Bayesian post-processing of Monte Carlo simulation in reliability analysis

Wolfgang Betz*, Iason Papaioannou, Daniel Straub

Engineering Risk Analysis Group, Technische Universität München, Germany

Abstract

In reliability analysis with Monte Carlo simulation, the uncertainty about the probability of failure can be formally quantified through Bayesian statistics. Credible intervals for the probability of failure can be derived analytically. This paper gives a detailed overview of Bayesian post-processing for Monte Carlo simulation. We investigate the influence of different weakly-informative prior assumptions on the resulting credible intervals. On this basis, we recommend to use a prior distribution on the probability of failure that follows from the principle of maximum information entropy. We also show that even if no failure sample occurs in a Monte Carlo simulation, Bayesian post-processing still allows to deduce useful information about the probability of failure. The presented Bayesian post-processing strategy can also be applied if Monte Carlo simulation is used for reliability updating; i.e., to evaluate the probability of failure conditional on data or observations. We derive expectations for credible intervals for this case.

Keywords: Monte Carlo simulation, Bayesian post-processing, credible intervals, reliability analysis, rare events, reliability updating

Nomenclature

CDF	cumulative distribution function
MCS	Monte Carlo simulation
PDF	probability density function
$\mathrm{E}[\cdot]$	expectation
$\mathcal{D}$	observation event
$f_{\mathbf{X}}(\mathbf{x})$	joint PDF of $\mathbf{X}$
$f_{P_f}(p)$	prior PDF for P_f
$f_{P_f\mid M,n}(p\mid M=m,n)$	posterior PDF for P_f
$\mathcal{F}$	failure event
$g(\mathbf{x})$	limit-state function
$I(\cdot)$	indicator function
M	number of observed failures in a MCS
m'	parameter of $f_{P_f}(p)$
n	number of samples in MCS
n'	parameter of $f_{P_f}(p)$
$n_{\mathbf{X}}$	dimension of $\mathbf{X}$
$p_f = \Pr[\mathcal{F}]$	probability of failure
$\hat{p}_{f,\mathrm{Bayes}}$	expectation of the posterior P_f
$\hat{p}_f$	point estimate of p_f
$\hat{p}_{f,\mathrm{MLE}}$	point estimate of MCS (see Eq. (5))
$p_{f\mid\mathcal{D}} = \Pr[\mathcal{F}\mid\mathcal{D}]$	probability of failure conditional on $\mathcal{D}$
P_f	random variable modeling the knowledge ab
$\mathrm{Var}[\cdot]$	variance
$\mathbf{x} \in \Omega$	outcome of $\mathbf{X}$
$\mathbf{X} = [X_1, X_2, \ldots, X_{n_{\mathbf{X}}}]$	vector of uncertain input quantities
γ_{c}	coverage probability of a γ confidence inter
δ	coefficient of variation
δ_{MCS}	coefficient of variation of $\hat{p}_{f,\mathrm{MLE}}$
$\Omega \subseteq \mathbb{R}^{n_{\mathbf{X}}}$	outcome space of $\mathbf{X}$
$\Omega_f \subseteq \Omega$	failure domain

*Corresponding author

Email addresses: wolfgang.betz@tum.de (Wolfgang Betz), iason.papaioannou@tum.de (Iason Papaioannou), straub@tum.de (Daniel Straub)

Manuscript, published in Reliability Engineering & System Safety, 227:108731, 2022. *DOI:* 10.1016/j.ress.2022.108731

1. Introduction

The aim of a *reliability analysis* is to evaluate the probability of failure of a system of interest. In this context, failure is defined as the system being in an undesired state, e.g.: admissible stresses are exceeded, the stability of a structure is no longer maintained, or water levels in a river that exceed a certain threshold will result in a flood event. Let such an undesired system response be denoted as event $\mathcal{F}$; the probability of failure p_f is defined as the probability that $\mathcal{F}$ occurs, i.e., $p_f = \Pr[\mathcal{F}]$.

Let $\mathbf{X} = [X_1, X_2, \ldots, X_{n_\mathbf{X}}]$ be the vector of $n_\mathbf{X}$ uncertain input quantities of the system of interest, and let $\mathbf{x} \in \Omega$ be a particular outcome of $\mathbf{X}$, where $\Omega \subseteq \mathbb{R}^{n_\mathbf{X}}$ denotes the outcome space of $\mathbf{X}$. The joint probability density function of $\mathbf{X}$ is denoted by $f_\mathbf{X}(\mathbf{x})$. Moreover, let $g(\mathbf{x})$ be a function such that $g(\mathbf{x}) \leq 0$ if and only if the system is in an undesired state for parameter values $\mathbf{x}$, and $g(\mathbf{x}) > 0$ otherwise. The function $g(\mathbf{x})$ is known as *limit-state function* or *performance function* in the literature [1]. The so-called failure domain $\Omega_f \subseteq \Omega$ is the subset of the outcome space Ω for which the condition $g(\mathbf{x}) \leq 0$ holds. The probability of failure $p_f \in [0, 1]$ can then be expressed as:

$$p_f = \Pr[\mathcal{F}] = \Pr[g(\mathbf{X}) \leq 0] = \int_{\Omega_f} f_\mathbf{X}(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{1}$$

In most cases, the integral in Eq. (1) cannot be solved analytically, because the boundary of the domain Ω_f is not known in an analytical manner. Instead, Eq. (1) is usually solved numerically, where the limit-state function is evaluated in a pointwise manner. The probabilities that have to be evaluated in a reliability analysis are typically rather small; i.e., $p_f \ll 10^{-2}$. That is, the event $\mathcal{F}$ is a *rare event*. This renders the numerical treatment of the integral in Eq. (1) difficult, because the failure domain Ω_f constitutes only a small subset of the outcome space Ω. The class of numerical methods specifically designed to solve Eq. (1) are referred to as *reliability methods*. *Monte Carlo simulation* (MCS) is a simple and straightforward method to solve the integral in Eq. (1). Other well known sampling-based reliability methods are *Subset Simulation* (SuS) [2], and *importance sampling methods* including *line sampling* [3, 4, 5, 6], *directional importance sampling* [7, 8], *sequential importance sampling* [9] and *cross-entropy-based importance sampling* [10, 11, 12].

For limit-state functions based on computationally demanding models, MCS is often not practical [13], as it requires a comparatively large number of limit-state function calls – in particular if p_f is small. However, compared to other reliability methods, the sampling process in MCS is not tailored to a specific limit-state function. Instead, an arbitrary number of limit-state functions that depend on the same input parameters can be assessed within a single MCS. Moreover, for each investigated limit-state function $g(\mathbf{X})$, the full distribution of $g(\mathbf{X})$ can be estimated additionally to $\Pr[g(\mathbf{X}) \leq 0]$ using the generated samples. Therefore, MCS is often the method of choice when a reasonably large number of limit-sate function evaluations is possible.

For sampling-based reliability methods, the point estimate $\hat{p}_f$ of the true underlying probability of failure p_f is a random variable; i.e., every new run of the method can result in a different value of $\hat{p}_f$. The larger the number of samples in a sampling-based approach, the smaller the coefficient of variation of $\hat{p}_f$. In practice, a sampling-based reliability method is typically executed only once for a chosen number of samples. To assess the quality of the estimated $\hat{p}_f$, it is helpful to obtain a measure for the uncertainty associated with $\hat{p}_f$ as a side-product of the simulation. This can be achieved within the framework of Bayesian inference [14, 15]. Thereby, a full probabilistic description of p_f can be obtained conditional on a single simulation run. The Bayesian approach is often applied to obtain credible intervals for p_f from a MCS [16, 17, 18]. Additionally, a Bayesian post-processing of Subset Simulation was proposed in [16], based on assuming a beta distribution for the conditional probability of failure and by neglecting the dependence of the conditional probabilities within Subset Simulation. However, the performance of credible intervals obtained by applying [16] can strongly depend on the shape of the employed limit-state function [19].

A strong advantage of rare event analysis with MCS is that the uncertainty on $\hat{p}_f$ can be quantified analytically and does not depend on the particular shape of the limit-state function. Among all sampling-based reliability methods, MCS is the only method for which a Bayesian post-processing with this property can be derived. In combination with a conjugate prior distribution, it is straightforward to derive the associated posterior distribution [16, 18, 20, 21]. The resulting Bayesian problem is closely related to Bayesian reliability assurance based on test data, which is discussed in, e.g., [17, 18]. In most applications of Bayesian post-processing of MCS, the choice of Jeffrey's prior is advocated due to its invariance under transformation, e.g., [17, 22, 23]. In this paper, we investigate different weakly-informative prior assumptions for the probability of failure obtained with MCS, and recommend to us a prior distribution that follows from the principle of maximum information entropy – and not the more commonly employed Jeffreys' prior. The obtained findings and the conducted investigations provide an addition to the discussion in [23]. Our studies concerning the outcome of a MCS without any failed sample complement the numerical demonstrations performed in [17]. Furthermore, we demonstrate that, with only minor modifications, the presented Bayesian post-processing strategy can also be applied if Monte Carlo simulation is used for reliability updating; i.e., to evaluate the probability of failure conditional on data or observations.

The paper is structured as follows: In Section 2, we briefly discuss the difference between *credible* and *confidence intervals* utilized to quantify the uncertainty on probability of failure estimates. Section 3 discusses a Bayesian approach to interpret the outcome of a MCS in the context of reliability analysis: Section 3.1 gives a brief overview of MCS and Section 3.2 presents the statistical uncertainty associated with a MCS outcome. Section 3.3 discusses a Bayesian approach to MCS and introduces the likelihood, prior and posterior for the problem at hand. Section 3.4 introduces credible intervals for MCS. Section 3.5 assesses the influence of modeling choices for the prior distribution on the posterior distribution and gives advice on what prior distributions to use in practice. Section 3.6 demonstrates what happens if not a single failed sample is observed in a MCS run. In Section 4, we show that it is straightforward to apply the Bayesian approach presented in Section 3 to *reliability updating*. In Section 5, we use an illustration example to demonstrate how the presented approaches can be applied and used in practice. In the annex, we discuss some commonly used confidence intervals for MCS.

2. Credible and confidence intervals

2.1. Introduction

Sampling-based reliability methods return point estimates $\hat{p}_f$ of the underlying probability of failure p_f. These estimates are associated with statistical uncertainty due to a finite number of samples used. For any sampling-based approach, this uncertainty is typically expressed in terms of *credible* or *confidence intervals*. The notion of *credible intervals* is mainly used in a Bayesian context, whereas *confidence intervals* are based on frequentist statistics [24]. In the Bayesian viewpoint, the uncertainty about the probability of failure is expressed by a random variable, denoted P_f. In this context, the results from a sampling-based reliability method are used to updated a prior belief on the distribution of P_f using Bayes' rule. Credible intervals are utilized to express the uncertainty on P_f. The interpretation of confidence intervals differs fundamentally from the interpretation of credible intervals [25]. Confidence intervals are based on the frequentist interpretation probability, according to which the uncertainty on the point estimate of $\hat{p}_f$ is quantified by analyzing the statistics of $\hat{p}_f$ conditional on the true (but unknown) p_f. In the following sections, we briefly introduce the two concepts before discussing their interpretation in the context of reliability analysis.

2.2. Credible intervals

Credible intervals are associated with a subjective interpretation of probability and give an interval that is believed to contain the true p_f with probability γ, based on the conducted simulation run and a prior belief. An interval associated with probability γ is called a γ credible interval. For example, a 95% credible interval states that conditional on the performed simulation run, the probability that the true p_f is within the interval bounds is *believed to be* 95%.

Credible intervals for MCS of rare events are discussed in Section 3.4.

2.3. Confidence intervals

Confidence intervals are associated with a frequentist interpretation of probability and state that if the simulation is repeated infinitely many times and the corresponding confidence interval is evaluated for each simulation run, the true p_f is contained within $\gamma_c \in [0, 1]$ of all obtained intervals. An interval qualifies as a γ confidence interval, if $\gamma_c \geq \gamma$ holds independent of the value of p_f, where γ_c is referred to as *coverage probability* of the associated γ confidence interval. An interval that satisfies this property is referred to as a *proper* confidence interval. For example, for a 95% confidence interval, the true value of p_f is contained in at least 95% of the confidence intervals obtained from repeated simulations.

Commonly used confidence intervals for MCS of rare events are discussed in Appendix A.

2.4. Discussion

To summarize, a credible interval quantifies the belief about p_f conditional on the performed simulation run and the available/infused prior information, whereas confidence intervals quantify the performance of the method for an assumed large number of repeated simulation runs. A pitfall of using confidence intervals in the context of reliability analysis is their nontrivial interpretation, which causes many to confuse them with the more intuitive credible intervals [25].

In a classical Bayesian setting, the observed data that enters the likelihood function is fixed. The experiment is conducted only once, and hence the long-run probability upon which confidence intervals are based is meaningless for the Bayesian interpretation. However, in this regard, the information collected during a simulation-based reliability analyis (e.g., a MCS) is special: The underlying experiment (the MCS) can, in theory, be repeated arbitrarily often; and one would probably expect that the true p_f is contained within at least γ of all obtained γ credible intervals. Thus, to avoid potential fallacies and misinterpretations, the γ credible interval used in a Bayesian post-processing of a MCS should ideally also qualify as a proper γ confidence interval. That is, any γ credible interval associated with a MCS should have a *coverage probability* γ_c at least as large as γ.

2.5. Computation of coverage probabilities

For a given underlying true failure probability p_f and total number of samples n, the coverage probability γ_c of a confidence interval C for MCS can be evaluated explicitly as follows:

$$\gamma_c(n, p_f) = \sum_{m=0}^{n} I_{C|M=m,n}\left(p_f\right) \cdot \mathcal{L}(p_f|m, n), \tag{2}$$

where $\mathcal{L}(p_f|m, n) = p_{\text{bin}}(m|n, p_f)$ is the probability mass function of a binomial random variable M with parameters n and p_f (compare Section 3.3). The quantity $I_{C|M=m,n}(m/n)$ is *one* if the true p_f is contained in the confidence interval of interest for m observed failures in n total samples, and *zero* otherwise.

The binomial distribution employed in Eq. (2) can often be well approximated by means of the Poisson distribution. This means that γ_c can be directly expressed as a function of $n \cdot p_f$ [46]. This relation is utilized for some of the plots shown in this paper.

3. Monte Carlo simulation of rare events

3.1. Overview of the method

Monte Carlo simulation (MCS) estimates the integral in Eq. (1) through generating samples from the input random variables $\mathbf{X}$. Consider an indicator function $I(\mathbf{x})$ that indicates membership of outcomes $\mathbf{x}$ to the failure domain Ω_f. The indicator function is by definition *one* if failure occurs and *zero* otherwise; i.e.:

$$I(\mathbf{x}) = \begin{cases} 1 & \text{if } g(\mathbf{x}) \leq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The integral over the failure region Ω_f in Eq. (1) can then be expressed as an integral over Ω:

$$p_f = \int_{\Omega} I(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x})\, \mathrm{d}\mathbf{x} = \mathrm{E}\left[I(\mathbf{x})\right]. \tag{4}$$

Eq. (4) shows that the probability of failure can be expressed as the expectation $\mathrm{E}\left[I(\mathbf{x})\right]$ of the indicator function. MCS approximates the expectation in Eq. (4) by a weighted sum over n samples $\mathbf{x}^{(i)}$, $i = 1, \ldots, n$, where the samples $\mathbf{x}^{(i)}$ are outcomes of $\mathbf{X}$:

$$p_f \approx \hat{p}_{f,\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n} I\left(\mathbf{x}^{(i)}\right). \tag{5}$$

Essentially, the unbiased estimate $\hat{p}_{f,\mathrm{MLE}}$ is obtained by dividing the number of observed failures $m = \sum_{i=1}^{n} I(\mathbf{x}^{(i)})$ by the total number of employed samples n; i.e. $\hat{p}_{f,\mathrm{MLE}} = m/n$. The MLE in $\hat{p}_{f,\mathrm{MLE}}$ signifies that the point estimate corresponds to the maximum likelihood estimate (MLE).

3.2. *Statistical uncertainty of the estimate* $\hat{p}_{f,\mathrm{MLE}}$

The sequence of samples $\left\{I\left(\mathbf{x}^{(i)}\right)\right\}$, $i = 1, \ldots, n$ can be perceived as the outcome of a Bernoulli process. Each element in this Bernoulli process of length n is an independent and identically distributed (iid) Bernoulli trial and follows a Bernoulli distribution with a mean value equal to p_f. Consequently, the number m of observed failures in a MCS with a total number of n samples is the outcome of a random variable M that follows a binomial distribution with parameter values equal to n and p_f. The mean of M is $n \cdot p_f$ and the variance of M is $n \cdot p_f \cdot \left(1 - p_f\right)$.

The estimate $\hat{p}_{f,\mathrm{MLE}}$ in Eq. (5) is obtained by dividing the outcome m of M by the total number of samples n. Thus, the variance of the estimate $\hat{p}_{f,\mathrm{MLE}}$ for repeated runs of the MCS with a fixed total number of samples n is:

$$\mathrm{Var}\left[\hat{p}_{f,\mathrm{MLE}}\right] = \mathrm{Var}\left[\frac{M}{n}\right] = \frac{p_f \cdot (1 - p_f)}{n}. \tag{6}$$

The coefficient of variation of the estimate $\hat{p}_{f,\mathrm{MLE}}$ can consequently be written as:

$$\delta_{\mathrm{MCS}} = \mathrm{Var}\left[\hat{p}_{f,\mathrm{MLE}}\right]/p_f = \sqrt{\frac{1 - p_f}{p_f \cdot n}}. \tag{7}$$

Even though δ_{MCS} depends on the (unknown) underlying p_f, Eq. (7) highlights the major strength and weakness of MCS: The weakness is that for small p_f, the total number of samples n must be large to achieve a reasonable coefficient of variation of the estimate. The strength of MCS is that δ_{MCS} does not depend on the number of stochastic variables $n_{\mathbf{X}}$, i.e., the dimension of the uncertain parameter vector $\mathbf{X}$. Moreover, MCS can be considered a very robust method, as its performance solely depends on the underlying true probability of failure – and not on the particular limit-state function $g(\mathbf{X})$, nor on the corresponding shape of the failure domain Ω_f.

The uncertainty associated with $\hat{p}_{f,\mathrm{MLE}}$ is often quantified in terms of confidence intervals based on a Normal approximation. In Appendix A.4 we discuss why such an approximate confidence interval should not be used in practice.

3.3. *Using Bayesian statistics to quantify the uncertainty on* p_f

Based on a MCS with m observed failures in n samples, the available information about the probability of failure can be quantified formally through Bayes' theorem:

$$f_{P_f|M,n}(p|M = m, n) = \frac{1}{c_{\mathrm{E}}} \cdot \mathcal{L}(p|m, n) \cdot f_{P_f}(p). \tag{8}$$

where:

$\mathcal{L}(p|m, n)$ is the likelihood function. Here it corresponds to the probability mass function $p_{\mathrm{bin}}(m|n, p)$ of the binomial distribution:

$$\mathcal{L}(p|m, n) = p_{\mathrm{bin}}(m|n, p) = \binom{n}{m} \cdot (p)^m \cdot (1 - p)^{n-m}. \tag{9}$$

$f_{P_f}(p)$ is the prior PDF. For the likelihood function stated in Eq. (9), the beta distribution serves as a conjugate prior [23, 26]. The beta-distributed prior can be expressed in terms of the prior parameters m' and n' as:

$$f_{P_f}(p) = \frac{p^{m'-1} \cdot (1 - p)^{n'-m'-1}}{\mathrm{B}(m', n' - m')}, \tag{10}$$

where $\mathrm{B}(\cdot,\cdot)$ denotes the beta function defined as $\mathrm{B}(\alpha,\beta) = \int_0^1 t^{\alpha-1}\,(1 - t)^{\beta-1}\,\mathrm{d}t$ with $\alpha, \beta > 0$.

c_{E} is a normalizing constant which ensures that $f_{P_f|M,n}(p|M = m, n)$ integrates to *one*.

$f_{P_f|M,n}(p|M = m, n)$ denotes the posterior probability density function (PDF) for P_f. For the problem at hand, it is a beta distribution with parameters $m + m'$ and $n - m + n' - m'$ [16, 18, 20, 21]:

$$f_{P_f|M,n}(p|M = m, n) = \frac{p^{m+m'-1} \cdot (1 - p)^{n-m+n'-m'-1}}{\mathrm{B}(m + m', n - m + n' - m')}. \tag{11}$$

The posterior mean of P_f is:

$$\mathrm{E}\left[P_f|M = m, n\right] = \frac{m + m'}{n + n'}. \tag{12}$$

Eq. (12) can be used as a point estimate of p_f. However, unlike the MLE estimate given in Eq. (5), the point estimate of Eq. (12) is biased (depending on the choice of m' and n'). The posterior variance is:

$$\mathrm{Var}\left[P_f|M = m, n\right] = \frac{(m + m') \cdot (n + n' - m - m')}{(n + n')^2 \cdot (n + n' + 1)}. \tag{13}$$

The coefficient of variation $\delta_{P_f|M=m,n}$ is:

$$\delta_{P_f|M=m,n} = \sqrt{\frac{n + n' - m - m'}{(m + m')\,(n + n' + 1)}}. \tag{14}$$

If n is large (i.e., $n \gg 10^3$), and $p_f \ll 1$, Eq. (14) can be approximated by $\delta_{P_f|M=m,n} \approx \sqrt{1/(m+m')}$.

The distribution of $P_f|M = m, n$ can be highly skewed, even for large n, especially if m is small compared to n [21]. Therefore, the interpretation of the coefficient of variation can be difficult and might not be very meaningful. Credible intervals (see Section 3.4) are typically a more robust measure to quantify the uncertainty on P_f.

3.4. *Credible intervals for Monte Carlo simulation of rare events*

3.4.1. *Introduction*

Instead of point estimates (e.g., $\hat{p}_{f,\mathrm{MLE}}$ in Eq. (5), $\mathrm{E}\left[P_f|M = m, n\right]$ in Eq. (12), or Λ in Eq. (**??**)), it is often more meaningful to express the information about the value of p_f conditional on the performed MCS in terms of credible intervals. Based on the distribution of $P_f|M = m, n$ given in Eq. (11), credible intervals can be obtained [18]:

$$\Pr\left[l \leq P_f \leq u | m, n, m', n'\right] = \gamma, \quad (15)$$

which states that with a probability of γ, the interval $[l, u] \subseteq [0, 1]$ contains the true p_f, if m failures in n samples were observed in a MCS run, and if the parameters of the prior were selected as m' and n'.

3.4.2. *Commonly used credible intervals*

For a given probability γ, there is an infinite number of γ credible intervals $[l, u]$ in $[0, 1]$ such that $\int_{[l,u]} f_{P_f|M,n}(p|M = m, n)\, \mathrm{d}\, p = \gamma$. Commonly used credible intervals are:

$C_{\gamma,\mathrm{up}}$ The γ *upper credible interval* gives the threshold u_{up}, for which there is a γ probability that the value of P_f is smaller than u_{up}; i.e., $\Pr\left[P_f \leq u_{\mathrm{up}} | m, n, m', n'\right] = \gamma$. The value of u_{up} can be calculated from the quantile function (the inverse of the cumulative distribution function, CDF) of the posterior for $P_f|M = m, n$ evaluated at γ.

The *upper credible interval* is of particular relevance for reliability analysis, as it quantifies how plausible it is that the true probability of failure p_f is smaller than an associated upper bound.

$C_{\gamma,\mathrm{HPD}}$ For unimodal distributions, the γ *highest posterior density (HPD) credible interval* is the narrowest interval out of all γ credible intervals. Any point within the interval has a higher density than any other point outside of the interval. In general, the interval cannot be calculated explicitly, but has to be obtained by solving an optimization problem.

For $m+m' > 1 \wedge n+n'-m-m' > 1$, the posterior distribution is unimodal and has a single mode within $(0, 1)$.

$C_{\gamma,\mathrm{eq.tail}}$ The γ *equal-tail credible interval* gives the interval $[l_{\mathrm{eq.tail}}, u_{\mathrm{eq.tail}}]$, for which there is a $(1-\gamma)/2$ probability that the value of P_f is smaller than $l_{\mathrm{eq.tail}}$ and a $(1-\gamma)/2$ probability that the value of P_f is larger than $u_{\mathrm{eq.tail}}$. If the distribution is symmetric and unimodal, the HPD and the equal-tail credible interval coincide. The values of the interval bounds $l_{\mathrm{eq.tail}}$ and $u_{\mathrm{eq.tail}}$ can be calculated from the quantile function of the posterior for $P_f|M = m, n$ evaluated at $(1-\gamma)/2$ and $1-(1-\gamma)/2$, respectively.

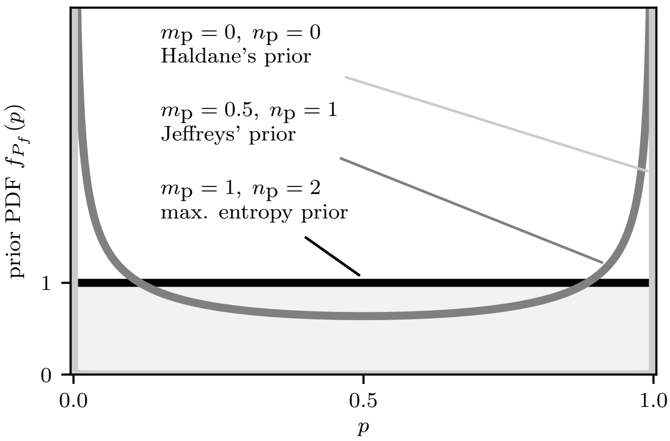

Figure 1: Different options for the weakly-informative prior PDF $f_{P_f}(p)$.

Unlike the upper credible interval and the equal-tail credible interval, the HPD credible interval is not invariant under transformation. This implies that the HPD interval with respect to the posterior for P_f is not identical to the HPD interval with respect to the posterior for the reliability index (see [27, 28] for a definition of the reliability index).

3.5. *Choice of the parameters m' and n' of the prior*

3.5.1. *No prior information about P_f is available*

If nothing is known about P_f in advance except that $P_f \in [0, 1]$, inevitably the question arises how to choose the parameters m' and n' of the prior [17]. Potential choices for weakly-informative prior distributions are [18, 23]:

Maximum entropy prior ($m' = 1, n' = 2$) The parameters are selected in accordance with the Principle of Maximum Information Entropy [16, 29, 30]. For this particular combination of parameter values, the beta distributed prior becomes a uniform distribution on $[0, 1]$. In this case, the beta distributed posterior is $\mathcal{B}(m+1, n-m+1)$.

Jeffreys' prior ($m' = 0.5, n' = 1$) The parameters are selected to give the so-called Jeffreys prior [31, 32] for the problem at hand [21, 23]. For this particular combination of parameter values, the beta distributed prior becomes the standard arcsine distribution on $[0, 1]$. The Jeffreys' prior is an noninformative prior that is invariant under reparameterization. In this case, the beta distributed posterior is $\mathcal{B}(m+0.5, n-m+0.5)$.

Haldane's prior ($m' = 0, n' = 0$) Haldane [33] proposed this choice of parameter values for the prior, which results in an improper prior, where the prior P_f is either *zero* or *one* with equal probability. In this case, the beta distributed posterior is $\mathcal{B}(m, n-m)$. This prior distribution is proportional to two Dirac delta functions placed at *zero* and *one*, and corresponds to a Bernoulli distribution with a mean of 0.5. If the posterior mean (Eq. (12)) obtained with Haldane's prior is used as a point estimate of p_f, the estimate corresponds to the MLE estimate given in Eq. (5), and is, thus, unbiased.

The shapes of the maximum entropy prior, the Jeffreys' prior and the Haldane's prior are depicted in Fig. 1. Any of the listed prior distributions could be classified as weakly-informative. Consequently, one could easily believe that its influence on the posterior distribution should be minimal. Nevertheless, it does matter which of the distributions listed above is selected as prior, even if n is large (as will be shown in the following paragraphs). This can be explained as follows: The number m of failures observed in a reliability analysis is usually small, even if n is large, as commonly $p_f \ll 10^{-3}$. As the probability of failure conditional on a performed MCS is characterized by $(m + m')$ and $(n + n')$, it matters whether m' is set to 1, 0.5 or 0. A discussion of the maximum entropy prior and Jeffreys' prior in the context of empirical Bayes can be found in [34].

In Section 2.4, we argue that a Monte Carlo γ *credible interval* should ideally qualify also as a confidence interval, and, thus, have a coverage probability γ_c that is at least as large as γ. In the following, we assess the coverage probabilities related to the above listed weakly-informative prior distributions. Coverage probabilities γ_c for 0.99, 0.95, 0.9 and 0.8 credible intervals are depicted in Figs. 2 to 4, as a function of $n \cdot p_f$. The assumption underlying the plots is that the Poisson approximation of the binomial distribution holds, such that γ_c can be expressed as a function of $n \cdot p_f$ [46]. We determined that this is the case for $p_f \le 10^{-3}$.

Coverage probabilities for the γ *upper credible interval* are shown in Fig. 2. Independent of the applied prior distribution, the coverage probability γ_c oscillates and has clear upward and downward spikes. For the maximum entropy prior, the coverage probability γ_c of the associated γ upper credible interval appears to be at least as large as γ. For a reliability-based assessment, such a behavior is desirable, as it leads to a conservative estimate. For both Jeffreys' prior and Haldane's prior, the coverage probability γ_c can be considerably below γ and, thus, the credible intervals obtained with these prior distributions do not qualify as proper confidence intervals. For Haldane's prior, γ acts as an upper bound for the coverage probability γ_c, which is clearly undesired. Thus, the maximum entropy prior exhibits the best performance, compared to Jeffreys' prior and Haldane's prior. The γ upper credible interval obtained using Jeffreys' prior is acceptable for $n \cdot p_f < 0.8$, as in this case $\gamma_c > \gamma$. For $n \cdot p_f > 0.8$, the coverage probability γ_c associated with Jeffreys' prior oscillates around γ.

For the maximum entropy prior, the downward spikes seem to exactly touch the γ values. However, a closer look reveals that at the downward spikes the coverage probability γ_c is indeed slightly smaller than γ. The relation $\gamma - \gamma_c$ at the downward spikes, obtained with the maximum entropy prior, is shown in Fig. 5. The figure reveals that the coverage probability γ_c at the downward spikes is smaller than γ in all cases. The difference between γ_c and γ increases with increasing $n \cdot p_f$ and decreasing γ. Thus, also the upper credible intervals based on the maximum entropy prior do not strictly qualify as proper confidence intervals. Nevertheless, for practical purposes, the associated error is negligible and only occurs if $n \cdot p_f$ is in the vicinity of a spike location.

Fig. 3 shows coverage probabilities for the γ *highest posterior density (HPD) credible interval*. In this case, the maximum entropy prior slightly outperforms Jeffreys' prior. The coverage probabilities γ_c associated with both prior distributions oscillates around γ. However, the coverage probability associated with Jeffreys' prior exhibits a single pronounced downward spike for $n \cdot p_f$ around one. For Haldane's prior, the coverage probability γ_c is in all cases smaller or equal than γ, which is an undesired behavior.

Fig. 4 shows coverage probabilities for the γ *equal-tail credible* interval. In this case, credible intervals based on Jeffreys' prior seem slightly more favorable than credible intervals based on the maximum entropy prior. For increasing $n \cdot p_f$, both the Jeffreys' prior and the maximum entropy prior lead to credible intervals with γ_c oscillating around γ. For $n \cdot p_f < 1$, the credible interval based on the maximum entropy prior has more pronounced downward spikes than the one obtained with Jeffreys' prior. However, a disadvantage of the Jeffreys' prior is that the associated credible interval has an unexpected strong downward spike for $n \cdot p_f$ between 1 and 10. Also for this type of credible intervals, Haldane's prior exhibits an undesired behavior, as the associated credible intervals have coverage probabilities γ_c much smaller than γ for smaller $n \cdot p_f$.

In summary, we recommend the maximum entropy prior as basis to obtain credible intervals. In particular, if upper credible intervals (Fig. 2) are of interest, the maximum entropy prior excels. For the HPD credible interval and the equal-tail credible interval, the maximum entropy prior and Jeffreys prior lead to credible intervals with a similar performance. Haldane's prior should not be used as basis to calculate γ credible intervals, as the associated coverage probabilities γ_c are usually clearly below γ.

3.5.2. Prior information about P_f is available

In some cases, prior information is available, e.g., form previous MCS of a similar system. One could try to express the available prior knowledge in terms of the expected prior probability of failure m'/n' and prior number of samples n' (compare also [23]), where n' can be interpreted as a measure for the prior confidence about the value of m'/n'. For the maximum entropy prior, we have $m'/n' = 0.5$ and $n' = 2$. Typically, for a reliability problem, one can safely suspect that the underlying true probability of failure p_f is closer to *zero* than to *one*; i.e., one could try to select a value smaller than 0.5 for m'/n'.

However, selecting a small m'/n' combined with a low n' gives a $m' \ll 1$. This results in a highly skewed prior distribution. Indeed, for $m' < 1$, the prior distribution has a mode at *zero*. As $m'/n' \to 0$ for a fixed n' that is small, the prior distribution converges to the *Dirac delta* function. This can lead to unexpected behavior of the prior distribution. Fig. 6 shows the

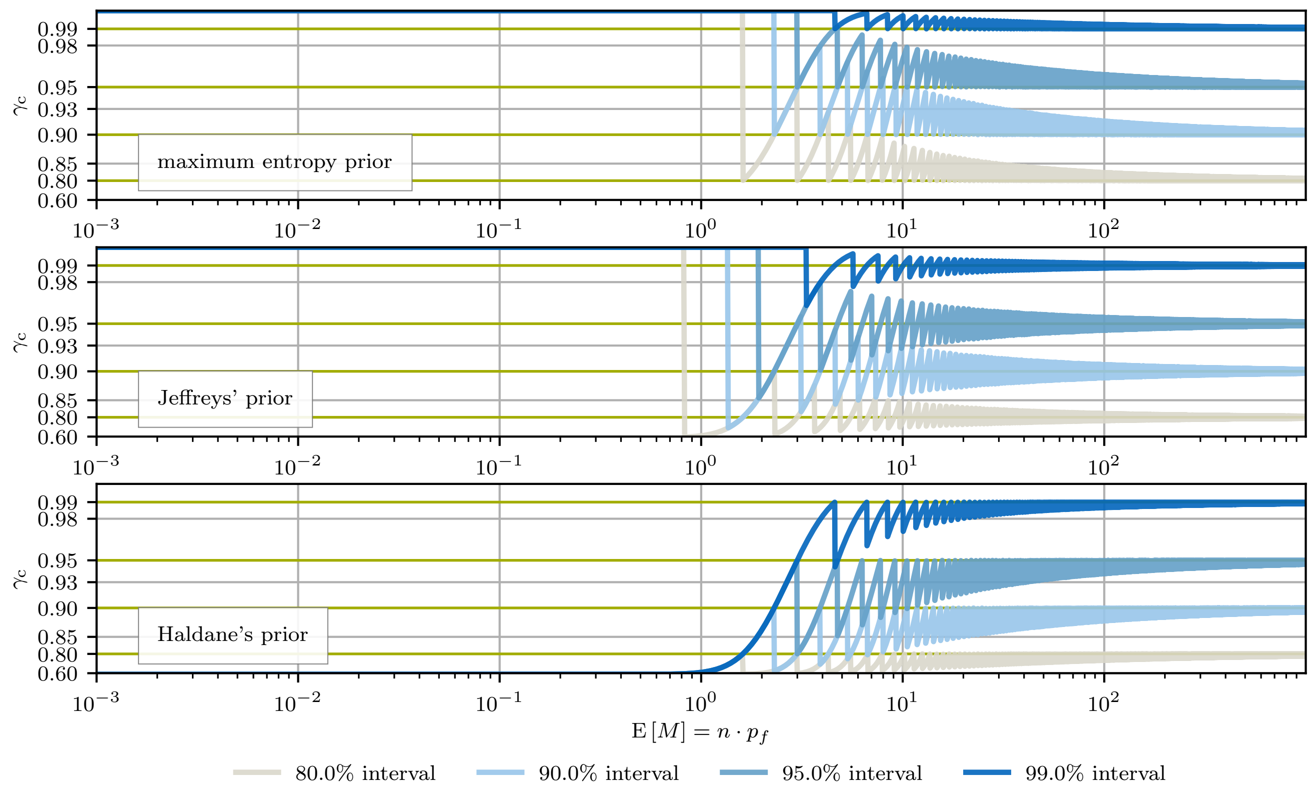

Figure 2: Coverage probability γ_c for the 0.99, 0.95, 0.9 and 0.8 ***upper credible intervals*** $C_{\gamma,\text{up}}$ as a function of the expected number of hits $\mathrm{E}[M] = n \cdot p_f$. The upper plot shows the corresponding credible intervals using the maximum entropy prior. The middle plot shows the corresponding credible intervals using Jeffreys' prior. The bottom plot shows the corresponding credible intervals using Haldane's prior. The plot is valid for $p_f \leq 10^{-3}$ and $n \in \mathbb{N}^*$.

threshold u_{up} of the 95% upper credible interval $C_{95\%,\text{up}}$ of the prior distribution for different values of m'/n' and n'. For small m'/n' combined with a low n', u_{up} becomes smaller than m'/n'.

Therefore, if the available prior knowledge is vague such that $m' \ll 1$, the maximum entropy prior is a more appropriate modeling choice: (i) It almost surely gives a conservative estimate of the γ upper credible interval $C_{\gamma,\text{up}}$ (compare Fig. 2). (ii) For $m \geq 1$, the posterior distribution obtained with the maximum entropy prior and the posterior distribution obtained with a prior for which $m' \ll 1$ and n' is small will be almost identical.

Constrained noninformative prior distributions (e.g., distributions with a given mean value) are discussed in [23]. For the conceptually similar problem of reliability estimation based on data collected from sampling tests, [18] discusses how to consider available prior information. Another relevant discussion of how to account for prior information in the more general context of Bayesian hypothesis testing can be found in [17].

3.6. Quantifying the uncertainty about p_f if $m = 0$

The unbiased estimate $\hat{p}_{f,\text{MLE}}$ for p_f is $\hat{p}_{f,\text{MLE}} = m/n$. Note that the value of $\hat{p}_{f,\text{MLE}}$ is *zero* if $m = 0$, independent of n.

A (biased) Bayesian estimate is the expected value of the posterior P_f. With the *maximum entropy prior* this estimate is (compare also Eq. (12)):

$$\hat{p}_{f,\text{Bayes}} = \mathrm{E}\left[P_f | M = m, n\right] = \frac{m+1}{n+2}. \tag{16}$$

In case $m = 0$, Eq. (16) reduces to

$$\hat{p}_{f,\text{Bayes}} = \mathrm{E}\left[P_f | M = 0, n\right] = \frac{1}{n+2}, \tag{17}$$

which is always larger than *zero*.

Structural reliability methods are typically applied to verify that the reliability of the system of interest is larger than a prespecified target reliability [17]. If this verification is based on $\hat{p}_{f,\text{MLE}}$, then $m = 0$ is not acceptable, as the point estimate $\hat{p}_{f,\text{MLE}} = 0/n = 0$ is misleading. Contrary to that, $m = 0$ is acceptable if Bayesian post-processing of MCS is applied. The posterior for P_f can be used to quantify how credible it is that the prespecified target reliability is maintained, even if $m = 0$. Therefore, Bayesian post-processing of MCS is of particular interest, if the reliability of the system of interest is much larger than the prespecified target reliability, as n can be significantly smaller than the number of samples needed to achieve an $m > 0$.

The posterior mean and selected credible intervals for $m = 0$ are illustrated in Fig. 7 for an increasing number of samples n. If, for example, the aim is to demonstrate that $p_f \leq 10^{-6}$, then $n = 3 \times 10^6$ samples are needed to verify that the posterior P_f is smaller than the target with probability 90%, $n = 4 \times 10^6$ samples are needed to verify that the posterior P_f is smaller than the target with probability 95%, and $n = 5 \times 10^6$ samples are needed to verify that the posterior P_f is smaller than the target with probability 99%. The posterior mean for P_f is already smaller than the target for at least 10^6 samples.

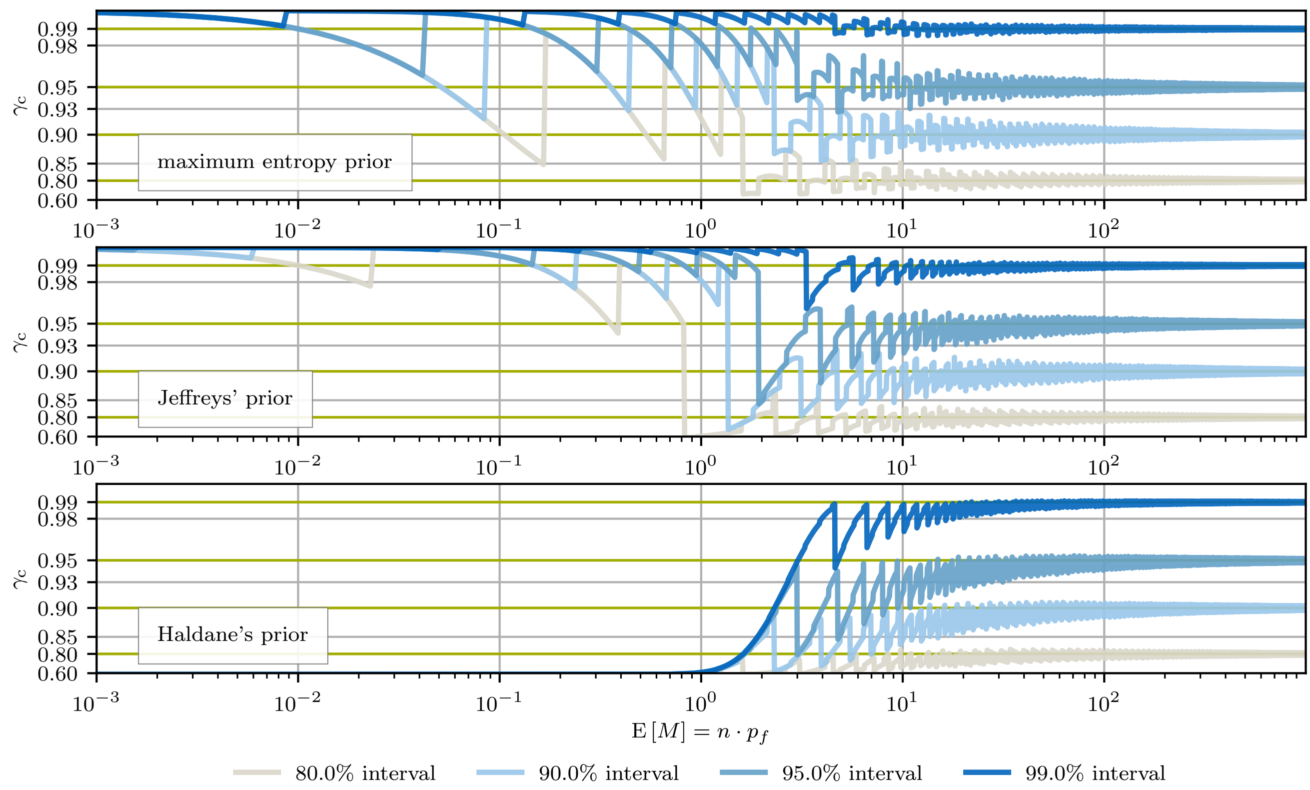

Figure 3: Coverage probability γ_c for the 0.99, 0.95, 0.9 and 0.8 ***highest posterior density (HPD) credible intervals*** $C_{\gamma,\mathrm{HPD}}$ as a function of the expected number of hits $\mathrm{E}[M] = n \cdot p_f$. The upper plot shows the corresponding credible intervals using the maximum entropy prior. The middle plot shows the corresponding credible intervals using Jeffreys' prior. The bottom plot shows the corresponding credible intervals using Haldane's prior. The plot is valid for $p_f \leq 10^{-3}$ and $n \in \mathbb{N}^*$.

4. Reliability updating with Monte Carlo simulation

4.1. Problem statement

For existing structures, additional data on parameters describing the system input or observations of system states might become available. Let such data and observations be denoted by the *observation event* $\mathcal{D}$. Accounting for $\mathcal{D}$ when assessing the probability of the system being in an undesired state $\mathcal{F}$ is referred to as *reliability updating*. The corresponding probability of failure $p_{f|\mathcal{D}}$ can be expressed in terms of Bayes' rule as:

$$p_{f|\mathcal{D}} = \Pr[\mathcal{F}|\mathcal{D}] = \frac{\Pr[\mathcal{F} \cap \mathcal{D}]}{\Pr[\mathcal{D}]}. \tag{18}$$

Without loss of generality [35, 36], one can introduce a function $g_\mathcal{D}(\mathbf{x})$ such that $g_\mathcal{D}(\mathbf{x}) \leq 0$ if and only if the state of the probabilistic system model matches the observation, and, thus, $\Pr[\mathcal{D}] = \Pr[g_\mathcal{D}(\mathbf{X}) \leq 0]$. We note that the latter applies also for cases where $\mathcal{D}$ consists of measurement data and, hence, the information is of equality type, see [36] for details. Employing $g_\mathcal{D}(\mathbf{x})$, Eq. (18) can be stated as:

$$p_{f|\mathcal{D}} = \frac{\Pr[g(\mathbf{X}) \leq 0 \wedge g_\mathcal{D}(\mathbf{X}) \leq 0]}{\Pr[g_\mathcal{D}(\mathbf{X}) \leq 0]}. \tag{19}$$

4.2. Monte Carlo simulation for reliability updating

The following two indicator functions are introduced:

$$I_\mathcal{D}(\mathbf{x}) = \begin{cases} 1 & \text{if } g_\mathcal{D}(\mathbf{x}) \leq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{20}$$

$$I_{\mathcal{F}\wedge\mathcal{D}}(\mathbf{x}) = \begin{cases} 1 & \text{if } g(\mathbf{x}) \leq 0 \wedge g_\mathcal{D}(\mathbf{x}) \leq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

Using $I_\mathcal{D}(\mathbf{x})$ and $I_{\mathcal{F}\wedge\mathcal{D}}(\mathbf{x})$, Eq. (19) can be expressed as:

$$p_{f|\mathcal{D}} = \frac{\mathrm{E}[I_{\mathcal{F}\wedge\mathcal{D}}(\mathbf{X})]}{\mathrm{E}[I_\mathcal{D}(\mathbf{X})]}, \tag{22}$$

where the two expectations in Eq. (22) can be estimated by means of MCS:

$$p_\mathcal{D} = \mathrm{E}[I_\mathcal{D}(\mathbf{X})] \approx \hat{p}_{\mathcal{D},\mathrm{MLE}} = \frac{m_\mathcal{D}}{n}, \tag{23}$$

$$p_{\mathcal{F}\wedge\mathcal{D}} = \mathrm{E}[I_{\mathcal{F}\wedge\mathcal{D}}(\mathbf{X})] \approx \hat{p}_{\mathcal{F}\wedge\mathcal{D},\mathrm{MLE}} = \frac{m_{\mathcal{F}\wedge\mathcal{D}}}{n}, \tag{24}$$

with $m_\mathcal{D} = \sum_{i=1}^n I_\mathcal{D}(\mathbf{x}^{(i)})$ and $m_{\mathcal{F}\wedge\mathcal{D}} = \sum_{i=1}^n I_{\mathcal{F}\wedge\mathcal{D}}(\mathbf{x}^{(i)})$. If the n in both Eqs. (23) and (24) is the same, Eq. (22) can be approximated as:

$$p_{f|\mathcal{D}} = \frac{p_{\mathcal{F}\wedge\mathcal{D}}}{p_\mathcal{D}} \approx \hat{p}_{f|\mathcal{D},\mathrm{MLE}} = \frac{\hat{p}_{\mathcal{F}\wedge\mathcal{D},\mathrm{MLE}}}{\hat{p}_{\mathcal{D},\mathrm{MLE}}} = \frac{m_{\mathcal{F}\wedge\mathcal{D}}}{m_\mathcal{D}}. \tag{25}$$

The quantities $\hat{p}_{\mathcal{D},\mathrm{MLE}}$ and $\hat{p}_{\mathcal{F}\wedge\mathcal{D},\mathrm{MLE}}$ in Eq. (25) can either be obtained based on two separate MCS runs, or jointly within

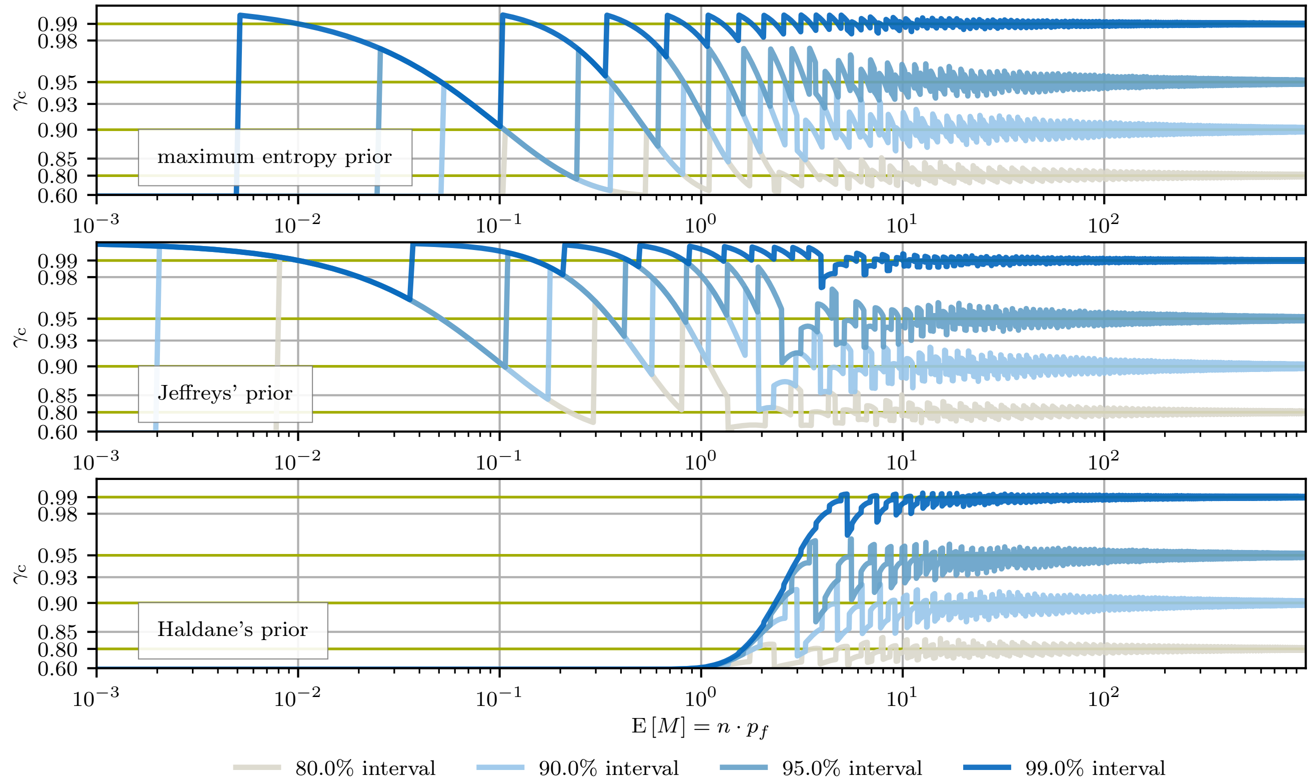

Figure 4: Coverage probability γ_c for the 0.99, 0.95, 0.9 and 0.8 ***equal-tail credible intervals*** $C_{\gamma,\mathrm{eq.tail}}$ as a function of the expected number of hits $\mathrm{E}[M] = n \cdot p_f$. The upper plot shows the corresponding credible intervals using the maximum entropy prior. The middle plot shows the corresponding credible intervals using Jeffreys' prior. The bottom plot shows the corresponding credible intervals using Haldane's prior. The plot is valid for $p_f \leq 10^{-3}$ and $n \in \mathbb{N}^*$.

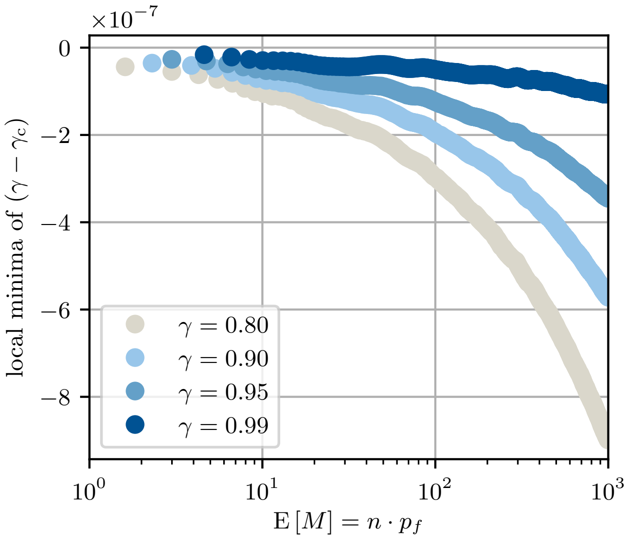

Figure 5: For the γ upper credible interval obtained by using the maximum entropy prior, this plot shows the relation $\gamma - \gamma_c$ relative to $n \cdot p_f$ at the downward spikes observed in Fig. 2.

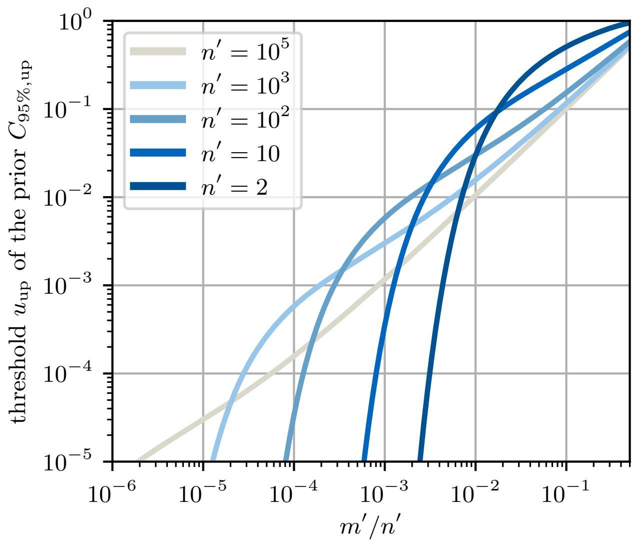

Figure 6: Threshold u_{up} of the 95% upper credible interval $C_{95\%,\mathrm{up}}$ of the prior distribution as a function of the parameters m' and n'.

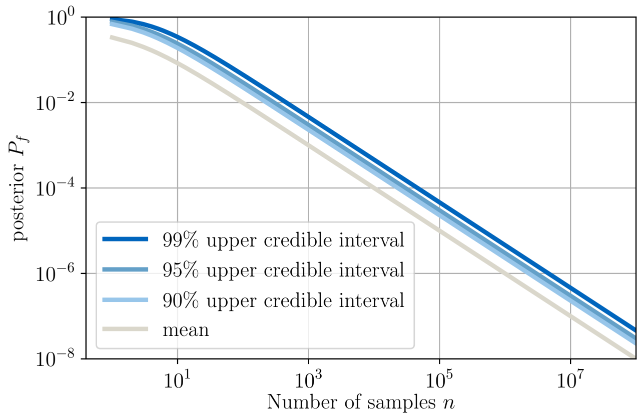

Figure 7: For $m = 0$, the plot shows the number of samples n needed to achieve the specified credibility for the value of P_f.

a single MCS run. Evaluating both $\hat{p}_{\mathfrak{D},\text{MLE}}$ and $\hat{p}_{\mathcal{F}\wedge\mathfrak{D},\text{MLE}}$ with the same MCS, samples will lead to reduced variance compared to using two independent simulations, due to the correlation between the two estimates. This can be shown trough a first-order Taylor series expansion; let δ be the coefficient of variation of $\hat{p}_{f|\mathfrak{D},\text{MLE}}$. It is shown in that Appendix B that:

$$\delta^2 \approx \delta_1^2 + \delta_2^2 - 2 \cdot \rho \cdot \delta_1 \cdot \delta_2, \tag{26}$$

where δ_1 and δ_2 are the coefficients of variation of the estimates in the numerator and denominator and ρ is their correlation coefficient.

Additionally, from a computational point of view, it is preferable to estimate $\hat{p}_{\mathfrak{D},\text{MLE}}$ and $\hat{p}_{\mathcal{F}\wedge\mathfrak{D},\text{MLE}}$ jointly within a single simulation run, as $g_{\mathfrak{D}}(\mathbf{x}^{(i)})$ needs to be evaluated for both $I_{\mathfrak{D}}(\mathbf{x}^{(i)})$ and $I_{\mathcal{F}\wedge\mathfrak{D}}(\mathbf{x}^{(i)})$, and the numerical effort to additionally evaluate $g(\mathbf{x}^{(i)})$ is often small. For example, if a finite element model is used, usually the computationally demanding task of solving the model needs to be performed only once per realization $\mathbf{x}^{(i)}$, and both $g_{\mathfrak{D}}(\mathbf{x}^{(i)})$ and $g(\mathbf{x}^{(i)})$ can be evaluated through comparatively cheap post-processing strategies. Therefore, in the following, only the case where both $I_{\mathfrak{D}}(\mathbf{x}^{(i)})$ and $I_{\mathcal{F}\wedge\mathfrak{D}}(\mathbf{x}^{(i)})$ are evaluated jointly within a single MCS run is considered.

4.3. *Bayesian estimates of $p_{f|\mathfrak{D}}$*

4.3.1. *Bayes' theorem in the context of reliability updating*

For reliability updating, a MCS with a total of n samples is conducted, and the quantities $m_{\mathfrak{D}}$ and $m_{\mathcal{F}\wedge\mathfrak{D}}$ defined in Section 4.2 are observed, as outcomes of binomial random variables $M_{\mathfrak{D}}$ and $M_{\mathcal{F}\wedge\mathfrak{D}}$, respectively. The joint posterior of $P_{f|\mathfrak{D}}$ and $P_{\mathfrak{D}}$ is:

$$f_{P_{f|\mathfrak{D}},P_{\mathfrak{D}}|m_{\mathcal{F}\wedge\mathfrak{D}},m_{\mathfrak{D}},n}(p_1, p_2) = \frac{\mathcal{L}(p_1, p_2|m_{\mathcal{F}\wedge\mathfrak{D}}, m_{\mathfrak{D}}, n) \cdot f_{P_{f|\mathfrak{D}}}(p_1) \cdot f_{P_{\mathfrak{D}}}(p_2)}{\Pr\left[M_{\mathcal{F}\wedge\mathfrak{D}} = m_{\mathcal{F}\wedge\mathfrak{D}}, M_{\mathfrak{D}} = m_{\mathfrak{D}}|n\right]} \tag{27}$$

where:

$f_{P_{f|\mathfrak{D}},P_{\mathfrak{D}}|m_{\mathcal{F}\wedge\mathfrak{D}},m_{\mathfrak{D}},n}(p_1, p_2)$ is the posterior joint PDF of $P_{f|\mathfrak{D}}$ and $P_{\mathfrak{D}}$.

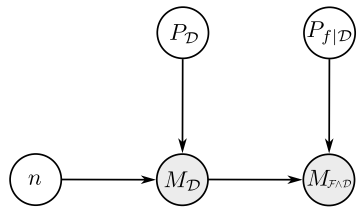

Figure 8: Representation of the dependency structure of the variable entering the reliability updating problem as a Bayesian network. After conducting the MCS, the states of the nodes highlighted in gray are known. A-priori $P_{f|\mathfrak{D}}$ is assumed to be independent of $P_{\mathfrak{D}}$.

$f_{P_{f|\mathfrak{D}}}(p_1)$ and $f_{P_{\mathfrak{D}}}(p_2)$ are the prior distributions. Eq. (27) implies that a-priori $P_{f|\mathfrak{D}}$ is independent of $P_{\mathfrak{D}}$.

$\mathcal{L}(p_1, p_2|m_{\mathcal{F}\wedge\mathfrak{D}}, m_{\mathfrak{D}}, n)$ is the likelihood function:

$$\mathcal{L}(p_1, p_2|m_{\mathcal{F}\wedge\mathfrak{D}}, m_{\mathfrak{D}}, n) = \Pr\Big[M_{\mathcal{F}\wedge\mathfrak{D}} = m_{\mathcal{F}\wedge\mathfrak{D}}, M_{\mathfrak{D}} = m_{\mathfrak{D}}|P_{f|\mathfrak{D}} = p_1, \tag{28}$$

which is discussed in Section 4.3.2 below.

$\Pr\left[M_{\mathcal{F}\wedge\mathfrak{D}} = m_{\mathcal{F}\wedge\mathfrak{D}}, M_{\mathfrak{D}} = m_{\mathfrak{D}}|n\right]$ is a normalizing constant which ensures that the posterior PDF integrates to *one*.

4.3.2. *Likelihood function*

The dependency structure of the variables that enter the definition of the reliability updating problem is depicted with a Bayesian network in Fig. 8. As the network shows, the outcome of $M_{\mathcal{F}\wedge\mathfrak{D}}$ is independent of $P_{\mathfrak{D}}$ and n conditional on the value of $M_{\mathfrak{D}}$. Thus, the likelihood function in Eq. (28) can be reformulated as:

$$\mathcal{L}(p_1, p_2|m_{\mathcal{F}\wedge\mathfrak{D}}, m_{\mathfrak{D}}, n) = \Pr\left[M_{\mathcal{F}\wedge\mathfrak{D}} = m_{\mathcal{F}\wedge\mathfrak{D}}|p_1, m_{\mathfrak{D}}\right] \cdot \Pr\left[M_{\mathfrak{D}} = m_{\mathfrak{D}}|p_2, \tag{29}$$

$\Pr\left[M_{\mathfrak{D}} = m_{\mathfrak{D}}|p_2, n\right]$ is the probability mass function of a binomial distribution with parameters n and p_2, as discussed in Section 3.3. For each of the $m_{\mathfrak{D}}$ out of n samples $\mathbf{x}^{(i)}$ for which $I_{\mathfrak{D}}(\mathbf{x}^{(i)}) = 1$, the indicator function $I_{\mathcal{F}\wedge\mathfrak{D}}(\mathbf{x}^{(i)})$ equals *one* with probability $P_{f|\mathfrak{D}} = p_1$. Therefore, $m_{\mathcal{F}\wedge\mathfrak{D}}$ is the outcome of a binomial distributed random variable $M_{\mathcal{F}\wedge\mathfrak{D}}$ with parameters equal to $m_{\mathfrak{D}}$ and p_1. Consequently, $\Pr\left[M_{\mathcal{F}\wedge\mathfrak{D}} = m_{\mathcal{F}\wedge\mathfrak{D}}|p_1, m_{\mathfrak{D}}\right]$ is the probability mass function of a binomial distribution with parameters $m_{\mathfrak{D}}$ and p_1.

4.3.3. *Posterior distribution for $P_{f|\mathfrak{D}}$*

As the the influences of $P_{f|\mathfrak{D}}$ and $P_{\mathfrak{D}}$ can be clearly separated in the likelihood function (compare Eq. (29)), the posterior distributions for $P_{f|\mathfrak{D}}$ and $P_{\mathfrak{D}}$ introduced in Eq. (27) can also be formulated separately for both $P_{f|\mathfrak{D}}$ and $P_{\mathfrak{D}}$. The posterior distribution for $P_{\mathfrak{D}}$ is based on a classical MCS (which is discussed in Section 3.3). The posterior distribution for $P_{f|\mathfrak{D}}$ can be expressed as:

$$f_{P_{f|\mathfrak{D}}|m_{\mathcal{F}\wedge\mathfrak{D}},m_{\mathfrak{D}}}(p_1) = \frac{\Pr\left[M_{\mathcal{F}\wedge\mathfrak{D}} = m_{\mathcal{F}\wedge\mathfrak{D}}|p_1, m_{\mathfrak{D}}\right] \cdot f_{P_{f|\mathfrak{D}}}(p_1)}{\Pr\left[M_{\mathcal{F}\wedge\mathfrak{D}} = m_{\mathcal{F}\wedge\mathfrak{D}}|M_{\mathfrak{D}} = m_{\mathfrak{D}}\right]}. \tag{30}$$

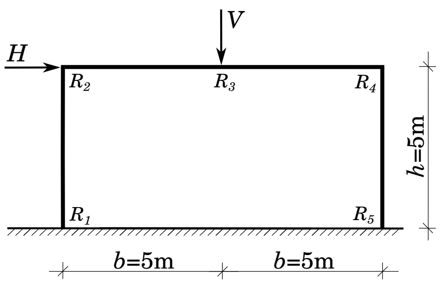

Figure 9: Steel frame structure investigated in Section 5.

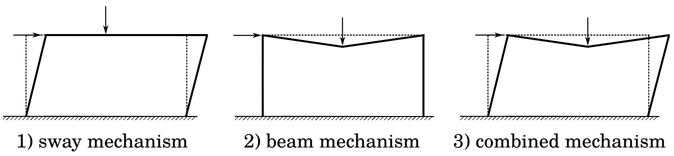

Figure 10: Failure mechanisms of the frame structure depicted in Fig. 9. The three failure mechanisms correspond to limit-state functions g_1, g_2 and g_3, defined in Eqs. (32)–(34).

As discussed in Section 3.3, the beta distribution is the conjugate prior for likelihood functions proportional to the probability mass function of a binomial distribution. Using such a prior model for $P_{f|\mathfrak{D}}$, Eq. (30) can be expressed as a beta distribution with parameters $m_{\mathcal{F}\wedge\mathfrak{D}} + m'_{\mathcal{F}\wedge\mathfrak{D}}$ and $m_{\mathfrak{D}} - m_{\mathcal{F}\wedge\mathfrak{D}} + m'_{\mathfrak{D}} - m'_{\mathcal{F}\wedge\mathfrak{D}}$ (compare Section 3.3):

$$f_{P_{f|\mathfrak{D}}|m_{\mathcal{F}\wedge\mathfrak{D}},m_{\mathfrak{D}}}(p_1) = \frac{{p_1}^{m_{\mathcal{F}\wedge\mathfrak{D}}+m'_{\mathcal{F}\wedge\mathfrak{D}}-1} \cdot (1-p_1)^{m_{\mathfrak{D}}-m_{\mathcal{F}\wedge\mathfrak{D}}+m'_{\mathfrak{D}}-m'_{\mathcal{F}\wedge\mathfrak{D}}-1}}{\mathrm{B}(m_{\mathcal{F}\wedge\mathfrak{D}} + m'_{\mathcal{F}\wedge\mathfrak{D}}, m_{\mathfrak{D}} - m_{\mathcal{F}\wedge\mathfrak{D}} + m'_{\mathfrak{D}} - m'_{\mathcal{F}\wedge\mathfrak{D}})}. \quad (31)$$

where $m'_{\mathcal{F}\wedge\mathfrak{D}}$ and $m'_{\mathfrak{D}}$ are the parameters used to describe the beta distribution of the prior $f_{P_{f|\mathfrak{D}}}(p_1)$. Following the discussion in Section 3.5, we recommend to use the maximum entropy prior for $f_{P_{f|\mathfrak{D}}}(p_1)$; i.e., $m'_{\mathcal{F}\wedge\mathfrak{D}} = 1$ and $m'_{\mathfrak{D}} = 2$.

5. Illustration example

5.1. Specification of the application example

By means of an application example, we demonstrate how Bayesian credible intervals can be evaluated based on a conducted MCS. We consider the plane frame structure depicted in Fig. 9 (motivated by the example given in [28]). Plastic hinge mechanisms leading to collapse of the structure are considered and analyzed by utilizing elastic-plastic stress-strain relations. Hinges are assumed to form at the end of elements or at points of load application only. The height and the half-width of the structure are constant values with $h = 5$m and $b = 5$m, respectively. The structure is loaded with horizontal load H and vertical load V, as shown in Fig. 9.

The failure mechanisms of the investigated frame structure are shown in Fig. 10. The corresponding component limit-state functions describing the three indicated failure mechanisms are:

$$g_1(\mathbf{X}) = R_1 + R_2 + R_4 + R_5 - h \cdot H, \quad (32)$$

$$g_2(\mathbf{X}) = R_2 + 2 \cdot R_3 + R_4 - b \cdot V, \quad (33)$$

$$g_3(\mathbf{X}) = R_1 + 2 \cdot R_3 + 2 \cdot R_4 + R_5 - h \cdot H - b \cdot V, \quad (34)$$

where $R_1, \ldots, R_5$ are plastic moments, H is a horizontal load and V is a vertical load. Each $R_1, \ldots, R_5$ follows a log-Normal distribution with mean m_R and standard deviation 5kNm, where m_R is modeled by a random variable (which all $R_1, \ldots, R_5$ have in common) that has a log-Normal distribution with mean 200kNm and standard deviation 40kNm. Conditional on the value of m_R, the $R_1, \ldots, R_5$ are independent. The uncertainty on H is modeled by a Gumbel distribution with mean 50kN and standard deviation 10kN. The uncertainty on V is modeled by a Gumbel distribution with mean 40kN and standard deviation 8kN. Thus, the vector $\mathbf{X}$ consists of $m_R, R_1, \ldots, R_5$, H and V.

The frame structure fails if at least one of the component limit-state functions indicates failure. Consequently, the limit-state function for system failure can be expressed as:

$$g(\mathbf{X}) = \min\left(g_1(\mathbf{X}), g_2(\mathbf{X}), g_3(\mathbf{X})\right). \quad (35)$$

After construction, the frame structure is successfully tested for a proof load P of 113.9kN applied at the loading point of the horizontal force H, where the value of P corresponds to 1.5 times the 98th percentile of H. The information obtained from successfully applying the proof load P can be expressed by the following condition:

$$g_{\mathfrak{D}}(\mathbf{X}) \leq 0, \quad (36)$$

where

$$g_{\mathfrak{D}}(\mathbf{X}) = h \cdot P - \min\left(R_1 + R_2 + R_4 + R_5, R_1 + 2R_3 + 2R_4 + R_5\right). \quad (37)$$

The goal of the conducted MCS is to learn about the values of $p_f = \Pr[g(\mathbf{X}) \leq 0]$ and $p_{f|\mathfrak{D}}$, where $p_{f|\mathfrak{D}}$ is defined according to Eq. (19).

5.2. Outcome of the conducted Monte Carlo simulation

A MCS with 10^{10} samples is conducted. In Table 1, the outcome of the MCS that utilizes only the first n of the 10^{10} samples is shown. Thus, the rows in Table 1 do not stem from independent MCS, but demonstrate convergence for an increasing number of samples within a single MCS. The notation used in Table 1 is as follows: m denotes the number of samples for which $g(\mathbf{X}) \leq 0$ (compare Section 3.1), $m_{\mathfrak{D}}$ denotes the number of samples for which $g_{\mathfrak{D}}(\mathbf{X}) \leq 0$ (compare Section 4.1), and $m_{\mathcal{F}\wedge\mathfrak{D}}$ denotes the number of samples for which $g(\mathbf{X}) \leq 0 \wedge g_{\mathfrak{D}}(\mathbf{X}) \leq 0$ (compare Section 4.1).

Bayesian post-processing is applied to quantify the uncertainty about the values of p_f and $p_{f|\mathfrak{D}}$, where the *maximum entropy prior* is used. Based on the results listed in Table 1, the posterior cumulative distribution functions for both P_f and $p_{f|\mathfrak{D}}$ are shown in Fig. 11. The posterior mean, as well as the 90%, 95% and 99% upper credible intervals for P_f and $P_{f|\mathfrak{D}}$ are shown in Fig. 12 for increasing n. For up to 10^3 samples, the credible intervals and the posterior mean for P_f and $P_{f|\mathfrak{D}}$

Table 1: Outcome of the conducted MCS. n is the number of samples considered. m is the number of samples for which $g(\mathbf{X}) \leq 0$. $m_{\mathfrak{D}}$ is the number of samples for which $g_{\mathfrak{D}}(\mathbf{X}) \leq 0$. $m_{\mathcal{F}\wedge\mathfrak{D}}$ is the number of samples for which $g(\mathbf{X}) \leq 0 \wedge g_{\mathfrak{D}}(\mathbf{X}) \leq 0$.
The sets of samples for smaller n are part of the all sets of samples for larger n. For example, the initial 10 samples are also part of the 100 samples, the 100 samples are part of the 10^3 samples, and so on.

n	m	$m_{\mathfrak{D}}$	$m_{\mathcal{F}\wedge\mathfrak{D}}$
10	0	10	0
100	0	96	0
10^3	0	949	0
10^4	3	$9,482$	1
10^5	21	$94,560$	1
10^6	243	$946,650$	29
10^7	2261	$9,466,042$	292
10^8	$22,277$	$94,653,620$	$2,855$
10^9	$219,283$	$946,538,911$	$28,208$
10^{10}	$2,187,289$	$9,465,429,813$	$281,234$

are similar. This is due to both m and $m_{\mathcal{F}\wedge\mathfrak{D}}$ being *zero* up to $n = 10^3$, and $m_{\mathfrak{D}}$ being close to n; i.e., $m_{\mathfrak{D}}/n$ converges to approximately 0.95 for large n. However, for $n > 10^3$, the posteriors for P_f and $P_{f|\mathfrak{D}}$ differ, as m and $m_{\mathcal{F}\wedge\mathfrak{D}}$ are larger than *zero* and take significantly different values. For increasing n the credible intervals become narrower. For large n, the posterior mean of P_f converges to $\approx 2.2\times10^{-4}$ and the posterior mean of $P_{f|\mathfrak{D}}$ converges to $\approx 3\times10^{-5}$. Thus, the information obtained from the conducted proof load test decreases the probability of failure by approximately one order of magnitude.

For $n < 10^5$, the posterior mean differs from the underlying true probability of failure. For n up to 10^3, m is *zero* (as is $m_{\mathcal{F}\wedge\mathfrak{D}}$). Consequently, the classical Monte Carlo estimate returns a value of *zero*; i.e., $\hat{p}_{f,\mathrm{MLE}} = m/n = 0/n = 0$. If Bayesian post-processing is applied, the uncertainty about P_f can objectively quantified. In this case, the value of n needs to be increased until there is sufficient credibility that the target reliability is maintained for the investigated system. If the reliability of the investigated system is much larger than the target reliability, then it is likely that the analysis is sufficiently accurate even if m is still *zero* (compare Fig. 7).

6. Concluding remarks

Bayesian post-processing of the outcome of a MCS is straightforward and inexpensive. Credible intervals for the posterior probability of failure can be evaluated based on the inverse cumulative distribution function of the beta distribution. The distribution of the posterior probability of failure can be highly skewed, even if the total number of samples is very large. Therefore, the interpretation of the coefficient of variation and variance of a Monte Carlo estimate can be difficult and misleading. Credible intervals are a more appropriate measure to quantify the uncertainty about the estimated probability of failure. With minor modifications, the Bayesian post-processing strategy can also be applied to quantify the uncertainty if MCS is used for reliability updating.

For practical application, upper credible intervals are of particular interest, as they quantify how probable it is that the probability of failure is smaller than a selected threshold value. In this context, the maximum entropy prior (a uniform distribution on [0; 1]) is recommended as prior distribution for the probability of failure when no prior information is available. Both Jeffreys' and Haldane's prior can considerably overestimate the probability of being smaller than the selected threshold value.

For MCS, the distribution for the posterior probability of failure can be derived analytically. The resulting posterior distribution depends on the probability of failure of the investigated system of interest, and is not influenced by the actual shape of the limit-state function of the problem at hand. In principle, the discussed Bayesian post-processing could also be extended to more advance sampling-based reliability methods. However, in this case one needs to make assumptions to obtain a distribution for the posterior probability of failure, as in e.g., [16]. Contrary to the discussed Bayesian post-processing approach for MCS, such assumptions limit the informative value of credible intervals obtained based on the approximate posterior distribution. In this case, the performance of the Bayesian post-processing step will depend not only on the probability of failure of the investigated system (as is the case for MCS) but additionally on the shape of the limit-state function.

Appendix A. Confidence intervals for Monte Carlo simulation of rare events

Appendix A.1. Introduction

In Section 2 we argue that for MCS, credible intervals should ideally also qualify as proper confidence intervals. For this reason, some commonly used confidence intervals related to a MCS are briefly discussed in the following.

For MCS, a γ confidence interval that has coverage probabilities γ_c equal to γ independent of the value of p_f does not exist [21, 37]. Nevertheless, proper γ confidence intervals exist for MCS. However, all of them exhibit coverage probabilities γ_c that can be larger than γ [21]. This is due to the discrete nature of the problem [37, 38]; the total number of potential states that $M = m$ can take for n samples is $n+1$ and thus finite, $m \in \{0, \ldots, K\}$.

Appendix A.2. The Clopper–Pearson confidence interval

The Clopper–Pearson (C-P) interval [39] gives a proper γ confidence interval for a MCS with m failure samples in n trials. Let $[l, u]$ denote the γ C-P confidence interval. The value of u is chosen such that the probability of observing an outcome less or equal than m for a binomial distribution with n trials and probability parameter u equals $\gamma/2$. Correspondingly, the value of l is chosen such that the probability of observing an outcome larger or equal than m for a binomial distribution with n trials and probability parameter l equals $\gamma/2$. Even though the C-P interval gives a proper confidence interval, it is often criticized as being overly conservative [21, 40]; i.e., the associated coverage probabilities γ_c are often much larger than the targeted confidence level γ. Moreover, the C-P interval is prone

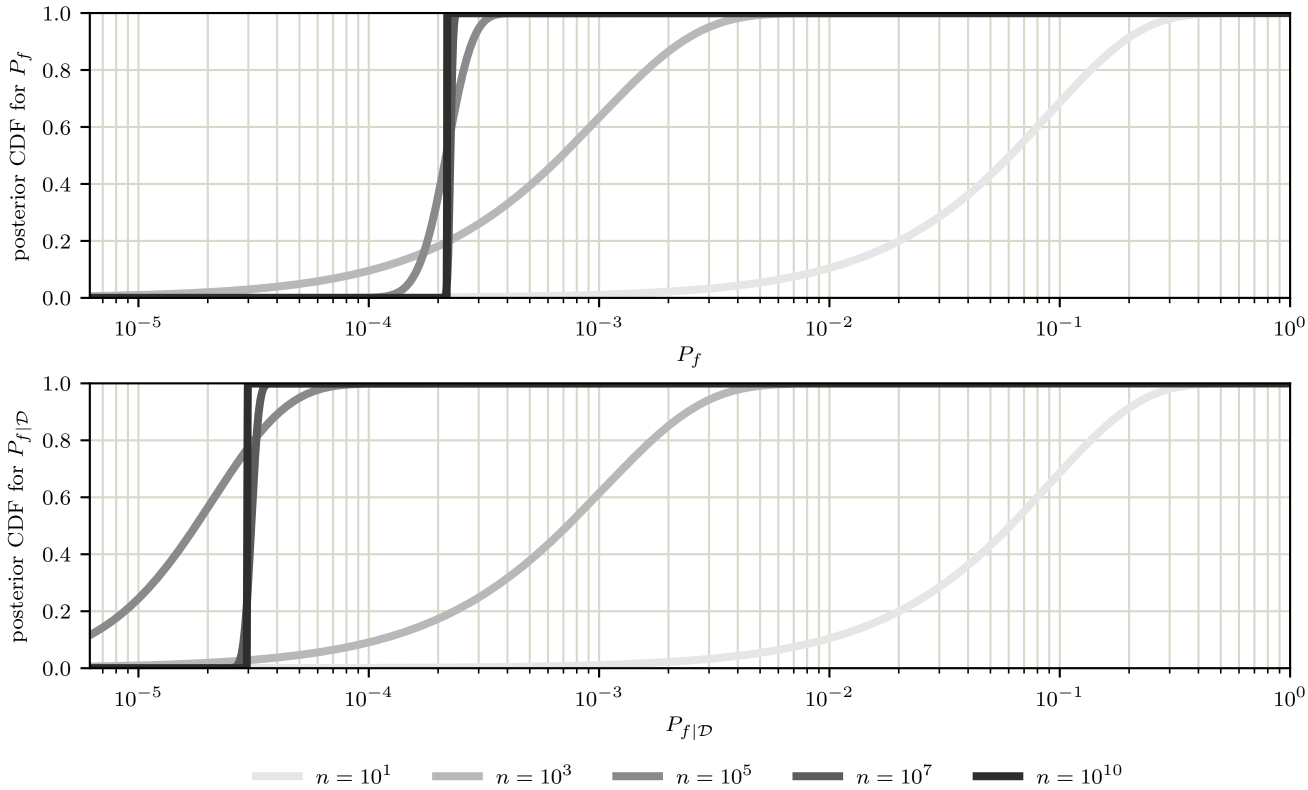

Figure 11: Convergence of the CDF of the posterior for P_f and $P_{f|\mathcal{D}}$ for a gradually increasing number of samples in the illustration example.

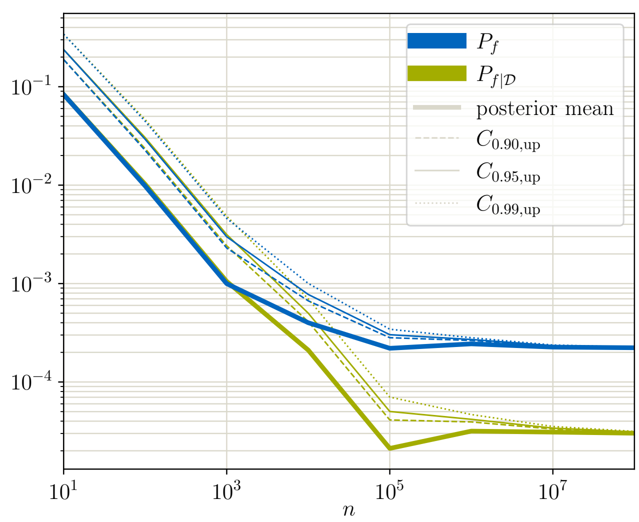

Figure 12: Convergence of P_f and $P_{f|\mathcal{D}}$ for a gradually increasing number of samples in the illustration example.

to misinterpretation, as the lower and upper interval bounds are computed based on different underlying distributions.

As the binomial distribution and the beta distribution are related, the quantities u and l can also be expressed in terms of the beta distribution [21]: Let $\mathcal{B}(\alpha, \beta)$ represent a beta distribution with parameters α and β. The value of l can then be obtained as the $\gamma/2$ quantile of $\mathcal{B}(m, n-m+1)$; the value of u is the $1-\gamma/2$ quantile of $\mathcal{B}(m+1, n-m)$. The above allows making a connection between the bounds of the C-P confidence interval and the quantiles of the posterior distribution of P_f obtained with Bayesian analysis. However, different posterior distributions are used to evaluate l and u (as both quantities are computed based on different prior distributions). The quantity l, can be associated with a beta-distributed prior with parameters $m' = 0$ and $n' = 1$. The quantity u, can be associated with a beta-distributed prior with parameters $m' = 1$ and $n' = 1$.

As the upper and lower bound of the C-P interval are evaluated based on different distribution models, a matching credible interval cannot be derived by applying Bayesian analysis. That is, one cannot claim that the interval has γ probability of containing the true p_f, conditional on a specific MCS. Such confidence intervals that do not have a corresponding Bayesian credible interval as counterpart are an easy source for misinterpretation [25].

Appendix A.3. Approximate confidence intervals for Monte Carlo simulation

As discussed in the previous section, even though the C-P interval is a proper confidence interval, it is often overly conservative. In practice, the confidence intervals typically used in the context of MCS are only approximate; i.e., they do not strictly maintain the confidence level γ in all cases. A comparison of different confidence intervals can be found in [21].

The Jeffreys' interval for this problem can be derived using a Bayesian approach in combination with the Jeffreys' prior. It is equivalent to the Bayesian equal-tail credible interval based on the Jeffreys' prior – of which the performance is illustrated in Fig. 4. Comparing the quantiles of the Beta distributions used in the Jeffreys' interval and in the C-P confidence interval, it can be seen that the Jeffreys' interval is always contained within the corresponding C-P confidence interval. Thus, one can interpret the Jeffreys' interval as a continuity corrected version of the C-P confidence interval [21]. However, note that the Jeffreys' interval does not qualify as a proper γ confidence interval, as the associated coverage probability γ_c can be smaller than γ.

Other approximate confidence intervals with acceptable performance properties include the Wilson interval [21, 41] and the Agresti–Coull interval [21, 40, 42], whereby the latter envelopes the former [21].

Appendix A.4. Confidence intervals based on a Normal approximation

The highly skewed shape of $P_f|M = m, n$ is discussed in Section 3.3. From this discussion, it can be concluded that confidence intervals based on assuming symmetry should be avoided, even if n is large. This is an important observation, as confidence intervals for p_f are frequently based on a Normal approximation if n is large [1, 43, 44, 45] (compare also to the discussion in [21]). The upper and lower bounds of the γ Normal approximation confidence interval C_N can be evaluated as [21]:

$$C_N = \frac{m}{n} \pm \frac{\Phi^{-1}\left(\frac{1+\gamma}{2}\right)}{\sqrt{n}} \cdot \sqrt{\frac{m}{n} \cdot \left(1 - \frac{m}{n}\right)}, \qquad \text{(A.1)}$$

where $\Phi^{-1}(\cdot)$ denotes the inverse cumulative distribution function of the standard Normal distribution.

Coverage probabilities for the Normal approximation confidence interval are shown in Fig. A.13. For $E[M] = n \cdot p_f < 10$, the Normal approximation gives coverage probabilities γ_c much smaller than γ. Only if the expected number of hits $E[M] = n \cdot p_f$ exceeds 10^2, the coverage probabilities γ_c obtained with the Normal approximation are close to γ. Thus, in practice, the Normal approximation confidence interval should not be used, as it can grossly underestimate the variability of the estimate for a given p_f.

Appendix B. Derivation of the first-order Taylor series expansion in Eq. (26)

Let $\hat{p} = \frac{\hat{p}_1}{\hat{p}_2}$, with $\hat{p}_1$ the estimate in the numerator and $\hat{p}_2$ the estimate in the denominator Eq. (25). A first-order Taylor expansion of $\hat{p}$ around of means of $\hat{p}_1$ ($E[\hat{p}_1] = p_1$) and $\hat{p}_2$ ($E[\hat{p}_2] = p_2$) reads:

$$\hat{p} \approx \frac{p_1}{p_2} + \frac{1}{p_2}(\hat{p}_1 - p_1) - \frac{p_1}{p_2^2}(\hat{p}_2 - p_2) \qquad \text{(B.1)}$$

The first-order approximation of the mean is $E[\hat{p}] \approx \frac{p_1}{p_2}$. The first-order approximation of the variance reads:

$$\mathrm{Var}[\hat{p}] \approx \frac{1}{p_2^2}\sigma_1^2 + \frac{p_1^2}{p_2^4}\sigma_2^2 - 2\frac{p_1}{p_2^3}\rho\sigma_1\sigma_2 \qquad \text{(B.2)}$$

where $\sigma_i^2 = \mathrm{Var}[\hat{p}_i]$ and ρ is the correlation coefficient of $\hat{p}_1$ and $\hat{p}_2$. Dividing the above with the square of the first-order approximation of the mean, $\frac{\hat{p}_1^2}{\hat{p}_2^2}$, we get the following approximation of the square of the coefficient of variation of $\hat{p}$:

$$\delta^2 \approx \frac{1}{p_1^2}\sigma_1^2 + \frac{1}{p_2^2}\sigma_2^2 - 2\frac{1}{p_1 p_2}\rho\sigma_1\sigma_2 = \delta_1^2 + \delta_2^2 - 2\rho\delta_1\delta_2 \qquad \text{(B.3)}$$

where δ_i is the coefficient of variation of $\hat{p}_i$.

References

[1] R. E. Melchers and A. T. Beck, *Structural reliability analysis and prediction*. John Wiley & Sons, 2018.

[2] S.-K. Au and J. L. Beck, "Estimation of small failure probabilities in high dimensions by Subset Simulation," *Probabilistic Engineering Mechanics*, vol. 16, no. 4, pp. 263–277, 2001.

[3] M. Hohenbichler and R. Rackwitz, "Improvement of second-order reliability estimates by importance sampling," *Journal of Engineering Mechanics*, vol. 114, no. 12, pp. 2195–2199, 1988.

[4] P. Koutsourelakis, H. Pradlwarter, and G. Schuëller, "Reliability of structures in high dimensions, part I: algorithms and applications," *Probabilistic Engineering Mechanics*, vol. 19, no. 4, pp. 409–417, 2004.

[5] M. de Angelis, E. Patelli, and M. Beer, "Advanced line sampling for efficient robust reliability analysis," *Structural safety*, vol. 52, pp. 170–182, 2015.

[6] I. Papaioannou and D. Straub, "Combination line sampling for structural reliability analysis," *Structural Safety*, vol. 88, p. 102025, 2021.

[7] P. Bjerager, "Probability integration by directional simulation," *Journal of Engineering Mechanics*, vol. 114, no. 8, pp. 1285–1302, 1988.

[8] O. Ditlevsen, R. E. Melchers, and H. Gluver, "General multi-dimensional probability integration by directional simulation," *Computers & Structures*, vol. 36, no. 2, pp. 355–368, 1990.

[9] I. Papaioannou, C. Papadimitriou, and D. Straub, "Sequential importance sampling for structural reliability analysis," *Structural safety*, vol. 62, pp. 66–75, 2016.

[10] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*, vol. 10. John Wiley & Sons, 2016.

[11] N. Kurtz and J. Song, "Cross-entropy-based adaptive importance sampling using gaussian mixture," *Structural Safety*, vol. 42, pp. 35–44, 2013.

[12] I. Papaioannou, S. Geyer, and D. Straub, "Improved cross entropy-based importance sampling with a flexible mixture model," *Reliability Engineering & System Safety*, vol. 191, p. 106564, 2019.

[13] H. Janssen, "Monte-carlo based uncertainty analysis: Sampling efficiency and sampling convergence," *Reliability Engineering & System Safety*, vol. 109, pp. 123–132, 2013.

[14] D. L. Kelly and C. L. Smith, "Bayesian inference in probabilistic risk assessment—the current state of the art," *Reliability Engineering & System Safety*, vol. 94, no. 2, pp. 628–643, 2009.

[15] N. O. Siu and D. L. Kelly, "Bayesian parameter estimation in probabilistic risk assessment," *Reliability Engineering & System Safety*, vol. 62, no. 1-2, pp. 89–116, 1998.

[16] K. M. Zuev, J. L. Beck, S.-K. Au, and L. S. Katafygiotis, "Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions," *Computers & Structures*, vol. 92, pp. 283–296, 2012.

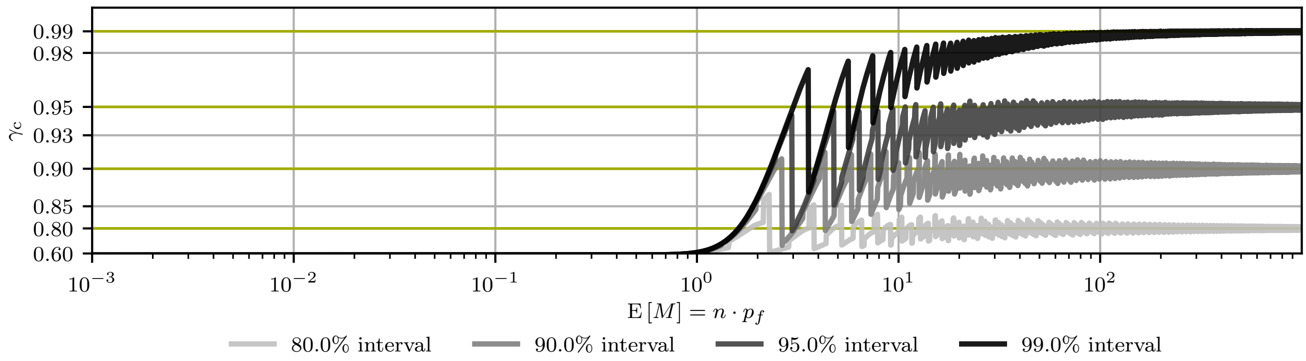

Figure A.13: Coverage probability γ_c for the 0.99, 0.95, 0.9 and 0.8 ***Normal approximation confidence interval*** C_N as a function of the expected number of hits $\mathrm{E}[M] = n \cdot p_f$. The plot is valid for $p_f \leq 10^{-3}$ and $n \in \mathbb{N}^*$.

[17] C. Smith, D. Kelly, and H. Dezfuli, "Probability-informed testing for reliability assurance through bayesian hypothesis methods," *Reliability Engineering & System Safety*, vol. 95, no. 4, pp. 361–368, 2010.

[18] F. Guérin, B. Dumon, and E. Usureau, "Reliability estimation by Bayesian method: definition of prior distribution using dependability study," *Reliability Engineering & System Safety*, vol. 82, no. 3, pp. 299–306, 2003.

[19] W. Betz, *Bayesian inference of engineering models*. PhD thesis, Technische Universität München, 2017.

[20] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[21] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval estimation for a binomial proportion," *Statistical Science*, pp. 101–117, 2001.

[22] J. Cho, Y. Kim, J. Kim, J. Park, and D.-S. Kim, "Realistic estimation of human error probability through Monte Carlo thermal-hydraulic simulation," *Reliability Engineering & System Safety*, vol. 193, p. 106673, 2020.

[23] C. L. Atwood, "Constrained noninformative priors in risk assessment," *Reliability Engineering & System Safety*, vol. 53, no. 1, pp. 37–46, 1996.

[24] P. M. Lee, *Bayesian statistics*. Arnold Publication, 1997.

[25] R. D. Morey, R. Hoekstra, J. N. Rouder, M. D. Lee, and E.-J. Wagenmakers, "The fallacy of placing confidence in confidence intervals," *Psychonomic bulletin & review*, vol. 23, no. 1, pp. 103–123, 2016.

[26] P. Diaconis, D. Ylvisaker, *et al.*, "Conjugate priors for exponential families," *The Annals of statistics*, vol. 7, no. 2, pp. 269–281, 1979.

[27] O. Ditlevsen, "Generalized second moment reliability index," *Journal of Structural Mechanics*, vol. 7, no. 4, pp. 435–451, 1979.

[28] O. Ditlevsen and H. O. Madsen, *Structural reliability methods*. Technical University of Denmark, 2007.

[29] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[30] E. T. Jaynes, "Information theory and statistical mechanics," *Physical review*, vol. 106, no. 4, p. 620, 1957.

[31] H. Jeffreys, *The theory of probability*. Oxford University Press, 1998.

[32] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 186, no. 1007, pp. 453–461, 1946.

[33] J. B. S. Haldane, "A note on inverse probability," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 28, pp. 55–61, Cambridge University Press, 1932.

[34] A. Gribok, V. Agarwal, and V. Yadav, "Performance of empirical Bayes estimation techniques used in probabilistic risk assessment," *Reliability Engineering & System Safety*, vol. 201, p. 106805, 2020.

[35] D. Straub, "Reliability updating with equality information," *Probabilistic Engineering Mechanics*, vol. 26, no. 2, pp. 254–258, 2011.

[36] D. Straub, I. Papaioannou, and W. Betz, "Bayesian analysis of rare events," *Journal of Computational Physics*, vol. 314, pp. 538–556, 2016.

[37] G. Casella, "Refining binomial confidence intervals," *Canadian Journal of Statistics*, vol. 14, no. 2, pp. 113–129, 1986.

[38] J. Neyman, "On the problem of confidence limits," *Annals of mathematical statistics*, vol. 6, no. 1, p. 1, 1935.

[39] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, vol. 26, no. 4, pp. 404–413, 1934.

[40] A. Agresti and B. A. Coull, "Approximate is better than "exact" for interval estimation of binomial proportions," *The American Statistician*, vol. 52, no. 2, pp. 119–126, 1998.

[41] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *Journal of the American Statistical Association*, vol. 22, no. 158, pp. 209–212, 1927.

[42] M. L. Samuels and J. A. Witmer, *Statistics for the life sciences*. Prentice Hall, Englewoo Cliffs, NJ, 2 ed., 1999.

[43] M. Lemaire, *Structural reliability*. John Wiley & Sons, 2009.

[44] C. A. Schenk and G. I. Schuëller, *Uncertainty assessment of large finite element systems*, vol. 24. Springer Science & Business Media, 2005.

[45] E. T. Jaynes and O. Kempthorne, "Confidence intervals vs bayesian intervals," in *Foundations of probability theory, statistical inference, and statistical theories of science*, pp. 175–257, Springer, 1976.

[46] J. L. Hodges and L. Le Cam, "The poisson approximation to the poisson binomial distribution," *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 737–740, 1960.