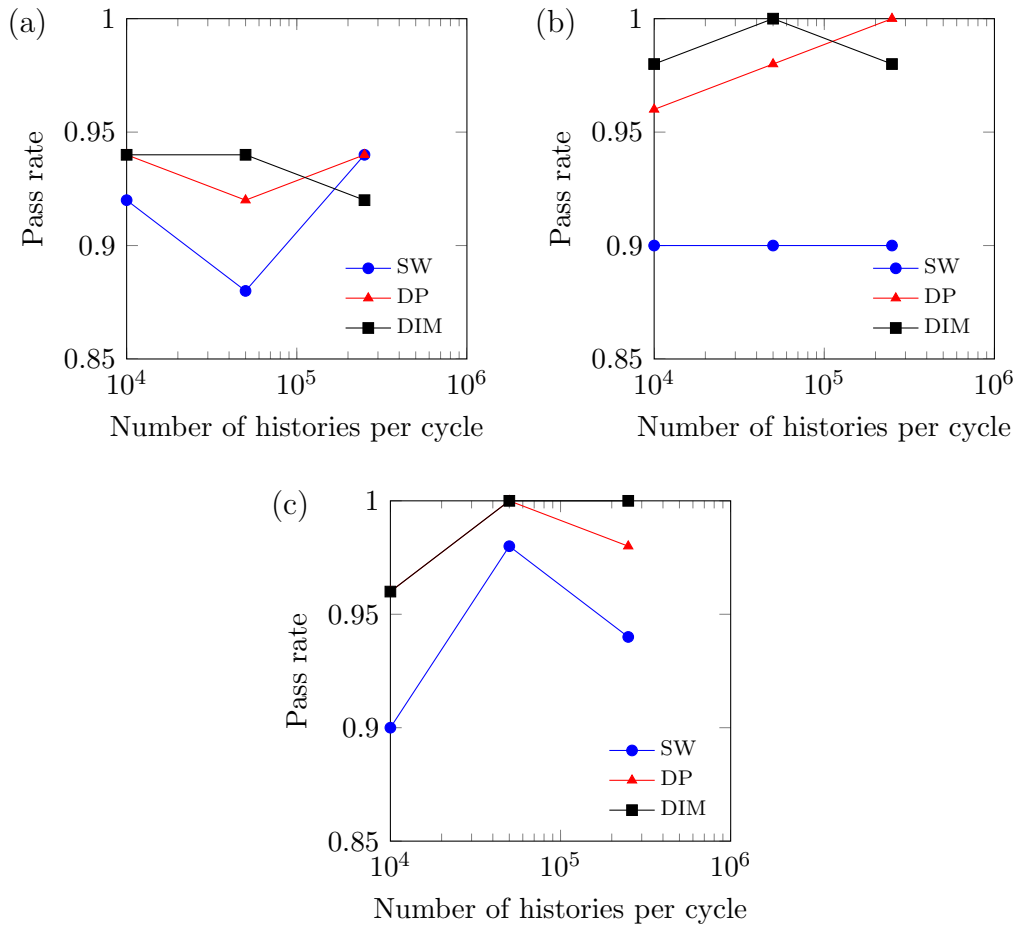


Figure C.3: Pass rates of the statistical tests for the (a) implicit, (b) collision and (c) analog estimate of k_{eff} .

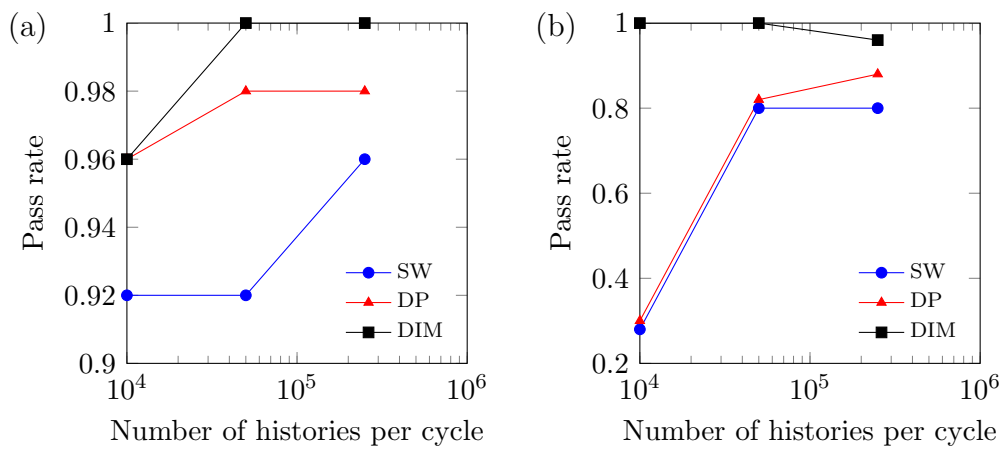


Figure C.4: Pass rates of the statistical tests for the analog estimates of (a) prompt neutron k_{eff} and (b) delayed neutron k_{eff} .