– Technical Report –

Registration No.	Date of Report
OPSEC# 7002	28/11/2022

Title: Characterization of Armor Plate Proof Velocity via Bayesian Inference

Author(s): James M. Gorman

Abstract: Ballistic testing has traditionally used a frequentist statistics approach to determine confidence in an armor design. Separately, there appears to have been a shift in the statistics community away from a frequentist approach towards a Bayesian approach for data analysis. This shift has been aided, in part, by the significant increase in computational capabilities over the last few decades that have made Bayesian problems more tractable. This report demonstrates how Bayesian inference uses additional information present in a ballistic event to determine a proof velocity prior to armor qualification.

U.S. Army Combat Capabilities Development Command
Ground Vehicle Systems Center
Detroit Arsenal
Warren, Michigan 48397-5000

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
28-11-2022	Final	Apr 21 - Nov 22

4. TITLE AND SUBTITLE

Characterization of Armor Plate Proof Velocity via Bayesian Inference

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

James M. Gorman

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

US Army DEVCOM Ground Vehicle System Center
6501 E 11 Mile Rd
Warren, MI 48397-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A: Approved for public release. Distribution is unlimited. OPSEC#: 7002

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Ballistic testing has traditionally used a frequentist statistics approach to determine confidence in an armor design. Separately, there appears to have been a shift in the statistics community away from a frequentist approach towards a Bayesian approach for data analysis. This shift has been aided, in part, by the significant increase in computational capabilities over the last few decades that have made Bayesian problems more tractable. This report demonstrates how Bayesian inference uses additional information present in a ballistic event to determine a proof velocity prior to armor qualification.

15. SUBJECT TERMS

Ballistic Testing, Bayesian Inference, Terminal Ballistics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	Same as Report (SAR)	36	James M. Gorman
Unclassified	Unclassified	Unclassified			19b. TELEPHONE NUMBER *(include area code)* 586-282-6580

Contents

List of Figures

List of Tables

1 Introduction

Ballistic testing is necessary in all cases where armor is used to protect human life. Despite the great increase in computational capabilities over the past few decades, the dynamic impact and penetration mechanics involved in ballistic events are challenging to accurately model, whether it is for kinetic energy (KE) threats (*e.g.* small arms) [1, 2, 3, 4] or chemical energy (CE) threats (*e.g.* shaped charge jets) [5, 6]. As a result, although numerical tools may assist in the development of armor solutions, eventually the armor design must be tested against the actual threat of interest or a representative surrogate.

Different organizations have different methods of qualifying an armor solution [7, 8]. However, one common thread in testing is that a frequentist statistics approach is used: that is, each shot outcome is treated as a binary response (partial penetration, PP, or complete penetration, CP)[1], and enough shots are taken around a velocity, with a given tolerance, until the appropriate confidence level is reached. In the case of a 90% probability of threat defeat at the 90% confidence level and assuming a binomial distribution, this generally requires 22 PPs out of 22 shots, 37 PPs out of 38 shots, or 50 PPs out of 52 shots [8]. While this 90-90 criterion could be reached by larger values, trying to qualify a target beyond these values may be considered too expensive.

Since qualification of an armor design can be time consuming and expensive (depending on the threat of interest), a report [8] detailed a way to reduce the number shots required to reach the 90-90 criterion to either 13 PP out of 13 shots or 15 PP out of 15 shots. However, the assumption of a binomial distribution was the most conservative of the distributions considered and, therefore, both of the new approaches would be less conservative. Since the design of armor is meant to protect lives and reduce injury, we disagree with the idea of using a less-conservative qualification method to save time and money.

Outside of armor qualification testing, another common way to analyze armor materials/designs is by determining the V_{50} [9, 11], which is the velocity at which there is an equal chance whether or not the threat perforates the witness plate behind the target. The V_{50} is typically determined by averaging 6-10 shots that are equally split between PP and CP. Additional specifications may exist for the maximum spread of these 6-10 shots [12, 13, 14], but this is not always the case.

[1]Nominally, a partial penetration is where the threat does not perforate the witness plate behind the armor, while a complete penetration is where the threat does perforate the witness plate behind the armor. A specific definition used by the United States military is given in Ref. [9], although this is not the only definition that is/has been used [10].

Using what might be referred to as a modified V_{50} test method, Refs. [15, 16, 17] use a wide range of threat velocities with a mixture of PP and CP results to characterize an armor plate/layup. Using the maximum likelihood estimate (MLE), a Gaussian cumulative distribution function (CDF) or probit is fit to the available data. This curve then provides the opportunity to provide a reliability estimate of threat defeat for a general velocity and not just one in particular. Therefore, it gives the engineer or program manager additional information for a given threat without necessarily requiring additional testing in case threat defeat requirements are changed at a later time.

While this appears useful, the simplification hides the complexity of terminal ballistic events, particularly when multi-material laminates are used. Multiple failure mechanisms may be present, especially as more composites are utilized in armor designs, and these failure mechanisms may be activated at different velocities [18]. These include, but are not limited to: brittle or radial fracture, fragmentation, ductile hole growth, plugging, and fiber shearing/rupture (in the case of fiber-reinforced composites). Therefore, the assumption of the normal CDF may not necessarily be true in practice due to multiple failure mechanisms, and, therefore, the use of the modified V_{50} test method may not actually provide the purported protection level for a particular velocity in general.

This does not, however, mean that the modified V_{50} test method is superfluous. Indeed, we will demonstrate how it may be used to provide an estimate of the "proof velocity", here defined as the highest velocity the plate or layup is reasonably expected to meet for a given design criterion (*e.g.* 90-90). The drawback of the method utilized in the above references is that MLE was used, and it will be shown below that this is not the best method for determining the parameters of the normal distribution. Rather, Bayesian inference ought to be used with minimally informative priors. Its use has increased significantly as of late, being utilized in numerous disparate fields [19, 20, 21].

In the context of terminal ballistics, most Bayesian methods have been utilized to improve material constitutive model characterization [22, 23, 24]. While this is certainly important for accurate impact calculations, there appears to be limited work on improving material characterization methods for armor protection methods. The present work distinguishes itself from previous material characterization work by focusing on extracting parameters that are useful in the design of protective armor solutions, not the specific constitutive parameters that model local deformation in the terminal ballistics event.

This report will proceed in the following manner. Section 2 will provide a review of probability and (Bayesian) statistics. Section 3 will provide the method of the Monte Carlo

calculations and goes through a set of exploratory scenarios and a discussion of the findings. Experimental test data is presented and discussed in Section 4. Potential future work is included in Section 5, and final conclusions are given in Section 6.

2 Review of Probability and Statistics

For those with a limited background in probability and/or statistics, a brief review will be provided here. $P(A)$ is to be read as the "probability of (event/parameter) A" and $P(A|D)$ as the "probability of (event/parameter) A given (data) D". $P(A|D)$ is a *conditional* probability, because it gives the probability of event A occurring having observed D. Inference methods care about conditional probabilities because they inform the individual of the probability of an event given the current data. In statistics, the two inference methods commonly used are often referred to as classical inference and Bayesian inference, as they are based off of frequentist/classical/orthodox statistics and Bayesian statistics.

The significant difference between frequentist and Bayesian statistics is how the parameters and data are treated and understood. In the frequentist approach, parameters are treated as absolute and data are random. The frequentist approach only makes sense in the case of numerous repeated events. After all, if there is a mayoral election held in a large city with the long-time incumbent against a young challenger, this match-up may not necessarily happen again (particularly if the incumbent loses). The frequentist approach does, however, seem reasonable in circumstances where many repeat occurrences can or will take place, such as flipping a coin or a ballistic engagement against an armor layup.

Therefore, in frequentist statistics, there is a nuance of how the confidence interval is defined, even though it may often be overlooked. If a "95% confidence interval for the normal mean μ is [-0.5, 1.0]...[, what] it means is that if the same procedure... was repeated very many times, for all kinds of different data sets, then in 95% of the cases would the true μ lie in the 95% confidence interval" [25]. Again, the parameters are treated as fixed while the data is random. Furthermore, because the frequentist approach does not always condition on the data, it can give the wrong confidence in light of the observed results [25]. This also means that, in some instances, frequentist statistics/inference may answer a different question than the practitioner is actually asking [26].

In Bayesian statistics, parameters are treated as random variables and data are absolute [27]. Therefore, probabilities must be assigned for different events using some prior assumptions, and it does not matter if an event may only occur once. In the election example given above, probabilities for who would win the election example given above could be readily

assigned given various data in the community (*e.g.* roads, school quality, *etc.*) and how well each candidate is liked by the public. The parameters, since they are random variables, are typically referenced with respect to a "credible interval", not a "confidence interval". Therefore, for the normal mean mentioned above, "this Bayesian interval conveys that there is a .95 probability that μ lies in [-0.5, 1.0]" [25], not an asymptotic result after many repeated measurements.

Since Bayesian statistics treats parameters as random variables, probability distributions must be specified. The starting estimate of the random variables is called the *prior* ($P(A)$), and this is updated with data to give the final distribution, called the *posterior* ($P(A|D)$). This is typically described using Bayes' rule [27]

$$P(A|D) = \frac{P(D|A)\,P(A)}{P(D)}, \tag{1}$$

where $P(D|A)$ is the *likelihood* and $P(D)$ is the model evidence (also commonly referred to as the normalization). This method of updating is similar to how humans often consider topics: there are initial assumptions given historical data/experience, after which this is updated as more data points are available/collected.

In many cases, the different treatments between frequentist and Bayesian statistics do not matter, because it can be demonstrated that in many/most cases the confidence intervals (from frequentists) and credible intervals (from Bayesians) overlap [27]. Nonetheless, this overlap cannot be guaranteed, and we would encourage readers to determine whether current analysis methods might be able to be updated using Bayesian statistics, as it is being adopted across many disciplines [28].

This article is not meant to fully flush out Bayesian statistics and all its nuances compared to frequentist statistics, as this has been done elsewhere [25, 26, 27]. At least one author [29] goes so far as to state that frequentist statistics should no longer be taught to anyone outside of the statistics "specialists" (*i.e.* Ph.D. students). This is perhaps a little extreme, but there are certainly benefits to using Bayesian statistics. Furthermore, there are many probabilistic programming languages (PPLs) already out there to make these calculations easier and computational capabilities are significantly better than they were decades ago, therefore these can hardly be used as excuses to put off the numerical calculations. Alas, this Bayesian vs. frequentist debate will not be settled in this article, nor is it our intention to try. We simply want to demonstrate the inference capabilities provided by Bayesian statistics to perform more-informed ballistic testing.

As mentioned above, there are a number of PPLs that make performing this numerical work easier. `Turing.jl` [30] was the PPL was used for this work, and it is just one of the PPLs that have been written using Julia [31]. Besides Turing.jl, other PPLs include: Gen [32] (also written in Julia), emcee [33] is one available in Python, and Stan [34] (available in multiple languages), among others.

For those in the terminal ballistics community, one might wonder why the additional information provided by Bayesian statistics/inference would be relevant. In a V_{50} test, generally only a single number is provided (the V_{50}), which may or may not include additional data such as the zone of mixed results (ZMR) [9]. The modified V_{50} test method provides enough information to characterize *two* parameters, and this second parameter allows one to more accurately estimate the proof velocity because it can account for changes in the ZMR. After all, the purpose of characterization testing is to assist in the development of armor solutions (*i.e.* armor qualification).

Now, up until this point, there is nothing distinctly Bayesian about using the modified V_{50} approach: any statistical inference method could be used, following the appropriate standards or conventions, to characterize the armor layup. After all, classical inference can be done as well, as was done in Ref. [17] via the maximum likelihood estimate. However, classical inference does not necessarily allow the easy use of outside information, while Bayesian inference can do this through the choice of a physics-informed prior distribution (explained below). These distributions are then used to calculate the estimated proof velocity given a perforation relation (*e.g.* logistic or probit curve). When data is expensive to acquire (such as via ballistic testing), this additional information is particularly helpful. With this in mind, it is now time to look at how this might be done in practice.

3 Theoretical Analysis

It should be noted that the work presented here will only consider KE threats that can be fired at velocities with relatively good repeatability. It will also be assumed that the threat only induces one failure mode. This does not always occur in experimental testing, and it is highly dependent on the projectile, armor layup, *etc.* [18]; therefore, multi-layer laminates may be more poorly characterized than monolithic solutions. It is beyond the scope of this work to consider the more challenging CE threats, although how this approach might be used will be considered in Future Work (Section 5).

The logistic curve is one common way of modeling data that has a binomial outcome [11, 35]; however, the scaling parameter (*i.e.* not the median value) does not appear to have physical

meaning with respect to ballistic events. Normal (Gaussian) distributions are commonly used across disciplines and both parameters typically having physical significance. Recall that the Gaussian distribution is given by

$$N(x;\mu,\,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \tag{2}$$

where μ is the mean and σ is the standard deviation. Unfortunately, the normal distribution has no analytic integral and must be integrated numerically in order to obtain cumulative values. This shortcoming has been recognized, and the error function [36]

$$\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\, dt \tag{3}$$

gives an expression that involves this numerically-evaluated integral. Using the error function, the CDF of the general Gaussian distribution is given by [37]

$$y = \frac{1}{2}\left(1 + \mathrm{Erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right), \tag{4}$$

and this function provides the probability input of a Bernouli trial. We suggest using Equation (4), instead of the logistic curve, because the parameters have physical meaning in the case of a terminal ballistic event. In particular, μ, since it is also the median, is the V_{50} value, or the point at which there is equal likelihood of a partial penetration or complete penetration. Similarly, σ gives a measure of the spread of the zone of mixed results. The ZMR is the region of data where it is reasonable to expect both PPs and CPs within a small change of velocity. The specifics of how wide the ZMR is a function of the choice of material and thickness, and can be seen in the example data shown in Figure 1.

Because of the way V_{50} tests are typically conducted, σ is generally not calculated. Furthermore, since the modified V_{50} test (described below) does not appear to be commonly used, there does not appear to be any representative value for σ for a given material type, *e.g.* metal, composite, *etc.* While not necessary if one is solely after the V_{50}, the additional data of this "nuisance" parameter provides necessary information for generating CDF curves for given ballistic data.

The probabilities of equation (1) are calculated via numerical integration, the specifics of this is outside the scope of this work. PPLs are used, in part, to simplify this process and its implementation in code. Those readers interested in the work behind the numerics are directed to Refs. [38, 39, 40] for additional reading on probability, numerical integration, and

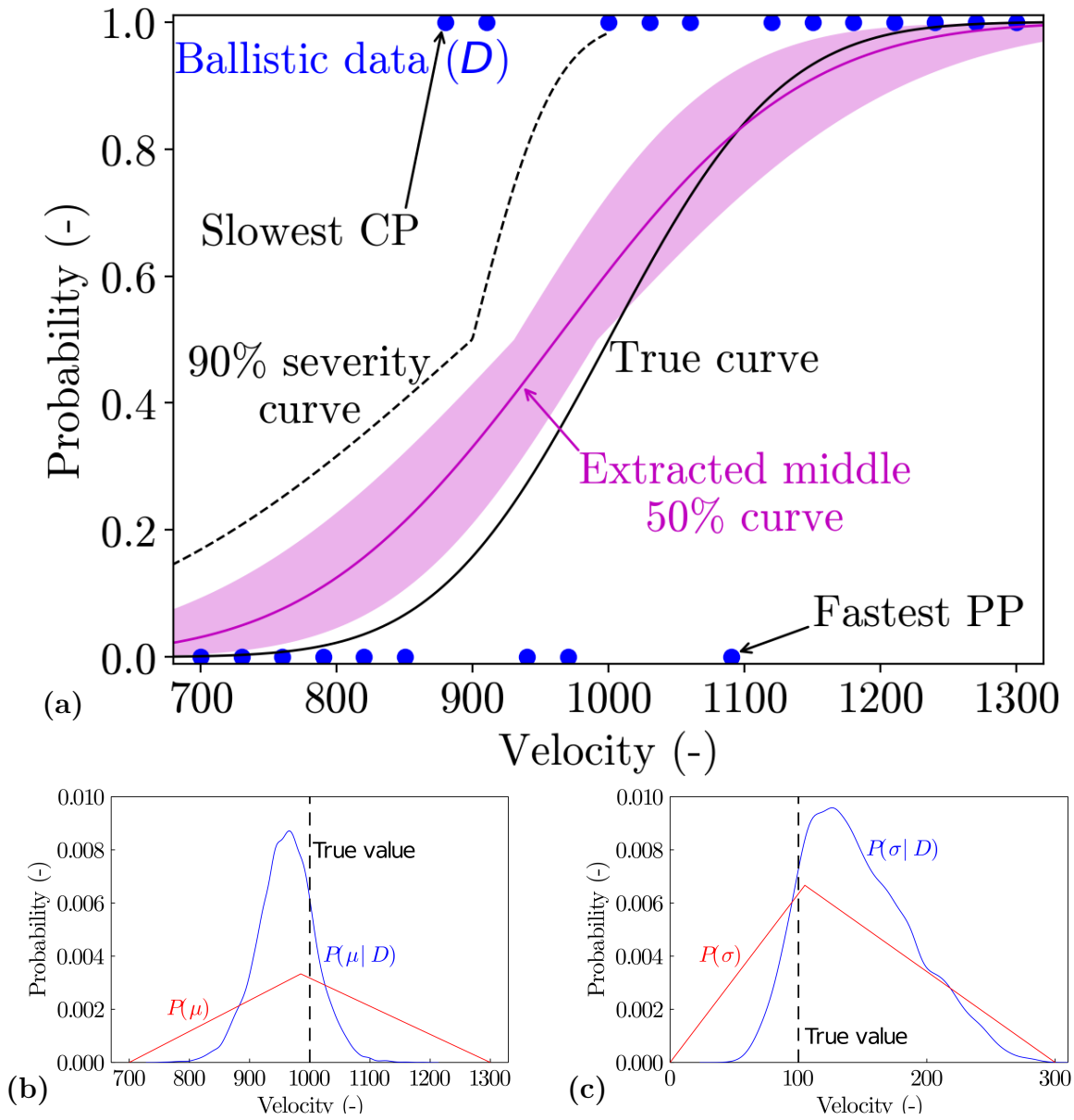

Figure 1: **(a)** Example numerical ballistic data and the prior and posterier distributions for **(b)** μ and **(c)** σ. In (a), the true curve from which the ballistic data is generated may be seen in black, while the extracted middle 50% curve is also plotted. The reader can see how the distributions significantly change with the ballistic data provided, although there does not appear to be enough data yet for convergence to the true parameter values, noted by the dashed lines in (b) and (c). Recall that the zone of mixed results (ZMR), if one exists, is the mixed-result data located between the slowest complete penetration (CP) and the fastest partial penetration (PP). Here, the ZMR is 880–1090, but that could change for a different random series. The 90% severity curve is what would be used to estimate the proof velocity for the 90-90 criterion of an armor layup [8], and, if it were to be calculated with the data available here, armor qualification would be reasonably expected to be met if the threat's velocity is ≤ 632. The specific tabular data generated for this example can be found in Table 1, while example extracted quantiles of the μ and σ posteriors can be found in Table 2.

statistical inference.

Hamiltonian Monte Carlo (HMC) [41] was the numerical method used for these calculations, with $\tau = 10^4$, $\varepsilon = 0.05$, and 10^4 iterations. Changing the solver did not appear to cause significant output differences compared to the values obtained via HMC, and therefore the solution was considered converged. During inference, half the data was randomly selected for training data and the other half for test data. This was likely sub-optimal for the larger datasets, but it was desired to be consistent throughout all calculations.

Two specific priors are considered here. The first is a uniform prior while the second is a triangular prior. In order to calculate the maximum likelihood estimate, a uniform/flat prior is assumed. The triangular prior is considered a physics-informed prior for the two parameters of interest. For μ, this means that the uniform prior was a flat distribution across the entire range of the data available (from the slowest PP to the fastest CP), while the triangular prior has the same range but a mode centered halfway between the fastest PP and slowest CP. For σ, the uniform prior entails a flat distribution from 0 to half the range of the velocities, while the triangular prior has the same range but with a mode at half the range between the fastest PP and slowest CP.[2] A plot with all the associated values for an example set of data using triangular priors, along with example data, may be found in Figure 1. The specific values used in this example are given in Table 1. Explanation of the 90% severity curve will take place later in this document.

The modified V_{50} approach is as follows: shots are taken above and below the estimated V_{50} value until "enough" PPs and CPs are present. The precise amount of data is subjective, and the exploratory calculations below intend to provide some guidelines. As with traditional V_{50} testing, it is convenient (although not required) if a zone of mixed results can be identified, but enough shots should be taken such that one is reasonably confident that data is present outside of the ZMR. These can be seen visually in Figure 1.

As can be seen in Figure 1a, the middle 50% curve does a fair job encompassing the "true" probit curve, particularly when upper and lower bounds include the entire variation of the parameters from 25%–75%. Specific quantile values are given in Table 2, which is a standard `Turing.jl` inference output (other values will be considered below). Of course, the velocity

[2]For these priors to be make sense, a wide enough range of data must be acquired such that one is reasonably confident to be outside of the ZMR. It would be unreasonable to use this assumption for a standard V_{50} test, *i.e.* 6–10 tightly grouped shots around a single velocity, because there is likely not enough information available to accurately determine σ in Equation (4). Similarly, we would argue it is unreasonable to assume μ and σ have an equal probability for all values across the aforementioned ranges given a wide enough set of ballistic data.

Velocity (-)	Result
700	0
730	0
760	0
790	0
820	0
850	0
880	1
910	1
940	0
970	0
1000	1
1030	1
1060	1
1090	0
1120	1
1150	1
1180	1
1210	1
1240	1
1270	1
1300	1

Table 1: Tabular data generated using the prescribed normal distribution ($\mu = 1000$, $\sigma = 100$) and plotted in Figure 1. Recall that partial penetrations are specified as a 0, while complete penetrations are documented as a 1.

Parameter/Quantile	2.5%	25%	50%	75%	97.5%
μ	862.9	929.9	961.4	991.2	1054.7
σ	78.6	114.1	140.5	175.0	240.3

Table 2: Quantile outputs, rounded to 1 decimal point, for μ and σ generated by the Bayesian inference from generated data in Table 1. This is standard quantile output from `Turing.jl`, although other quantiles may be extracted as well (as demonstrated below). The middle 50% data here (*i.e.* 25%–75%) was used to generate the shaded region around median curve in Figure 1.

data presented here was chosen so that the values would be easy to work with. More data obviously could have been generated to increase confidence, but the principles would not have changed.

As an exploratory measure, this work will consider what can be done to more accurately extract μ and σ, as these parameters, together, will be needed to characterize an armor solution. The following numerical examples will include Monte Carlo calculations that range from 13–500 data points spread across a wide range of velocities. Note that the (unit-less) values for μ (mean/median) and σ (standard deviation) are 1000 and 100 unless otherwise stated. Additionally, data will be extracted uniformly-placed data points in the range of $\mu \pm 3\sigma$ unless stated otherwise. Later in this exploratory work we will consider the effect of changing the size of σ *while keeping the data sample space constant*. It will be noted that the qualitative findings did not significantly change when additional calculations were done using other point distributions (*e.g.* the integration points of chebyshev polynomials), even though those results are not included here.

The first thing to consider is how many data points are necessary to accurately extract the parameters, and this is plotted in Figure 2. As mentioned above, both uniform and triangular priors were used for data extraction. When less data is present, the physics-informed triangular prior does a slightly better job at extracting the true value: here, defined as within 5% of the specified μ or within 10% of the specified σ. As seen in Figure 2b and elsewhere below, σ is a nuisance parameter that is challenging to extract unless a lot of data is present. Not surprising, when enough data is present for the full Bayesian inference, the true values are converged to regardless of the starting assumptions.

Another aspect to point out is that the two point-approximations considered here, the uniform-prior MLE and the triangle-prior *maximum a posteriori* (MAP), both performed worse than the median extracted value. For those with detailed knowledge of statistical inference, this result was expected as the minimizer to the expected loss function is the posterior median [40]. However, many scientists have moved away from a thorough study of statistics [26], particularly when only limited data are available [42], and therefore it was worthwhile to reinforce that the 50% value of the posterior distribution is the appropriate value to use for computing credible interval curves and *not* the MLE or MAP.

Data was fit to curves to assist the reader in observing the trends of the data. The shaded region provides an estimate of the overall uncertainty of the curves. Repeat calculations using different random number seeds, not included here, observed the expected trend that less variation in the random variable distributions was typically seen when a greater number

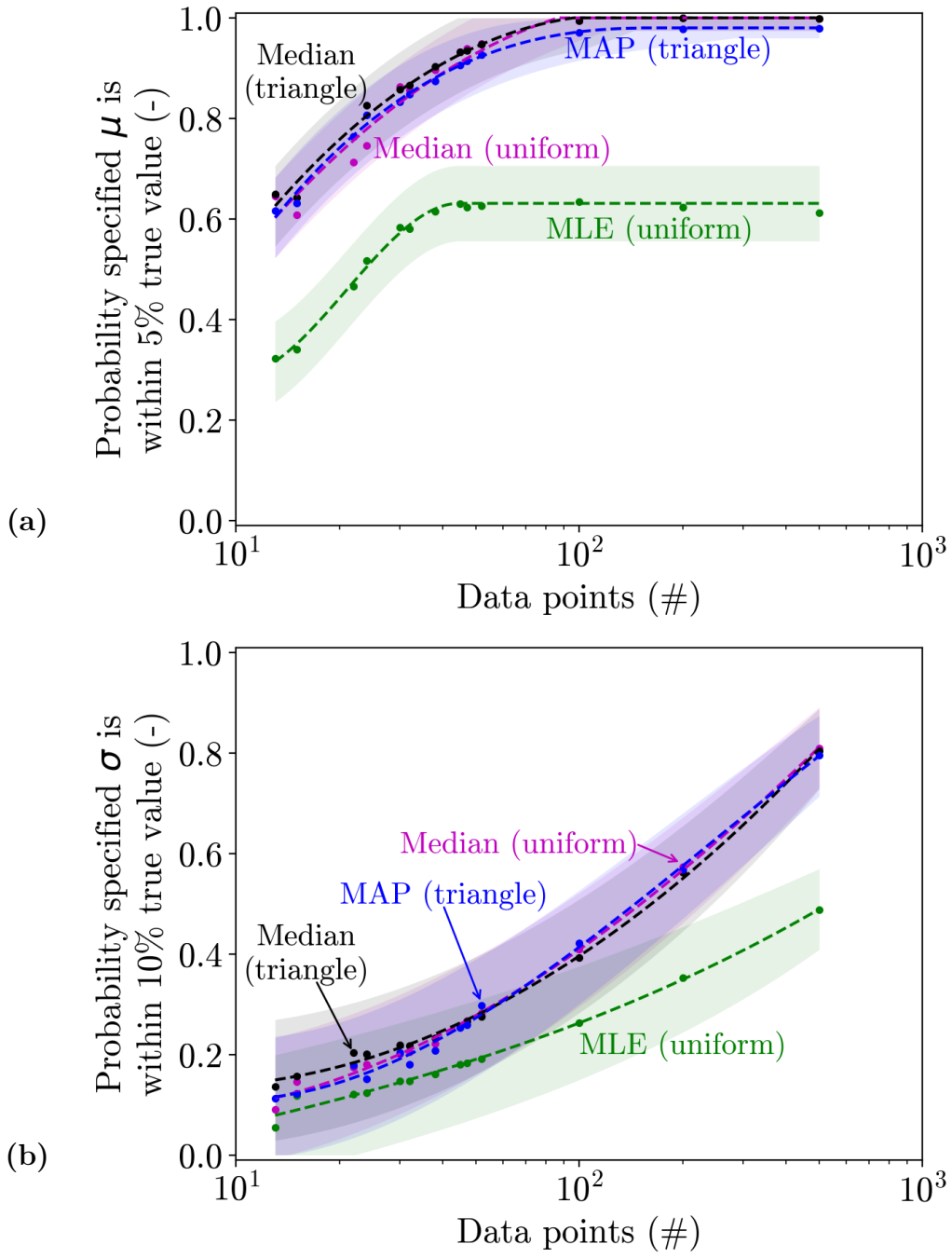

Figure 2: Convergence of the extracted parameters **(a)** μ and **(b)** σ with increasing number of data points. The data was uniformly distributed across an interval with range $\mu \pm 3\,\sigma$. It is obvious that μ is easier to extract accurately compared to σ. Furthermore, it is observed that the median (50%) value of the posterior provides the most accurate parameter estimate compared to the MLE or MAP, although this is slightly more clear for the μ data compared to the σ data. The mean parameter value closely followed the median value, and therefore was left off of these plots. Intuitively, it can be observed that, for a large amount of data, the prior distribution used becomes negligible because the vast amount of data becomes more important compared to the starting assumptions. Other sampling distributions (*e.g.* raised cosine or nodes of legendre polynomials) provided a similar trend for convergence.

of data points available for inference.

A word ought to be said about the point estimators in Figure 2. Given the posterior, the optimization package `Optim.jl` [43] calculated either the MLE or MAP. Unfortunately, convergence to a point estimate was not always achieved, which, perhaps, may have been due to a relatively flat posterior. In one instance, $\sim 1.5\%$ of the MAP values failed to converge while $\sim 27\%$ of the MLE values failed to converge. It was beyond the scope of this work to diagnose why these convergences failed, and so these instances were removed from consideration.

Depending on the desired accuracy of the parameter extraction, a more or less stringent accuracy may need to be specified. In Figure 3, we consider the parameter convergence curves when the median μ is within 2-10% of the true value and when the median σ is between 5-20%. Triangular priors were assumed for these calculations, and the anticipated result is found: the wider the range, the more likely it is that the true value is found within the inference. Similar trends were observed when using a uniform prior.

Given the asymptotic convergence of the probit curve's parameters, we wondered whether or not it was more beneficial to conduct "repeat shots" of data or if it was better to have them equally distributed across the interval. Obviously this is an approximation, because in real ballistic testing it is unlikely to conduct tests and get two shots in a series with the *exact* same velocity, much less all of them required to be repeat-velocity shots. Nonetheless, this was considered a useful investigation because this can be approximated in practice. 24 data points were chosen for this calculation in order to easily divide the total number, and to allow multiple overlap locations with every unique velocity repeated the same number of times. These results are presented in Figure 4.

Looking at the repeat-shot data presented in Figure 4, there does not appear to be any significant correlation between the unique velocity points and the probability of ascertaining the correct value for μ or σ. Similar results (*i.e.* no correlation between parameter extraction and unique velocity points) were found when considering other possible shot distributions using the same data range. The main observation from the other distribution data not included here was that there was no discernible pattern that determined which unique velocity point might be slightly higher or lower than the others. Therefore, there does not appear to be any benefit for completing uniformly-spaced "repeat" shots instead of simply uniformly-spaced unique velocity shots.

Next, let us consider the case where σ is varied while the data inference range is held constant. Since the standard deviation is extracted so infrequently, the significance of this nuisance

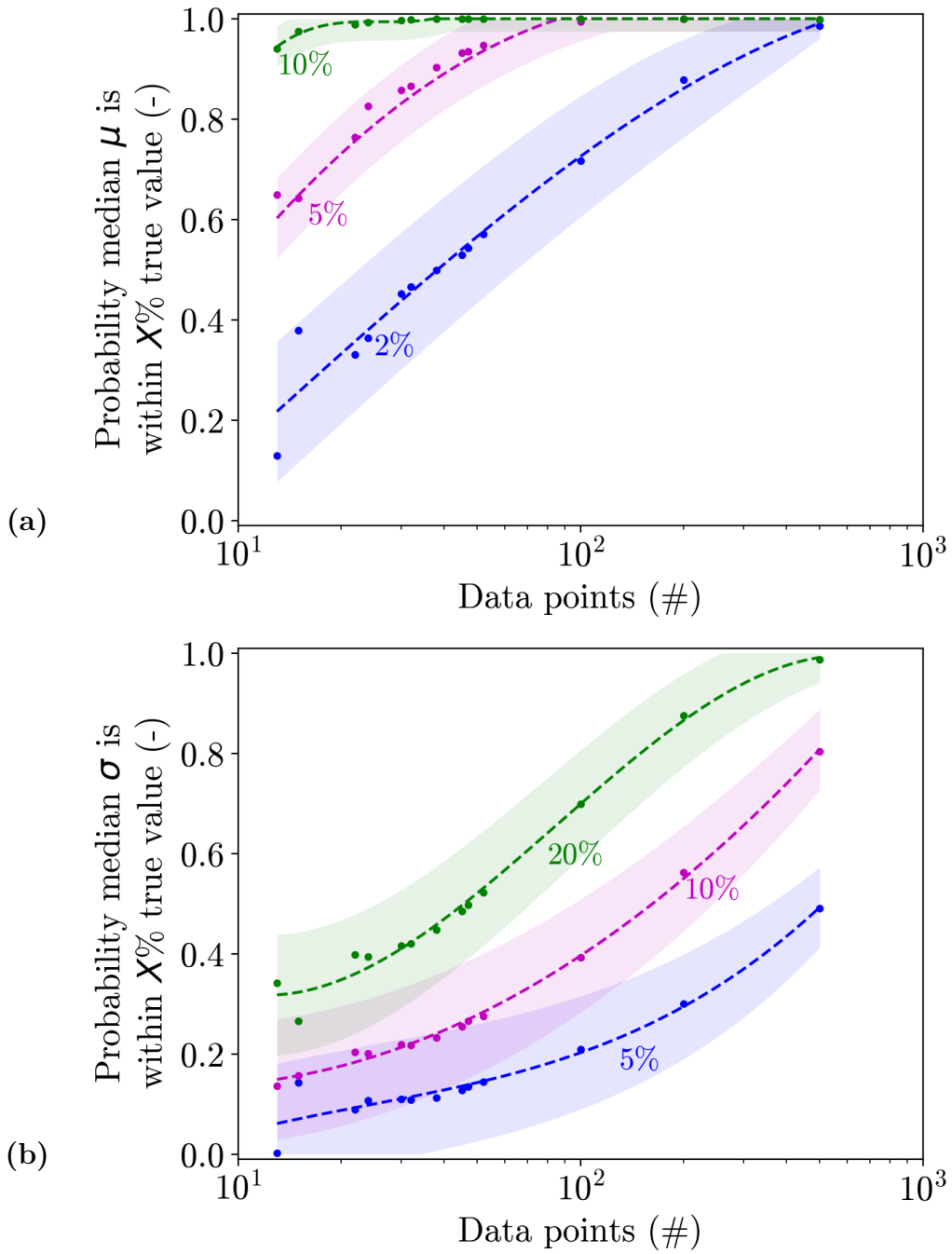

Figure 3: Convergence of the median (50%) extracted parameters **(a)** μ and **(b)** σ with increasing number of data points when widening the tolerance of the extracted parameter. The data was uniformly distributed across an interval with range $\mu \pm 3\,\sigma$, and a physics-informed triangular prior was utilized. As expected, the more stringent the requirement for the extracted 50% value be to the true value, the more data points are required to approach 100% likelihood of extracting the "true" value. Similar trends were seen when using a uniform prior.

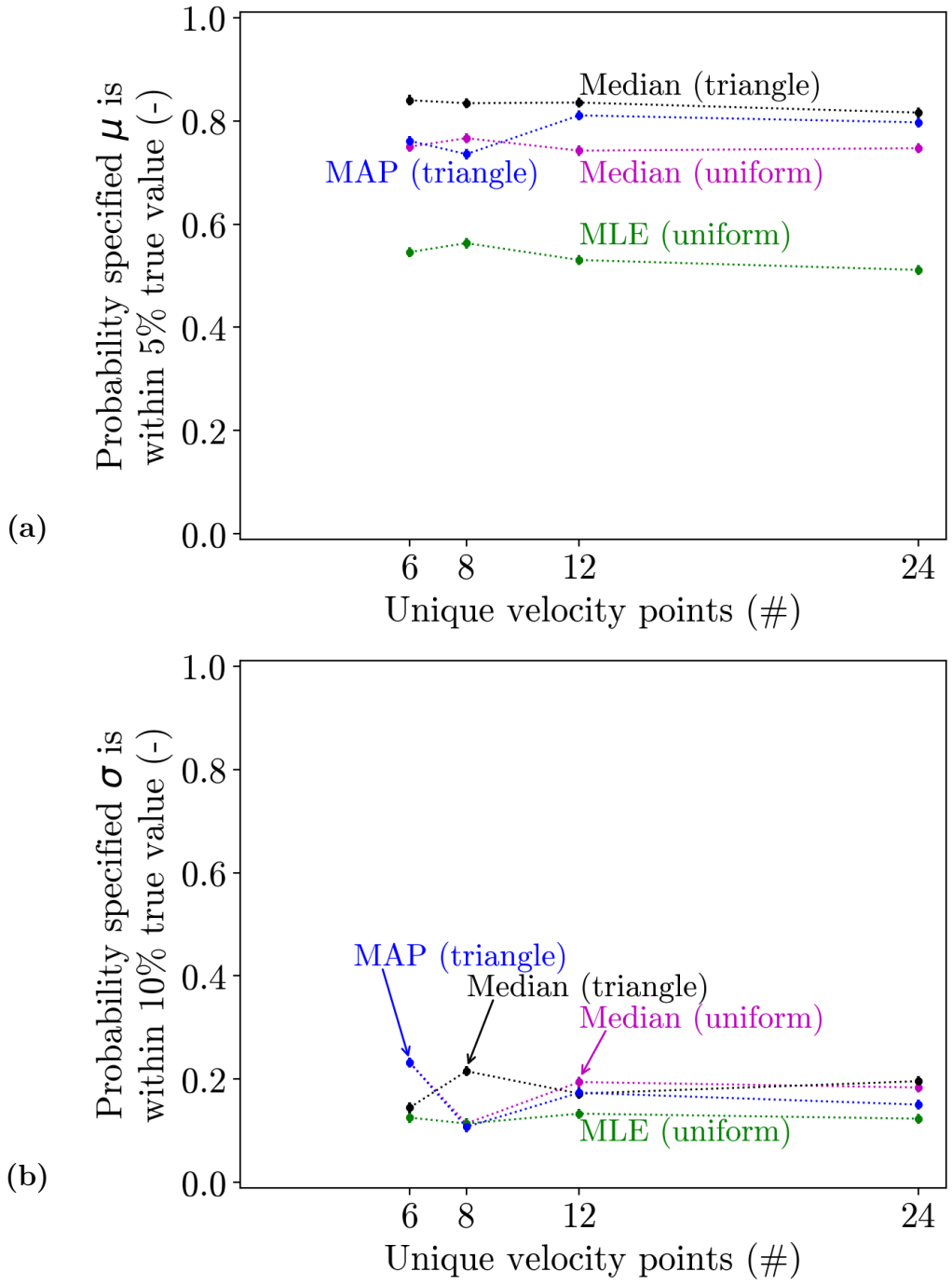

Figure 4: Effect of the presence of unique velocity points (but constant overall data points) on the extracted parameters **(a)** μ and **(b)** σ. When considering the data above, and data from other distributions not included here, there does not appear to be any correlation between the number of unique velocity data points with increased confidence in parameter extraction (when keeping the total number of data points constant).

parameter needs to be explored if the modified V_{50} method for ballistic characterization is to become common. The results from calculations using a uniform prior can be found in Figure 5. Note that $\sigma_D \equiv (\max(D) - \min(D))/2\sigma$: that is, the number of standard deviations of data present above or below the median.

With respect to μ, there is clearly a benefit if σ is small, because it is much more accurately extracted. This makes sense intuitively, because if there is a sharp delineation between PPs and CPs (*i.e.* no ZMR) then μ is probably about halfway between the fastest partial penetration and slowest complete penetration. For varying the standard deviation, the σ results were slightly surprising. Despite the differences in σ_D, essentially all the data lie right on top of each other. Therefore, uniform spacing alone may not be the best solution for accurately extracting σ.

Shifting of the true mean from the center of acquired data is an example of what might occur when the true parameters are unknown. When testing a new material or armor layup, one may be able to estimate the V_{50}, but the true result could also be far off. Therefore, it is important to know how significant the effect on parameter inference estimates when the more data is present significantly above, or below, than the true μ. These results are presented in Figure 6. From the data observed, it appears as though there is not a significant difference between the data that was centered around the V_{50} and the data that was off-center. The small change may be due, in part, to the assumption of the symmetric normal distribution, which is perhaps why there is not a significant loss in inference capability.

To some extent, all of the data and analyses presented up to this point have been prelude. For armor design, what is desired is a confidence in some level of protection, one common criterion being the 90-90 criterion [8] mentioned in the introduction: 90% chance of threat defeat (10% of threat perforation) at the 90% confidence level. Estimating the proof velocity, *i.e.* the maximum equivalent frequentist velocity, can be done using the upper/lower bounds of the extracted μ and σ values and moving towards the appropriate extrema.

As a starting point, let us return to the example data from Figure 1. Looking at the data that was generated and assumed priors, we can base the 90% (credible) severity curves on how each parameter should be treated in a worst-case terminal ballistics event. In the case of μ, the most severe case is obviously the lowest V_{50} value, and so the 10% and 100% values bound the μ variation. On the other hand, a larger standard deviation is more severe because it widens the ZMR, so the 0% and 90% values will bound the σ variation. If these overlapping regions were to be plotted, it would look similar to the middle 50% region seen in Figure 1a, except the region would be wider. The upper bound of this region is labeled

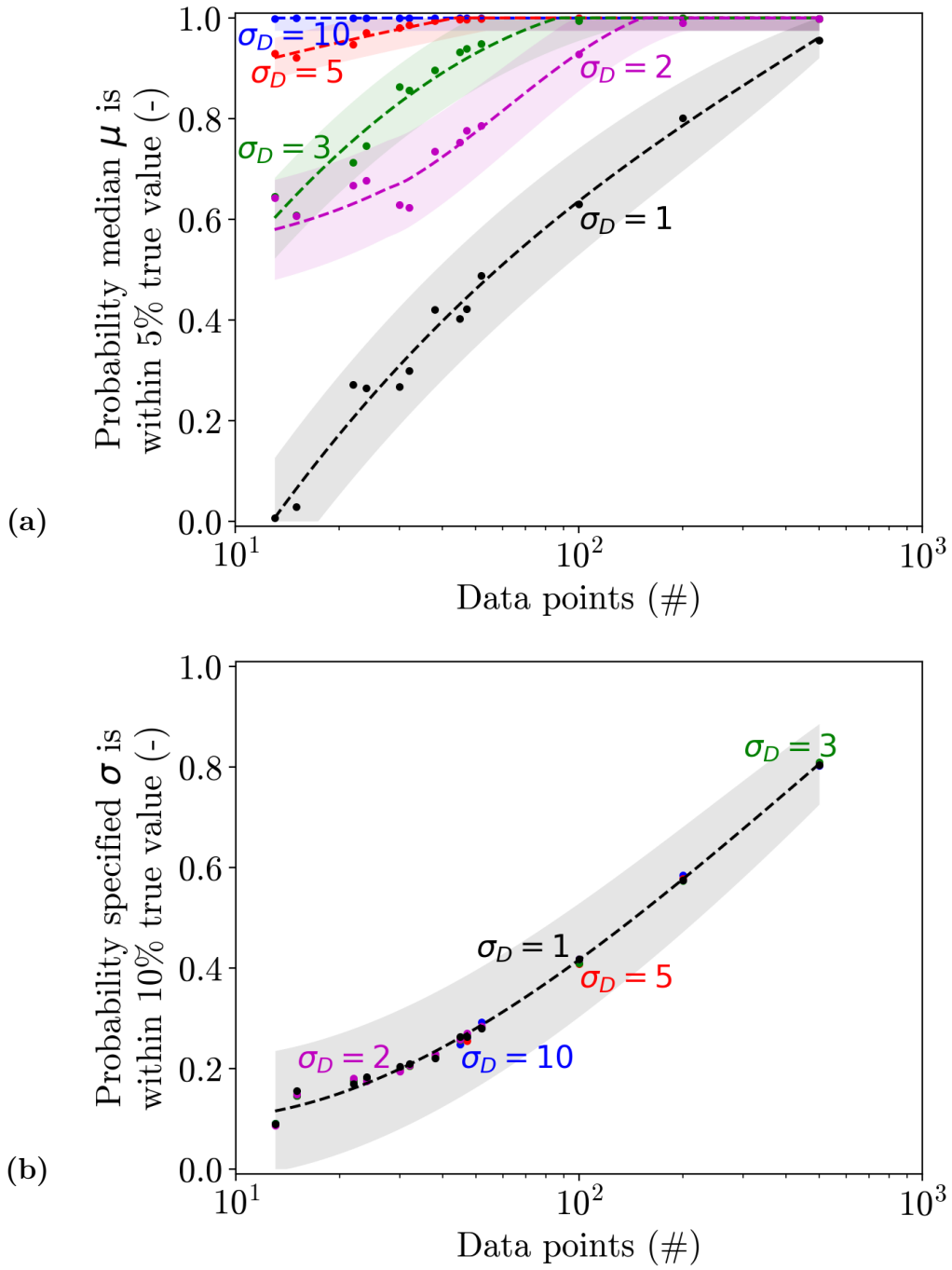

Figure 5: Effect of varying σ on the extracted parameters **(a)** μ and **(b)** σ while keeping the sampling range constant. Note that σ_D is the number of standard deviations of one-half the data range. Looking at the data plotted, the smaller the standard deviation relative to the size of the data, the easier it is to extract μ accurately. This seems reasonable since there will be a small ZMR (*i.e.* sharp transition from partial penetrations to complete penetrations). On the other hand, the relative value of σ appears to have little effect correctly extracting its value, as almost all of the data lies right on top of each other. Note that since the data in (b) is so tightly grouped, only one curve fit and uncertainty estimate is given. This suggests that more data is necessary in the ZMR in order to get greater confidence in σ, and uniform spacing may not be the best approach. Similar trends were seen when using a triangular prior.

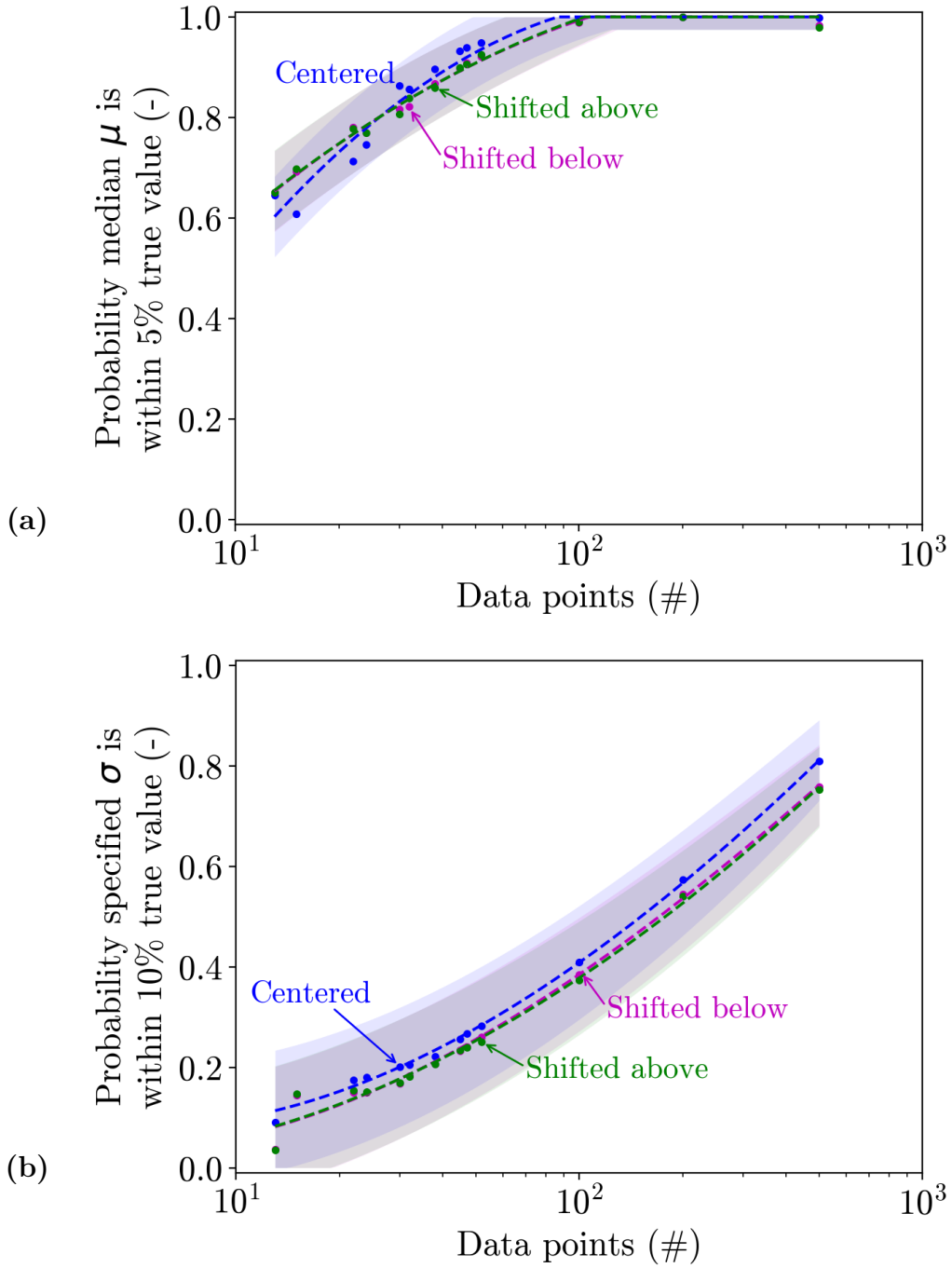

Figure 6: Effect of an off-center μ on the extracted parameters **(a)** μ and **(b)** σ. As can be seen from the data above, there is not much difference between the inference curves between the centered-μ inference and when μ is shifted above/below by one standard deviation. Note that in all cases, there was the normal 6-σ range of uniformly-spaced data: this range was just shifted up or down one standard deviation for the off-center inference. A similar trend was observed when using a triangular prior.

the 90% severity curve and is denoted with the dashed line in Figure 1a. If one were to follow the dashed line out further out to the left, the V_{10} value of this curve would be 632. This is the estimated proof velocity, which is the highest value the design is expected to pass qualification testing (here, assumed to be the 90-90 criterion) because there is the 90% "confidence" (read: credible interval in the Bayesian sense) in 90% threat defeat (or 10% chance of threat perforation).

While discussion could end there, we were curious how the rest of the data would look if it was normalized. What is the distribution of the proof velocity for the 90-90 criterion? That is, what is the estimated equivalency between the classic "22 out of 22" qualification testing and the modified V_{50} test using Bayesian inference based solely on μ and σ?

To answer this question, let us take all of the posterior data from Figure 6 and perform the same extraction that was completed above. However, this time we will convert the velocities to the standard normal: $\bar{v} = (v - \mu)/\sigma$, with $\bar{v}$ as the new normalized velocity. From there, we get the 90-90 "random variable" from the function of two other random variables (*i.e.* μ, σ). The same normalization can be done using data from Figure 5. Selected velocity distributions are plotted in Figure 7 for (a) $\sigma_D = 3$ and (b) $\sigma_D = 10$.

The 500 data point calculations utilized the upper limit of the inference data analyzed here since this is a number far greater than what would likely be used to characterize a single armor solution. In the limit where μ and σ are known to only an infinitesimal uncertainty, the 90-90 severity curve will collapse onto the "true" normal cumulative distribution function. Using the standard normal distribution of $N(x; \mu = 0, \sigma^2 = 1)$, we can set Equation (4) equal to 0.1 and solve to see that $\bar{v} \approx -1.28155$. This convergence can be seen as the number of data points increased, although this limit is more readily seen in Figure 8a when more data is present. In Figure 8b, 80-80, 90-90, and 95-95 design criteria are plotted using the same method. It can be calculated that the normalized velocities in the instance of infinite data for the 80-80 and 95-95 criteria are $\bar{v} \approx -0.84162$ and $\bar{v} \approx -1.64485$, respectively. It should be noted that the 95% credible interval uncertainty bars in Figure 8 are uneven to reflect the asymmetric nature of the random variables (some of which are plotted in Figure 7).

Since 52 shots is typically the most that will be taken for a traditional 90-90 qualification test, this was chosen as the upper limit of the plotted data in Figure 7, minus of the 500-point data to check the limiting case. The median $\bar{v}$ values lie are -3.505 and -2.351 (rounded to 3 decimal places) for 13 and 52 data points, respectively, when $\sigma_D = 3$, which correspond to probabilities of threat perforation of 0.00023 and 0.0094 using our assumption of normal distribution. These values are shifted up or down if the design criterion is changed to

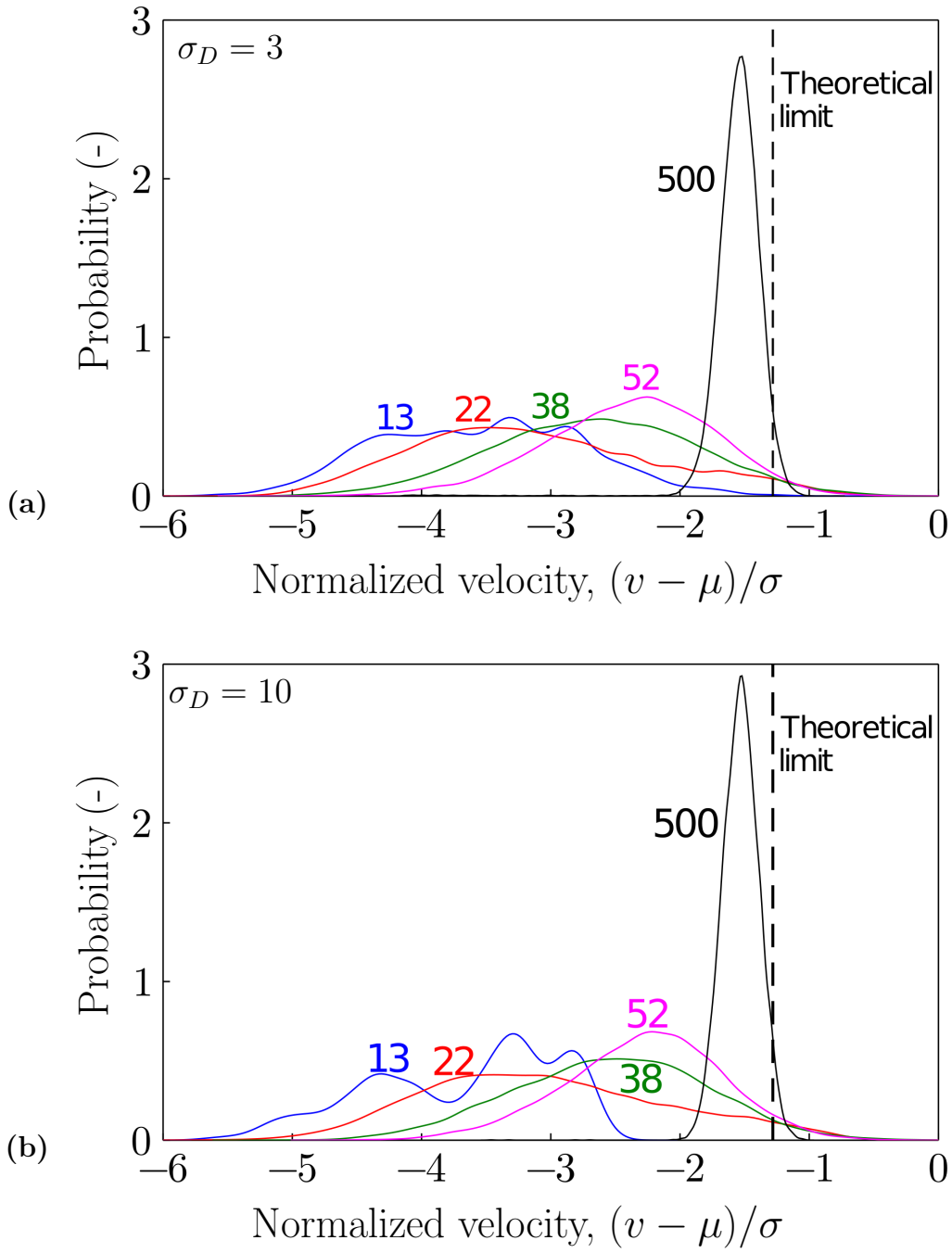

Figure 7: Plots of the normalized 90-90 velocities for **(a)** $\sigma_D = 3$ and **(b)** $\sigma_D = 10$. (a) was generated using all the data from Figure 6, while (b) was generated using the data from $\sigma_D = 10$ in Figure 5. All of the associated posterior distributions of μ and σ were used to calculate the 90-90 criterion, and then this was normalized. The distributions are not significantly different from each other, even if the median μ is more accurately known for $\sigma_D = 10$. However, with such few data points, there likely is not enough information available to sharpen the distribution's peak. In the case where μ and σ are known to within an infinitesimal uncertainty, the normalized velocity is approximately -1.28155. The 95% credible intervals of the normalized 90-90 velocities are plotted in Figure 8a.

95-95 or 80-80. Strictly speaking, this proof velocity itself is a random variable, and the distribution can be populated by running the Monte Carlo calculation many times. However, repeat calculations using different random number seeds did not significantly change the proof velocities for the different design criteria, and the proof velocity was changed more by different selections of randomly-generated datasets.

The data and analysis leading up to Figures 7 and 8 demonstrate how this method could be used to estimate the proof velocity of an armor solution. Given a set of ballistic test data (velocities and PPs/CPs), a calculation can be run to estimate the equivalent of the design criterion (*e.g.* 90-90). If this proof velocity is greater than or equal to a threat's specified velocity for armor qualification, then the solution has a reasonable chance of meeting the design criterion. Obviously, this method would be more useful for new designs against a threat when little historical data available. Since there is no limit to the number of data points that can be collected using the modified V_{50} test method (in an attempt to increase the estimated velocity), a good rule of thumb might be to stop testing after 52 fair shots, perhaps with a higher concentration of shots in/around the zone of mixed results.

Intuitively, one would expect that there should be less certainty in a design that underwent only 13 shots of testing compared to 22 shots, and this is clearly observed in Figure 8. And while this can be done and a proof velocity determined, this runs the risk of a determining an overly-conservative proof velocity when additional mass could still be removed while meeting the design criterion. Indeed, it is observed in Figure 8 that most of the time the proof velocity will be *underestimated* given our assumptions, and this ought to be taken into consideration when analyzing experimental data.

A couple comments should be made concerning the use of point estimates (*e.g.* MLE [17] or MAP) when calculating ballistic protection. As was stated above, in most instances the confidence/credible intervals overlap, and, in the case of classical inference, this would mean the MLE would be fine. However, since we are using Bayesian inference, and this appears to be the first time using this method for (modified) V_{50} ballistic testing, we are throwing away the opportunity to provide a physics-informed prior and gain more confidence in our results, since MLE assumes an unreasonable uniform prior distribution when a large range of velocities is present.

The other concern when using a point estimate is how to specify a credible interval. In frequentist statistics, this can be done using various starting assumptions [8], but this method involves repeat shots at the same velocity. If a Bayesian approach is used, then a credible interval cannot be applied if only a point estimate is given. Additionally, since only one set

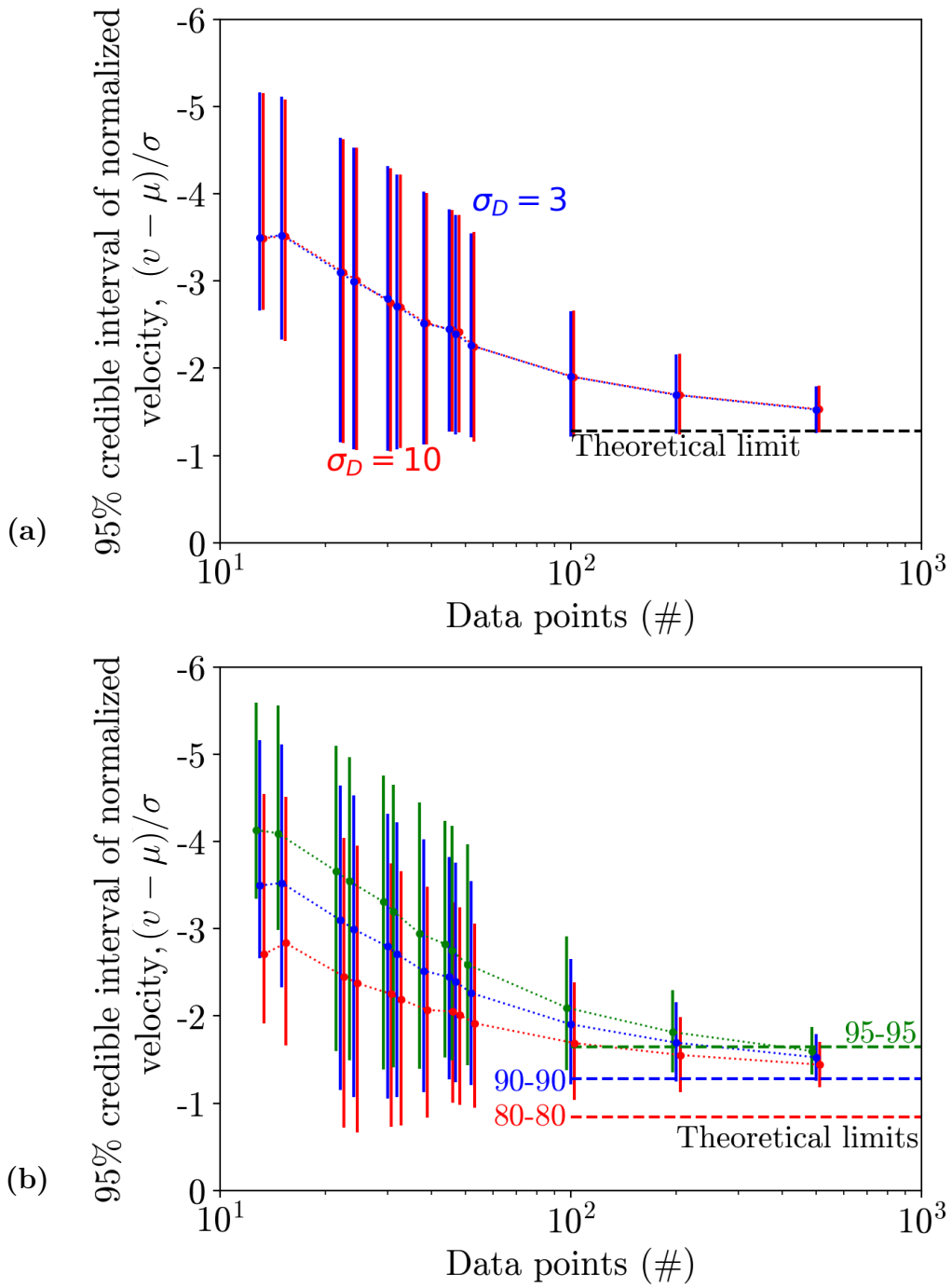

Figure 8: The 95% credible intervals of **(a)** the 90-90 criterion from data plotted in Figure 7 and **(b)** the 80-80, 90-90, and 95-95 criteria plotted using the $\sigma_d = 3$ data. Note that data is included here that could not be included in Figure 7 due to space limitations. While all data were extracted using the same number of data points, some of the sets are offset by a small percentage so the uncertainties could be readily compared. As the reader can see in (a), there is not much difference between the data, and both sets of data approach the theoretical limit of $\bar{v} \approx -1.28155$. For the corresponding 80-80 and 95-95 design criteria, the asymptotic limits are $\bar{v} \approx -0.84162$ and $\bar{v} \approx -1.64485$, respectively. In (b), it is observed that the 80-80 and 95-95 criteria are approached in the limit as well, although at different rates.

of data at a time would be analyzed in practice (although the set would perhaps be updated with more shots), it is not expected that the additional time required to perform the full Bayesian inference would significantly slow down ballistic testing.

The use of the physics-informed triangular prior, as seen in Figure 2, can provide a slight improvement of extracting the correct value for μ (V_{50}) and σ compared to the uniform prior for small amounts of data. This seems reasonable, because, given enough data, the prior distribution becomes irrelevant and the posterior distribution is swamped by the information provided by the data. Unfortunately, the use of the triangular prior requires approximately 2.5x the computational time of the uniform distribution. For running single calculations with a slowly-expanding list of shots, this is not expected to be much of an issue. The gain for experimental analysis may be worth the small additional cost in time, particularly because "small" sets of data will likely be the only ones available. However, for exploratory calculations like the ones presented here, this limited the calculations that could be completed in a reasonable time-frame. Additional commentary on the computational time of this analysis will be provided when experimental data is analyzed in Section 4.

A comment will be made concerning changes in protection requirements after testing is complete. In the event a design criterion wants to be changed from 90-90 to, say, 95-95, no new data needs to be collected. The μ and σ posteriors can be re-evaluated to generate a new proof velocity for the updated design criterion. This was done in Figure 8b.

One final suggestion will be given concerning the amount of data present for characterizing plates/armor solutions. The analyses above, with the exception of Figure 5, always had 6-σ of data present. Due the uncertainty of σ for materials and layups, it would probably be better to use the value as a lower bound for data acquisition, and it might be better to try and acquire data that spans approximately 7-σ or 8-σ. Having a large enough range ensures that the bounds instilled by the priors are reasonable and μ and σ.

4 Experimental Investigation

Having observed some trends from the numerical examples, let us consider some experimental data. Unfortunately, many of the specifics of ballistic testing must be redacted due to the sensitive nature of the test results. Both the KE threat and monolithic target material are relevant to defense applications, and therefore they are not specified. The velocities given below have been normalized[3] by an experimentally-verified 90-90 proof velocity. It is

[3]The velocities were normalized to 1, and not 1000, to preclude potential confusion between this value and the dimensionless velocity of 1000 utilized in the numerical examples in Section 3.

emphasized that this proof velocity is *not* to be considered an upper bound, but a known datum to be used for comparison. It should be noted that these two sets of ballistic testing (*i.e.* 90-90 and the modified V_{50} testing) were conducted independently of each other, but the same target coupon was used to generate both data sets. The target was large enough to accommodate many shots on target but coupon size and shot spacing are withheld. There did not appear to be any interaction between shots or the edges of the target. Testing was conducted at normal obliquity.

Experimental testing requires more nuance than the numerical results presented above. From a practical perspective, multiple barrels and cartridge sizes may need to be used achieve the velocity range required to characterize a plate or layup, as was the case here. It is also harder to target a specific velocity, because shot-to-shot variation occurs even with the same measured propellant mass and nominally identical projectiles. Furthermore, all data points should be "fair", with a root-mean-square (rms) pitch/yaw measurement that is less than 5° [9][4], since highly yawed rounds typically penetrate far less than aligned threats. A 0.51 mm (0.020") thick aluminum sheet was used as a witness plate to define a PP or CP [9].

The experimental shot data is plotted in Figure 9, and the tabular data may be found in Table 3. It can be seen in the tabular data that the pitch and yaw were well within the aforementioned guideline. A total of 51 shots, 23 PPs and 28 CPs, were taken on target to maximize the amount of data generated while adhering to the shot spacing and edge separation requirements. Exploratory shots were first taken to gauge the span of the zone of mixed results before more shots were taken within the ZMR in order to increase the confidence of σ. The ballistic data is not as evenly distributed within the ZMR as one might prefer, but this is to be expected from an experiment.

Using the present case as a representative instance, the analysis needed to generate the μ and σ posteriors using all the data (Table 3) usually took less than 90 seconds on a laptop computer. Considering the fact that time between shots in ballistic testing can easily take 2 minutes to declare the range safe, look at the target, adjust the shotline, and reload, this computational time will likely not measurably slow down testing. Note that even more time between shots may be required if pitch and yaw measurements are taken or if the KE threats utilized require a screw-on breach, which suggests that this method could be readily introduced as an updated armor plate characterization method even if the computation would need to be re-run after every shot (although this would likely not be necessary).

[4]It is noted that this guideline was only used as an example from a testing standard. Better results may be found if a smaller value of maximum rms total yaw of 3° was used so there is less shot-to-shot variability.

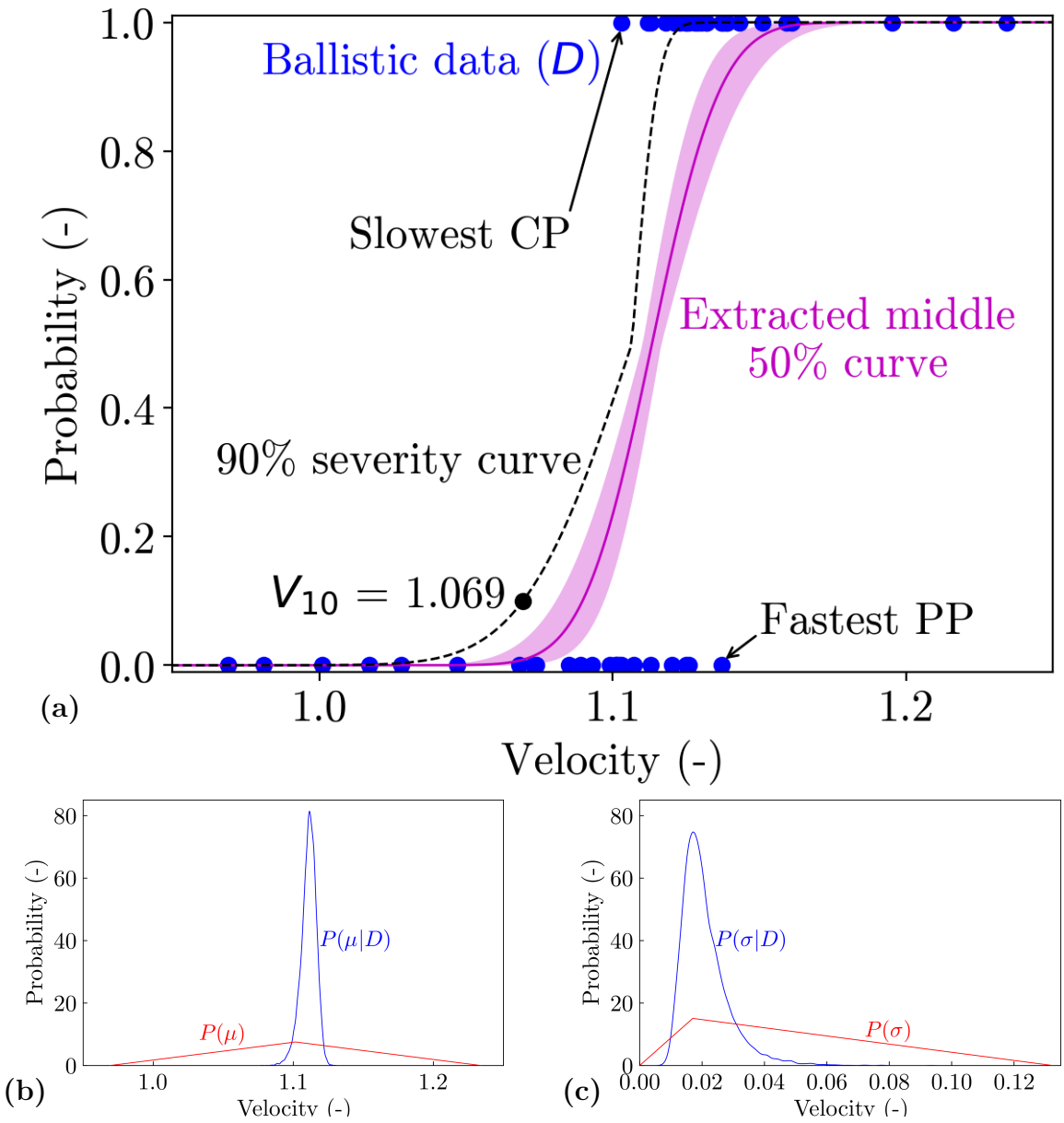

Figure 9: **(a)** Experimental ballistic data and the prior and posterier distributions for **(b)** μ and **(c)** σ. In (a), the extracted middle 50% curve is plotted. Note that all velocities have been normalized by an experimentally-verified 90-90 proof velocity. The zone of mixed results is 1.103–1.137. The experimental data presented provides a proof velocity of ≤ 1.069, or 6.9% above the verified value. This suggests that the armor thickness would likely be able to decreased slightly and still maintain the desired threat defeat capability. The specific tabular data generated for this example can be found in Table 3, while example extracted quantiles of the μ and σ posteriors can be found in Table 4.

Shot	Velocity (-)	Pitch (°)	Yaw (°)	Result	Shot	Velocity (-)	Pitch (°)	Yaw (°)	Result
22284-3	0.969	0.2	0.4	0	22284-34	1.121	0.1	0.2	1
22284-4	0.981	0.6	0.7	0	22285-9	1.123	0.5	0.6	1
22284-5	1.001	0.4	0.7	0	22284-27	1.125	0.6	0.8	0
22284-6	1.017	0.3	0.7	0	22284-23	1.125	0.3	0.7	1
22284-7	1.028	0.5	0.6	0	22284-31	1.125	0.1	0.3	1
22284-8	1.047	0.6	0.7	0	22284-26	1.126	0.5	0.8	0
22284-39	1.068	0.8	0.6	0	22284-47	1.126	0.5	0.4	1
22284-9	1.073	0.2	0.8	0	22284-32	1.128	0.3	0.7	1
22284-12	1.074	0.7	0.6	0	22285-6	1.129	0.3	0.2	1
22284-49	1.085	0.9	0.6	0	22285-4	1.130	0.3	0.6	1
22284-24	1.089	0.4	1.0	0	22284-43	1.132	1.0	0.7	1
22284-37	1.093	0.3	0.7	0	22284-45	1.132	1.1	0.4	1
22284-38	1.099	0.7	0.6	0	22284-14	1.137	0.4	0.6	0
22284-25	1.101	0.9	0.7	0	22284-46	1.137	0.3	0.6	1
22285-3	1.101	0.7	0.4	0	22284-48	1.137	0.3	0.6	1
22284-41	1.102	0.2	0.4	0	22284-22	1.138	0.1	0.4	1
22284-30	1.103	1.3	0.8	0	22284-40	1.139	0.5	0.9	1
22284-36	1.103	0.2	0.3	1	22285-8	1.139	0.6	0.5	1
22284-29	1.107	0.5	0.1	0	22284-28	1.143	0.8	0.3	1
22285-2	1.112	0.4	0.9	1	22284-44	1.151	0.7	0.3	1
22284-42	1.113	0.7	0.4	0	22284-19	1.159	0.3	0.8	1
22284-35	1.113	0.9	0.7	1	22285-5	1.161	0.5	0.8	1
22285-7	1.118	0.5	0.4	1	22284-15	1.195	0.3	0.3	1
22284-13	1.120	0.1	0.1	0	22284-20	1.216	0.6	0.2	1
22284-18	1.121	1.2	0.7	1	22284-21	1.234	0.2	0.4	1
22284-33	1.121	0.5	0.5	1					

Table 3: Experimental data that has been normalized by an experimentally-determined 90-90 proof velocity. This data is plotted in Figure 9. Velocities that were slightly offset (*e.g.* 10^{-8}) from identical values did not significantly change inference results. Recall that PPs are specified as a 0 and CPs are 1. The extracted quantile data for μ and σ may be found in Table 4.

Parameter/Quantile	**2.5%**	**25%**	**50%**	**75%**	**97.5%**
μ	1.1016	1.1100	1.1133	1.1164	1.1219
σ	0.0108	0.0153	0.0185	0.0231	0.0392

Table 4: Quantile outputs, rounded to 4 decimal points, for μ and σ generated by the Bayesian inference from the data in Table 3. The middle 50% data here (*i.e.* 25%–75%) was used to generate the shaded region around the extracted median curve in Figure 9.

Quantile data for the extracted posteriors of μ and σ may be found in Table 4. As seen in the numerical examples, the confidence in μ is greater than σ, where the 95% credible interval spanned only about 0.02 for the former and almost 0.03 for the latter. That said, the additional data provided by data concentrated in the ZMR appears to have helped increase confidence in σ. The μ and σ distributions sharpened considerably when considering the available data. The σ posterior still has a distinct tail trailing towards higher values. Assuming the median value of σ (0.0185), the ballistic data spanned approximately 14-σ, which is significantly more than what was considered in the numerical examples above.

The estimated 90-90 proof velocity was 1.069, or almost 7% faster than the experimentally-validated value. To some extent this is expected, as armor engineers may utilize a conservative design in order to ensure the solution will defeat the threat. Previous analysis, not shared here, suggested this thickness should be adequate to defeat the threat, but it does not appear to have been verified. This test provides additional evidence that the armor gives more than the required minimum protection against the KE threat. Unfortunately, it should be noted that the weight savings from choosing a thinner plate may not actually be realized if it is not readily manufacturable, since the machining cost may be prohibitively high.

At the end of Section 3, it was suggested that the upper bound of 52 data points should be enough to adequately characterize an armor layup. That number is probably still a good rule of thumb, but one might be able to get with even fewer if the majority of data is concentrated within and around the ZMR. For comparison, using the data from 43 shots calculated a 90-90 proof velocity of 1.065, not that different from the value of 1.069 determined from the full 51 shot dataset.

It is noted that the results above were generated without the foreknowledge that a (normalized) velocity of 1 satisfied the 90-90 criterion. When the bayesian analysis was re-run with this information, the estimated 90-90 proof velocity did not change much[5]. This seems

[5]When thousands of Monte Carlo calculations were run to generate the random variables, the middle 95% credible interval without the 22 out of 22 shots was 1.0673–1.0711, while this interval increased slightly to 1.0684–1.0716 when the 22 shots were included.

reasonable, because we aren't including much additional information with repeat shots at this velocity when we "know" the anticipated outcome. The same could be said if there were consistent shots taken at a velocity where the threat overmatched the target: not much additional information will be gathered. Therefore, this finding reinforces the observation from the numerical calculations that it appears as though it is best to have most of the data concentrated within or around the zone of mixed results.

5 Future Work

One important question that cannot be addressed here is the effect of manufacturing variability of materials. Specifically, do heat treatments and other manufacturing processes vary the elastic modulus, hardness, or toughness enough to cause a measurable variation of the V_{50}? We do not know of any data that closely measured these values. Furthermore, even if careful, systematic measurements were made, it is unknown whether or not the V_{50} measurements were done with enough data points to pick up on the changes, as the standard test method involves a small number of data points [9].

Better prior distributions ought to be realized to assist in using the modified V_{50} test method estimating armor design criteria. Due to the nature of ballistic testing and the knowledge of how the parameters behave (*i.e.* μ and σ can both be bounded by extrema and all values are *not* equally likely), we desired a probability density functions to account for this outside information along with compact support. The triangle distribution was chosen because it fit this requirement and was readily available from `Distributions.jl` [44]. While it was based on the physics of the ballistic events and the data available, both prior distributions for μ and σ probably have tails that drop off too quickly near the initial modal value and too slowly near the upper and lower bounds (*e.g.* Figure 9). A slightly more complicated alternative might be a piecewise-continuous function that concentrates 90-95% of the area in the regions of high probability. In the case of μ, this would be the central $\sim 90\%$ region bound within the ZMR, with the mode at the ZMR midpoint and the rest of the $\sim 10\%$ left for regions outside of the ZMR. For σ, this could mean $\sim 95\%$ of the probability between zero and the spread of the ZMR velocity range, centered at half that range, with the remaining $\sim 5\%$ left for values higher than the ZMR range. This would constrain the distributions to be within physical limitations of the available data, as was done here, but concentrating the distributions within reasonable regions before the application of Bayes' rule.

These results are most relevant for KE threats because the amount of propellant (*e.g.* gunpowder or compressed gas) can be varied relatively easily to change the velocity of the threat.

It is currently being considered how this new test framework might be used in the case of CE threats, which tend to be more expensive than KE threats and take longer to assemble. One possibility is that the repeated test data could be used within the Bayesian framework presented above to increase test confidence in the generated data. An issue would be determining the appropriate metric that will be considered the random variable, and this metric might change in different circumstances. Therefore, such an analysis may require multiple simultaneous Bayesian regressions. Another challenge would be determining a realistic (preferably physics-informed) prior for this scenario, and, if additional progress is to be made, further discussion will need to be had within the armor community.

One item not addressed in this work is the instance of a shatter gap [35], which is a scenario wherein a threat perforates an armor solution at a slower velocity while it is defeated at a higher velocity before ultimately perforating the armor, once again, at an even higher velocity. With the use of a random variables, we imagine an instance where another one or two random variables, perhaps termed the shatter-gap random variable(s), would modify the cumulative distribution function such that the perforation probability increases over the selected range. As with other testing methods, if there was concern of an armor solution might suffer from a shatter gap vulnerability, additional ballistic testing would need to take place to determine whether or not it is present. Of course, due to the distributed nature of the modified V_{50} characterization test, less testing may have to take place because the test method explores part of the region where a shatter gap might exist.

6 Conclusions

The work presented here provides an updated approach to ballistic characterization testing using Bayesian inference that does not appear to have been done before. The strength behind the modified V_{50} test method, coupled with Bayesian inference, is the ability to provide reasonable, physics-informed assumptions through the use of a specified prior distribution. Here, a triangular prior was assumed, although it was observed a better physics-informed prior could be used.

Exploratory calculations provided insight on how this might be used in a design scenario. It was observed that shifting the data collection off-center from the true μ value did not significantly hinder parameter inference. It was also observed that it is easier to extract μ when σ is relatively small. Taking repeat shots at specific velocities does not appear to correlate with improving parameter inference. Overall, σ inference remains challenging, and improving its extraction will likely require non-uniform data across the sample space. Experimentally, the

test results suggested that one armor solution performs better than required and might be able to be decreased slightly while still maintaining the same protection level.

Finally, a method was presented for converting the μ and σ random variables into a proof velocity that is estimated to be maximum value to meet a design criterion for armor qualification. This is most useful for armor layups utilizing novel materials or new threats to be analyzed, because limited or no data prevents easy extrapolation from historical results. In the limit of the 90-90 criterion, the proof velocities were converging to the appropriate limit value of approximately $\mu - 1.28155\sigma$, although it took many data points to approach this value. This method provides more relevant information than that which is provided using the traditional V_{50} characterization test because it allows reasonable, outside information to be included in the analysis.

7 Acknowledgments

The author appreciates helpful comments made by Mr. Matthew Magner and Mr. Rick Rickert during the write-up of this report and thanks Dr. Christopher Gorman, Mr. Nathan Hwangbo, Dr. Thomas Meitzler, and Prof. Peter Thompson for their helpful reviews. The author also thanks the Survivability Armor Ballistics Laboratory at the US Army DEVCOM Ground Vehicle Systems Center for conducting the ballistic testing.

References

[1] Gangolu, J., Kumar, A., Bhuyan, K., and Sharma, H. Performance-based probabilistic capacity models for reinforced concrete and prestressed concrete protective structures subjected to missile impact. *International Journal of Impact Engineering*, 2022. 164:104207.

[2] Talladay, T. G. and Templeton, D. W. A computational comparison of high strain rate strength and failure models for glass. *Defence Technical Information Center*, 2012. ADA571731.

[3] Talladay, T. G. and Rickert, F. C. Impact on Ceramic Armor Seams. *Defence Technical Information Center*, 2016. AD1049172.

[4] Terranova, B., Whittaker, A., and Schwer, L. Simulation of wind-borne missile impact using lagrangian and smooth particle hydrodynamics formulations. *International Journal of Impact Engineering*, 2018. 117:1–12.

[5] Guo, H., Zheng, Y., Yu, Q., Ge, C., and Wang, H. Penetration behavior of reactive liner shaped charge jet impacting thick steel plates. *International Journal of Impact Engineering*, 2019. 126:76–84.

[6] feng Zhu, Q., xiang Huang, Z., qiang Xiao, Q., dong Zu, X., and Jia, X. Theoretical and experimental study of shaped charge jet penetration into high and ultra-high strength concrete targets. *International Journal of Impact Engineering*, 2018. 122:431–438.

[7] NIJ Standard-0101.06: Ballistic Resistance of Personal Body Armor. 2008.

[8] Webb, D. W. A comparison of various methods used to determine the sample size requirements for meeting a 90/90 reliability specification. *Defense Technical Information Center*, 2011. ADA540716.

[9] MIL-STD-662F: Test Method Standard V_{50} Ballistic Test for Armor, 1997.

[10] Smith, P. and Hetherington, J. *Blast and Ballistic Loading of Structures*. Butterworth-Heinemann Ltd, Oxford, 1994. ISBN 0-7506-2024-2.

[11] Andres, C. and Boughers, W. An Analysis of V_{50} Ballistic Limit Results Adjusting 1st Shot Velocity, Step-Up Step-Down Increments , Truth Characteristics and Velocity Control Distributions. *Defense Technical Information Center*, 2012. ADA624217.

[12] MIL-DTL-12560K with Amendment 3: Detail Specification Armor Plate, Steel, Wrought, Homogeneous (for use in Combat-Vehicles and for Ammunition Testing), 2020.

[13] MIL-DTL-46100E with Amendment 1: Detail Specification Armor Plate, Steel, Wrought, High-Hardness, 2008.

[14] MIL-DTL-46027M (MR): Detail Specification Armor Plate, Aluminum Alloy, Weldable 5083, 5456 & 5059, 2021.

[15] Collins, J. C. Quantal Response: Practical Sensitivity Testing. *Defence Technical Information Center*, 2012. ADA568728.

[16] Collins, J. C. Quantal Response: Estimation and Inference. *Defence Technical Information Center*, 2014. ADA611092.

[17] Eridon, J. and Mishler, S. Ballistic Validation Test Statistics and Confidence Intervals. In *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS)*. Novi, MI, 2020 .

[18] Zukas, J., Nicholas, T., Swift, H., Greszczuk, L., and Curran, D. *Impact Dynamics*. John Wiley & Sons, Inc., New York, NY, 1982. ISBN 0-471-08677-0.

[19] Angelino, E., Johnson, M. J., and Adams, R. P. Patterns of Scalable Bayesian Inference. *Foundations and Trends in Machine Learning*, 2016. 9(2-3):119–247.

[20] Sethurajan, A., Krachkovskiy, S., Goward, G., and Protas, B. Bayesian uncertainty quantification in inverse modeling of electrochemical systems. *Journal of Computational Chemistry*, 2019. 40(5):740–752.

[21] Vigliotti, A., Csányi, G., and Deshpande, V. S. Bayesian inference of the spatial distributions of material properties. *Journal of the Mechanics and Physics of Solids*, 2018. 118:74–97.

[22] Gao, Y., Li, D., Zhang, W., Guo, Z., Yi, C., and Deng, Y. Constitutive modelling of the tib2–b4c composite by experiments, simulation and neutral network. *International Journal of Impact Engineering*, 2019. 132:103310.

[23] Portone, T., Niederhaus, J., Sanchez, J., and Swiler, L. Bayesian model selection for metal yield models in high-velocity impact. *International Journal of Impact Engineering*, 2020. 137:103459.

[24] Zhai, Y., Zhao, R., Li, Y., Li, Y., Meng, F., and Wang, T. Stochastic inversion method for dynamic constitutive model of rock materials based on improved dream. *International Journal of Impact Engineering*, 2021. 147:103739.

[25] Wagenmakers, E.-J., Lee, M., Lodewyckx, T., and Iverson, G. J. Bayesian Versus

Frequentist Inference. In *Bayesian Evaluation of Informative Hypotheses*, pages 181–207. Springer, New York, NY, 2008.

[26] Jaynes, E. T. Confidence Intervals vs Bayesian Intervals. In *Foundations of Probability THeory, Statistical Inference, and Statistical Theories of Science*, volume 2, pages 175–257. 1976.

[27] VanderPlas, J. Frequentism and Bayesianism: A Python-driven Primer. *Proceedings of the 13th Python in Science Conference*, 2014. pages 85–93.

[28] van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., and Yau, C. Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 2021. 1:1.

[29] Briggs, W. M. It is Time to Stop Teaching Frequentism to Non-statisticians. 2012.

[30] Ge, H., Xu, K., and Ghahramani, Z. Turing: A language for flexible probabilistic inference. *International Conference on Artificial Intelligence and Statistics, AISTATS*, 2018.

[31] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 2017. 59(1):65–98.

[32] Cusumano-Towner, M. F., Lew, A. K., Saad, F. A., and Mansinghka, V. K. Gen: A general-purpose probabilistic programming system with programmable inference. *Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, 2019. pages 221–236.

[33] Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, 2013. 125:306–312.

[34] Stan Modeling Language Users Guide and Reference Manual, 2.28, 2019.

[35] Riley, M. A., Rice, K. D., and Forster, A. L. Assessment of Uncertainty in Ballistic Response Estimates Obtained from Ballistic Limit Testing. *Personal Armour Systems Symposium*, 2012. pages 1–10.

[36] Lehtinen, N. Error Functions. Technical report, Stanford University, Stanford, CA, 2010.

[37] *The Book of Statistical Proofs*. 2012.

[38] Briol, F. X., Oates, C. J., Girolami, M., Osborne, M. A., and Sejdinovic, D. Probabilistic integration: A role in statistical computation? *Statistical Science*, 2019. 34(1):1–22.

[39] Cockayne, J., Oates, C. J., Sullivan, T. J., and Girolami, M. Bayesian probabilistic numerical methods. *SIAM Review*, 2019. 61(4):756–789.

[40] Li, B. and Babu, G. *A Graduate Course in Statical Inference*. Springer, New York, NY, 2019. ISBN 978-1-4939-9759-6.

[41] Duane, S., Kennedy, A., Pendleton, B., and Roweth, D. Hybrid Monte Carlo. *Physics Letters B*, 1987. 195(2):216–222.

[42] MacKay, D. J. C. Course on Information Theory, Pattern Recognition, and Neural Networks, 2012.

[43] Mogensen, P. and Riseth, A. Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 2018. 3(24):615.

[44] Besançon, M., Papamarkou, T., Anthoff, D., Arslan, A., Byrne, S., Lin, D., and Pearson, J. Distributions.jl: Definition and Modeling of Probability Distributions in the JuliaStats Ecosystem. *Journal of Statistical Software*, 2021. 98(16):1–30.