Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

Bayesian Approaches to Quantifying Uncertainty in Sport and Exercise Measurements: A Guide for Practitioners and Applied Researchers

Paul A. Swinton

Doi: 10.51224/SRXIV.261

SportRχiv hosted preprint version 1

20/02/2023

School of Health Sciences, Robert Gordon University, Aberdeen, UK.

Corresponding Author

Dr. Paul Swinton

School of Health Sciences, Robert Gordon University

Garthdee Road

Aberdeen, UK,

AB10 7QG

p.swinton@rgu.ac.uk, +44 (0) 1224 262 3361

Please cite as: Swinton, PA. Bayesian approaches to quantifying uncertainty in sport and exercise measurement: A guide for practitioner and applied researchers. Pre-print available from SportRχiv. https://doi.org/10.51224/SRXIV.261

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

Abstract

Measurement, analysis and communication of data are key requirements of practitioners and applied researchers in sport and exercise. Many challenges exist, however, in the analysis and communication of data due primarily to noisy measurements and the use of traditional analysis approaches designed for large data acquisitions and long run perspectives. The purpose of this review is to provide an introduction and worked examples to Bayesian statistics that present an alternative to traditional frequentist approaches and explore the means by which they may ameliorate some of the aforementioned limitations. The review focusses on the importance and methods used to include expert knowledge within Bayesian analyses through informative prior distributions. Additionally, the review focusses on the use of subjective probabilities and the subsequent ability to communicate results in an intuitive and informative manner. The review explores both conjugate and sample-based approaches to Bayesian analyses, highlighting their different strengths and the contexts where they may best be used. Worked examples show where substantive prior information can be incorporated, uncertainty can be decreased to obtain more relevant information and thereby underpin appropriate decision making. Given the flexibility afforded by Bayesian approaches and the advantages they can provide, there is a need for more resource development and interdisciplinary work between researchers and practitioners to increase their use in sport and exercise.

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

1.0 Introduction

The need for practitioners and applied researchers to consider uncertainty in the data they collect to better select, monitor and evaluate interventions is of key importance and has previously been discussed (Swinton et al, 2018). Two primary challenges when considering uncertainty include: 1) the amount of noise that generally exists when making measurements in sport and exercise; and 2) appropriate interpretations of uncertainty and reconciling these with probabilistic intuitions of both those that collect and provide the data. Approaches to reduce uncertainty when dealing with noisy measurements were covered in a previous review (Swinton et al, 2023) and focussed on collection of multiple measurements across days and throughout any intervention. It was shown that even with noisy measurements, higher frequency data collection could potentially reduce uncertainty to obtain relevant information and thereby underpin appropriate decision making. There may be situations, however, where high frequency data collection is not possible. In addition, data analysis methods covered previously (Swinton et al, 2023), are limited by interpretations that are generally non-intuitive and challenging to communicate accurately (Gelman and Greenland 2019).

Challenges in interpretation occur as almost all approaches to analyse data in sport and exercise adopt frequentist approaches. Here, probability is defined by the proportion of times an event occurs across a long sequence of repetitions. This definition applies well to games of chance (e.g. rolling dice) and situations where we can take large numbers of measurements either within or across individuals. The proportions we observe across a large series of measurements (e.g. the proportion of individuals with a body fat measurement below a given percentage, or the proportion of times an individual's blood pressure is beyond a given value) begin to align with a frequentist long run perspective and therein the associated probabilities. In sport and exercise, however, we generally do not have access to entire populations, let alone large samples. Where typically we have access to groups of only seven to ten individuals (Swinton et al, 2022), or a single case measured infrequently, a compromise is usually enacted whereby we consider the sample as one in a long run of samples that theoretically could be investigated. From this perspective, we can calculate statistics such as the mean or an effect size and use frequentist methods to quantify uncertainty that provides insights based on the assumption that we will carry out the same process repeatedly on future samples. The main tool to quantify uncertainty under a frequentist framework is the confidence interval (CI). Here, an interval such as a 95% CI is constructed around statistics such as the mean to indicate that if the procedure was repeated over and over on new samples, then in 95% of instances the interval would include the true population value (e.g. the population mean). In principle, the CI provides limited information regarding the specific case measured which is the real focus of interest, and is frequently misinterpreted by acting as though there is a 95% probability that any given interval contains the true value (Gelman and Greenland 2019). Furthermore, in sport and exercise where CI's are generally wide because of noisy data and small samples, it becomes challenging to interpret if values in the middle of the interval are in some sense different from values at the ends of the interval. Interpretations partly depend on which terms (e.g. probable, likely, plausible, consistent) are used to discuss the values.

An alternative to a frequentist perspective of probability, is subjective, also referred to as personal probability. From a subjective perspective, probability represents someone's degree of belief in an uncertain proposition (O'hagan et al, 2006). The terms subjective and personal reinforce the notion that individuals are likely to differ in their probability assessments. One major difference is that frequentist probabilities apply to repeatable events whereas subjective probabilities can be stated for any uncertain event. When first encountering subjective probability, there may be hesitancy to embrace the notion of subjectivity and instead favour traditional approaches that appear to offer less ambiguity and greater consistency. It can be argued, however, that all statistical approaches are subjective and in any analysis of the real world we do not have access to the 'true model' generating the data such that subjective judgements always play a role. It is the purpose of Bayesian statistics to make inferences through subjective probabilities (O'hagan et al, 2006). With Bayesian approaches, statistical models are selected such that parameters (e.g. the mean, standard deviation or effect size) in a model are described by probability distributions. The process is achieved through Bayes theorem (eq.1), which tells us how to derive a density function for the unknown parameters $\boldsymbol{\theta}$ after we have observed data $\boldsymbol{x}$.

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \pi(\boldsymbol{\theta})f(\boldsymbol{x}|\boldsymbol{\theta})/f(\boldsymbol{x}) \qquad eq.1$$

The theorem presented in eq.1 involves four different functions: 1) the prior distribution $\pi(\boldsymbol{\theta})$, which summarises everything an individual believes about the parameters apart from the data; 2) the likelihood $f(\boldsymbol{x}|\boldsymbol{\theta})$ which captures the data generating mechanism conditional on the parameters; 3) the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ which summarises belief regarding the parameters after combining the prior, likelihood and data; and 4) the proportionality constant $f(\boldsymbol{x})$ which is simply a normalising constant that ensures the posterior distribution integrates to 1 as all probability distributions must.

Much of the subjectivity of Bayesian inference comes from the fact that a prior distribution that describes the probability of parameter values must be selected by each individual, and that there is no 'correct' distribution. As far as possible, the prior distribution selected should accurately quantify an individual's subjective probabilities before observing the data (although a temporal distinction is not necessary). If two individuals agree on the data generating mechanism $f(\boldsymbol{x}|\boldsymbol{\theta})$ but select different prior distributions, once the data are observed they are then likely to arrive at different probabilities regarding the parameters. Similarly, a single individual may select multiple prior distributions for the same parameters which will lead to multiple posterior distributions. Such analyses are commonly referred to as sensitivity analyses and are frequently conducted as developing appropriate prior distributions can be challenging (see section 3).

The purpose of this review is to discuss and provide examples of how to use Bayesian methods to quantify uncertainty in measurements commonly used by practitioners and applied researchers in sport and exercise. The review will highlight how Bayesian approaches can be used to effectively quantify uncertainty and express this in intuitive and easy to understand ways. The review will also highlight how practitioners and applied researchers can incorporate their expertise and subjective judgements to better interpret baseline values and intervention effectiveness to further deal with the challenges of noisy measurements in sport and exercise. The review introduces two classes of methods for performing Bayesian analyses, namely conjugate methods that can be performed using simple spreadsheets, and more flexible sample-based methods that require specialist software. The main text focusses on the broad details and overarching concepts. Further mathematical detail and statistical code to implement the analyses in R are provided in the appendices.

2.0 Establishing plausible baseline scores

As identified previously (Swinton et al, 2018), one of the key requirements for practitioners is to establish plausible baseline values of their client so that they can select appropriate interventions and obtain benchmarks to evaluate subsequent intervention effectiveness. In classical test theory, it is assumed that if it were possible to conduct a large number of tests on the same individual then the values observed would follow a normal (Gaussian) distribution, with mean equal to the true score and standard deviation (σ) describing variability around this mean. In mathematical notation we have:

$$x_i = \mu + \epsilon_i \qquad eq.2$$

where x_i is the observed score, μ is the mean which is the hypothetical true score, ϵ_i is the measurement error described by a normal distribution with mean 0 and standard deviation σ, that is $\epsilon \sim N(0, \sigma^2)$.

Note from a frequentist perspective the true score is something that is "fixed", that is, it does not make sense to describe the value probabilistically (if we create a CI it is either in the interval [100%] or not [0%]). From this perspective the true value is never known but as discussed in the introduction, we can view the measurement process as one instance and create CI's such that over the long run the true value will reside within the interval according to the percentage set. In contrast, with the Bayesian perspective, we treat the true score as a random variable that we estimate by combining our prior beliefs, our selected data generating model, and the actual data collected in the form of a probability distribution. Note, in the model selected above there are two parameters $\boldsymbol{\theta}$ in our data generating model including the mean (true score μ) and the measurement error standard deviation σ. With Bayesian methods we generally provide a prior distribution that covers all parameters in the data generating model. We start our discussions, however, with the simplest models where we consider σ to be known and only specify a prior distribution for μ. That is, we assume no uncertainty in the measurement error standard deviation which is an unrealistic simplification in almost all cases except potentially those where we have extensive test-retest data on our own measurement processes with a specific population.

2.1 Model 1: Single data point with conjugate prior and known measurement error

The simplest model is where we have a single data point x, which from eq.2 we assume comes from a normal distribution and σ is known. We also select a normal distribution for our prior beliefs of μ, which we express as $\pi(\mu) \sim N(m, v)$. Here m represents what we think is the most likely value for μ, and v describes our uncertainty, with larger v giving greater probability to values further from m. For this model, the normal distribution is referred to as a conjugate prior distribution, as the posterior distribution for μ ($\pi(\mu|x, \sigma)$) will also be normally distributed (the derivation can be found Appendix 1A). Conjugate models enable us to avoid using specialist software and we can obtain the posterior distribution using direct calculation with visuals and probabilities assessed using simple software such as MS Excel. In section 3.1 we consider in more detail how the mean (m) and variance (v) for the prior distribution can be selected. Once these values are selected, however, the posterior mean (m^*) and posterior variance (v^*) of μ are calculated as:

$$v^* = \frac{1}{\frac{1}{v} + \frac{1}{\sigma^2}}$$

$$m^* = v^* \left(\frac{m}{v} + \frac{x}{\sigma^2}\right). \qquad eq.3$$

From eq.3 we can see that the posterior mean (m^*) of μ is a weighted average of the prior mean and the observed data point, such that if we have low measurement error then more weight is given to the data point. Contrastingly, if we express a strong prior belief such that v is small, then more weight will be given to the prior mean of μ. To see how this works in practice, we consider example data where we measure body fat percentage with a value of 17.5% obtained, and state that measurement error standard deviation is σ=3%. We consider two different prior distributions, both with the same mean (m) of 20%, but one with a standard deviation of 4% (v=16%), and one where we express greater uncertainty (8% [v=64%]). Using the first prior distribution, we update and obtain a posterior standard deviation of 2.4%:

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

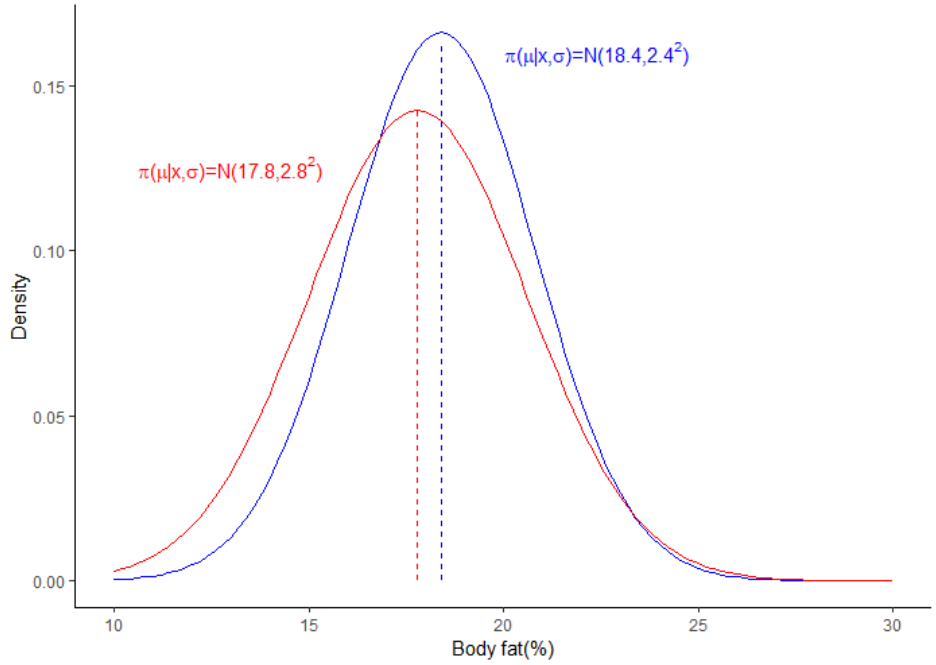
$$\sqrt{v^*} = \sqrt{\frac{1}{\frac{1}{v}+\frac{1}{\sigma^2}}} = \sqrt{\frac{1}{\frac{1}{4^2}+\frac{1}{3^2}}} \approx 2.4.$$

We use this value to update and obtain our posterior mean of 18.4%:

$$m^* = v^*\left(\frac{m}{v}+\frac{x}{\sigma^2}\right) = 2.4^2 \ldots \left(\frac{20}{4^2}+\frac{17.5}{3^2}\right) \approx 18.4.$$

Using the second prior distribution and the same process as above, we update and obtain a posterior standard deviation of 2.8% and a posterior mean of 17.8%. Note in both cases how the posterior mean gets pulled towards the prior mean, and the posterior standard deviation is less than the standard deviation of our known measurement error. Also note how the posterior mean gets pulled further towards the prior mean with the first distribution where there is less uncertainty for μ. The two posterior distributions $\pi(\mu|x,\sigma)$ generated with the different prior distributions are visualised in Figure 1. It is also important to remember that we interpret these distributions probabilistically and thus can address probabilistic questions. For example, using the first prior distribution we can calculate the (subjective) probability that the true score is less than 25%: $\left(P(\mu < 25) = P\left(Z < \frac{25-18.4}{2.4}\right) = \Phi(2.75) \approx 0.997\right)$, or for example, the probability that the true score is between 20 and 25%: $\left(P(\mu \leq 25) - P(\mu \leq 20) = P\left(Z < \frac{25-18.4}{2.4}\right) - P\left(Z < \frac{20-18.4}{2.4}\right) = \Phi(2.75) - \Phi(0.667) \approx 0.250\right)$. Further details are provided in Appendix 2A.

Figure 1: Posterior distributions of true score estimates from a single measurement of body fat percentage with known measurement error and two different prior distributions.



Posterior distributions given the conjugate priors {blue curve was obtained with $N(20, 4^2)$ as prior distribution; red curve: was obtained with $N(20, 8^2)$ as prior distribution}.

2.2 Model 2: Mean of multiple data points with conjugate prior and known measurement error

As described in a previous review (Swinton *et al*, 2023), an effective method to reduce uncertainty with noisy data is to take the mean of repeated measurements. Where we assume the data generating process is normally distributed $\left(f(\boldsymbol{x}|\boldsymbol{\theta}) = N(\mu, \sigma^2)\right)$, taking the mean ($\bar{x}$) of n independent measurements results in the following $\bar{x} \sim N(\mu, \sigma^2/n)$ (see Appendix 1B for details). Combining this distribution and the results from the previous section, we can deduce that the

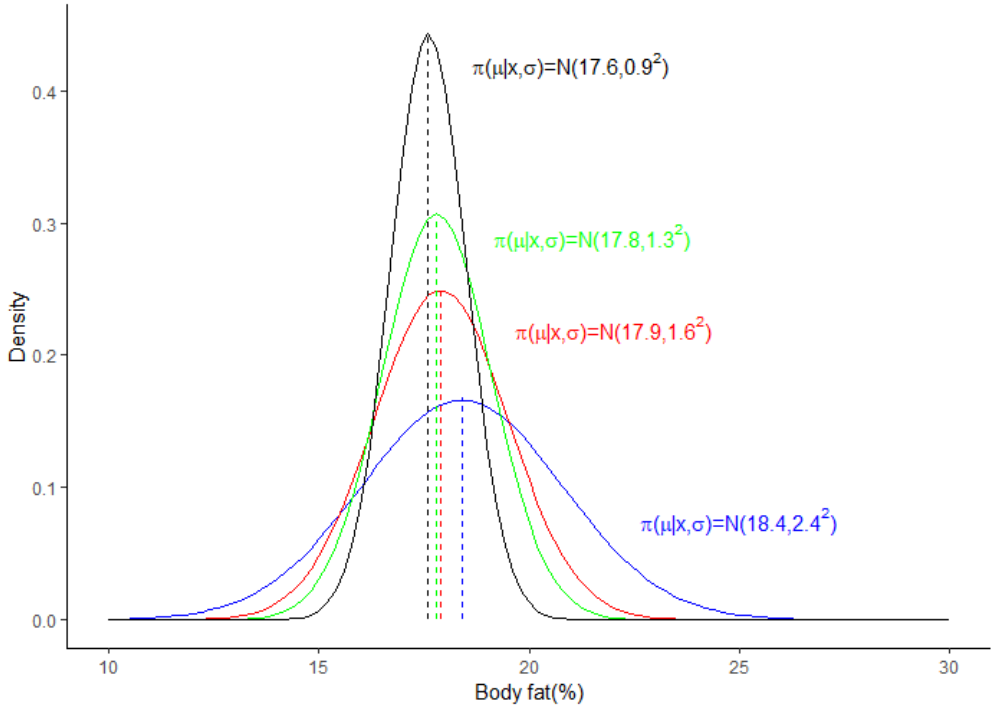
posterior distribution of μ conditional on $\bar{x}$ is normally distributed, but with v and m updated with σ^2 replaced by σ^2/n, and x replaced with $\bar{x}$:

$$v^* = \frac{1}{\frac{1}{v} + \frac{1}{\sigma^2/n}}$$

$$m^* = v^*\left(\frac{m}{v} + \frac{\bar{x}}{\sigma^2/n}\right). \qquad eq.4$$

Note, a more formal derivation is provided in Appendix 1B. Returning to our previous example, if we select our initial prior distribution ($N(20,4^2)$) for our true score μ, and calculate for example, that the mean of five data points is 17.5%, then we find that the posterior distribution of μ is $N(17.8, 1.3^2)$ (see Appendix 2B for full worked example). Note that the posterior mean is closer to the observed value compared to the example with only one data point (17.8 vs. 18.4%), which highlights a general feature of Bayesian analyses which is with more data prior distributions becomes less influential. Figure 2 illustrates the posterior distributions obtained when the same mean is calculated with progressively more data.

Figure 2: Posterior distributions of true score estimates from mean of multiple measurements of body fat percentage with known measurement error and increasing amounts of data.



Posterior distributions given the same conjugate prior $\{N(20, 4^2)\}$ and the mean of n measurements {blue curve: $n = 1$, red curve: $n = 3$, green curve: $n = 5$, black curve: $n = 10$}.

3.0 Prior distributions

Whist prior distributions can be considered to exist on a spectrum, for simplicity they are often characterised as either informative or non-informative. The latter characterisation attempts to express ignorance of parameter values that are more probable than others, and is often used in attempts to 'let the data speak for themselves'. Because most non-informative priors contain some information either on the original scale or following some transformation, a range of terms including vague, diffuse, flat and weak are often used interchangeably. Importantly, we frequently deal with cases where the selected data generating model includes multiple parameters, some of which provide insight into features of interest, and others that influence the data we observe but are not of interest. These latter class of parameters are referred to as nuisance parameters and in a Bayesian analysis we may choose to apply non-informative prior distributions (Spiegelhalter et al, 2004). In contrast, within the context of this review where a practitioner or applied researcher has decided to use somewhat sophisticated methods to analyse their data, it is unrealistic that they will not have knowledge or access to information that can be used to develop informative priors for parameters of actual interest. It may be challenging to appropriately capture knowledge and information in the form of probability distributions, however, given the perspective of subject probabilities it is clear that attempts should be made. For example, when measuring a client's body fat percentage, experienced practitioners will have knowledge of typical values of different populations and are likely to be able to refine such knowledge by simply observing the client. Similar population data and initial impressions based on previous measurements or simple initial assessments are likely to exit for most domains that interest practitioners and applied researchers in sport and exercise. It can be argued, therefore, that a full embrace of Bayesian approaches should seek to use informative priors where possible and adopt procedures that enable these to be developed as rigorously as possible.

3.1 Methods of prior elicitation

The process of obtaining knowledge from a source to form a prior distribution in a Bayesian analysis is referred to as prior elicitation (Falconer et al, 2022). Three general methods are used to elicit prior distributions and include: 1) historical information methods; 2) interrogation methods; and 3) graphical/visual methods. Historical information methods can include or refine previous posterior distributions as the next prior distribution, and in this way, we are continually updating our estimates in a sequential fashion. The information could be obtained from the client themselves (e.g. from previous measurements), similar clients, or from published research. Many studies in sport and exercise report means and standard deviations from the participants involved. If we wished to select a normally distributed prior for the mean as shown in sections 2.1 and 2.2, we could use these published values. It is important to note, however, that values presented in a study are only estimates of the population values. Given recruitment strategies are rarely randomised (Swinton et al, 2022), standard deviations may not capture the actual variability in the population and so potentially this value may be inflated for the prior distribution. Where multiple studies provide summary data on a similar population, the values may be pooled and weighted with greater weights given to studies with larger sample sizes. For example, with w studies of sample size $n_1, n_2, \dots, n_w$ the pooled mean and variance would equal:

$$m = \frac{\sum_{i=1}^{w} n_i m_i}{\sum_{i=1}^{k} n_i}$$

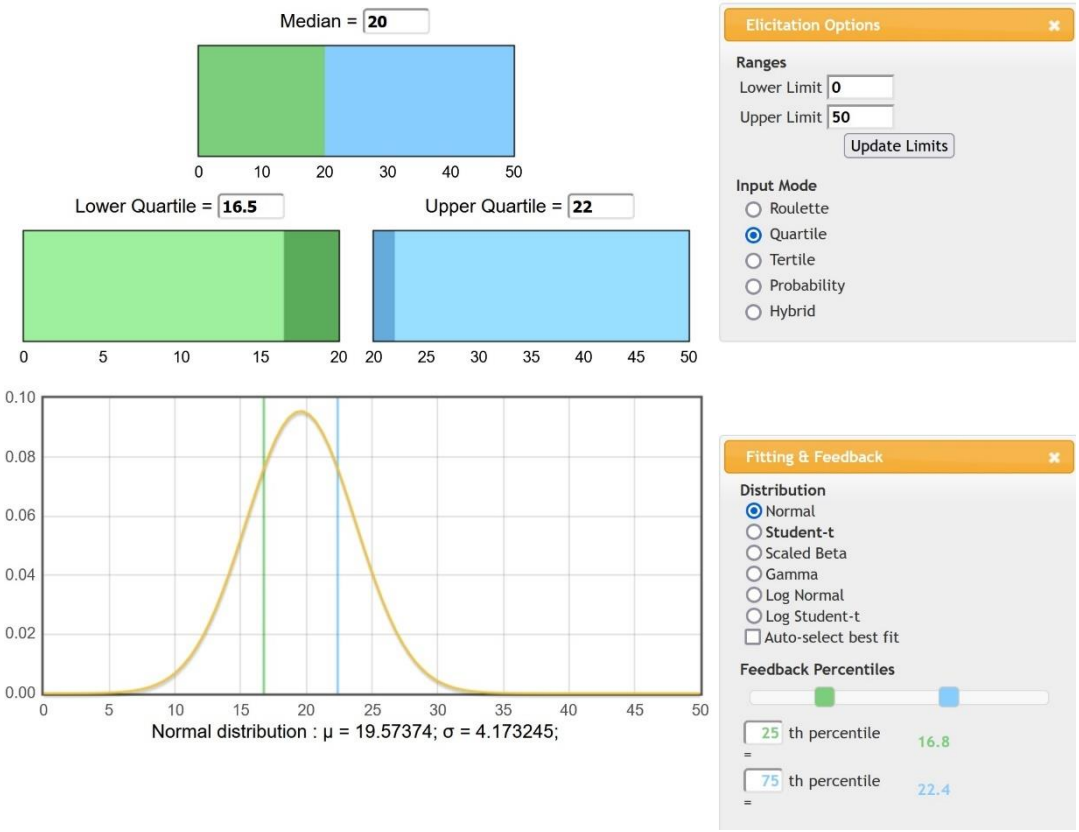
$$v = \frac{\sum_{i=1}^{w} (n_i - 1) v_i}{\sum_{i=1}^{w} n_i - w}. \qquad eq.5$$

An alternative approach to the use of historical information is to use so called interrogation methods to elicit experts' knowledge about uncertain quantities in probabilistic form. In many disciplines this takes the form of interviews where an analyst with experience in statistics and prior elicitation uses a range of techniques to capture the experts' knowledge. Most often, the analyst uses distribution or probability based questions focusing on observed values and/or the parameters of interest. Previous sources have been devoted to the topic of elicitation of expert knowledge (O'Hagan et al, 2006). Additionally, prior elicitation tools such as SHELF (Gosling 2018) and MATCH (Morris et al, 2014), have been developed to assist in elicitation with instructions and calculations included within easy to use graphical interfaces. Many tools also include graphical/visual methods to assist with expert prior elicitation and provide feedback to check

probability distributions align with the expert's beliefs (Falconer et al, 2022). In the case of practitioners in sport and exercise using Bayesian methods and informative priors, it is likely that they will take on both the role of analyst and expert. That is, the goal will generally be for the practitioner to capture their own subjective probabilities regarding a client's capacity, which entails quantifying their informed prior distributions. With applied researchers in sport and exercise there may be a focus on eliciting prior distributions from practitioners and coaches. Alternatively, researchers may seek to obtain prior distributions that summarise the beliefs of different 'types' of researchers. This may include, for example, development of representative sceptical (e.g. assuming a novel intervention has a negative effect), enthusiastic (e.g. assuming an intervention has a positive effect), and neutral (e.g. centring an intervention on zero) prior distributions.

An intuitive prior elicitation method that is often used is the quartile method (O'Hagan et al, 2006), which can be implemented with the MATCH web-based tool (Morris et al, 2014). For the models discussed in section 2, the practitioner would consider the true score of their client and express their uncertainty with a normal distribution. Use of the quartile method would require the practitioner to determine their median estimate, and then subsequently their lower and upper quartile estimates. These values can be elicited by the following sequence of questions: 1) What do you believe is the most likely value for the true score? Select a value such that the actual true score is equally likely to lie above or below your value; 2) Suppose you are informed the actual true score is below your initial value selected, now specify an updated value such that actual true score is equally likely to lie above or below your updated value; and 3) Suppose you are informed the actual true score is above your initial value selected, now specify an updated value such that actual true score is equally likely to lie above or below your updated value. These quartile values can then be inputted into the MATCH tool and the percentile sliders used to provide feedback and further assess whether the values provided are in accordance with an individual's subjective beliefs. Note, that if the quartiles are not symmetric around the median, the values of the elicited distribution will shift to best match the overall beliefs. Returning to the body fat percentage example, Figure 3 shows that the practitioner has stated that their 1st, 2nd and 3rd quartile values are 16.5, 20 and 22%, respectively. As a result, the distribution elicited has a mean of 19.6 and standard deviation of 4.2%, generating 1st and 3rd quartile values of 16.8 and 22.4, respectively (Figure 3).

Figure 3: Example using MATCH web based prior elicitation tool to elicit a normal distribution using the quartile method.



Prior elicitation tool (Morris et al, 2014).

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

4.0 More complex models for establishing plausible baseline scores

In the models presented in section 2, we assumed that the measurement error standard deviation was known and simply plugged in this value. In general, this process will underestimate our uncertainty in the true score as there is additional uncertainty in error magnitudes that will be influenced by a range of factors including the individual being measured and differences in protocols and equipment. How we express our uncertainty in measurement errors will likely be a compromise between what we actually believe and our ability to find a representative distributional form, whilst also fits with the approach we wish to use (e.g. conjugate or sample-based methods). Here we consider our final conjugate model and also highlight why more flexible sample-based approaches are often required.

4.1 Model 3: Mean of multiple data points with conjugate prior and unknown measurement error

If we wish to estimate both the true score and the measurement error standard deviation, we will have to obtain a joint prior distribution that describes the probabilities of both parameters. Using Bayes theorem, we often express joint probabilities of two parameters as the product of a conditional distribution and a marginal distribution. The conjugate prior distribution commonly used is the normal inverse gamma (NIG) which is expressed as:

$$\pi(\boldsymbol{\theta}) = \pi(\mu, \sigma^2) = NIG\left(\mu, \sigma^2 | m, \frac{1}{k}, a, b\right) = N\left(\mu | m, \frac{1}{k}\sigma^2\right) IG(\sigma^2 | a, b). \qquad eq.6$$

First, we note that the conjugate distribution is expressed in terms of the measurement error variance and not the standard deviation. The distribution comprises the marginal inverse gamma distribution for σ^2, and somewhat counterintuitively, the notion that the uncertainty of our prior knowledge of the true score μ depends on the unknown measurement variance (Cowles 2013). In practice, eliciting the four parameters $(m, 1/k, a, b)$ would begin by considering the unknown standard deviation σ. When data are generated from the process described in eq.2, the standard deviation provides information on the median absolute difference between any two observations (Elfadaly and Garthwaite, 2015). Using this feature of the data generating process, an individual can use their knowledge to provide the most likely value for σ and express their uncertainty (e.g. provide 1st and 3rd quartiles). The values obtained would then be squared and the parameters a, b selected from the inverse gamma distribution to match this uncertainty. Checks can be made using the distribution (e.g. probability σ^2 lies between two values) with the square root taken to transform values back to the standard deviation scale. Once a typical value for σ and thereby σ^2 has been established, this can be plugged in to the normal distribution and k set so that the variance for the true score prior distribution matches the individual's beliefs. A final check can then be made on how the uncertainty for the true score changes across a range of values for σ^2 that are reasonable given the selection of parameters a, b (an example illustrating this process in more detail is provided in Appendix 2D). Once we have elicited the normal inverse gamma prior distribution, the posterior distribution is also normal inverse gamma with the following updates:

$$m^* = \frac{km + n\bar{x}}{k + n}$$

$$k^* = k + n$$

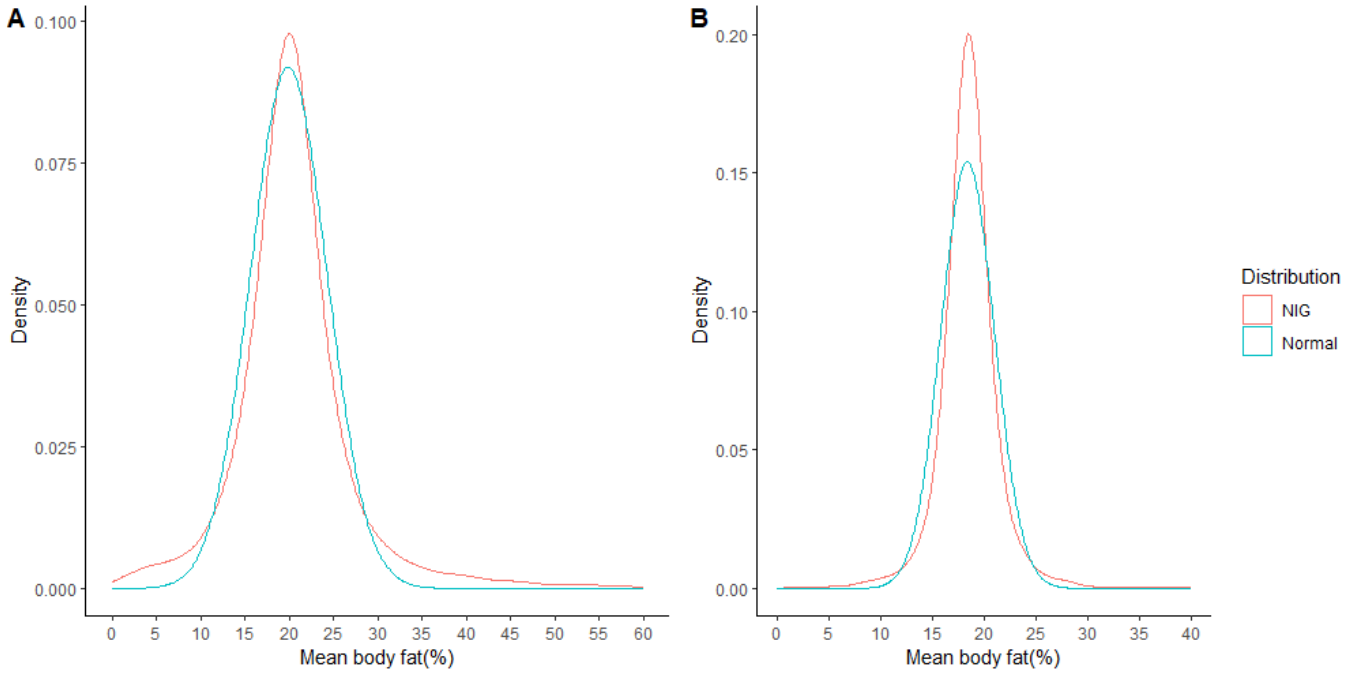
$$a^* = a + \frac{n}{2}$$

$$b^* = b + \frac{1}{2}\left[\sum_{i=1}^{n}(x_i - \bar{x}^2) + \frac{kn(\bar{x} - m)^2}{k + n}\right]. \qquad eq.7$$

The derivation of the updating process is provided in Appendix 1C. Returning to our previous example for body fat percentage where we initially believe that the majority of the true score distribution is likely to lie around 17 to 23% (e.g. the 1st and 3rd quartiles from $N(20, 4^2)$). If we also believe that the measurement error standard deviation is most likely to be 3%, with the 1st and 3rd quartiles around 2 and 6%, respectively, then as shown in Figure 4A, we may select $NIG(\mu, \sigma^2 | m = 20, k = 0.625, a = 0.6, b = 3.1)$ as our prior distribution. Note the heavier tails for the marginal prior distribution of μ (Figure 4A). If we observe a single measurement of 17.5%, then using eq.7, our joint posterior

distribution is $NIG(\mu, \sigma^2 | m = 18.46, k = 1.62, a = 1.1, b = 4.3)$. To describe the marginal posterior distribution of μ (Figure 4B), we first sample from $IG(\sigma^2 | a = 1.1, b = 4.3)$, and then input σ^2 to $N\left(\mu | 18.46, \frac{1}{1.62}\sigma^2\right)$. Alternatively, we can use a marginal distribution for μ which takes the form of a t-distribution (see Appendix 1c for details). Both approaches show similar posterior distribution results between the normal and normal inverse gamma distributions, except the latter exhibiting heavier tails. Note, if we take repeated measurements then we will obtain more precise estimates of the true score, and if of interest, we will obtain updated information regarding the measurement error standard deviation (which could also be used for more precise prior distributions of σ^2 in the future). As can be seen, however, the process of selecting the conjugate normal inverse gamma prior distribution is rather involved and lacks flexibility. It is at this point in general, that conjugate methods are forgone and sample-based methods used so that more natural prior distributions can be implemented and analyses more easily interpreted.

Figure 4: Comparison of true score prior (A) and posterior (B) distributions using normal prior with known variance, and normal inverse gamma distribution with unknown variance.



Blue curves represent distributions obtained with $N(20, 4^2)$ prior for μ and $\sigma = 3\%$. Red curves represent distributions obtained with $NIG(\mu, \sigma^2 | m = 20, k = 0.625, a = 0.6, b = 3.1)$ prior for μ and σ^2.

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

5.0 Sampling-based methods

As we progress to more complex Bayesian models including the use of different prior distributions, we are unable to perform exact posterior inferences analytically. Alternative approaches include the use of numerical and simulation techniques to perform the required integration steps, however, these generally have limitations when dealing with complex models (Cowles 2013). Because of these limitations, Bayesian analyses are most commonly performed using sampling-based methods with specialist software. The approach allows us to specify a data generating model, independent prior distributions for each parameter, and then sample from the joint posterior distribution. We can then use the samples from each parameter to estimate characteristics of the distributions including means, standard deviations, quantiles, and skewness. These samples are generally large and we can also use them to calculate probabilities, by for example, quantifying the proportion of the sample that lies between two values. Different algorithms are used to draw samples from joint posterior distributions and readers are directed towards Betancourt (2013) and Wang and Park (2020) for further details. Multiple specialist software now exists to perform Bayesian analyses, with the most popular for analyses in sport (Santos-Fernandez et al ,2019) including WinBUGS (Lunn et al, 2000), Stan (Stan development team), MATLAB (MATLAB 2017), and various R packages (R Core Team 2017). In this review, we use the popular R package brms (Bürkner 2017), which uses the probabilistic programming language Stan. Further details and resources for the brms package can be obtained from the following sources (Bürkner 2017; Nalborczyk et al, 2019).

5.1 Use of brms to check previous examples

To introduce the brms package and sampling-based methods in general, we repeat the analyses of sections 2 (see appendix 2G for further details). Analyses in brms can be considered within a regression framework, where we input our data, and specify the data generating model and the prior distributions. Where we have measurements from a participant, we can consider these our response variable which is typically expressed as y. As we wish to estimate the mean which generated these observed values, we consider the most basic intercept only regression $y \sim 1$.To run the model in R, we load the brms package, we incorporate the data into a data frame, and use the `brm` function. We use the following commands:

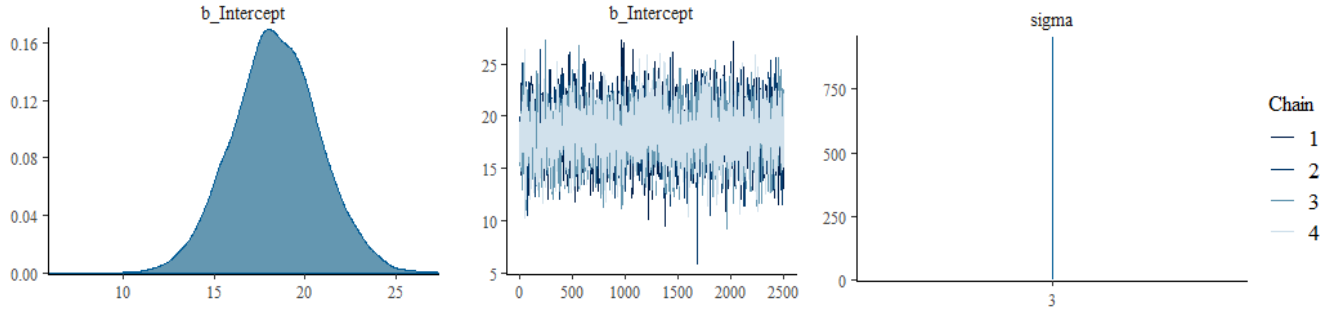
```
library(brms)

Data = data.frame(y=17.5)

Prior1 = c(
         prior(normal(20,4), class = Intercept),
         prior(constant(3),  class = sigma))

set.seed(123)

Model1 = brm(y~1, data = Data, prior = Prior1, family = gaussian,
                  chains = 4, iter = 5000, warmup = 2500)
```

In order, the commands above: 1) load the brms package; 2) create a data frame with a single variable y and single entry of 17.5; 3) set the prior distributions for the two parameters which include a normal distribution for μ with mean 20 and standard deviation 4, and fix σ at 3; 4) set the pseudo random number generator to provide reproducible results; and 5) run the sampling-based Bayesian model assuming measurement error follows a normal distribution.

Figure 5 shows the results of the analysis which is obtained with the `plot` command. The results show the posterior distribution for μ (labelled as b_Intercept) and the fixed value of σ. The plot also shows the mixing of the samples from four different chains which are obtained using different starting values. For further details on important issues such as the number of chains, sample iterations, burn-in and convergence see Roy (2020).

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

Figure 5: Visual output of true score posterior distribution using informative prior for the mean and known measurement error standard deviation with brms.



b_intercept: True score μ. Sigma: known measurement error standard deviation σ. Plots are generated with samples from four different chains.

To perform analyses on our posterior distribution for μ and check the mean, standard deviation, and probabilities (as was done in section 2.1), we use the following commands and obtain approximately the same results:

```
library(posterior)

Posterior1 = as_draws_df(Model1)
Posterior1Mu = Posterior1$b_Intercept
round(mean(Posterior1Mu),1)
[1] 18.4
round(sd(Posterior1Mu),1)
[1] 2.4
round(mean(Posterior1Mu<25),3)
[1]0.996
round(mean(Posterior1Mu>20&Posterior1Mu<25),3)
[1]0.245
```

To repeat the analysis of section 2.2 which was conducted on the mean of five data points, we can again include a single value and reduce the known standard deviation to $\sigma = 3/\sqrt{5}$, or we can create a new data frame with five entries with mean 17.5. The commands for this latter approach are:

```
Data2 = data.frame(y=rep(17.5,5))

set.seed(123)

Model2 = brm(y~1, data = Data2, prior = Prior1, family = gaussian,
                  chains = 4, iter = 5000, warmup = 2500)

Posterior2 = as_draws_df(Model2)
Posterior2Mu = Posterior2$b_Intercept
round(mean(Posterior2Mu),1)
[1] 17.7
round(sd(Posterior2Mu),1)
[1] 1.3
```

5.2 Sampling-based methods with alternative prior distributions

Now that we are using sampling-based methods we can consider a range of more intuitive prior distributions. One common starting point would be to consider a uniform prior distribution for σ. This type of prior distribution ($U(a,b)$) places flat probabilities across a range of values (a to b) and then zero probability for all other values. A practitioner may choose this type of prior distribution when they are very confident that the true value is in a specific range, but have no real sense of which values within the range are more probable. For example, if a practitioner is very confident that the measurement error standard deviation is somewhere between 2 and 6%, and wanted to combine the $U(2,6)$ prior distribution with say five measurements, then we could obtain the posterior samples and make inferences regarding both μ and σ. Unlike frequentist approaches where CI's are centred on a point estimate, Bayesian sample-based analyses lend themselves to several methods to summarise credible values. A common approach is the highest density interval (HDI) which summarises the distribution by specifying a width (e.g. 75 or 95%), such that every point inside the interval has higher credibility than any point outside the interval. The following commands are used to make inferences regarding both μ and σ using HDI's:

```
set.seed(123)
Data3 = data.frame(y=rnorm(5,19,3))

Prior2 = c(
         prior(normal(20,4), class = Intercept),
         prior(uniform(2,6),lb=2,ub=6, class = sigma))

Model3 = brm(y~1, data = Data3, prior = Prior2, family = gaussian,
                  chains = 4, iter = 5000, warmup = 2500)

library(HDInterval)
Posterior3 = as_draws_df(Model3)
Posterior3Mu = Posterior3$b_Intercept
Posterior3sigma = Posterior3$sigma
hdi(Posterior3Mu,credMass = 0.75)
[1] lower 18.0 upper 21.1
hdi(Posterior3sigma,credMass = 0.75)
[1] lower 2.0 upper 3.9
```

One of the most common distributions used to model uncertainty in measurement error standard deviations is the gamma distribution. In section 4.1 we used the inverse gamma distribution for measurement error variance so that the joint posterior distribution could be obtained analytically. The gamma distribution, however, is more commonly used as a prior distribution for the standard deviation and multiple elicitation methods (Garthwaite and Dickey 1988; Elfadaly and Garthwaite 2015) and software (Garthwaite et al, 2013) have been developed to assist. In addition, these elicitation methods attempt to express probabilities through the user responding to generated sample data such that revisions in parameter estimates after seeing the data are used to reveal underlying uncertainties (Garthwaite and Dickey 1988). Such approaches may be more effective than considering parameter values only. Appendix 2H provides examples comparing differen prior distributions for the standard deviation measurement error.

Sampling-based methods also offer greater flexibility regarding the data generating process. For example, we may use a technology or perform a test in sport and exercise that infrequently generates very large errors. In this case, a normal distribution for observed scores may not be appropriate, and instead a t-distribution centred on the true score but with heavier tails may be more realistic. Using sampling-based methods it is simple to maintain the same prior distribution for the mean, but select either informative or non-informative priors distributions for measurement errors that are now modelled according to the t-distribution (see Appendix 2H for worked examples). Note here the important difference between stating that our prior beliefs for μ do not require a distribution with heavy tails, but that the data generating process itself, and therefore the observed scores may follow such a distribution.

6.0 Establishing plausible change scores

We conclude this review by highlighting where Bayesian approaches may have the strongest advantage over frequentist approaches, which is through interpreting and expressing uncertainty in intervention effects. As has been shown previously (Swinton et al, 2023), interpreting effectiveness can be challenging with traditional methods in sport and exercise where we only have a single measurement pre- and post-intervention, and often are faced with relatively low improvements and large measurement errors. As we must consider measurement error both pre and post-intervention, true score change CI's are wider than true score baseline CI's (see Appendix 1D). Given the data generating mechanism expressed in eq.2, the distribution of observed change scores can be expressed as:

$$\delta_i = \Delta + \omega_i \qquad eq.8$$

where δ_i is the observed change score, Δ is the mean which is the hypothetical true change score, ω_i is the measurement error described by a normal distribution with mean 0 and standard deviation $\sigma\sqrt{2}$, that is $\omega_i \sim N(0, 2\sigma^2)$. CI's for change scores are thus expanded by a factor of $\sqrt{2}$, which can as a result, severely limit the ability to interpret intervention effectiveness using frequentist methods. To judge if an intervention is effective, we may choose to compare the observed change score and associated uncertainty to some threshold value that we denote the smallest worthwhile change (SWC). A typical approach would be to calculate for example, a 50 or 75% true score change CI around the observed score and state that an intervention is effective if the bounds of the interval lie beyond the SWC (Swinton et al, 2018). From a frequentist perspective, we are unable to make probabilistic claims regarding the specific instance we are investigating, but simply state that over the long run and adopting the same procedures, the true change score will exceed the SWC on the proportion of occasions that match the CI percentage. Here there are limitations in both interpretation and ability to appropriately judge intervention effectiveness. Firstly, we deal with the ability to pass judgement on whether or not an intervention has been effective.

6.1 Simple Bayesian approach to quantifying uncertainty in change scores

Returning to our body fat percentage example, previous studies show that over the course of a 12-week resistance training intervention, body fat percentage can be expected to reduce by approximately 1.5% (Wewege et al, 2022). If we assume that σ=3%, then a 75% true score change interval would equal $\delta_i \pm 1.15 \times \sqrt{2} \times 3$ (where 1.15 is the 0.875-quantile of the standard normal distribution). If we set the SWC to 0.5%, then we would require an observed change score of at least $0.5 + 4.9 = 5.4\%$ to conclude the intervention was effective. Given that most individuals only experience a 1.5% improvement, this example highlights the problem faced by practitioners. We could reduce the CI, but this would mean that over the long run we would be "incorrect" more frequently. In a previous review (Swinton et al, 2023), we discussed approaches to reduce uncertainty which involved collection of more data. Where this is appropriate, collection of more data should be completed, including when adopting Bayesian approaches. A key limitation of the frequentist approach outlined is that it fails to include relevant information such as expected changes and also what the practitioner believes the change has been based on their experience and observing the individual across the intervention. When adopting a Bayesian approach, we can in a principled way incorporate this information. Additionally, we can address the other important issue of interpretation and avoid a binary distinction of whether an intervention was effective or not through expressing the (subjective) probability that the true score change exceeds the SWC after observing the data.

To begin the process, we are required to specify our prior distribution for the true score change. As outlined in section 3, we may choose to consult previous research and use published data from an appropriate population. The challenge here, is that unlike the baseline case where research routinely presents means and standard deviations, in many areas it is rare for researchers to present the standard deviation of change scores. Most studies present pre- and post-intervention data but from this we are often unable to appropriately estimate the degree of variation in change scores. Where this information is presented, we can incorporate it within a normal distribution for true score change, and where it is presented in multiple studies we can use eq.5 to pool data. For body composition values, means and standard deviations of change scores are often presented with results frequently reporting standard deviation of change scores of approximately 3% (Wewege et al, 2022). We also require a prior distribution for the standard deviation of observed change scores (ω). Here, however, we can use knowledge regarding measurement error standard deviation (σ) and simply multiply by $\sqrt{2}$. If we use the previous example where the practitioner was confident that σ was between 2 and 6 but was

not confident of which values within the range were more probable, then we could use the uniform prior distribution $U(2\sqrt{2}, 6\sqrt{2})$ for ω. Finally, if we observed a pre- to post-intervention change score of -3.5%, then the analysis would be complete using the following commands:

```
Data4 = data.frame(y=-3.5)

Prior3 = c(
         prior(normal(-1.5,3), class = Intercept),
         prior(uniform(sqrt(2)*2, sqrt(2)*6),lb= sqrt(2)*2,ub= sqrt(2)*6,
         class = sigma))
set.seed(123)
Model4 = brm(y~1, data = Data4, prior = Prior3, family = gaussian,
                  chains = 4, iter = 5000, warmup = 2500)

Posterior4 = as_draws_df(Model4)
Posterior4Mu = Posterior4$b_Intercept
hdi(Posterior4Mu,credMass = 0.75)
[1] lower -5.07 upper 0.75
mean(Posterior4Mu<=-0.5)
[1]0.733
```

The interpretation of this analysis would be that a 75% credible interval for the true score change would be between -4.0 and 0.3%%, and the probability that the true change score exceeded the SWC of -0.5% is 0.755. This example highlights how much more informative and easier it can be to interpret a Bayesian approach compared to traditional approaches. If we have access to more data, then we can also be more precise regarding the likely true score change. To highlight this, we consider two scenarios, the first where we have access to a limited amount of additional data, and the second where we have access to high frequency data. In the first scenario, we consider a situation where a practitioner is able to repeat body fat measurements on four occasions both pre- and post-intervention. Using a simple analysis approach, we take the average of the four measurements pre-intervention as our baseline score, and the average of the four measurements post-intervention as our final score. With this simple analysis approach, the measurement error standard deviation of the change scores (ω) reduces by half ($\sqrt{n}$) and therefore, equals $\sigma\sqrt{2}/2$. We therefore adjust our prior distribution for ω accordingly. If we obtained the following data – pre-intervention: 20.2, 21.4, 18.2, 20.9; post-intervention: 15.2,17.1,16.4,18.0; and carry forward our initial belief regarding σ, then we use the following commands:

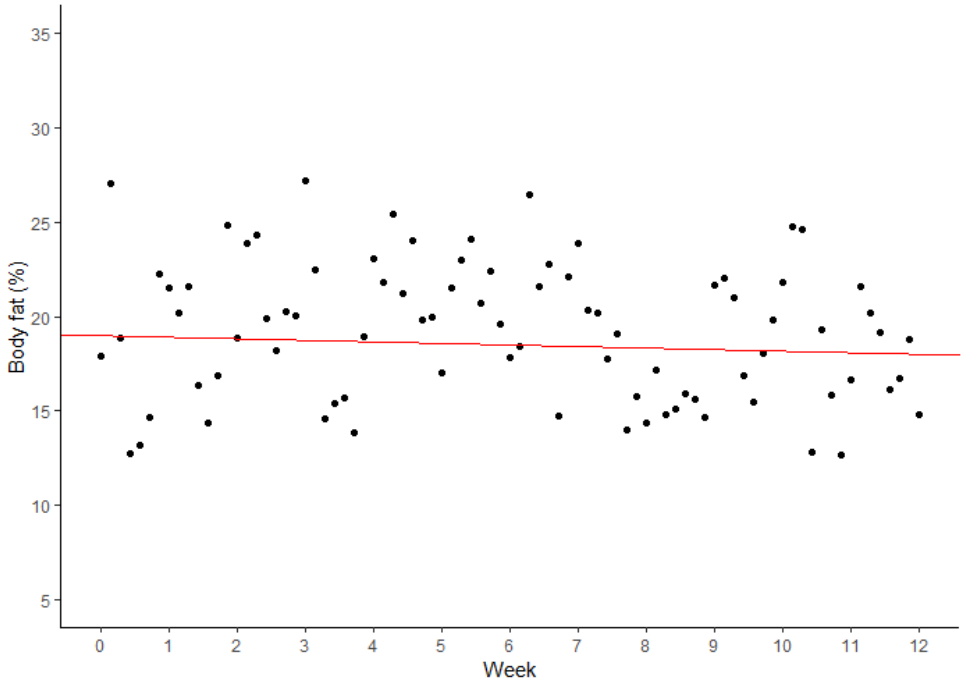
```
Data5 =data.frame(y=mean(c(15.2,17.1,16.4,18.0))-mean(c(20.2, 21.4, 18.2, 20.9)))

Prior4 = c(
         prior(normal(-1.5,3), class = Intercept),
         prior(uniform(sqrt(2)*2/2, sqrt(2)*6/2),lb= sqrt(2)*2/2,ub= sqrt(2)*6/2,
         class = sigma))
set.seed(123)
Model5 = brm(y~1, data = Data5, prior = Prior4, family = gaussian,
                  chains = 4, iter = 5000, warmup = 2500)

Posterior5 = as_draws_df(Model5)
Posterior5Mu = Posterior5$b_Intercept
hdi(Posterior5Mu,credMass = 0.75)
[1] lower -5.03 upper -0.45
mean(Posterior5Mu<=-0.5)
[1]0.855
```

Here we can see that the upper bound of the 75% credible interval for the true score change has further reduced and the probability that the true score change lies beyond the SWC of -0.5% has increased.

6.2 Bayesian autoregressive modelling to quantifying uncertainty in change scores

We finish this review by highlighting a second scenario where we have access to high frequency data. Rather than restricting ourselves to gold standard methods such as Dual-energy X-ray absorptiometry or well-trained anthropometrists, we could measure body fat percentage with a method such as bioelectrical impedance analysis (BIA). This is likely to have higher measurement errors, but can be used on a daily basis to provide unobtrusive measurements. Importantly, access to the high-frequency data is likely to make up for increased measurement error and result in improved inferences particularly where improvements across interventions are relatively small. Previously we have highlighted that an effective model that may be used to track an individual over a standard intervention period is linear regression with an autoregressive AR(1) process used for measurement errors (Swinton et al, 2023). The slope of the regression line represents the weekly change in the outcome, with the inherent assumption that this remains constant throughout the intervention enabling us to use this value to estimate the overall change. As this is a Bayesian analysis, we are required to specify prior distributions for all parameters including: 1) the intercept (baseline score covered in sections 2); 2) the slope of the regression line; 3) the measurement error standard deviation; and 4) the serial correlation parameter which causes measurement errors closer together in time to be stronger associated than those further apart in time (see Appendix 1E for further details). A full worked example with explanations of the mock data is presented in Appendix 2J. We assume that a measurement is made each day with a device that we believe the measurement error standard deviation can be described by a gamma distribution with parameters a=3.9, $b = 1$, such that the 75% HDI lies between 1.4 and 5.4%. We also assume that the BIA values are compatible with standard anthropometric measurements that are available at the beginning of the intervention. This allows us to perform a full Bayesian analysis outlined in section 5.1 to produce a posterior distribution that we can use as a prior distribution for the intercept. We also convert our knowledge that over the course of a 12-week intervention true score changes can be described by $N(-1.5,3^2)$, to develop a prior distribution for the weekly regression slope which equals $N(-0.125,0.25^2)$. Finally, we express some knowledge of the serial correlation parameter and apply the uniform prior distribution $U(0.2,0.5)$. Figure 6 illustrates the mock data which were developed based on the individual actually decreasing their body fat percentage by only 1% across the intervention. The 75% HDI for the true change score across the intervention is found to be -3.8 to -0.01%, with a 0.802 probability that value is below the SWC of -0.5%. We can contrast this with a frequentist analysis using the `arima` function in R, where the 75% true score change CI is -2.2 (-5.8 to 1.6%), such that we would generally conclude the intervention was not successful.

Figure 6: Mock data illustrating daily body fat percentage across an intervention as generated from a linear regression with AR(1) process.



Red line illustrates the true change across the intervention which is modelled by an intercept of 19% and a weekly decrease of -1/12% leading to a decrease of only 1% over the entire intervention.

7.0 Conclusion

The purpose of this review was to introduce, discuss, and provide examples of how to conduct Bayesian analyses to analyse commonly used data from single individuals in sport and exercise. The review highlighted the value of Bayesian analyses and therein the importance in developing informative prior distributions that leverage the expertise and knowledge contained by practitioners, researchers, and within the broader field. Through principled use of previously gained knowledge, Bayesian analyses can provide inferences that better underpin decision making. In addition, the review highlights how subjective probabilities can be used to communicate analyses that are easier to understand and fit with our probabilistic intuitions. The review introduced both conjugate and more common sample-based analyses. Whilst conjugate methods are limited, they may provide simple, and user-friendly analyses that can be implemented in standard software used to monitor and provide feedback to clients. In addition, where the same measurements are routinely collected and practitioners have accurately quantified measurement error standard deviations, similar results may be achieved between conjugate methods and more complex sample-based methods. Where practitioners wish to employ more complex models or develop more realistic prior distributions, sample-based analyses provide an extensive range of options. It is here that sensitivity analyses and development with regards to technical issues such as software languages, how distributions and parameters are expressed, and diagnostic tools (Depaoli, and Van de Schoot 2017) become important. Given the limited use of Bayesian analyses in sport and exercise, more interdisciplinary research is required between researchers and practitioners to develop models whose parameters are insightful, and determine how best to obtain prior distributions and thresholds values (e.g. SWC) that can be inputted to generate more actionable insight.

References

Betancourt, M., 2019. The convergence of Markov chain Monte Carlo methods: from the Metropolis method to Hamiltonian Monte Carlo. *Annalen der Physik*, *531*(3), p.1700214. https://doi.org/10.1002/andp.201700214

Bürkner, P.C., 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, *80*, pp.1-28. https://doi.org/10.18637/jss.v080.i01

Cowles, M.K., 2013. Applied Bayesian statistics: with R and OpenBUGS examples (Vol. 98). Springer Science & Business Media. https://doi.org/10.1007/978-1-4614-5696-4

Depaoli, S. and Van de Schoot, R., 2017. Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological methods*, *22*(2), p.240. https://doi.org/10.1037/met0000065

Elfadaly, F.G. and Garthwaite, P.H., 2015. Eliciting prior distributions for extra parameters in some generalized linear models. *Statistical Modelling*, *15*(4), pp.345-365. https://doi.org/10.1177/1471082X14553374

Falconer, J.R., Frank, E., Polaschek, D.L. and Joshi, C., 2022. Methods for Eliciting Informative Prior Distributions: A Critical Review. *Decision Analysis*, *19*(3), pp.189-204. https://doi.org/10.1287/deca.2022.0451

Garthwaite, P.H. and Dickey, J.M., 1988. Quantifying expert opinion in linear regression problems. *Journal of the Royal Statistical Society: Series B (Methodological)*, *50*(3), pp.462-474. https://doi.org/10.1111/j.2517-6161.1988.tb01741.x

Garthwaite, P.H., Al-Awadhi, S.A., Elfadaly, F.G. and Jenkinson, D.J., 2013. Prior distribution elicitation for generalized linear and piecewise-linear models. *Journal of Applied Statistics*, *40*(1), pp.59-75. https://doi.org/10.1080/02664763.2012.734794

Gelman, A. and Greenland, S., 2019. Are confidence intervals better termed "uncertainty intervals"?. *BMJ*, *366*. https://doi.org/10.1136/bmj.l5381

Gosling, J. P. (2018) SHELF: The Sheffield Elicitation Framework. In Dias, L., Morton, A., Quigley, J. (eds.) *Elicitation. International series in operations research and management science, vol. 261. Springer, Cham.* https://doi.org/10.1007/978-3-319-65052-4_4

Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D., 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, *10*, pp.325-337. https://doi.org/10.1023/A:1008929526011

MATLAB. 2017. "MATLAB and Statistics Toolbox Release." The MathWorks, Natick, MA, USA

Morris, D.E., Oakley, J.E., and. Crowe, J.A., 2014. A web-based tool for eliciting probability distributions from experts, Environmental Modelling & Software, Volume 52, Pages 1-4, ISSN 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2013.10.010.

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A. and Bürkner, P.C., 2019. An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), pp.1225-1242. https://doi.org/10.1044/2018_JSLHR-S-18-0006

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E. and Rakow, T., 2006. Uncertain judgements: eliciting experts' probabilities. https://doi.org/10.1002/0470033312

Roy, V., 2020. Convergence diagnostics for markov chain monte carlo. *Annual Review of Statistics and Its Application*, *7*, pp.387-412. https://doi.org/10.1146/annurev-statistics-031219-041300

Santos-Fernandez, E., Wu, P. and Mengersen, K.L., 2019. Bayesian statistics meets sports: a comprehensive review. *Journal of Quantitative Analysis in Sports*, *15*(4), pp.289-312. https://doi.org/10.1515/jqas-2018-0106

Spiegelhalter, D.J., Abrams, K.R. and Myles, J.P., 2004. Bayesian approaches to clinical trials and health-care evaluation (Vol. 13). John Wiley & Sons. https://doi.org/10.1002/0470092602

Stan Development Team. 2017. The Stan Core Library. http://mc-stan.org.

Swinton, P.A., Hemingway, B.S., Saunders, B., Gualano, B. and Dolan, E., 2018. A statistical framework to interpret individual response to intervention: paving the way for personalized nutrition and exercise prescription. *Frontiers in nutrition*, *5*, p.41. https://doi.org/10.3389/fnut.2018.00041

Swinton, P.A., Burgess, K., Hall, A., Greig, L., Psyllas, J., Aspe, R., Maughan, P. and Murphy, A., 2022. Interpreting magnitude of change in strength and conditioning: Effect size selection, threshold values and Bayesian updating. *Journal of sports sciences*, *40*(18), pp.2047-2054. https://doi.org/10.1080/02640414.2022.2128548

Swinton, PA. Stephens Hemingway B, Gallagher IJ, Dolan E. 2023. Statistical methods to reduce the effects of measurement error in sport and exercise: A guide for practitioners and applied researchers. Pre-print available from SportRχiv. https://doi.org/10.51224/SRXIV.247

Team, R.C., 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. *See https://www. R-project. org*.

Wang, M.Y. and Park, T., 2020. A brief tour of Bayesian sampling methods. *Bayesian inference on complicated data*, *17*. https://doi.org/10.5772/intechopen.91451

Wewege, M.A., Desai, I., Honey, C., Coorie, B., Jones, M.D., Clifford, B.K., Leake, H.B. and Hagstrom, A.D., 2022. The effect of resistance training in healthy adults on body fat percentage, fat mass and visceral fat: a systematic review and meta-analysis. *Sports Medicine*, *52*(2), pp.287-300. https://doi.org/10.1007/s40279-021-01562-2

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

Bayesian Approaches to Quantifying Uncertainty in Sport and Exercise Measurements: A Guide for Practitioners and Applied Researchers. Appendices.

Paul A. Swinton

Appendix 1: Additional mathematical detail

Appendix 1A (Section 2.1 - Model 1: Single data point with conjugate prior and known measurement error)

Derivation of posterior distribution for μ

Given the probability distribution function of the normal distribution

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

and prior beliefs about μ represented by a normal distribution

$$\pi(\mu) = \frac{1}{\sqrt{2\pi v}}\exp\left[-\frac{1}{2v}(\mu-m)^2\right],$$

Then the posterior distribution for a single data point and known σ^2 can be expressed as

$$\pi(\mu|x) = \frac{\pi(\mu)f(x|\mu)}{f(x)}.$$

We derive the posterior up to proportionality such that

$$\begin{aligned}\pi(\mu|x,\sigma) &\propto \frac{\pi(\mu)f(x|\mu)}{f(x)}\\ &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}(x-\mu)^2+\frac{1}{v}(\mu-m)^2\right)\right]\\ &= \exp\left[-\frac{Q}{2}\right],\end{aligned}$$

where

$$Q = \frac{1}{\sigma^2}(x-\mu)^2+\frac{1}{v}(\mu-m)^2.$$

If Q can be re-arranged to be in the form $\frac{(\mu-a)^2}{b}+c$, with c not depending on μ, then the posterior distribution of μ $\left(\pi(\mu|x)\right)$ must be $N(\theta|a,b)$.

$$Q = \frac{\mu^2 - 2\mu x + x^2}{\sigma^2} + \frac{\mu^2 - 2\mu m + m^2}{v}$$

$$= \mu^2\left(\frac{1}{\sigma^2} + \frac{1}{v}\right) - 2\mu\left(\frac{x}{\sigma^2} + \frac{m}{v}\right) + \frac{x^2}{\sigma^2} + \frac{m^2}{v}$$

$$= \mu^2 / \left(\frac{1}{\frac{1}{\sigma^2} + \frac{1}{v}}\right) - 2\mu\left(\frac{x}{\sigma^2} + \frac{m}{v}\right) + \frac{x^2}{\sigma^2} + \frac{m^2}{v}.$$

.

Completing the square with $c_1\mu^2 + c_2\mu + c_3 = c_1(\mu + h)^2 + c_4$, where $h = \frac{c_2}{2c_1}$ and $c_4 = c_3 - \frac{c_2^2}{4c_1}$, gives

$$Q = \left(\frac{1}{\frac{1}{\sigma^2} + \frac{1}{v}}\right)\left(\mu - \left(\left(\frac{x}{\sigma^2} + \frac{m}{v}\right)\left(\frac{1}{\frac{1}{\sigma^2} + \frac{1}{v}}\right)\right)\right)^2 + c_4.$$

Therefore

$\pi(\mu|x) = N(m^*, v^*)$, where

$v^* = \frac{1}{\frac{1}{v} + \frac{1}{\sigma^2}}$ and $m^* = v^*\left(\frac{m}{v} + \frac{x}{\sigma^2}\right)$.

Appendix 1B (Section 2.2 - Model 2: Mean of multiple data points with conjugate prior and known measurement error)

Derivation of distribution of $\bar{x}$

We show that given $x \sim N(\mu, \sigma^2)$ then $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \sim N(\mu, \sigma^2/n)$.

The expectation of $\bar{x}$ is

$$E(\bar{x}) = E\left(\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right) = \frac{1}{n}\left(E(x_1) + E(x_2) + \cdots + E(x_n)\right)$$

$$= \frac{1}{n}(n\mu) = \mu.$$

As each x_i are assumed independent, then the variance of $\bar{x}$ is

$$Var(\bar{x}) = Var\left(\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right) = \frac{1}{n^2}\left(Var(x_1) + Var(x_2) + \cdots + Var(x_n)\right)$$

$$= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

Therefore, we have $\bar{x} \sim N(\mu, \sigma^2/n)$.

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

Derivation of posterior distribution for μ

We now provide the derivation of the posterior distribution of μ given the mean of n independent observations and assuming normal distributions for the likelihood (with σ^2 known) and the prior for μ. For n normally distributed independent measurements x_i, the likelihood is given by

$$f(\boldsymbol{x}|\mu) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right].$$

Therefore,

$$\pi(\mu|\boldsymbol{x}) \propto \pi(\mu) f(\boldsymbol{x}|\mu)$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2+\frac{1}{v}(\mu-m)^2\right)\right]$$

$$= \exp\left[-\frac{\bar{Q}}{2}\right].$$

We follow similar steps as with the previous derivation, but first noting that

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\bar{x}-\mu+\bar{x})^2, \bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i, s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2, \text{we have}$$

$$\bar{Q} = \frac{\sum_{i=1}^{n}\left((x_i-\bar{x})-(\mu-\bar{x})\right)^2}{\sigma^2}+\frac{\mu^2-2\mu m+m^2}{v}$$

$$= \frac{\sum_{i=1}^{n}\left((x_i-\bar{x})^2+(\mu-\bar{x})^2-2(x_i-\bar{x})(\mu-\bar{x})\right)}{\sigma^2}+\frac{\mu^2-2\mu m+m^2}{v}$$

$$= \frac{ns^2+n(\mu-\bar{x})^2-2(n\bar{x}-n\bar{x})(n\mu-n\bar{x})}{\sigma^2}+\frac{\mu^2-2\mu m+m^2}{v}$$

$$= \mu^2\left(\frac{n}{\sigma^2}+\frac{1}{v}\right)-2\mu\left(\frac{\bar{x}}{\sigma^2}+\frac{m}{v}\right)+\frac{n(\bar{x}^2+s^2)}{\sigma^2}+\frac{m^2}{v}$$

$$= \mu^2/\left(\frac{1}{\frac{n}{\sigma^2}+\frac{1}{v}}\right)-2\mu\left(\frac{\bar{x}}{\sigma^2}+\frac{m}{v}\right)+\frac{n(\bar{x}^2+s^2)}{\sigma^2}+\frac{m^2}{v}.$$

As previous completing the square gives

$$\bar{Q} = \left(\frac{1}{\frac{n}{\sigma^2}+\frac{1}{v}}\right)\left(\mu-\left(\left(\frac{\bar{x}}{\sigma^2}+\frac{m}{v}\right)\left(\frac{1}{\frac{n}{\sigma^2}+\frac{1}{v}}\right)\right)\right)^2+c_5.$$

Therefore

$$\pi(\mu|\boldsymbol{x}) = N(m^*, v^*), \text{where}$$

$$v^* = \frac{1}{\frac{1}{v}+\frac{n}{\sigma^2}} \text{ and } m^* = v^*\left(\frac{m}{v}+\frac{\bar{x}}{\sigma^2}\right).$$

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

Appendix 1C (Section 4.1 - Model 3: Mean of multiple data points with conjugate prior and unknown measurement error)

Derivation of posterior distribution for μ *and* σ^2

The inverse gamma (IG) probability distribution can be expressed as

$$IG(\sigma^2|a,b) = \frac{b^a}{\Gamma(a)}\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left[-\frac{b}{\sigma^2}\right].$$

The normal inverse gamma (NIG) probability distribution is a joint distribution used to express prior probabilities for the mean μ and variance σ^2. It is the product of the marginal inverse gamma distribution $IG(\sigma^2|a,b)$ and conditional normal distribution $N(\mu|m,\sigma^2/k)$. The normal inverse gamma distribution is therefore

$$NIG(\mu,\sigma^2|a,b,m,1/k) = \frac{b^a}{\Gamma(a)}\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left[-\frac{b}{\sigma^2}\right]\frac{\sqrt{k}}{\sigma\sqrt{2\pi}}\exp\left[-\frac{k}{2\sigma^2}(\mu-m)^2\right].$$

For n normally distributed independent measurements x_i we have

$$f(\boldsymbol{x}|\mu,\sigma^2) \propto \frac{1}{\sigma^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right],$$

therefore, the posterior distribution using the NIG prior distribution can be expressed as

$$\pi(\mu,\sigma^2|\boldsymbol{x}) \propto \pi(\mu,\sigma^2)f(\boldsymbol{x}|\mu,\sigma^2)$$

$$\propto \frac{b^a}{\Gamma(a)}\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left[-\frac{b}{\sigma^2}\right]\frac{\sqrt{k}}{\sigma\sqrt{2\pi}}\exp\left[-\frac{k}{2\sigma^2}(\mu-m)^2\right]\frac{1}{\sigma^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right].$$

$$\propto \frac{1}{(\sigma^2)^{a+\frac{n}{2}+1}}\exp\left[-\frac{1}{\sigma^2}\left(b+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right]\frac{\sqrt{k}}{\sigma}\exp\left[-\frac{1}{2\sigma^2}(k(\mu-m)^2)\right]$$

As before with $\sum_{i=1}^{n}(x_i-\mu)^2 = ns^2 + n(\mu-\bar{x})^2$, we have

$$\pi(\mu,\sigma^2|\boldsymbol{x}) \propto \frac{1}{(\sigma^2)^{a+\frac{n}{2}+1}}\exp\left[-\frac{1}{\sigma^2}\left(b+\frac{ns^2}{2}\right)\right]\frac{\sqrt{k}}{\sigma}\exp\left[-\frac{1}{2\sigma^2}(k(\mu-m)^2+n(\mu-\bar{x})^2)\right].$$

We also have that

$$k(\mu-m)^2 + n(\mu-\bar{x})^2 = (k+n)\mu^2 - 2\mu(km+n\bar{x}) + km^2 + n\bar{x}^2.$$

Completing the square gives

$$(k+n)\left(\mu-\frac{km+n\bar{x}}{k+n}\right)^2 + km^2 + n\bar{x}^2 - \frac{(km+n\bar{x})^2}{k+n} =$$

$$(k+n)\left(\mu-\frac{km+n\bar{x}}{k+n}\right)^2 + \frac{(k^2m^2+knm^2+kn\bar{x}^2+n^2\bar{x}^2)-(km+n\bar{x})^2}{k+n} =$$

$$(k+n)\left(\mu-\frac{km+n\bar{x}}{k+n}\right)^2 + \frac{knm^2+kn\bar{x}^2-2kmn\bar{x}}{k+n} =$$

$$(k+n)\left(\mu-\frac{km+n\bar{x}}{k+n}\right)^2 + \frac{kn(m^2+\bar{x}^2-2m\bar{x})}{k+n} =$$

$$(k+n)\left(\mu-\frac{km+n\bar{x}}{k+n}\right)^2+\frac{kn(\bar{x}-m)^2}{k+n}.$$

Inserting this expression back into the previous statement gives

$$\pi(\mu,\sigma^2|\boldsymbol{x}) \propto \frac{1}{(\sigma^2)^{a+\frac{n}{2}+1}}\exp\left[-\frac{1}{\sigma^2}\left(b+\frac{ns^2}{2}\right)\right]\frac{\sqrt{k}}{\sigma}\exp\left[-\frac{1}{2\sigma^2}\left((k+n)\left(\mu-\frac{km+n\bar{x}}{k+n}\right)^2+\frac{kn(\bar{x}-m)^2}{k+n}\right)\right]$$

$$\propto \frac{1}{(\sigma^2)^{a+\frac{n}{2}+1}}\exp\left[-\frac{1}{\sigma^2}\left(b+\frac{ns^2}{2}+\frac{kn(\bar{x}-m)^2}{2(k+n)}\right)\right]\frac{\sqrt{k}}{\sigma}\exp\left[-\frac{1}{2\sigma^2}\left((k+n)\left(\mu-\frac{km+n\bar{x}}{k+n}\right)^2\right)\right]$$

We note that this is the normal inverse gamma distribution with

$$a^*=a+\frac{n}{2}, b^*=b+\frac{ns^2}{2}+\frac{kn(\bar{x}-m)^2}{2(k+n)}, k^*=k+n, m^*=\frac{km+n\bar{x}}{k+n}.$$

Derivation of marginal posterior distribution for μ

The t-distribution with a single parameter ν is a symmetric distribution centred on 0, with probability distribution function

$$f(\boldsymbol{x}|\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}.$$

For $\nu>2$ the variance equals $\frac{\nu}{\nu-2}$, such that with greater ν the variance approaches one and resembles the standard normal distribution. The t-distribution can be generalised to a three parameter location-scale family introducing a location parameter $\hat{u}$ and scale parameter $\hat{\sigma}$ through

$X=\hat{u}+\hat{\sigma}T$, which has the probability distribution function

$$f(\boldsymbol{x}|\nu,\hat{u},\hat{\sigma})=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}\hat{\sigma}}\left(1+\frac{1}{\nu}\left(\frac{x-\hat{u}}{\hat{\sigma}}\right)^2\right)^{-(\nu+1)/2}.$$

The generalised t-distribution has mean equal to $\hat{u}$ (for ν>1) and variance $\hat{\sigma}^2\frac{\nu}{\nu-2}$ (for ν>2).

We can show that the marginal posterior distribution for μ when using the normal inverse gamma as a prior distribution, is a generalised t-distribution. To show this, first we note that

$$\Gamma(z)=\int_0^\infty x^{z-1}e^{-x}dx.$$

To obtain the marginal posterior distribution for μ we must perform the following integration

$$\int_0^\infty N(\mu,\sigma^2|\boldsymbol{x})\,d\sigma^2=\int_0^\infty\frac{(b^*)^{a^*}}{\Gamma(a^*)}\left(\frac{1}{\sigma^2}\right)^{a^*+1}\exp\left[-\frac{b^*}{\sigma^2}\right]\frac{\sqrt{k^*}}{\sigma\sqrt{2\pi}}\exp\left[-\frac{k^*}{2\sigma^2}(\mu-m^*)^2\right]d\sigma^2$$

$$=\frac{(b^*)^{a^*}}{\Gamma(a^*)}\frac{\sqrt{k^*}}{\sqrt{2\pi}}\int_0^\infty\frac{1}{(\sigma^2)^{a^*+3/2}}\exp\left[-\left(\frac{1}{2\sigma^2}(k^*(\mu-m^*)^2+2b^*)\right)\right]d\sigma^2.$$

If we make the substitution $\lambda = (\sigma^2)^{-1}$, then $\frac{d\lambda}{d\sigma^2} = -(\sigma^2)^{-2} \rightarrow d\sigma^2 = -d\lambda(\sigma^2)^2$, then we have

$$\int_0^\infty N(\mu, \sigma^2|\boldsymbol{x}) = \frac{(b^*)^{a^*}}{\Gamma(a^*)}\frac{\sqrt{k^*}}{\sqrt{2\pi}}\int_0^\infty \lambda^{(a^*+1/2)-1}\exp\left[-\left(\frac{\lambda}{2}(k^*(\mu-m^*)^2+2b^*)\right)\right]d\lambda.$$

If we make a second substitution $\tau = \frac{\lambda}{2}(k^*(\mu-m^*)^2+2b^*)$, then $d\tau = \frac{1}{2}(k^*(\mu-m^*)^2+2b^*)d\lambda$, then we have

$$\int_0^\infty N(\mu, \sigma^2|\boldsymbol{x}) = \frac{(b^*)^{a^*}}{\Gamma(a^*)}\frac{\sqrt{k^*}}{\sqrt{2\pi}}\int_0^\infty \left(\frac{\tau}{\frac{k^*(\mu-m^*)^2+2b^*}{2}}\right)^{(a^*+1/2)-1} \exp(-\tau)\frac{d\tau}{\frac{k^*(\mu-m^*)^2+2b^*}{2}}$$

$$= \frac{(b^*)^{a^*}}{\Gamma(a^*)}\frac{\sqrt{k^*}}{\sqrt{2\pi}\left[\frac{k^*(\mu-m^*)^2+2b^*}{2}\right]^{a^*+\frac{1}{2}}}\int_0^\infty \tau^{\left(a^*+\frac{1}{2}\right)-1}\exp(-\tau)\,d\tau$$

$$= \frac{(b^*)^{a^*}}{\Gamma(a^*)}\frac{\sqrt{k^*}}{\sqrt{2\pi}}\frac{\Gamma\left(a^*+\frac{1}{2}\right)}{\left[\frac{k^*(\mu-m^*)^2+2b^*}{2}\right]^{a^*+\frac{1}{2}}}.$$

Further rearranging gives

$$\int_0^\infty N(\mu, \sigma^2|\boldsymbol{x}) = \frac{(b^*)^{a^*}}{\Gamma(a^*)}\frac{\sqrt{k^*}}{\sqrt{2\pi}}\frac{\Gamma\left(a^*+\frac{1}{2}\right)}{\left(\frac{1}{2}\right)^{a^*+\frac{1}{2}}(2b^*)^{a^*+\frac{1}{2}}\left[1+\frac{k^*}{2b^*}(\mu-m^*)^2\right]^{a^*+\frac{1}{2}}}$$

$$= \frac{\Gamma\left(a^*+\frac{1}{2}\right)}{\Gamma(a^*)\sqrt{2}\sqrt{b^*/k^*}\left[1+\left(\frac{\mu-m^*}{\sqrt{2}\sqrt{b^*/k^*}}\right)^2\right]^{a^*+\frac{1}{2}}}.$$

We can see that the above is a t-distribution with mean m^*, degrees of freedom $2a^*$, and scale parameter given by $\sqrt{\nu}\hat{\sigma} = \sqrt{2}\sqrt{b^*/k^*}$, such that $\hat{\sigma} = \sqrt{\frac{b^*}{a^*k^*}}$.

Appendix 1D (Section 6.0 – Establishing plausible change scores)

If we have observed scores pre-intervention of the form $x_{pre_i} = \mu_{pre} + \epsilon_{pre_i}$, and post-intervention of the form $x_{post_i} = \mu_{post} + \epsilon_{post_i}$, where ϵ_{pre} and ϵ_{post} are independent and both come from the same distribution $N(0, \sigma^2)$; then our change score $\delta_i = x_{post_i} - x_{pre_i}$ is also normally distributed with mean

$$E\left(x_{post_i} - x_{pre_i}\right) = E\left(x_{post_i}\right) - E\left(x_{pre_i}\right) = \mu_{post} - \mu_{pre} = \Delta,$$

and standard deviation

$$\sqrt{Var\left(x_{post_i} - x_{pre_i}\right)} = \sqrt{Var\left(x_{post_i}\right) + Var\left(x_{pre_i}\right)} = \sqrt{2\sigma^2} = \sqrt{2}\sigma.$$

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

Appendix 1E (Section 6.2 – Bayesian autoregressive modelling to quantifying uncertainty in change scores)

The final model discussed in this review is simple linear regression where the measurement errors follow an AR(1) process. The model can be expressed as $y_i = \beta_0 + \beta_1 x_i + \xi_i$, with errors ξ_i autocorrelated such that $\xi_i = \theta\xi_{i-1} + \psi$, where $|\theta| < 1$ is the serial correlation parameter, ψ is normally distributed with mean 0 and variance ζ^2. To derive the properties of the AR(1) errors it is best to express the series in-terms of an infinite-order moving average process $MA(\infty)$ where $\xi_i = \theta\xi_{i-1} + \psi_i = \theta(\theta\xi_{i-2} + \psi_{i-1}) + \psi_i = \theta(\theta[\theta\xi_{i-3} + \psi_{i-2}] + \psi_{i-1}) + \psi_i = \cdots = \theta^i\xi_0 + \sum_{j=0}^{i-1}\theta^j\psi_{i-j} \to \sum_{j=0}^{\infty}\theta^i\psi_{i-j}$ as $i \to \infty$ if $|\theta| < 1$ and ξ_0 is finite. By expressing the AR(1) series in this way we have $Var(\xi_i) = \sum_{j=0}^{\infty}\theta^{2i}\ \zeta^2$ which from an infinite geometric series $\sum_{j=0}^{\infty} ar^k = \frac{a}{1-r}$ for $|r| < 1$ and for series containing only even powers of r, $\sum_{j=0}^{\infty} ar^{2k} = \frac{a}{1-r^2}$. Thus $Var(\xi_i) = \sum_{j=0}^{\infty}\theta^{2i}\,\zeta^2 = \frac{\zeta^2}{1-\theta^2}$, $|\theta| < 1$.

If we set $\sigma^2 = \frac{\zeta^2}{1-\theta^2}$, then the linear regression model with AR(1) errors $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\xi}$ is distributed as $N_n(\boldsymbol{X\beta}, \boldsymbol{R}\sigma^2)$, where $\boldsymbol{R}$ is a $n \times n$ correlation matrix with entry $i, j = \theta^{|j-i|}$.

Given data we believe may be reasonably modelled with an AR(1) process, we can use the `arima` function in R. Using a frequentist approach, the function will estimate the intercept, slope, θ and ζ^2 and produce standard errors which can be used to generate confidence intervals around the point estimates. From a Bayesian perspective we must place prior distributions around each of the parameters ($\beta_0, \beta_1, \zeta, \theta$), and then when we combine with our data and use the appropriate commands in brms (see Appendix 2J), we will then obtain posterior distributions for each of the parameters.

Doi: 10.51224/SRXIV.261 SportRχiv Preprint version 1

Appendix 2: Example problems with R code

```
# load R packages
library(ggplot2)
library(invgamma)
library(cowplot)
```

 # Appendix 2A

```
# Worked Example 1: Section 2.1 Single data point with conjugate prior and known measurement error

# Write function for Bayesian updating, which takes as argument:
# single data point,
# sigma = known measurement error standard deviation
# mean = prior mean for mu
# variance= prior variance for mu

# Function returns posterior distribution for mu which is normally distributed
# with mean and sd.

Singlenormalf = function(datapoint,sigma,mean,variance){
  vstar = 1/((1/variance)+(1/(sigma^2)))
  mstar = vstar*((mean/variance)+(datapoint/(sigma^2)))
  Out = c(round(mstar,1),round(sqrt(vstar),1))
  names(Out)=c("Posterior Mean","Posterior Sd")
  return(Out)}

# Complete the examples in text
# Example 1: Data point = 17.5, known sd = 3, prior mean = 20, prior variance = 16

Singlenormalf(17.5,3,20,16)

# Posterior Mean   Posterior Sd
#      18.4            2.4

# Example 2: Data point = 17.5, known sd = 5, prior mean = 20, prior variance = 64

Singlenormalf(17.5,3,20,64)

# Posterior Mean   Posterior Sd
#      17.8          2.8

# Probabilities: Calculate probability that with the first prior distribution,
# mu is less than 25%, and that mu is between 20 and 25%

round(pnorm(25,18.4,2.4),3)

# 0.997

round(pnorm(25,18.4,2.4)-pnorm(20,18.4,2.4),3)

# 0.250

# Draw Figure 1 which illustrates the different posterior distributions

ggplot(data.frame(x = c(10, 30)), aes(x = x)) +
  stat_function(fun = dnorm, args = list(18.4, 2.4),colour="blue") +
  stat_function(fun = dnorm, args = list(17.8, 2.8),colour="red")+theme_classic() +
  labs(x="Body fat(%)",y="Density")+
  geom_segment(aes(x = 18.4, y = 0, xend = 18.4, yend = 0.163), colour =
"blue",linetype="dashed")+
  geom_segment(aes(x = 17.8, y = 0, xend = 17.8, yend = 0.142), colour =
"red",linetype="dashed")+
  annotate(geom="text", x=22.4, y=0.16,color="blue",
```

```
          label=expression(paste(pi*"(",mu*"|x,",sigma*")=N(18.4,",2.4^2,")")))+
  annotate(geom="text", x=13, y=0.125, color="red",
          label=expression(paste(pi*"(",mu*"|x,",sigma*")=N(17.8,",2.8^2,")")))
```

 # Appendix 2B

```
# Worked Example 2: Section 2.2 Mean of multiple data points with conjugate prior and known measurement error
# Create similar function as to that previously, except now we input a vector of data

Multiplenormalf = function(data,sigma,mean,variance){
  n = length(data)
  xbar = mean(data)
  vstar = 1/((1/variance)+(1/((sigma^2)/n)))
  mstar = vstar*((mean/variance)+(xbar/((sigma^2)/n)))
  Out = c(round(mstar,1),round(sqrt(vstar),1))
  names(Out)=c("Posterior Mean","Posterior Sd")
  return(Out)}


# five values with mean of 17.5, sigma = 3, and prior distribution N(20,4^2)

Multiplenormalf(c(17.5,18.5,19.5,16.5,15.5),3,20,16)

# Posterior Mean   Posterior Sd
#     17.8             1.3

# Draw figure 2, where posterior distribution is calculated with mean of
# 1, 3, 5, and 10 data points.

Mean1 = Multiplenormalf(17.5,3,20,16)
Mean3 = Multiplenormalf(rep(17.5,3),3,20,16)
Mean5 = Multiplenormalf(rep(17.5,5),3,20,16)
Mean10 = Multiplenormalf(rep(17.5,10),3,20,16)

ggplot(data.frame(x = c(10, 30)), aes(x = x)) +
  stat_function(fun = dnorm, args = list(Mean1[1], Mean1[2]),colour="blue") +
  stat_function(fun = dnorm, args = list(Mean3[1], Mean3[2]),colour="red") +
  stat_function(fun = dnorm, args = list(Mean5[1], Mean5[2]),colour="green") +
  stat_function(fun = dnorm, args = list(Mean10[1], Mean10[2]),colour="black") +
  labs(x="Body fat(%)",y="Density")+ theme_classic()+
  geom_segment(aes(x = Mean1[1], y = 0, xend = Mean1[1], yend = 0.168), colour =
"blue",linetype="dashed")+
  geom_segment(aes(x = Mean3[1], y = 0, xend = Mean3[1], yend = 0.25), colour =
"red",linetype="dashed")+
  geom_segment(aes(x = Mean5[1], y = 0, xend = Mean5[1], yend = 0.305), colour =
"green",linetype="dashed")+
  geom_segment(aes(x = Mean10[1], y = 0, xend = Mean10[1], yend = 0.44), colour =
"black",linetype="dashed")+
  annotate(geom="text", x=25, y=0.075,color="blue",
          label=expression(paste(pi*"(",mu*"|x,",sigma*")=N(18.4,",2.4^2,")")))+
  annotate(geom="text", x=22, y=0.22,color="red",
          label=expression(paste(pi*"(",mu*"|x,",sigma*")=N(17.9,",1.6^2,")")))+
  annotate(geom="text", x=21.5, y=0.29,color="green",
          label=expression(paste(pi*"(",mu*"|x,",sigma*")=N(17.8,",1.3^2,")")))+
  annotate(geom="text", x=21, y=0.42,color="black",
          label=expression(paste(pi*"(",mu*"|x,",sigma*")=N(17.6,",0.9^2,")")))

qnorm(0.75,20,4)
```

 # Appendix 2C

```
# Here we calculate the pooled mean and standard deviation from a hypothetical data
# set comprising different sample sizes.
# The sample sizes are

SampleN = c(7,12,20,40,75)
```

```
# The sample means are

SampleMeans = c(22.4, 26.9, 20.8, 23.5, 22.9)

# The sample standard deviations are

SampleSds = c(3.4, 4.1, 3.6, 2.9,3.0)

# Function to calculate pooled mean, variance, sd

PooledMVarf = function(N,means,sds){
  w = length(means)
  PooledM = sum(N*means)/sum(N)
  PooledVar = sum((N-1)*sds^2)/(sum(N)-w)
  PooledSd = sqrt(PooledVar)
  Out = c(round(PooledM,1),round(PooledVar,1),round(PooledSd,1))
  names(Out)=c("Pooled Mean","Pooled Variance","Pooled Sd")
  return(Out)}

PooledMVarf(SampleN,SampleMeans,SampleSds)

# Pooled Mean     Pooled Variance       Pooled Sd
#      23.1              10.0                3.2
```

 # Appendix 2D

```
# Here we develop a normal inverse gamma prior distribution for mu and sigma^2

# We start be identifying that we believe sigma is likely to lie between 2 and 6,
# with median at 3. This equates to variances of 4, 9, and 36 respectively.
# So we want to fit an inverse gamma distribution where the 0.25-, 0.5-, and 0.75-quantiles
# are 4, 9 and 36.

# As a relatively simple means of identifying parameter values which create this distribution
# we create a grid and fit sum of squares to find the lowest values.

# Create grid

IGGrid = expand.grid(seq(0.2,3,0.1),seq(1,20,0.1))

IGGridSS = c(NULL)
for(i in 1:length(IGGrid[,1])){
  IGGridSS[i] = sqrt((pinvgamma(4,shape=IGGrid[i,1],rate=IGGrid[i,2])-0.25)^2 +
                       (pinvgamma(9,shape=IGGrid[i,1],rate=IGGrid[i,2])-0.5)^2 +
                       (pinvgamma(36,shape=IGGrid[i,1],rate=IGGrid[i,2])-0.75)^2)
}

# Find minimum

IGGridSSMin = IGGrid[order(IGGridSS)[1],]
IGGridSSMin

# a = 0.6  b= 3.1

# Check actual quantiles

pinvgamma(4,shape=IGGridSSMin[[1]],rate=IGGridSSMin[[2]])
pinvgamma(9,shape=IGGridSSMin[[1]],rate=IGGridSSMin[[2]])
pinvgamma(36,shape=IGGridSSMin[[1]],rate=IGGridSSMin[[2]])


# Plot curve for variance

set.seed(123)
GammaVardf=data.frame(x=rinvgamma(10000,shape=IGGridSSMin[[1]],rate=IGGridSSMin[[2]])
)
```

```
ggplot(GammaVardf, aes(x=x)) +  geom_density(adjust=1.5) +
  scale_x_continuous(limits=c(0,100),breaks=seq(0,100,25))+ theme_classic()

# Plot curve for sd

set.seed(123)
SdDist=data.frame(x=sqrt(rinvgamma(10000,shape=IGGridSSMin[[1]],rate=IGGridSSMin[[2]]
)))

ggplot(SdDist, aes(x=x)) +  geom_density(adjust=2) +
  scale_x_continuous(limits=c(0,20),breaks=seq(0,20,5))+ theme_classic()

quantile(SdDist$x,seq(0.1,0.9,0.1))

# Now that we are happy with the distribution for sigma, we want to select a typical value
# to scale k. We investigate the quantiles

quantile(GammaVardf$x,seq(0.4,0.6,0.01))

# We select median value of 10, and set 10/k = 16, hence k =

10/16

# 0.625

# We compare our marginal prior for mu using the NIG, with our original normal prior distribution N(20,4^2)
# First we sample from the marginal

set.seed(123)
NIGMu = rnorm(10000,20,sqrt(GammaVardf$x/0.625))
NormalNIGDF = data.frame(Mu=c(NIGMu,rnorm(10000,20,4)),
                         Distribution=c(rep("NIG",10000),rep("Normal",10000)))

NIGPrior = ggplot(NormalNIGDF, aes(x=Mu,color=Distribution)) +  geom_density(adjust=3)
+
  scale_x_continuous(limits=c(0,60),breaks=seq(0,60,5))+
  labs(x="Mean body fat(%)",y="Density")+ theme_classic()
NIGPrior

# We can see a large amount of overlap but heavier tails with NIG compared to Normal

# Appendix 2E
# Given the NIG prior and data we update for NIG posterior
# First we write our update function

# NIG Update

NIGUpdate = function(a,b,m,k,data){
  n = length(data)
  xbar = mean(data)
  aupdate = a + (n/2)
  bupdate = b + 0.5*(sum((data-xbar)^2) + ((k*n*(xbar-m)*(xbar-m))/(k+n)))
  mupdate = (k*m + n*xbar)/(k+n)
  kupdate = k + n
  Out = c(aupdate,bupdate,mupdate,kupdate)
  names(Out)=c("a*","b*","m*","k*")
  return(Out)}


NIGU = NIGUpdate(a=0.6,b=3.1,m=20,k=0.625,data=17.5)
round(NIGU,2)

#  a*    b*    m*    k*
# 1.10  4.30  18.46  1.62
```

```
# We compare our marginal posterior for mu, with our original normal posterior distribution N(18.4,2.4^2)
# First we sample from the marginal

set.seed(10000)
GammaVarpost = rinvgamma(10000,shape=NIGU[[1]],rate=NIGU[[2]])
NIGPostMu = rnorm(10000,NIGU[[3]],sqrt(GammaVarpost/NIGU[[4]]))
NormalNIGPostDF = data.frame(Mu=c(NIGPostMu,rnorm(10000,18.4,2.4)),
                             Distribution=c(rep("NIG",10000),rep("Normal",10000)))

NIGPosterior      =ggplot(NormalNIGPostDF,      aes(x=Mu,color=Distribution))      +
geom_density(adjust=3) +
  scale_x_continuous(limits=c(0,40),breaks=seq(0,40,5))+
  labs(x="Mean body fat(%)",y="Density")+ theme_classic()

NIGPosterior

# Combine prior and posterior comparison plots

plot_grid(NIGPrior+theme(legend.position  =  "none"),  NIGPosterior,  labels  =  c('A',
'B'))
```

 # Appendix 2F

```
# Here we show that the marginal for mu with a NIG prior is a t-distribution
# with parameters mean = m*, df = 2a*, scale = sqrt(b*/(a*k*))
# We compare with the data above (NIGPostMu)

# First an intro to the generalised t-distribution which has a mean (mu), degrees of freedom (df)
# and scale parameter (sigma_hat) such that sd = sigma_hat*sqrt(df/df-2)

# demonstrate the above

set.seed(123)
t3 = rt(100000,3)
round(mean(t3),1)
round(sd(t3),1)
round(sqrt(3/(3-2)),1)

# Shift t distribution by 100 and multiply by sigma_hat = 10

t3.100.10 = 100+ 10*t3
round(mean(t3.100.10),1)
round(sd(t3.100.10),1)
round(10*sqrt(3/(3-2)),1)

# Show this is the same from brms t-distribution sampler

library(brms)
set.seed(123)
t3.100.10.sim = rstudent_t(100000, df=3, mu = 100, sigma = 10)
round(mean(t3.100.10.sim),1)
round(sd(t3.100.10.sim),1)

# Calculate the marginal t-distribution given updated NIG parameters

Marginalt= function(a,b,m,k,data){
  Posterior = NIGUpdate(a,b,m,k,data)
  mean = Posterior[[3]]
  df = 2*Posterior[[1]]
  scale = sqrt(Posterior[[2]]/
                 (Posterior[[1]]*Posterior[[4]]))
  Out = c(mean,df,scale)
  names(Out)=c("Mean","df","Scale")
  return(Out)}

TMarginat = Marginalt(a=0.6,b=3.1,m=20,k=0.625,data=17.5)
```

```
set.seed(123)
TMarginatsim = rstudent_t(10000, df=TMarginat[[2]], mu = TMarginat[[1]], sigma =
TMarginat[[3]])


# compare distributions

NIGPostMarginalTDF = data.frame(Mu=c(NIGPostMu,TMarginatsim),
                                Distribution=c(rep("NIG",10000),rep("T",10000)))

ggplot(NIGPostMarginalTDF, aes(x=Mu,color=Distribution)) +  geom_density(adjust=3) +
  scale_x_continuous(limits=c(0,40),breaks=seq(0,40,5))+
  labs(x="Mean body fat(%)",y="Density")+ theme_classic()
```

 # Appendix 2G

```
# Here we use brms to recalculate results from section 2.
# Single data point known measurement error
# Create data frame

Data = data.frame(y=17.5)

# create normal prior for mu and set sigma to known 3%

Prior1 = c(
  prior(normal(20,4), class = Intercept),
  prior(constant(3),  class = sigma))

# Run model

set.seed(123)
Model1 = brm(y~1, data = Data, prior = Prior1, family = gaussian,
             chains = 4, iter = 5000, warmup = 2500)
# Plot results
plot(Model1)

# Load library to obtain draws from posterior distribution

library(posterior)

# obtain posterior sample

Posterior1 = as_draws_df(Model1)
Posterior1Mu = Posterior1$b_Intercept

# calculate mean of mu

round(mean(Posterior1Mu),1)

# 18.4
# calculate sd of mu

round(sd(Posterior1Mu),1)

# 2.4
# calculate probability mu is less than 25

round(mean(Posterior1Mu<25),3)

# 0.996
# calculate probability mu is between 20 and 25

round(mean(Posterior1Mu>20&Posterior1Mu<25),3)

# Next example with multiple data points and known measurement error
```

```
Data2 = data.frame(y=rep(17.5,5))

set.seed(111)
Model2 = brm(y~1, data = Data2, prior = Prior1, family = gaussian,
             chains = 4, iter = 5000, warmup = 2500)

Posterior2 = as_draws_df(Model2)
Posterior2Mu = Posterior2$b_Intercept
round(mean(Posterior2Mu),1)

# 17.7

round(sd(Posterior2Mu),1)

# 1.3

# Next example with multiple data points and uniform prior for sigma
# First we load HDInterval to calculate credible intervals

library(HDInterval)
set.seed(123)
Data3 = data.frame(y=rnorm(5,19,3))

Prior2 = c(
  prior(normal(20,4), class = Intercept),
  prior(uniform(2,6),lb=2,ub=6, class = sigma))

Model3 = brm(y~1, data = Data3, prior = Prior2, family = gaussian,
             chains = 4, iter = 5000, warmup = 2500)

Posterior3 = as_draws_df(Model3)
Posterior3Mu = Posterior3$b_Intercept
Posterior3sigma = Posterior3$sigma
hdi(Posterior3Mu,credMass = 0.75)

#   lower    upper
# 18.00440 21.10619

hdi(Posterior3sigma,credMass = 0.75)
#lower     upper
# 2.000478  3.947395
```

 # Appendix 2H

```
# Here we give multiple examples with different prior elicitations and different data generating mechanisms
# (e.g. errors distributed according to Gaussian or t-distributions).
# First we start with the gamma prior for measurement error standard deviation.
# Software exists to elicit this parameter and can be  downloaded from
#https://www.open.ac.uk/stem/mathematics-and-statistics/research/research-groups/statistics/bayesian-statistics/elicitation-prior-distributions
# Here we start with a simple example where we believe the 1st, 2nd and 3rd quartiles are 1.5, 3, 6
# We find suitable parameters with a grid search with a = 1 to 6; b = 0.1 to  5
# We use the pgamma function and calculate sum of squares

# Create grid

IGGrid2 = expand.grid(seq(1,6,0.1),seq(0.1,5,0.1))

IGGrid2SS = c(NULL)
for(i in 1:length(IGGrid2[,1])){
  IGGrid2SS[i] = sqrt((pgamma(1.5,shape=IGGrid[i,1],rate=IGGrid[i,2])-0.25)^2 +
                      (pgamma(3,shape=IGGrid[i,1],rate=IGGrid[i,2])-0.5)^2 +
                      (pgamma(6,shape=IGGrid[i,1],rate=IGGrid[i,2])-0.75)^2)
}
```

```
# Find minimum

IGGrid2SSMin = IGGrid2[order(IGGrid2SS)[1],]
IGGrid2SSMin

# a = 1.3, b = 0.3

# Now we actually check the quantile values with the parameters and
# they should be close to 0.25, 0.5 and 0.75.

pgamma(1.5,shape=IGGrid2SSMin[[1]],rate=IGGrid2SSMin[[2]])
pgamma(3,shape=IGGrid2SSMin[[1]],rate=IGGrid2SSMin[[2]])
pgamma(6,shape=IGGrid2SSMin[[1]],rate=IGGrid2SSMin[[2]])

# We can see that the parameter values give us the desired quartiles.
# We can then graph the distribution and check further quantiles for feedback

set.seed(123)
GammaSDdf=data.frame(x=rgamma(10000,shape=IGGrid2SSMin[[1]],rate=IGGrid2SSMin[[2]]))
ggplot(GammaSDdf, aes(x=x)) +  geom_density(adjust=1.5) +
  scale_x_continuous(limits=c(0,20),breaks=seq(0,20,2))+ theme_classic()

# Checking toward the ends of the distribution

qgamma(0.1,shape=IGGrid2SSMin[[1]],rate=IGGrid2SSMin[[2]])
qgamma(0.9,shape=IGGrid2SSMin[[1]],rate=IGGrid2SSMin[[2]])

# Based on this feedback, say we want to pull the upper end in slightly then we
# can refit the distribution but set the value of 6% sd to 0.85 and not 0.75
# refit

IGGrid2 = expand.grid(seq(1,6,0.1),seq(0.1,5,0.1))

IGGrid2aSS = c(NULL)
for(i in 1:length(IGGrid2[,1])){
  IGGrid2aSS[i] = sqrt((pgamma(1.5,shape=IGGrid[i,1],rate=IGGrid[i,2])-0.25)^2 +
                         (pgamma(3,shape=IGGrid[i,1],rate=IGGrid[i,2])-0.5)^2 +
                         (pgamma(6,shape=IGGrid[i,1],rate=IGGrid[i,2])-0.85)^2)
}



# Find minimum
IGGrid2aSSMin = IGGrid2[order(IGGrid2aSS)[1],]
IGGrid2aSSMin

# a = 1.8, b = 0.5

# Now we actually check the quantile values with the parameters and
# they should be close to 0.25, 0.5 and 0.85.

pgamma(1.5,shape=IGGrid2aSSMin[[1]],rate=IGGrid2aSSMin[[2]])
pgamma(3,shape=IGGrid2aSSMin[[1]],rate=IGGrid2aSSMin[[2]])
pgamma(6,shape=IGGrid2aSSMin[[1]],rate=IGGrid2aSSMin[[2]])

# We can see that the parameter values give us the desired quantiles.
# We can then graph the distribution and check further quantiles for feedback

set.seed(123)
GammaSDadf=data.frame(x=rgamma(10000,shape=IGGrid2aSSMin[[1]],rate=IGGrid2aSSMin[[2]]
))
ggplot(GammaSDadf, aes(x=x)) +  geom_density(adjust=1.5) +
  scale_x_continuous(limits=c(0,20),breaks=seq(0,20,2))+ theme_classic()
```

```
# Checking toward the ends of the distribution

qgamma(0.1,shape=IGGrid2aSSMin[[1]],rate=IGGrid2aSSMin[[2]])
qgamma(0.9,shape=IGGrid2aSSMin[[1]],rate=IGGrid2aSSMin[[2]])

# We now use this distribution and combine with repeated data to obtain inferences
# for both true score and measurement error standard deviation

set.seed(123)
Data4 = data.frame(y=rnorm(5,19,3))

Prior3 = c(
  prior(normal(20,4), class = Intercept),
  prior(gamma(1.8,0.5),class = sigma))

Model4 = brm(y~1, data = Data4, prior = Prior3, family = gaussian,
             chains = 4, iter = 5000, warmup = 2500)

Posterior4 = as_draws_df(Model4)
Posterior4Mu = Posterior4$b_Intercept
hdi(Posterior4Mu,credMass = 0.75)

#  lower  upper
#   18.31  20.97

Posterior4sigma = Posterior4$sigma
hdi(Posterior4sigma,credMass = 0.75)

#  lower  upper
#  1.62   3.64

# As a second example we now look at observational data where we obtain some large errors

set.seed(123)
Data5 = data.frame(y=c(16.5,17.5,17.5,18.5,29,30))
Model5 = brm(y~1, data = Data5, prior = Prior2, family = gaussian,
             chains = 4, iter = 5000, warmup = 2500)

Posterior5 = as_draws_df(Model5)
Posterior5Mu = Posterior5$b_Intercept
hdi(Posterior5Mu,credMass = 0.75)

#  lower  upper
#   19.1  23.1

# Because of these large errors, we want to update our data generating mechanism from
# a normal distribution to a t-distribution. This is achieved by changing the family
# from gaussian to student. We also have to update our prior distributions.
# However, first we use the default weakly informative priors for the degrees of
freedom and scale.

Prior4 = c(
  prior(normal(20,4), class = Intercept),
  prior(gamma(2,0.1),class = nu),
  prior(student_t(3, 0, 3.1),class = sigma))
set.seed(123)
Model6 = brm(y~1, data = Data5, prior = Prior4, family = student,
             chains = 4, iter = 5000, warmup = 2500)

Posterior6 = as_draws_df(Model6)
Posterior6Mu = Posterior6$b_Intercept
hdi(Posterior6Mu,credMass = 0.75)

#  lower  upper
#   18.1  23.1
```

```
Posterior6sigma = Posterior6$sigma
hdi(Posterior6sigma,credMass = 0.75)

#  lower  upper
#  3.5   7.2

Posterior6df = Posterior6$nu
hdi(Posterior6df,credMass = 0.75)

#  lower upper
#  1.7   27.4
```

Finally we include an informative prior distribution for sigma, the scale parameter

```
Prior5 = c(
  prior(normal(20,4), class = Intercept),
  prior(gamma(2,0.1),class = nu),
  prior(gamma(1.8,0.5),class = sigma))
set.seed(123)
Model7 = brm(y~1, data = Data5, prior = Prior5, family = student,
             chains = 4, iter = 5000, warmup = 2500)

Posterior7 = as_draws_df(Model7)
Posterior7Mu = Posterior7$b_Intercept
hdi(Posterior7Mu,credMass = 0.75)

#  lower  upper
#   18.4  23.3

Posterior7sigma = Posterior7$sigma
hdi(Posterior7sigma,credMass = 0.75)

#  lower  upper
#  3.7   7.2

Posterior7df = Posterior7$nu
hdi(Posterior7df,credMass = 0.75)

#  lower upper
#  1.1   26.7
```

 # Appendix 2I

Here we give multiple examples to model true change scores.
Simple Bayesian approach to quantifying uncertainty in change scores

observed change score of -3.5%

```
Data7 = data.frame(y=-3.5)
```

Informative priors that on average participants reduce body fat percentage by
#1.5% and the standard deviation of the change score is 3%
Also we believe that the measurement error standard deviation is somewhere between
2 and 6, such that the change score standard deviation is between sqrt(2)*2 and sqrt(2)*6

```
Prior6 = c(
  prior(normal(-1.5,3), class = Intercept),
  prior(uniform(sqrt(2)*2, sqrt(2)*6),lb= sqrt(2)*2,ub= sqrt(2)*6,
        class = sigma))
set.seed(123)
Model8 = brm(y~1, data = Data7, prior = Prior6, family = gaussian,
             chains = 4, iter = 5000, warmup = 2500)

Posterior8 = as_draws_df(Model8)
Posterior8Mu = Posterior8$b_Intercept
```

```
hdi(Posterior8Mu,credMass = 0.75)

#     lower     upper
# -5.0695250  0.7481412

mean(Posterior8Mu<=-0.5)

# 0.7334

# We now perform the above analysis but with more data and adjusting the informative prior
# on the change standard deviation

Data8 =data.frame(y=mean(c(15.2,17.1,16.4,18.0))-mean(c(20.2, 21.4, 18.2, 20.9)))

Prior7 = c(
  prior(normal(-1.5,3), class = Intercept),
  prior(uniform(sqrt(2)*2/2, sqrt(2)*6/2),lb= sqrt(2)*2/2,ub= sqrt(2)*6/2,
        class = sigma))
set.seed(123)
Model9 = brm(y~1, data = Data8, prior = Prior7, family = gaussian,
             chains = 4, iter = 5000, warmup = 2500)

Posterior9 = as_draws_df(Model9)
Posterior9Mu = Posterior9$b_Intercept
hdi(Posterior9Mu,credMass = 0.75)

#     lower     upper
# -5.0281567 -0.4545785

mean(Posterior9Mu<=-0.5)

# 0.8548
```

 # Appendix 2J

```
# Autoregressive Modelling

# We assume that a single measurement is made each day for 12 weeks and we model errors from
# AR(1) with autocorrelation of 0.3 and sd equal to 3.5
# Note that this causes an overall variation of

sqrt(3.5^2/(1-0.3^2))

# 3.668997

# Build AR(1)

set.seed(112)
UBF0.3 = WBF0.3 =rnorm(85,0,3.5)
for(i in 2:85){
  UBF0.3[i] = 0.3*UBF0.3[i-1] + WBF0.3[i]}
sd(UBF0.3)

# Add linear regression to AR(1) Errors
# We assume that the individual starts with a BF of 19% and over the
# course of the 12 weeks looses 1%.

BFPercent0.3 = c(NULL)

for(i in 1:85){
  BFPercent0.3[i]=19 - ((i-1)*(1/84)) + UBF0.3[i]
}

# Plot data

BFPercentDF = data.frame(Obs0.3 = BFPercent0.3,
```

```
                          Week = seq(0,12,1/7),
                          True = seq(19,18,-(1/84)))

ggplot(BFPercentDF, aes(x=Week,Obs0.3)) +
  geom_point() + theme_classic() +
  geom_abline(intercept = 19, slope = -1/12,color="red")+
  scale_x_continuous(limit = c(0,12), breaks = seq(0,12,1))+
  scale_y_continuous(limit = c(5,35),breaks = seq(5,35,5))+
  theme_classic()+ xlab("Week") + ylab("Body fat (%)")

# First we analyse the data using a frequentist approach and the arima function in R for AR(1) models

arima(BFPercentDF$Obs0.3, xreg=seq(0,12,1/7), order=c(1,0,0))
# Coefficients:
#       ar1        intercept      seq(0, 12, 1/7)
#     0.3610       20.2548         -0.1796
# s.e.  0.1011     1.1076           0.1589

# sigma^2 estimated as 11.2

# Using the point estimate and standard error we calculate a 95% true score change CI

-0.1796*12
12*(-0.1796-(qnorm(0.975)*0.1589))
12*(-0.1796+(qnorm(0.975)*0.1589))

# We now complete our Bayesian analysis. We believe that the intercept is normally distributed according to
# N(18.4,2.4^2). We believe that the weekly change can be described by the normal distribution N(-0.125, 0.25^2)
# We believe that the measurement error standard deviation can be described by gamma(a=3.9, b=1)
# Finally we believe that the serial correlation parameter lies between 0.2 and 0.5 and use a uniform prior.


Prior8 = c(
  prior(normal(18.4,2.4), class = Intercept),
  prior(normal(-0.125,0.25), class = "b", coef="Week"),
  prior(gamma(3.9, 1),class = sigma),
  prior(uniform(0.2, 0.5),lb= 0.2,ub= 0.5,class="ar"))

set.seed(123)
Model10 = brm(Obs0.3 ~Week + ar(p = 1),data=BFPercentDF,
              prior = Prior8, family = gaussian,
              chains = 4, iter = 5000, warmup = 2500)

plot(Model10)
Posterior10 = as_draws_df(Model10)
Posterior10Mu = Posterior10$b_Intercept
hdi(Posterior10Mu,credMass = 0.75)

#   lower   upper
#   19.0   21.3

Posterior10Week = Posterior10$b_Week
hdi(Posterior10Week,credMass = 0.75)

#   lower      upper
# -0.319       -0.000

Posterior10Total = 12*Posterior10Week
hdi(Posterior10Total,credMass = 0.75)

#    lower     upper
#   -3.83     -0.004
mean(Posterior10Total<=-0.5)

# 0.8018
```

```
hdi(Posterior10$sigma,credMass = 0.75)

#   lower   upper
#  3.092    3.703

hdi(Posterior10$`ar[1]`,credMass = 0.75)

#  lower    upper
# 0.294    0.473
```