Neuroscience Informatics 5 (2025) 100195

Contents lists available at ScienceDirect

Neuroscience Informatics

journal homepage: www.elsevier.com/locate/neuri

Original article

Bayesian Inference General Procedures for A Single-subject Test study

Jie Li [a,b,*], Gary Green [b,c], Sarah J.A. Carr [b], Peng Liu [d], Jian Zhang [a]

[a] School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, UK
[b] Innovision-IP Ltd., 50 Seymour Street, London, W1H 7JG, UK
[c] York Neuroimaging Centre, University of York, Innovation Way, York, YO10 5NY, UK
[d] Department of Mathematical Sciences, Loughborough University, Loughborough, LE11 3TU, UK

ARTICLE INFO

Article history:
Received 29 January 2025
Received in revised form 6 March 2025
Accepted 7 March 2025

Keywords:
Bayesian inference
Skewed Student t distribution
Single-subject test
Magnetoencephalography (MEG)
Jeffreys prior

ABSTRACT

Abnormality detection in identifying a single-subject which deviates from the majority of a control group dataset is a fundamental problem. Typically, the control group is characterised using standard Normal statistics, and the detection of a single abnormal subject is in that context. However, in many situations, the control group cannot be described by Normal statistics, making standard statistical methods inappropriate. This paper presents a Bayesian Inference General Procedures for A Single-subject Test (BIGPAST) designed to mitigate the effects of skewness under the assumption that the dataset of the control group comes from the skewed Student t distribution. BIGPAST operates under the null hypothesis that the single-subject follows the same distribution as the control group. We assess BIGPAST's performance against other methods through simulation studies. The results demonstrate that BIGPAST is robust against deviations from normality and outperforms the existing approaches in accuracy, nearest to the nominal accuracy 0.95. BIGPAST can reduce model misspecification errors under the skewed Student t assumption by up to 12 times, as demonstrated in Section 3.3. We apply BIGPAST to a Magnetoencephalography (MEG) dataset consisting of an individual with mild traumatic brain injury and an age and gender-matched control group. For example, the previous method failed to detect abnormalities in 8 brain areas, whereas BIGPAST successfully identified them, demonstrating its effectiveness in detecting abnormalities in a single-subject.

1. Introduction

Abnormality detection is a fundamental task in many scientific fields, including Medicine, Psychology, Econometrics, Telehealthcare, Neuroscience as well as other fields for Industry 5.0 [1–6]. One example of such abnormality detection lies in the detection of traumatic brain injury (TBI), which is a fundamental problem, with over 2.8 million individuals a year in America presenting at an Emergency room [7]. We exemplify the use of skewed Student t statistics in the study of brain activity as a potential biomarker for the identification of a brain injury. We use magnetoencephalography (MEG) to measure brain activity and compare a single individual to normal controls from age and gender-matched cohorts. Existing literature typically assumes that the dataset from the control group follows a Normal distribution. However, extensive practical evidence shows that the Normal distribution assumption is often violated. This violation is frequently observed in brain Magnetoencephalography (MEG) datasets. Therefore, the current approaches using Normal statistics could not be directly applied. Unlike the existing literature, we assume that the dataset from the control group belongs to a more general distributional family - the skewed Student t distribution, which includes the Normal distribution and Student t as special cases. The benefit of this skewed Student t distributional assumption is that it could cover a variety of empirical behaviours and is thus more suitable for practical needs. This generalisation is essential since many brain Magnetoencephalography (MEG) datasets show a significant level of asymmetry in the observation distribution and hence are highly skewed. It is also the case that the high skewness not only appears in a MEG dataset but also in other fields, such as Cybersecurity [4], Medicine [5], Econometric [3], Neuroscience [8], Telehealthcare [2] and Internet of Thing (IoT) [6] to name only a few.

The main focus of this research is the adaptability of statistical models under the skewed student t distribution. Specifically, the statistical challenges arise from the interplay between the mechanism of data observation from a single-subject, the type of hypothesis (two-sided, less, greater), and the sign of the skewness parameter, which we will discuss one by one in the following paragraphs.

* Corresponding author.
E-mail address: jl725@kent.ac.uk (J. Li).

https://doi.org/10.1016/j.neuri.2025.100195

Table 1
The data generation mechanism of single-subject observations in Crawford et al. [10], Crawford and Garthwaite [11] respectively. 'NA' means that the observations in the cell are not available. 'Positive' means the single-subject is claimed as abnormal, while 'Negative' means the single-subject is claimed as normal.

		Predicated	
		Positive	Negative
Actual	Positive	NA	NA
	Negative	n_{01}	n_{00}

(a) Type I error evaluation in Crawford et al. [10]

		Predicated	
		Positive	Negative
Actual	Positive	n_{11}	n_{10}
	Negative	NA	NA

(b) Power evaluation in Crawford and Garthwaite [11]

The mechanism of data observation in Type I error and Power test. Under the assumption that the dataset observed from the control group follows a Normal distribution, several methods exist to compare a single-subject against a control group. Crawford and Howell [9] proposed a t-score test defined as follows:

$$t = \frac{x^* - \bar{x}}{s/\sqrt{(n+1)/n}},$$

where x^* represents the observation of the single-subject, $\bar{x}$ denotes the average of n observations from the control group, and s stands for the standard deviation of the dataset from the control group. The test score t follows a Student t distribution with $n-1$ degrees of freedom. Let β be the prespecified significance level, define the p value as $\mathbb{P}(t > |t_{1-\beta/2}(n-1)|)$ where $t_{1-\beta/2}(n-1)$ is the $(1-\beta/2)$-th critical value of Student t distribution with $n-1$ degrees of freedom. If the p value is less than β, then one can claim that x^* has an abnormality. Moreover, under the assumption of a skewed Student t distribution, Crawford et al. [10] examined the effects of departures from normality on testing for a deficit in a single-subject via using a Leptokurtic distribution with different skewness parameters. They asserted that the t score [9] outperforms the z score in terms of Type I error, and the t score shows robustness as good as the result under the Normal assumption. Additionally, Crawford and Garthwaite [11] investigated the statistical power of testing for an abnormality in a single-subject using Monte Carlo simulations. They found that the power of t score performs slightly worse than the Normal assumption results.

The simulation results for Type I error in Crawford et al. [10] and power in Crawford and Garthwaite [11] led to the assertion that the t score [9] is robust even when the underlying distribution is skewed Student t distribution. However, the simulation settings for generating the samples for a single-subject in Crawford et al. [10], Crawford and Garthwaite [11] are not comprehensive to evaluate the performance of the t score under skewed Student t distribution. The samples of the single-subject may come from either the distribution of the control group or a different distribution. Let $c : d$ be the ratio of the samples from the control group to those from a different distribution. In Crawford et al. [10], the ratio is $100 : 0$; and in Crawford and Garthwaite [11], the ratio is $0 : 100$. Both are special cases for evaluating the Type I error and Power, respectively.

For $\mathbf{c} : \mathbf{d} = \mathbf{100} : \mathbf{0}$, Crawford et al. [10] generates all single-subject observations from the distribution of the control group to evaluate the Type I error only, see Table 1a. In Table 1a, n_{01} is the number of false positives (FP) while n_{00} is the number of true negatives (TN). The false positive rate (FPR, i.e., Type I error) is estimated as $n_{01}/(n_{01}+n_{00})$. As there are no actual positive observations in Table 1a, the true positive rate (TPR, i.e., power) can not be estimated given the simulation settings in Crawford et al. [10]. Therefore, the comparison of Type I error rates in Crawford et al. [10] is not thoroughly discussed, as the study did not control for statistical power. Conversely, **for $\mathbf{c} : \mathbf{d} = \mathbf{0} : \mathbf{100}$**, Crawford and Garthwaite [11] generate all single-subject observations from a distribution that is different from the distribution of the control group to evaluate the power only, see Table 1b. In Table 1b, n_{11} is the number of true positives (TP) while n_{01} is the number of false negatives (FN). The true positive rate (TPR, i.e., power) is estimated as $n_{11}/(n_{11}+n_{10})$. However, the Type I error is unavailable based on the observations. Again, Crawford and Garthwaite [11] did not fully discuss the comparison of the power evaluation due to the absence of Type I error. Neither Crawford et al. [10] nor Crawford and Garthwaite [11] discussed that the ground truth observations of a single-subject are well-mixed. For instance, if 50% of the single-subject observations are generated from the underlying distribution of the control group data and the other 50% of observations come from a different distribution, the Type I error and power of the t score test would differ significantly. Direct evidence illustrated in Fig. 2c shows that the median of powers of the Crawford-Garthwaite (CG) method and BIGPAST under two-sided test are around 0.62 and 0.96, respectively. To address this issue, all the simulation settings (unless specified) of the single-subject in this paper are well-mixed, i.e., the single-subject observations are generated from a mixture of the null and alternative distribution. The well-mixed setting is more realistic and can provide a comprehensive evaluation of the performance of BIGPAST test under skewed Student t assumption.

Before delving into the rationale behind the assumption of the skewed Student t distribution, it is essential to review the definitions of heavy tails and light tails within the context of the skewed Student t distribution. Let $F(x)$ be the cumulative distribution function of a random variable X. The right tail distribution is denoted as $\overline{F}(x) := \Pr(X > x)$ while the left tail distribution is denoted as $\underline{F}(x) := \Pr(X \le x)$. F is said to have a right heavy tail if $\lim_{x\to+\infty} \exp(tx)\overline{F}(x) = \infty$ for all $t > 0$, and have a right light tail if $\exp(tx)\overline{F}(x) = O(1)$ for large positive x and some constant $t > 0$. Similarly, F is said to have a left heavy tail if $\lim_{x\to-\infty} \exp(tx)\underline{F}(x) = \infty$ for all $t > 0$, and have a left light tail if $\exp(tx)\underline{F}(x) = O(1)$ for small negative x and some constant $t > 0$. In other words, light tails decay to zero much faster than the exponential distribution, whereas heavy tails decay much slower, see Fig. 1.

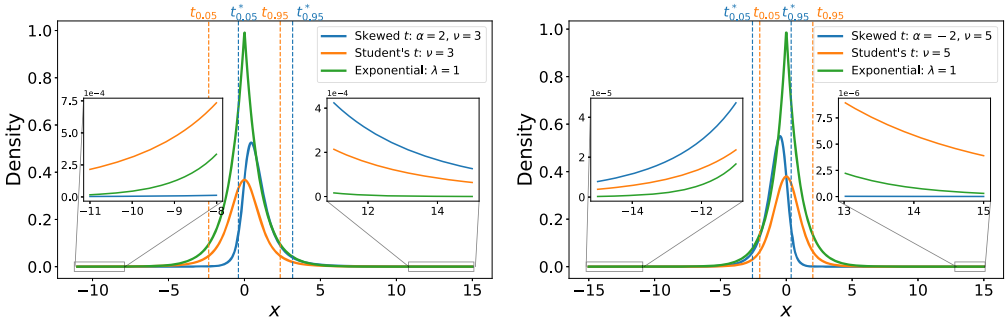
It is well-known that the Student t distribution is symmetric and has two heavy tails. The skewed Student t distribution is a generalisation of the Student t distribution [12]. Let $f(z|\alpha,\nu)$ be the skewed Student t density function [13], defined as follows:

$$f(z|\alpha,\nu) := 2t(z|\nu)T(w|\nu+1), \quad z \in \mathbb{R}, \alpha \in \mathbb{R}, \nu > 0, \tag{1}$$

where $r(z,\nu) = \sqrt{(\nu+1)/(\nu+z^2)}$ and $w = \alpha z r(z,\nu)$, $t(\cdot|\nu)$ represents the density function of the t distribution with degrees of freedom ν, and $T(\cdot|\nu+1)$ represents the cumulative density function of the t distribution with degrees of freedom $\nu+1$. Let ξ and ω be the location and scale parameters, with the substitution $z = (x-\xi)/\omega$ in Equation (1), then the skewed Student t density function is given by

$$f(x|\alpha,\nu,\xi,\omega) := \frac{1}{\omega} f\left(\frac{x-\xi}{\omega}|\alpha,\nu\right). \tag{2}$$

The skewness parameter α changes the symmetry of Student t distribution given the degrees of freedom ν. When $\alpha = 0$, ν is finite, the skewed Student t distribution reduces to Student t distribution which has heavy tails. When $\alpha = 0$ and $\nu \to +\infty$, the skewed Student t distribution reduces to the Normal distribution



(a) Positive skewness: $\alpha = 2$ with $\nu = 3$ (b) Negative skewness: $\alpha = -2$ with $\nu = 5$

Fig. 1. The heavy and light tails of the standard skewed Student t distribution for both positive and negative skewness parameters. The orange lines represent the standard Student t density function, the green lines represent the curve of $\exp(-|x|)$. The vertical lines are quantiles: $t_{0.05}, t_{0.95}$ (resp. $t^*_{0.05}, t^*_{0.95}$) are 5% and 95% quantiles of the standard Student t distribution (resp. the standard skewed Student t distribution). In (a), $t_{0.05} = -2.353, t_{0.95} = 2.353$, $t^*_{0.05} = -0.398, t^*_{0.95} = 3.169$. In (b), $t_{0.05} = -2.015, t_{0.95} = 2.015$, $t^*_{0.05} = -2.568, t^*_{0.95} = 0.370$.

which has lighter tails; When $|\alpha| \to \infty$ and $\nu < +\infty$, the skewed Student t distribution reduces to the half Student t distribution which can be tuned to allow for a heavy tail ($\nu \to 1$) or a light tail (large ν). Therefore, for other settings of (α, ν), the skewed Student t distribution may have one light tail and one heavy tail. The sign of α dictates whether the heavy tail is on the left or right side of the distribution, see Fig. 1. In Fig. 1a, the skewness is positive, and the standard skewed Student t distribution has a right heavy tail and a left light tail. Conversely, in Fig. 1b, the skewness is negative, the standard skewed Student t distribution has a right light tail and a left heavy tail, see the zoomed insets in Figs. 1a and 1b respectively.

The effect of hypothesis (two-sided, less, greater). In this paper, we consider three kinds of alternative hypotheses: "less than" ($<$), "greater than" ($>$) and "two-sided". For clarity, we refer to the alternative hypothesis with symbol ($<$) as the *one-sided test (less)*, and the alternative hypothesis with symbol ($>$) as the *one-sided test (greater)*. Under the assumption of symmetric distributions (e.g., Normal or Student t), one-sided tests, either "less" or "greater", share the same absolute critical value given the same significance level. However, under the assumption of the skewed Student t distribution, the absolute critical value of the one-sided test (less) is different from the absolute critical value of the one-sided test (greater) due to the relationship between skewness and heavy tails. Hence, under the assumption of a skewed Student t distribution, there are six combinations of hypothesis tests: the one-sided test (less), the one-sided test (greater), and the two-sided test, each with either a positive or negative skewness parameter.

The interaction effect of hypothesis type and skewness sign. Let us look into the parameter settings in Crawford et al. [10]. They set $\alpha < 0$ and only considered the one-sided test (less). The observations of the control group and single-subject are generated from the same skewed Student t distribution with the same skewness parameter. Furthermore, the observations of the single-subject are all negative, as described in Table 1a. They used the t statistic to test the abnormalities of the single-subject. Fig. 1b describes a similar skewness parameter setting as in Crawford et al. [10]. Given the significance level 0.05, the critical value of the one-sided test (less) for the t test is $t_{0.05} = -2.015$ while the critical value of the one-sided test (less) for the skewed t test is $t^*_{0.05} = -2.568$. Apparently, $t_{0.05} > t^*_{0.05}$, which implies that if some single-subject observation x^* is claimed as abnormal by $t^*_{0.05}$, then x^* must be claimed as abnormal by $t_{0.05}$.

Motivation to reduce the model misspecification error. Let's define the *model misspecification error* for the one-sided test (less) as the probability $\Pr_t(t^*_{0.05} \le x \le t_{0.05})$ where $\Pr_t(\cdot)$ is the probability under the Student t distribution. In Fig. 1b, the model misspecification error for the one-sided test (less) is $\Pr_t(-2.568 \le x \le -2.015) \approx 0.0249$. However, if we consider the one-sided test (greater) under the same settings, the model misspecification error for the one-sided test (greater) is $\Pr_t(0.0370 \le x \le 2.015) \approx 0.3131$. The model misspecification error for the one-sided test (greater) is almost 12 times higher than that for the one-sided test (less). This means that the one-sided test (less) is more robust than the one-sided test (greater) when $\alpha < 0$. Under the settings of $\alpha < 0$, one-sided test (less) and not well-mixed observations of single-subject, Crawford et al. [10] claimed that the t score is robust in terms of Type I error even under skewed Student t distribution. However, this assertion does not hold if the one-sided test (less) is replaced with the one-sided test (greater). For the power comparison, Crawford and Garthwaite [11] make a similar statement. However, by using a similar discussion to the above, we can find that the robustness statement for the power only holds for the specific settings in their paper.

Moreover, we also investigate the model misspecification error between Normal and skewed Student t distributions in Fig. 3. The z score will always have model misspecification error under the skewed Student t distribution assumption. Neither the z score test, the t score test, nor even the Crawford-Garthwaite Bayesian method [1] can eliminate the model misspecification error.

The rationale for this paper's assumption of the skewed Student t distribution for the underlying proposal can be attributed to the fact that it can help eliminate model misspecification errors, provided that the underlying distribution is a special case of the skewed Student t distribution. These special cases encompass the regular Student t distribution ($\alpha = 0$), skewed Normal distribution ($\nu \to \infty$), half Normal distribution ($\alpha \to \pm\infty, \nu \to \infty$), and Normal distribution ($\alpha = 0, \nu \to \infty$) as special cases. To show the elimination of model misspecification error, simulation results with a Normal assumption but using BIGPAST method can be found in Table S4 of the online supplementary material.

We can conclude that the interplay between skewness (whether positive or negative), the type of hypothesis ('two-sided', 'less', or 'greater'), and the single-subject observations (whether well-mixed or not) significantly influences both the Type I error rate and the power of the t score test under the assumption of the skewed Student t distribution. The key contributions are:

- This paper introduces a Bayesian Inference General Procedures for A Single-subject Test (BIGPAST) framework, which employs the nested sampling technique under the skewed Student t assumption to fully discuss the test performance in different settings (i.e., the combinations of the sign of α, type of hypothesis and well-mixed single-subject observations).

- We propose a Jeffreys prior, which is defined by explicit mathematical formulas. We evaluate BIGPAST's performance against the existing approaches through simulation studies. The results demonstrate that BIGPAST is robust against departures from normality and surpasses the existing methods based on the assumption of a Normal distribution in terms of Type I error, power and accuracy.

The remainder of the paper is organised as follows. In Section 2, we introduce the Bayesian inference procedures for comparing a single-subject with a control group under the skewed Student t distribution. In Section 3, we conduct a comprehensive set of simulation studies to evaluate the performance of the proposed BIGPAST method. In Section 4, we apply the BIGPAST method to a MEG dataset comprising an individual with mild traumatic brain injury and a control group. Finally, we summarise our findings and discuss future research directions in Section 5.

2. Bayesian inference procedures for single-subject's abnormality detection

In this section, we introduce Bayesian inference procedures to compare a single-subject with a control group, assuming that the control group observations follow a skewed Student t distribution. First, we should verify whether the assumption holds. For practical applications, we recommend employing a modified version of the general goodness-of-fit test [14] tailored explicitly for the skewed Student t distribution in the goodness-of-fit section of the online supplementary material. From now on, we assume that the control group data follows a skewed Student t distribution in the subsequent sections.

2.1. Bayesian Inference General Procedures for a Single-subject Test (BIGPAST)

Let x^* denote the observation of a single-subject. We aim to test whether x^* is a random sample from the same distribution as the control group. Let $f_c(\alpha, \nu, \xi, \omega)$ represent the skewed Student t density function of $x_1, x_2, \ldots, x_n$. The hypothesis testing can be formulated as follows:

$$\mathcal{H}_0 : x^* \sim f_c(\alpha, \nu, \xi, \omega) \quad v.s. \quad \mathcal{H}_1 : x^* \not\sim f_c(\alpha, \nu, \xi, \omega),$$

where $\not\sim$ represents that x^* is not an observation from $f_c(\alpha, \nu, \xi, \omega)$. We propose the Bayesian inference general procedures for a single-subject test under the null hypothesis $\mathcal{H}_0$. The procedures are as follows:

1. Choose the proposed Jeffery prior $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$ (defined below) as the joint prior of α, ν, ξ, ω.
2. Compute the posterior distribution of α, ν, ξ, ω based on the control group sample $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and the prior $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$. By Bayes's Theorem, we have

   $$\boldsymbol{\pi}(\alpha, \nu, \xi, \omega|\mathbf{x}) \propto l(\mathbf{x}|\alpha, \nu, \xi, \omega)\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega),$$

   where $l(\mathbf{x}|\alpha, \nu, \xi, \omega) := \prod_{i=1}^{n} f_c(x_i|\alpha, \nu, \xi, \omega)$ is the likelihood function.
3. Draw m samples of α, ν, ξ, ω from the posterior distribution $\boldsymbol{\pi}(\alpha, \nu, \xi, \omega|\mathbf{x})$ using the Metropolis-Hastings algorithm [15]. Algorithm 1 lists the details of sampling, the samples are denoted as $\alpha_j, \nu_j, \xi_j, \omega_j,\ j = 1, 2, \ldots, m$.
4. For each $\alpha_j, \nu_j, \xi_j, \omega_j$, draw s random samples from the skewed Student t distribution, denoted as $x_{1j}, x_{2j}, \ldots, x_{sj}$, $j = 1, 2, \ldots, m$. Sort the observations $x_{1j}, x_{2j}, \ldots, x_{sj},\ j = 1, 2, \ldots, m$, in ascending order and label them as $x_{(1)}, x_{(2)}, \ldots, x_{(B)}$ where $B = ms$.
5. Given the significance level β and $\mathbf{x}$, the credible interval of x^* is $[x_{(\lfloor B\beta/2 \rfloor)}, x_{(\lceil B(1-\beta/2) \rceil)}]$, where $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are the floor and ceiling functions, respectively. If x^* is not in the credible interval, then reject the null hypothesis $\mathcal{H}_0$ at the significance level β.

Algorithm 1: Metropolis-Hastings Algorithm.

Data: Initial parameters: $(\alpha_0, \nu_0, \xi_0, \omega_0)$, data: $\mathbf{x}$, size $m_0 = 10000$, burn-in: $b = 0.4$, step size: $\delta = 0.5$.
Result: Samples from the posterior distribution

1 $n_0 \leftarrow 4$;
2 $\Theta \leftarrow$ zeros matrix with dimension:(m_0, n_0);
3 $(\alpha^{\text{old}}, \nu^{\text{old}}, \xi^{\text{old}}, \omega^{\text{old}}) \leftarrow (\alpha_0, \nu_0, \xi_0, \omega_0)$;
4 $i \leftarrow 0$;
5 **while** $i < m_0$ **do**
6 $\quad \nu^{\text{new}} \leftarrow \nu^{\text{old}} + \delta\Phi^{-1}[\Phi(-\nu^{\text{old}}/\delta) + u_\nu\Phi(\nu^{\text{old}}/\delta)]$;
7 $\quad p_\nu \leftarrow \log\Phi(\nu^{\text{old}}/\delta) - \log\Phi(\nu^{\text{new}}/\delta)$;
8 $\quad \omega^{\text{new}} \leftarrow \omega^{\text{old}} + \delta\Phi^{-1}[\Phi(-\omega^{\text{old}}/\delta) + u_\omega\Phi(\omega^{\text{old}}/\delta)]$;
9 $\quad p_\omega \leftarrow \log\Phi(\omega^{\text{old}}/\delta) - \log\Phi(\omega^{\text{new}}/\delta)$;
10 $\quad \alpha^{\text{new}} \leftarrow \alpha^{\text{old}} + \epsilon_\alpha$;
11 $\quad \xi^{\text{new}} \leftarrow \xi^{\text{old}} + \epsilon_\xi$;
12 $\quad p_{\text{new}} \leftarrow \log(\boldsymbol{\pi}(\alpha^{\text{new}}, \nu^{\text{new}}, \xi^{\text{new}}, \omega^{\text{new}}|\mathbf{x}))$;
13 $\quad p_{\text{old}} \leftarrow \log(\boldsymbol{\pi}(\alpha^{\text{old}}, \nu^{\text{old}}, \xi^{\text{old}}, \omega^{\text{old}}|\mathbf{x}))$;
14 $\quad A \leftarrow p_{\text{new}} - p_{\text{old}} + p_\nu + p_\omega$;
15 $\quad$ **if** $\log(u) \le A$ **then**
16 $\quad\quad (\alpha^{\text{old}}, \nu^{\text{old}}, \xi^{\text{old}}, \omega^{\text{old}}) \leftarrow (\alpha^{\text{new}}, \nu^{\text{new}}, \xi^{\text{new}}, \omega^{\text{new}})$;
17 $\quad$ **end**
18 $\quad \Theta[i, :] \leftarrow (\alpha^{\text{old}}, \nu^{\text{old}}, \xi^{\text{old}}, \omega^{\text{old}})$;
19 $\quad i \leftarrow i + 1$;
20 **end**
21 **return** $\Theta[\lfloor m_0 * b \rfloor + 1 :, :]$;

In Algorithm 1, the Metropolis-Hastings algorithm draws samples from the posterior distribution. The step size δ is set to 0.5, and the burn-in rate b is set to 0.4. The algorithm returns the samples after the burn-in period. u_ν, u_ω, u are random numbers drawn from the uniform distribution $U(0, 1)$. ϵ_α and ϵ_ξ are drawn from the Normal distribution with mean zero and standard deviation δ. A brief explanation of Algorithm 1 is as follows:

- Lines 6-9: Draw the new parameters ν^{new} and ω^{new} from the proposal distributions. As ν and ω are positive, the proposal distributions for ν and ω are the truncated normal distributions with mean ν^{old}, ω^{old} and standard deviation δ, respectively.
- Lines 10-11: Draw the new parameters α^{new} and ξ^{new} from the proposal distributions. The proposal distributions for α and ξ are normal distributions with mean α^{old}, ξ^{old} and standard deviation δ, respectively.
- Lines 12-13: Compute the logarithm of the posterior distribution for the new and old parameters.
- Lines 14-19: Compute the acceptance rate A and update the parameters. The truncated proposal distributions do not have symmetric densities, so we have to adjust the acceptance rate A by adding the logarithm of the ratios, i.e., p_ν and p_ω to the acceptance rate A in line 14.

Steps 3, 4 and 5 employ the nested sampling method to determine the critical interval. One popular way is to use the following steps 3′ and 4′:

3′ Obtain the maximum posterior estimators of α, ν, ξ, ω by maximising the posterior distribution $\boldsymbol{\pi}(\alpha, \nu, \xi, \omega|\mathbf{x})$, denoted as $\hat{\alpha}_{\text{MAP}}, \hat{\nu}_{\text{MAP}}, \hat{\xi}_{\text{MAP}}, \hat{\omega}_{\text{MAP}}$.
4′ Given the significance level β, the credible interval of x^* given $\mathbf{x}$ is $[x^{\text{MAP}}_{\beta/2}, x^{\text{MAP}}_{1-\beta/2}]$, where $x^{\text{MAP}}_{\beta/2}$ and $x^{\text{MAP}}_{1-\beta/2}$ are the $\beta/2$ and

$1-\beta/2$ quantiles of $\boldsymbol{\pi}(\hat{\alpha}_{\text{MAP}}, \hat{\nu}_{\text{MAP}}, \hat{\xi}_{\text{MAP}}, \hat{\omega}_{\text{MAP}}|\mathbf{x})$, respectively. If x^* does not fall within this credible interval, then the null hypothesis $\mathcal{H}_0$ can be rejected at the significance level β.

However, in practice, we found that the maximum posterior estimation is time-consuming, and the steps 3′ and 4′ do not perform better than the steps 3, 4 and 5, see the comparison study in Section 3.4. Therefore, we recommend using the nested sampling method to determine the critical interval.

Moreover, when $\alpha \to +\infty$ or $-\infty$, the skewed Student t distribution degenerates to half Student t distribution. Then, the two-sided credible interval in Step 5 above should be slightly modified as $[x_{(1)}, x_{(\lceil B(1-\beta)\rceil)}]$ if $\alpha \to +\infty$ or $[x_{(\lfloor B\beta\rfloor)}, x_{(B)}]$ if $\alpha \to -\infty$.

2.2. Jeffreys priors

In Bayesian probability, Jeffreys prior [16] refers to a prior density function proportional to the square root of the determinant of the Fisher information matrix. In this section, we proposed an explicit Jeffery prior for the parameters of the skewed Student t distribution and compared it to existing priors.

In the literature, two types of priors are commonly used for ξ and ω in Equation (2): the partial information prior [17,18] and the independence Jeffreys prior [19,20]. Both of these priors assume that the joint prior of ξ and ω, denoted as $\boldsymbol{\pi}(\xi, \omega)$, is proportional to ω^{-1}. Consequently, the joint prior of α, ν, ξ, ω can be expressed as follows:

$$\boldsymbol{\pi}(\alpha, \nu, \xi, \omega) \propto \omega^{-1}\boldsymbol{\pi}(\alpha, \nu).$$

The joint prior of α and ν, $\boldsymbol{\pi}(\alpha, \nu)$, may be specified by the Jeffreys prior [21] or the natural informative prior [22]. Branco et al. [21] proposed the following prior for the general skewed Student t density function:

$$\boldsymbol{\pi}^B(\alpha, \nu, \xi, \omega) \propto \omega^{-1}\left(8\alpha^2 + \pi^2\right)^{-3/4}\boldsymbol{\pi}(\nu),$$

where $\boldsymbol{\pi}(\nu)$ is proposed by [23] for the Student t distribution and given by

$$\boldsymbol{\pi}(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{1/2}\left\{\psi^{(1)}\left(\frac{\nu}{2}\right) - \psi^{(1)}\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{1/2},$$

where $\psi^{(1)}(\cdot)$ is the trigamma function, i.e., the second derivative of $\log(\Gamma(\cdot))$. Here, $\Gamma(\cdot)$ is the Gamma function. When $\alpha \to +\infty$, the order of $\boldsymbol{\pi}^B(\alpha, \nu, \xi, \omega)$ with respect to α is $O(\alpha^{-3/2})$ [21].

Dette et al. [22] proposed the so-called natural informative prior of the parameter α based on the total variation distance, which is given by

$$BTV(\alpha|\theta_1, \theta_2) = \frac{1}{B(\theta_1, \theta_2)}\left(\frac{\tan^{-1}(\alpha)}{\pi} + \frac{1}{2}\right)^{\theta_1 - 1} \times \left(\frac{-\tan^{-1}(\alpha)}{\pi} + \frac{1}{2}\right)^{\theta_2 - 1}\frac{1}{\pi(1+\alpha^2)}$$

where $B(\cdot,\cdot)$ represents the Beta function. When $\theta_1 = 1, \theta_2 = 1$, the prior $BTV(\alpha|1,1)$ reduces to the standard Cauchy prior. Similarly, Dette et al. [22] constructed the joint prior of α, ν, ξ, ω based on $BTV(\alpha|1,1)$ as

$$\boldsymbol{\pi}^D(\alpha, \nu, \xi, \omega) \propto \omega^{-1}\pi^{-1}\left(1+\alpha^2\right)^{-1}\boldsymbol{\pi}(\nu),$$

which is order of $O(\alpha^{-2})$ when $\alpha \to +\infty$.

A notable concern with $\boldsymbol{\pi}^B(\alpha, \nu, \xi, \omega)$ and $\boldsymbol{\pi}^D(\alpha, \nu, \xi, \omega)$ is that the marginal prior of ν is independent of α. However, given that the skewed Student t distribution is a generalisation of the Student t distribution, the marginal prior of ν should ideally depend on α. We propose the Jeffreys prior for the skewed Student t density function, as detailed in Theorem 1.

Theorem 1. *Let $f(z|\alpha, \nu)$ be the skewed Student t density function, defined as follows:*

$$f(z|\alpha, \nu) := 2t(z|\nu)T(w|\nu+1), \quad z \in \mathbb{R}, \alpha \in \mathbb{R}, \nu > 0,$$

where $r(z, \nu) = \sqrt{(\nu+1)/(\nu+z^2)}$ and $w = \alpha z r(z, \nu)$. Given any $\nu > 0$, let $\boldsymbol{\pi}^J(\alpha, \nu)$ represent the joint Jeffreys prior for α and ν. We have

$$\boldsymbol{\pi}^J(\alpha, \nu) \propto \sqrt{I_{\alpha\alpha}I_{\nu\nu} - I_{\alpha\nu}^2}, \tag{3}$$

and the joint prior of α, ν, ξ, ω is given by

$$\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega) \propto \omega^{-1}\boldsymbol{\pi}^J(\alpha, \nu), \tag{4}$$

where

$$I_{\alpha\alpha} \approx \frac{\pi\Gamma\left(\frac{\nu}{2}+1\right)^2\left({}_2F_1\left(\frac{1}{2}, \nu+1; \frac{\nu+1}{2}; -\frac{\alpha^2}{\sigma_{\nu+1}^2}\right) - {}_2F_1\left(\frac{1}{2}, \nu+2; \frac{\nu+1}{2}; -\frac{\alpha^2}{\sigma_{\nu+1}^2}\right)\right)}{\alpha^2\Gamma\left(\frac{\nu+1}{2}\right)^2},$$

$$I_{\nu\nu} = \mathbb{E}\left[h^2(w)\frac{\alpha^2z^2(\nu+1)}{4\left(\nu+z^2\right)^3}\right] + \frac{1}{4}\mathbb{E}\left[\left(\log\frac{\nu}{\nu+z^2}\right)^2\right] + \frac{1}{4}\mathbb{E}\left[\left(\frac{z^2-1}{\nu+z^2}\right)^2\right] + d_\nu^2 + d_\nu\mathbb{E}\left[\log\frac{\nu}{\nu+z^2}\right] + \frac{1}{2}\mathbb{E}\left[\left(\log\frac{\nu}{\nu+z^2}\right)\left(\frac{z^2-1}{\nu+z^2}\right)\right],$$

$$I_{\alpha\nu} = -\frac{\pi\alpha\Gamma\left(\frac{\nu}{2}+1\right)^2}{8\left(\alpha^2+\sigma_{\nu+1}^2\right)\Gamma\left(\frac{\nu+1}{2}\right)\Gamma\left(\frac{\nu+5}{2}\right)} \times \left((\nu+4)H_5 - \frac{\left(\alpha^2(2\nu+3) + (\nu+3)\sigma_{\nu+1}^2\right)H_6}{\sigma_{\nu+1}^2}\right),$$

and

$$\mathbb{E}\left[\log\frac{\nu}{\nu+z^2}\right] = \psi\left(\frac{\nu}{2}\right) - \psi\left(\frac{\nu+1}{2}\right),$$

$$\mathbb{E}\left[\left(\log\frac{\nu}{\nu+z^2}\right)\left(\frac{z^2-1}{\nu+z^2}\right)\right] = -\frac{1}{\nu^2+\nu},$$

$$\mathbb{E}\left[\left(\frac{z^2-1}{\nu+z^2}\right)^2\right] = \frac{1}{\nu^2+3\nu},$$

$$\mathbb{E}\left[\left(\log\frac{\nu}{\nu+z^2}\right)^2\right] = \left(\psi\left(\frac{\nu}{2}\right) - \psi\left(\frac{\nu+1}{2}\right)\right)^2 \times \psi^{(1)}\left(\frac{\nu}{2}\right) - \psi^{(1)}\left(\frac{\nu+1}{2}\right),$$

$$\mathbb{E}\left[h^2(w)\frac{\alpha^2z^2(\nu+1)}{4\left(\nu+z^2\right)^3}\right] = \frac{\pi^{5/2}\Gamma\left(\frac{\nu}{2}+1\right)}{8\nu^2\Gamma\left(\frac{\nu+5}{2}\right)B\left(\frac{\nu}{2}, \frac{1}{2}\right)B\left(\frac{\nu+1}{2}, \frac{1}{2}\right)^2} \times \left((\nu+3)H_1 - (\nu+3)H_2 - H_3 + H_4\right).$$

Furthermore, we define $d_\nu := -\frac{1}{2}\psi\left(\frac{\nu}{2}\right) + \frac{1}{2}\psi\left(\frac{\nu+2}{2}\right) + 2c_\nu * g_1(\nu)$, where c_ν is a constant that depends solely on ν. The hypergeometric functions, denoted as $H_1, H_2, H_3, H_4, H_5, H_6$, along with $h^2(w)$ and $g_1(\nu)$ are defined in the proof. ${}_2F_1(a, b; c; d)$ represents the hypergeometric function. The term $\sigma_{\nu+1}^2$ is defined in Lemma S1 of the online supplementary material. The functions $\psi(\cdot)$ and $\psi^{(1)}(\cdot)$ represent the digamma and trigamma function respectively.

Table 2
Comparison of the mean absolute deviation (MAD) among $\boldsymbol{\pi}^J$ with $c_\nu = 0$ and $c_\nu = 1$, $\boldsymbol{\pi}^B$, $\boldsymbol{\pi}^D$, and the Uniform prior. The estimation with a Uniform prior distribution is generally equivalent to the maximum likelihood estimation (MLE). Each entry in the table represents the MAD, calculated based on 1000 replications.

Parameters	Priors	(α, ν)					
		$(-1, 1)$	$(-10, 10)$	$(-30, 30)$	$(-50, 50)$	$(-1, 50)$	$(-50, 1)$
α	$\boldsymbol{\pi}^J, c_\nu = 0$	0.156	1.989	9.402	11.839	0.471	8.415
	$\boldsymbol{\pi}^J, c_\nu = 1$	0.162	2.009	9.105	11.072	0.457	7.703
	$\boldsymbol{\pi}^B$	0.142	1.901	9.399	14.267	0.497	11.489
	$\boldsymbol{\pi}^D$	0.142	1.913	9.122	14.362	0.532	11.807
	Uniform	0.146	2.451	236.974	375.183	0.385	612.510
ν	$\boldsymbol{\pi}^J, c_\nu = 0$	0.079	2.612	6.130	6.656	14.321	0.064
	$\boldsymbol{\pi}^J, c_\nu = 1$	0.085	2.418	5.474	6.210	8.961	0.063
	$\boldsymbol{\pi}^B$	0.064	2.642	15.143	20.924	35.187	0.061
	$\boldsymbol{\pi}^D$	0.063	2.671	15.247	21.587	35.319	0.061
	Uniform	0.064	13.619	82.848	50.030	27.068	0.062
ξ	$\boldsymbol{\pi}^J, c_\nu = 0$	0.136	0.035	0.017	0.012	0.344	0.017
	$\boldsymbol{\pi}^J, c_\nu = 1$	0.148	0.035	0.016	0.011	0.333	0.013
	$\boldsymbol{\pi}^B$	0.104	0.028	0.018	0.013	0.373	0.012
	$\boldsymbol{\pi}^D$	0.105	0.029	0.019	0.014	0.401	0.013
	Uniform	0.105	0.026	0.016	0.010	0.261	0.011
ω	$\boldsymbol{\pi}^J, c_\nu = 0$	0.120	0.071	0.049	0.041	0.156	0.088
	$\boldsymbol{\pi}^J, c_\nu = 1$	0.128	0.067	0.046	0.039	0.143	0.085
	$\boldsymbol{\pi}^B$	0.099	0.076	0.076	0.064	0.199	0.085
	$\boldsymbol{\pi}^D$	0.099	0.078	0.079	0.068	0.211	0.086
	Uniform	0.099	0.070	0.047	0.040	0.109	0.083

The proof of Theorem 1 can be found in the online supplementary material. Upon applying a Taylor expansion, we find that as $\alpha \to 0$, the order of $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$ with respect to α is $O(1 + \alpha^2)$. Conversely, as $\alpha \to +\infty$, the order becomes $O(\alpha^{-3/2})$, aligning with the findings of Branco et al. [21]. The computation of $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$ relies solely on the basic functions provided by the Python packages *scipy* and *numpy*. As such, the formulas in Theorem 1 can be directly implemented to define a function for computing the Jeffreys prior. Additionally, we have developed a Python package, `skewt-scipy` (https://pypi.org/project/skewt-scipy/), for the skewed Student t distribution.

For the Bayesian estimation of ν, we anticipate that $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$ will outperform $\boldsymbol{\pi}^B(\alpha, \nu, \xi, \omega)$. Because $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$ depends on α, unlike $\boldsymbol{\pi}(\nu)$ in $\boldsymbol{\pi}^B(\alpha, \nu, \xi, \omega)$ is independent of α. This is verified by the comparison in Section 3.1. The Jeffreys prior $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$ is a suitable choice for the skewed Student t density function, see Table 2. Table 2 also shows that the BIGPAST based on the Jeffreys prior $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$ is robust to detecting the departures from normality and outperforms the existing approaches based on the assumption that the data come from a Normal distribution.

3. Simulation

3.1. Jeffreys prior comparison study

In this section, we conduct a comparative analysis of our proposed prior $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$ against the priors $\boldsymbol{\pi}^B(\alpha, \nu, \xi, \omega)$ and $\boldsymbol{\pi}^D(\alpha, \nu, \xi, \omega)$. We evaluate these priors under six different parameter settings for (α, ν): $(-1, 1)$, $(-10, 10)$, $(-30, 30)$, $(-50, 50)$, $(-1, 50)$, and $(-50, 1)$. For each setting, the location and scale parameters are fixed at $\xi = -2, \omega = \sqrt{2}$. To simplify the notation, we use $\boldsymbol{\pi}^J$, $\boldsymbol{\pi}^B$, and $\boldsymbol{\pi}^D$ as shorthand for $\boldsymbol{\pi}^J(\alpha, \nu, \xi, \omega)$, $\boldsymbol{\pi}^B(\alpha, \nu, \xi, \omega)$, and $\boldsymbol{\pi}^D(\alpha, \nu, \xi, \omega)$, respectively. For each parameter setting, we generate $N = 1000$ samples, each with a sample size of $n = 500$, from the skewed Student t density function. We then numerically minimise the negative logarithm of the posterior distribution to obtain the maximum posterior estimators of α, ν, ξ, ω. The performance of different priors is evaluated using the mean absolute deviation (MAD) from the true parameter. Alternative metric distances may also be applicable in this context; however, a detailed discussion of these alternatives is beyond the scope of this paper. The MAD is calculated as $1/N \sum_{i=1}^{N} |\hat{\theta}_i - \theta|$, where θ is the true parameter and $\hat{\theta}_i$ is the i-th maximum posterior estimator of θ. Our comparative study also includes the maximum likelihood estimators (MLE). The results are presented in Table 2.

As demonstrated in Table 2, when $|\alpha|$ and ν are less than 30, the performance of $\boldsymbol{\pi}^J$ aligns closely with that of $\boldsymbol{\pi}^B$ and $\boldsymbol{\pi}^D$ in terms of estimating α and ν. However, for larger values of $|\alpha|$ and ν, $\boldsymbol{\pi}^J$ outperforms both $\boldsymbol{\pi}^B$ and $\boldsymbol{\pi}^D$. Furthermore, $\boldsymbol{\pi}^J$, $\boldsymbol{\pi}^B$, and $\boldsymbol{\pi}^D$ all outperform the maximum likelihood estimators (MLE) for larger $|\alpha|$ and ν. In terms of estimating the location ξ and scale ω, the mean absolute deviations (MADs) of all priors are comparable to the MLE. These results underscore the efficacy of $\boldsymbol{\pi}^J$ for the skewed Student t density function, particularly when $|\alpha|$ and ν exceed 30.

3.2. Comparison with existing approaches

To evaluate the performance of BIGPAST in comparison to the z-score, t-score [9], Crawford-Garthwaite Bayesian approach [1], and Anderson-Darling non-parametric approach [24] for abnormality detection in a single-subject against a control group, we adopt the "*severe skew* ($\gamma_1 = -0.7$)" settings from Section 3.1.1 in Crawford et al. [10]. The parameter γ_1 is a measure of skewness in Crawford et al. [10]. This setting is equivalent to parameter settings: $\alpha = -3.23$, $\nu = 7$, $\xi = 0$, and $\omega = 1$ in the definition of our skewed Student t distribution. The single-subject and control group observations are generated from $f(x| -3.23, 7, 0, 1)$. We consider sample sizes of 50, 100, 200, and 400 for the control group. The simulation code in this section can be found in GitHub repository: BIGPAST. Given a significance level of 0.05, a one-sided test (less) and $\mathbf{c} : \mathbf{d} = \mathbf{100} : \mathbf{0}$, we record the abnormality detection results using the z-score, t-score, hyperbolic arcsine transformation ($\log(x + \sqrt{x^2 + 1}), x \in \mathbb{R}$), Crawford-Garthwaite Bayesian method, Anderson-Darling method, and BIGPAST. Other transformations for $x \in \mathbb{R}$ are applicable; however, determining the optimal transformation is beyond the scope of this paper. Therefore, we will only consider the hyperbolic arcsine transformation in this study. This

Table 3
The table presents the simulation results for false positive rate (FPR) and accuracy (ACC), derived from single-subject observations that are exclusively negative, with a one-sided test (less). Each entry in the table is computed based on one million test outcomes. The abbreviations 'CG', 'CG-HA' and 'AD' denote Crawford-Garthwaite Bayesian method, Crawford-Garthwaite Bayesian method based on hyperbolic arcsine transformation, and Anderson-Darling methods, respectively.

	n	z	t	CG-HA	CG	AD	BIGPAST
FPR	50	0.0700	0.0664	0.0544	0.0664	0.0733	0.0519
	100	0.0696	0.0678	0.0550	0.0678	0.0566	0.0537
	200	0.0666	0.0657	0.0567	0.0657	0.0576	0.0526
	400	0.0679	0.0674	0.0554	0.0674	0.0562	0.0521
ACC	50	0.9300	0.9336	0.9456	0.9336	0.9267	0.9481
	100	0.9304	0.9322	0.9450	0.9322	0.9434	0.9463
	200	0.9334	0.9343	0.9433	0.9343	0.9424	0.9474
	400	0.9321	0.9326	0.9446	0.9326	0.9438	0.9479

procedure is repeated one million times as Crawford et al. [10]. The results, presented in Table 3, indicate that BIGPAST outperforms the other methods in terms of false positive rate (FPR) and accuracy (i.e., ACC=(TP+TN)/(TP+TN+FP+FN).) when all single-subject observations are from the ground truth distribution.

Given that all single-subject observations are from ground truth distribution, implying a negative underlying result, the true positive rate and false negative rate (Type II error) are undefined. To address this, we modify the generation of single-subject observations such that 50% yield negative results and the remaining 50% yield positive results, i.e., $c : d = 50 : 50$. Even in the case of $c : d = 100 : 0$, Table 3 demonstrates that BIGPAST outperforms the other methods. To balance the trade-off between the false positive rate (FPR) and the true positive rate (TPR), we will focus on the case of $c : d = 50 : 50$ for the remainder of this section. Additionally, results for the case of $c : d = 80 : 20$ are provided in the online supplementary material, see Table S3.

The observations are generated using the method described in Section 3.4. The true positive rate and false positive rate for the z score, t score, hyperbolic arcsine transformation, Crawford-Garthwaite Bayesian method, Anderson-Darling method, and BIGPAST, pertaining to a two-sided test, are presented in Table 4. Similar simulations have been conducted for the one-sided test (greater) and one-sided test (less), with the results detailed in the online supplementary material, see Tables S1 and S2.

From Table 4, we can conclude that BIGPAST outperforms the t-score and Crawford-Garthwaite methods in terms of false positive rate (FPR) and accuracy (ACC) when the single-subject observations are mixed. As Normal distribution is a special case of skewed Student t distribution ($\alpha = 0, \nu \to +\infty$), the performance of BIGPAST is still as good as those of t-score and Crawford-Garthwaite methods when the underlying distribution is the Normal distribution, see Table S4 of the online supplementary material. The results are not surprising because the errors in this context originate from two primary sources: model misspecification and inherent randomness. Notably, BIGPAST can effectively mitigate the errors arising from model misspecification; see Fig. 3 in the next section.

3.3. Model misspecification error

In this section, we conduct a comparative analysis of the BIGPAST, CG, and AD methodologies under varying alternative hypotheses and parameter configurations. The predetermined significance level is 0.05. The skewness parameter α and degrees of freedom ν are assigned values from the sets $(1, 10)$, $(1, 5)$, $(2, 5)$, and $(3, 5)$, respectively. The location and scale of the skewed Student t distribution are set to 0 and 1, respectively. Other sample settings include a sample size $n = 100$, $c_\nu = 1$, and a Metropolis-Hastings sampling size of 2000. The comparison procedures are as follows: Firstly, $N = 200$ independent samples are randomly generated from the skewed Student t distribution with parameters α and ν. Secondly, $m = 1000$ single-subjects are randomly drawn, and the BIGPAST, CG, and AD tests are conducted on each of the N samples. Thirdly, for each of N samples, the Type I error (false positive rate), power (true positive rate), and accuracy are calculated based on the $m = 1000$ single-subjects. Finally, the Type I error, power, and accuracy are reported in the form of box plots for the four skewed Student t parameter settings, as shown in Fig. 2 and S5 to S7 in online supplementary material.

Table 4
The simulation results of false positive rate (FPR), true positive rate (TPR) and accuracy (ACC), derived from single-subject observations that consist of 50% positive and 50% negative, are presented under a two-sided test. Each cell in the resulting table is calculated based on one million test outcomes. 'CG', 'CG-HA' and 'AD' denote the Crawford-Garthwaite Bayesian method, the Crawford-Garthwaite Bayesian method based on hyperbolic arcsine transformation and the Anderson-Darling methods, respectively.

	n	z	t	CG-HA	CG	AD	BIGPAST
FPR	50	0.1320	0.1182	0.0450	0.1182	0.1107	0.0428
	100	0.1269	0.1199	0.0569	0.1199	0.0782	0.0349
	200	0.1207	0.1166	0.0333	0.1167	0.0644	0.0334
	400	0.1236	0.1215	0.0338	0.1215	0.0368	0.0187
TPR	50	0.6908	0.6502	0.8595	0.6502	0.9101	0.8327
	100	0.6597	0.6415	0.8917	0.6416	0.9356	0.8992
	200	0.6420	0.6312	0.9196	0.6311	0.9655	0.9552
	400	0.6316	0.6264	0.9229	0.6264	0.9660	0.9413
ACC	50	0.7794	0.7660	0.9072	0.7660	0.8997	0.8949
	100	0.7664	0.7608	0.9174	0.7609	0.9287	0.9322
	200	0.7606	0.7573	0.9432	0.7572	0.9505	0.9609
	400	0.7540	0.7525	0.9445	0.7524	0.9646	0.9613

Under the mechanism of these parameter settings, the model is misspecified for the CG test. The theoretical assumption of CG test is the Normal distribution, the mean and variance can be directly computed from the skewed Student t distribution by $\mu = \delta b_\nu, \nu > 1$ and $\sigma^2 = \nu/(\nu-2) - (b_\nu\delta)^2, \nu > 2$, where

$$b_\nu = \frac{\sqrt{\nu}\Gamma\left(\frac{1}{2}(\nu-1)\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}, \quad \nu > 1; \quad \delta = \frac{\alpha}{\sqrt{1+\alpha^2}}.$$

We utilise the total variation distance to quantify the disparity between skewed Student t (for BIGPAST) and Normal (for CG) distributions. The total variation distances of the aforementioned four settings are: 0.057, 0.111, 0.155 and 0.186, respectively, as illustrated in Fig. 2, S5a, S6a and S7a of the online supplementary material. In the context of a two-sided hypothesis test, the divergence between the BIGPAST and CG methodologies becomes larger as the total variation distance increases, as shown in Fig. 2 and S5 to S7 of the online supplementary material. The result of the AD approach demonstrates robustness for different total variation distances but performs significantly worse than BIGPAST in terms of Type I error, as seen in Fig. 2b. The AD approach also performs slightly worse than BIGPAST in terms of accuracy, as depicted in Fig. 2d.

The complexity of the one-sided test (either 'greater' or 'less') compared to the two-sided test due to the skewed Student t distribution increases due to the skewed Student t distribution having light and heavy tails. The model misspecification error impacts the tails of the skewed Student t distribution differently. When $\alpha > 0$ (resp. $\alpha < 0$), the skewed Student t distribution has a right (resp. left) heavy tail. In this paragraph, we explore the scenario where the direction of the alternative hypothesis is 'greater', i.e., one-sided test (greater). If we implement the CG test, the 95% quantile of the Normal distribution is less than the 95% quantile of the skewed Student t distribution. Consequently, the Type I error of the CG test exceeds that of the BIGPAST test, leading to a higher power for the CG test, as shown in Figs. 2b and 2c. This can be explained by the light-red region between the 95% quantile of skewed Student t distribution and the 95% quantile of Normal

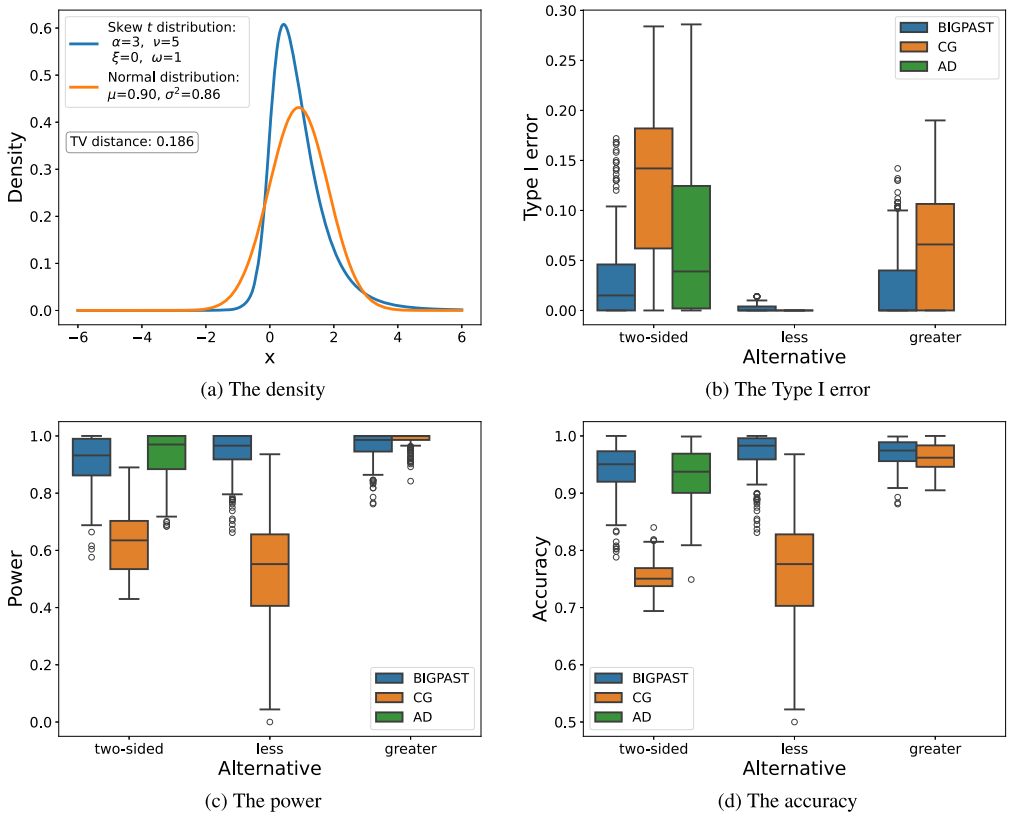


Fig. 2. The comparison results of BIGPAST, CG and AD approaches under three alternative hypotheses: 'two-sided', 'less' and 'greater' respectively when $\alpha = 3$ and $\nu = 5$. Fig. 2a shows the densities of skewed Student t and Normal distributions, the μ and σ^2 of Normal distribution are equal to the mean and variance of skewed Student t distribution. In fact, the Normal distribution is the theoretical assumption of the CG test when the sample comes from the skewed Student t distribution. TV distance is short for total variation distance. Each box plot summarises over 200 independent replications.

distribution, as depicted in Fig. 3b. Theoretically, all single-case observations from this region will be classified as positive by the CG test and as negative by the BIGPAST test, leading us to refer to this region as the false positive region. The error generated by the false positive region arises from model misspecification rather than sampling randomness, as illustrated in Fig. 3.

Conversely, for the light tail of the skewed Student t distribution, the 5% quantile of the Normal distribution is less than the 5% quantile of the skewed Student t distribution. Given that the alternative is 'less' and $\alpha > 0$, the region between the 5% quantile of skewed Student t distribution and the 5% quantile of the Normal distribution is referred to as the false negative region other than the false positive region. Because all the single-case observations from this region will be classified as negative by the CG test but as positive by the BIGPAST test. Consequently, the theoretical Type I error is zero, excluding the random error of sampling, as depicted in Fig. 2b. The power of the CG test, i.e., the actual positive rate, is lower than that of the BIGPAST test, as illustrated in Fig. 2c.

For the left skewed Student t distribution ($\alpha < 0$), we can obtain similar results just by reflecting the densities in line $x = 0$. This will not be detailed here. From the preceding discussion, we can conclude that (1) The difference between the BIGPAST and CG methodologies becomes more significant in terms of Type I error, power, and accuracy as the total variation distance increases; (2) Assuming a skewed Student t distribution, the BIGPAST approach outperforms the CG approach in terms of accuracy due to the presence of model misspecification error; (3) Given the asymmetry between the light and heavy tails of the skewed Student t distribution, we recommend using a two-sided test instead of a one-sided test. This is because the combinations of a one-sided test ('less' or 'greater') and the sign of α significantly influence the evaluation of detection.

For the left skewed Student t distribution ($\alpha < 0$), we can obtain similar results just by reflecting the densities in line $x = 0$. This will not be detailed here. From the preceding discussion, we can conclude that (1) The difference between the BIGPAST and CG methodologies becomes more significant in terms of Type I error, power, and accuracy as the total variation distance increases; (2) Assuming a skewed Student t distribution, the BIGPAST approach outperforms the CG approach in terms of accuracy due to the presence of model misspecification error; (3) Given the asymmetry between the light and heavy tails of the skewed Student t distribution, we recommend using a two-sided test instead of a one-sided test. This is because the combinations of a one-sided test ('less' or 'greater') and the sign of α significantly influence the evaluation of detection.

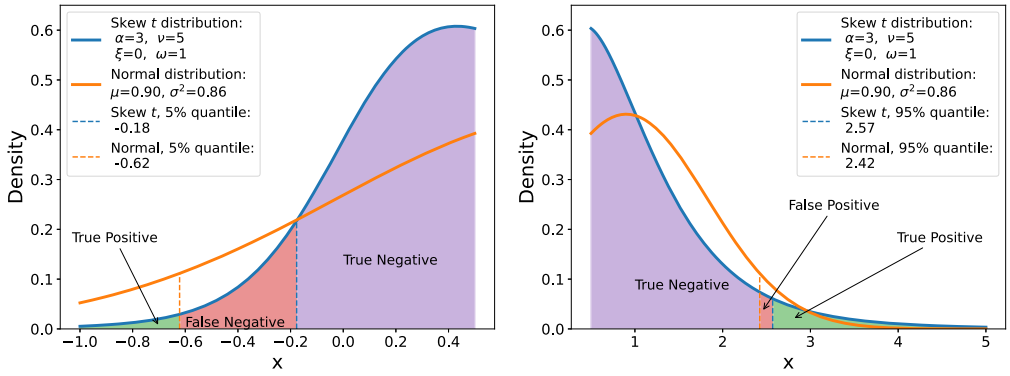
3.4. Comparison study with other frameworks

This section is dedicated to assessing the performance of the proposed BIGPAST methodology and existing approaches when a control group is present. The first approach is a variant of BIGPAST based on the maximum a posteriori (MAP) estimation displayed in steps 3′ and 4′.

The second approach is a non-Bayesian approach that relies on maximum likelihood estimation (MLE). This approach assumes that the control group follows a skewed Student t distribution. The MLE procedures can be outlined as follows:

1′ Obtain the maximum likelihood estimators of α, ν, ξ, ω by maximising the likelihood function $l(\mathbf{x}|\alpha, \nu, \xi, \omega)$. These estimators are denoted as $\hat{\alpha}_{\text{MLE}}, \hat{\nu}_{\text{MLE}}, \hat{\xi}_{\text{MLE}}, \hat{\omega}_{\text{MLE}}$.

2′ Given the significance level β, the confidence interval of x^* given $\mathbf{x}$ is $[x^{\text{MLE}}_{\beta/2}, x^{\text{MLE}}_{1-\beta/2}]$, where $x^{\text{MLE}}_{\beta/2}$ and $x^{\text{MLE}}_{1-\beta/2}$ are the $\beta/2$



(a) Alternative: 'less' (b) Alternative: 'greater'

Fig. 3. Model misspecification error on the light and heavy tails for one-sided test.

and $1-\beta/2$ quantiles of $f_c(\hat{\alpha}_{\text{MLE}}, \hat{\nu}_{\text{MLE}}, \hat{\xi}_{\text{MLE}}, \hat{\omega}_{\text{MLE}})$, respectively. If x^* does not fall within this confidence interval, then the null hypothesis $\mathcal{H}_0$ can be rejected at the significance level β.

The third approach is a non-parametric (NP) approach. Given group data $\mathbf{x}$ and a significance level β, we can derive the empirical confidence interval $[x^{\text{NP}}_{(\lfloor n*\beta/2 \rfloor)}, x^{\text{NP}}_{(\lceil n*(1-\beta/2) \rceil)}]$ where $x^{\text{NP}}_{(\lfloor n*\beta/2 \rfloor)}$ and $x^{\text{NP}}_{(\lceil n*(1-\beta/2) \rceil)}$ are the empirical quantiles of $\mathbf{x}$ at $\beta/2$ and $1-\beta/2$, respectively. If x^* does not fall within this empirical confidence interval, then the null hypothesis $\mathcal{H}_0$ can be rejected at the significance level β. The NP approach does not rely on any distributional assumptions. Given that our BIGPAST methodology employs the Metropolis-Hastings sampler in Step 3, we will refer to the four approaches above: MH, MAP, MLE, and NP, respectively.

We now outline the data generation mechanism for a single-subject and control group. Given the parameters α, ν, ξ, ω, we randomly draw N independent control groups $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ from $f_c(\alpha, \nu, \xi, \omega)$. The following is a straightforward method for generating single-subject data. Define $S_1 := (-\infty, q_{2.5\%}] \cup [q_{97.5\%}, \infty)$ and $S_2 := (q_{2.5\%}, q_{5\%}] \cup [q_{95\%}, q_{97.5\%})$ where $q_{2.5\%}, q_{5\%}, q_{95\%}, q_{97.5\%}$ are 0.025, 0.05, 0.95, 0.975 quantiles of $f_c(\alpha, \nu, \xi, \omega)$. We draw k_1 and k_2 random copies of single-subjects $x^*_1 \ldots, x^*_{k_1}, x^*_{k_1+1}, \ldots, x^*_{k_1+k_2}$ respectively from the uniform distributions defined on S_1 and S_2. Let $K = k_1 + k_2$. At the significance level of 0.05, the ground truth for the first k_1 single-subjects is negative, while the ground truth for the remaining k_2 single-subjects is positive. In the simulation, we first apply the MH, MAP, MLE, and NP approaches to $\mathbf{x}_i, i = 1, 2, \ldots, N$ to obtain the four types of intervals. Next, we conduct the hypothesis tests by comparing $\mathbf{x}^* = (x^*_1, x^*_2, \ldots, x^*_K)$ with each type of intervals. Finally, we compute the false discovery rate (FDR) and accuracy (ACC) defined by

$$\text{FDR}_i^{\mathcal{M}} := \frac{\text{FP}_i^{\mathcal{M}}}{\text{TP}_i^{\mathcal{M}} + \text{FP}_i^{\mathcal{M}}}, \quad \text{ACC}_i^{\mathcal{M}} := \frac{\text{TP}_i^{\mathcal{M}} + \text{TN}_i^{\mathcal{M}}}{K},$$

where $i = 1, 2, \ldots, N$, TP, FP, and TN represent the number of true positive, false positive and true negative, respectively. $\mathcal{M}$ could be one of MH, MAP, MLE, NP.

The parameters settings are as follows: we consider ten combinations of (α, ν): (0,3), (1,3), (3,3), (5,5), (10,10), (20,20), (30,30), (50,50), (5,50) and (50,5) with fixed $\xi = -2$ and $\omega = 2$. In MH approach, we let $m = 5000$, $s = 100$ and $c_\nu = 1$. The other parameter settings are $N = 400$, $n = 100$, $k_1 = k_2 = 1000$, $\beta = 0.05$. The results are presented in Table 5 and Fig. 4.

Table 5 displays the FDR and ACC of the four approaches. The MH and MAP methods consistently outperform the NP and MLE methods in terms of FDR and ACC. When $|\alpha| \leq 3$ and $\nu \leq 3$, both

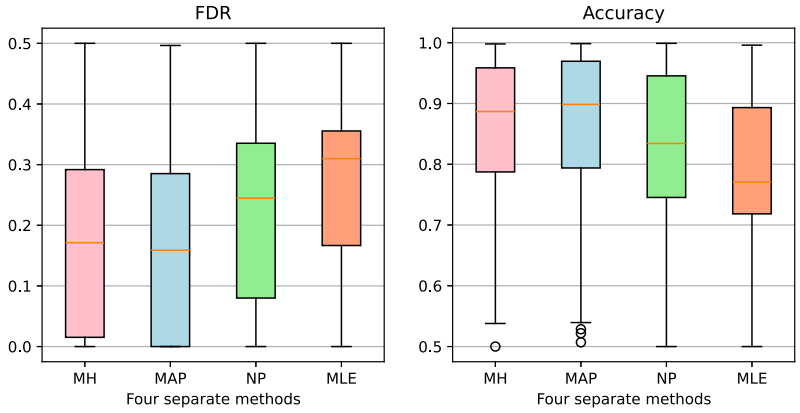
Table 5
The average false discovery rate (FDR) and accuracy (ACC) for the MH, MAP, MLE, and NP methods were calculated over 400 repetitions. In this context, MH refers to the original BIGPAST method, MAP denotes the Bayesian inference framework based on maximum a posteriori estimation, MLE represents the method based on maximum likelihood estimation, and NP signifies the non-parametric method.

(α, ν)	FDR				ACC			
	MH	MAP	NP	MLE	MH	MAP	NP	MLE
(0,3)	0.174	0.165	0.216	0.266	0.865	0.872	0.833	0.787
(1,3)	0.158	0.145	0.201	0.257	0.877	0.888	0.845	0.797
(3,3)	0.142	0.129	0.197	0.254	0.891	0.901	0.849	0.798
(5,5)	0.131	0.136	0.206	0.255	0.905	0.899	0.846	0.803
(10,10)	0.118	0.145	0.200	0.259	0.915	0.894	0.851	0.797
(20,20)	0.132	0.157	0.190	0.255	0.907	0.886	0.861	0.803
(30,30)	0.132	0.171	0.188	0.261	0.907	0.875	0.862	0.796
(50,50)	0.134	0.184	0.193	0.254	0.906	0.864	0.859	0.801
(5,50)	0.144	0.158	0.185	0.261	0.898	0.885	0.865	0.796
(50,5)	0.139	0.165	0.218	0.249	0.899	0.876	0.834	0.805

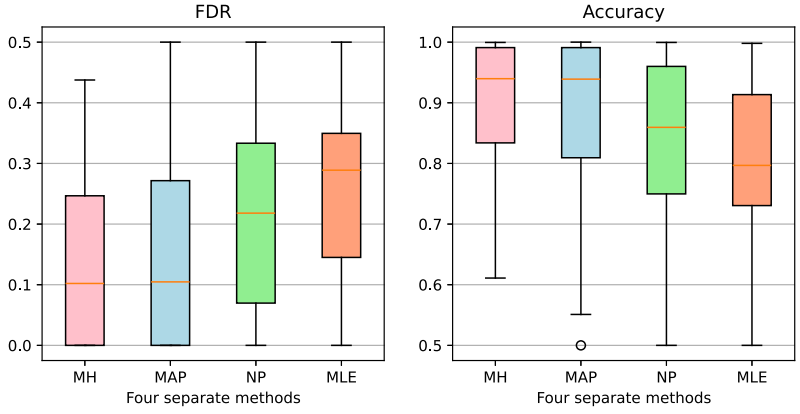
the FDR and ACC of the MAP approach surpass those of the MH approach. This is expected, as the MAP approach provides more accurate estimation of the underlying parameters of the skewed Student t distribution when $|\alpha|$ and ν are small, as is shown in column 3 of Table 2. When $|\alpha|$ or ν is large, the FDR and ACC of the MH approach perform better than those of the MAP approach. This can be attributed to the randomness of parameters in the Metropolis-Hastings sampler. However, the MAP approach fails to estimate the underlying parameters when α or ν is large, as is shown in column 4 of Table 2, and lacks randomness of parameters. The box plots in Fig. 4 clearly illustrate the comparison between the MH and MAP approaches. The results demonstrate that the FDR of MH (i.e. BIGPAST) is generally lower than others, and the ACC of MH (i.e. BIGPAST) is generally greater than other methods in conducting hypothesis tests for single-subject data.

4. Real data analysis

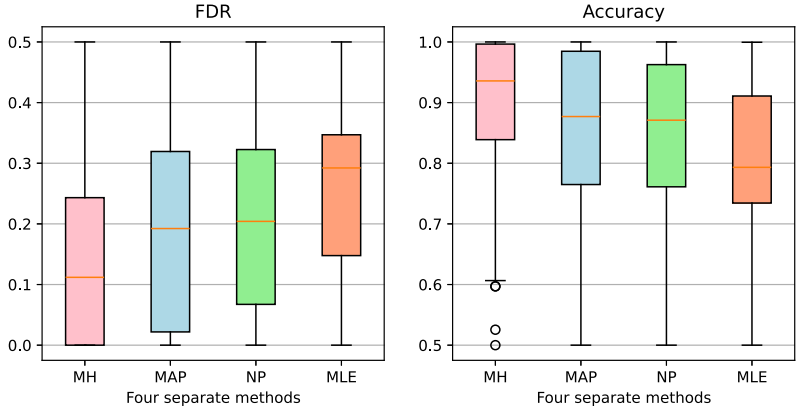
This dataset is derived from a mild traumatic brain injury (mTBI) study conducted by Innovision-IP Ltd. It comprises a functional time series of Magnetoencephalography (MEG) data and an anatomical T1-weighted Magnetic Resonance Imaging (MRI) scan for 103 healthy subjects and one patient with mild traumatic brain injury. The MEG data, recorded from 306 sensors, was pre-processed using the 'MNE-Python' package [25]. The MEG data were analysed to reconstruct the source generators using a minimum norm inverse method. The MRI data were pre-processed with the 'FreeSurfer' package [26]. The standard FreeSurfer pipeline (detailed on their website) was followed and includes motion correction, intensity normalisation, skull-stripping, normalisation to template space and segmentation of the cortex and subcortical structures using the Desikan-Killiany atlas. This results in 68 re-



(a) $\alpha = 0, \nu = 3$



(b) $\alpha = 5, \nu = 5$



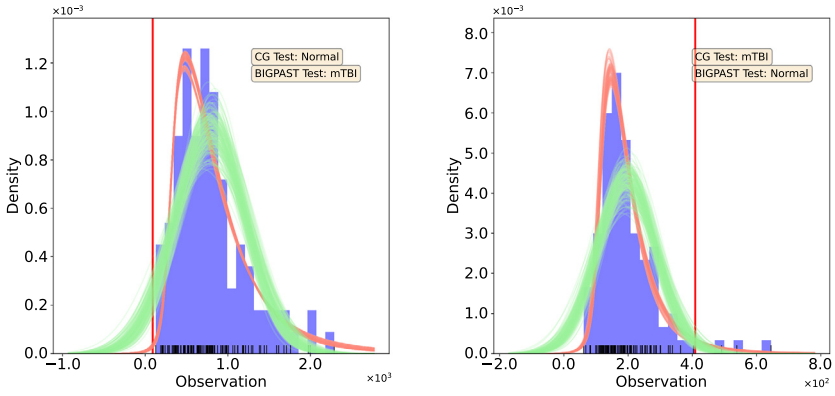
(c) $\alpha = 50, \nu = 50$

Fig. 4. The box plots for false discovery rate (FDR) and accuracy (ACC) of the MH, MAP, MLE, and NP approaches for three distinct skewness parameters: low ($\alpha = 0$), medium ($\alpha = 5$), and high ($\alpha = 50$). In this context, MH refers to the original BIGPAST method, MAP denotes the Bayesian inference framework based on maximum a posteriori estimation, MLE represents the method based on maximum likelihood estimation, and NP signifies the non-parametric method.

gions covering the entire cortex. For each cortical region, we calculate the average of source powers within this region. This process yields a control group dataset with dimensions $103 \times 68 \times 100$ and a patient dataset with dimensions 68×100. To simplify, we treat the epochs as repeated observations and take averages over them to reduce the data's complexity. Consequently, the control group dataset has dimensions 103×68, and the patient dataset is reduced to a vector of length 68.

For each brain region of the Desikan-Killiany cortical atlas, we apply the BIGPAST method to the control group and patient dataset to test whether the patient significantly differs from the control group. This real data analysis shows that the single-subject has mTBI. However, in practice, we can calculate the observation of a new client and conduct the hypothesis test using BIGPAST to determine whether the client significantly differs from the control group.

We further explore the impact of waveband frequency on the comparison between a control group and a single-subject. The wavebands are arranged in ascending order of frequency: delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-45 Hz). The subsequent results also include those of the Crawford-Garthwaite method, denoted as CG for brevity. The prespecified significance level is 0.05. The test results of Crawford-Garthwaite (CG) and BIGPAST are presented in Table 6, where the bracket *(Normal, mTBI)* in the third row represents that the result of CG is Normal, but the result of BIGPAST is mTBI. The numbers in



(a) Delta Band, $\bar{\tilde{\alpha}} \approx 3.95$, Left Caudal Anterior Cingulate

(b) Gamma Band, $\bar{\tilde{\alpha}} \approx 2.69$, Right Fusiform

Fig. 5. The comparison results between the Crawford-Garthwaite (CG) and BIGPAST approaches in terms of density estimations. The light-red lines represent the skewed Student t densities, with parameters drawn from the BIGPAST approach's posterior distribution. The light green lines depict the Normal densities, with parameters drawn from the CG approach's posterior distribution. The CG test is based on 10,000 densities, while the BIGPAST test is based on 5,000 densities (the first 5,000 densities are discarded for burn-in). For clarity, each panel only displays 200 densities drawn from the CG and BIGPAST posterior distribution. The red vertical line on the x axis represents the observation of a single-subject (i.e., a patient with mTBI). A histogram is used to demonstrate the goodness of fit for both the CG and BIGPAST approaches.

Table 6
The test results of the Crawford-Garthwaite (CG) based on Gaussian assumption and BIGPAST approaches based on the skewed Student t assumption. The numbers in cells corresponding to specific brain cortical regions are 1: lh-caudal anterior cingulate, 2: lh-caudal middle frontal, 4: lh-entorhinal, 6: lh-fusiform, 10: lh-isthums cingulate, 12: lh-lateral orbitofrontal, 14: lh-medial orbitofrontal, 17: lh-parahippocampal, 22: lh-postcentral, 24: lh-precentral, 26: lh-rostral anterior cingulate, 35: rh-caudal anterior cingulate, 38: rh-entorhinal, 40: rh-fusiform, 47: rh-lingual, 50: rh-paracentral, 51: rh-parahippocampal, 58: rh-precentral, 59: rh-precuneus, 62: rh-superior frontal. The prefixes 'lh-' and 'rh-' indicate the left and right hemispheres of the brain, respectively. 'NA' stands for 'not available', and 'Others' represents the remaining brain cortical indices not listed in the column but can be found in the indices in Table S5 of the online supplementary material.

The Results of (CG, BIGPAST)	Wave bands				
	delta	theta	alpha	beta	gamma
(mTBI, mTBI)	6, 17, 40, 51	4, 6, 17, 48	6, 14, 17, 40	6, 12, 17, 38, 40	17
(Normal, mTBI)	1, 10, 16, 26, 35, 50, 59, 62	1, 2, 24, 35, 62	2, 24, 35, 58, 59	1, 2, 10, 22, 24, 35, 59	1, 2, 35
(mTBI, Normal)	NA	14, 40	NA	14, 47	6, 40
(Normal, Normal)	Others	Others	Others	Others	Others

cells are the indices of human brain regions; for example, number 1 in the "delta" column represents the caudal anterior cingulate in the left brain hemisphere. For this brain region, CG's result is Normal, while the result of BIGPAST is mTBI. We did not consider the false discovery rate adjustment here as (1) steps 3 and 4 in BIGPAST employ the nested sampling technique, and the p-values are hard to obtain. (2) The aim of Table 6 is to provide a comparison between the CG and BIGPAST methods, and the FDR adjustment is not necessary for this purpose.

The null hypothesis in this real data analysis is two-sided, with one heavy tail and one light tail in the skewed Student t distribution. The cortical region indices in row 3 of Table 6 exhibit a positive skewness parameter, i.e., $\alpha > 0$. Consequently, the light tail is on the left, and the heavy tail is on the right. For all cortical region indices in row 3 of Table 6, the single-subject observation is located at the light tail of the skewed Student t distribution. This positioning suggests a high probability of rejecting the null hypothesis, see Fig. 5a. Similarly, for cortical regions 6 (lh-fusion, gamma band), 14 (lh-medial orbitofrontal, theta and beta bands), 40 (rh-fusiform, theta and gamma bands), and 47 (rh-lingual, beta band), the single-subject observation is located at the heavy tail of the skewed Student t distribution, as shown in Fig. 5b. The skewed Student t assumption is more appropriate for these cortical regions than the assumption of Normal distribution. The BIGPAST approach outperforms the CG approach in detecting the single-subject observation at the heavy tail of the skewed Student t distribution, as demonstrated in Fig. 5. Because the BIGPAST approach corrects the model misspecification error at both tails compared to the CG approach. By combining rows 2 and 3 of Table 6 and Fig. 5, we conclude that the BIGPAST approach is more reliable than the CG approach in hypothesis tests for single-subject data when the control group data exhibits skewness.

5. Conclusion

In this paper, we proposed a novel Bayesian inference framework for the abnormality detection on a single-subject versus a control group. Under the assumption of a Normal distribution, the t-score [9] is capable of managing scenarios where the sample size is fewer than 30 [27]. However, in neuroscience, the number of epochs often exceeds 100, sometimes even reaching 1000, and the distributions, whether over epochs or subjects, exhibit skewness. Under these conditions, methods predicated on the assumption of normality cease to be effective. Our proposed methodology, BIGPAST, leverages the skewed Student t distribution, effectively capturing the inherent asymmetry of the data.

Our proposed BIGPAST methodology outperforms the existing Crawford-Garthwaite (CG) approach in terms of accuracy, power, and Type I error. Particularly when the control group data exhibits skewness, BIGPAST provides a more reliable solution for hypothesis testing on single-subject data. Furthermore, BIGPAST demonstrates superior robustness to model misspecification errors compared to the CG approach. The performance of the BIGPAST and CG methodologies is significantly influenced by the total variation distance between the skewed Student t distribution and the Normal distribution. When this distance is large, the BIGPAST approach exhibits superior accuracy compared to the CG method. The skewed Student t distribution's asymmetry, characterised by its light and heavy tails, further emphasises the necessity of employing a two-sided test over a one-sided test.

The BIGPAST framework can be readily adapted to accommodate other distributions, such as Sub-normal, Sub-Weibull, or Extreme Value distributions. The key requirement is to identify an appropriate prior distribution for the parameters of the chosen distribution. A further extension of BIGPAST could encompass multivariate distributions, thereby enabling the application of the BIGPAST methodology to high-dimensional data in future.

CRediT authorship contribution statement

JL: Conceptualisation, Methodology, Programming and Software, Investigation, and Writing—Original Draft. GG: Conceptualisation, Supervision, Methodology, Resources, Investigation, Project Administration and Writing—Review & Editing. SC: Methodology and Writing—Review & Editing. PL: Supervision, Methodology, Project Administration and Writing—Review & Editing. JZ: Conceptualisation, Supervision, Methodology, Investigation, Project Administration and Writing—Review & Editing.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patient(s) and/or volunteers.

Data and code availability

The calculation of skewed Student t distribution is implemented in the Python package skewt-scipy. The code for simulation studies can be found in GitHub repository: BIGPAST. The real data comes from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) dataset [28].

Human and animal rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans as well as in accordance with the EU Directive 2010/63/EU for animal experiments.

Funding

This research received financial support jointly from UKRI through Innovate UK and Innovision-IP Ltd. (Grant No: 10053865).

Declaration of competing interest

The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

Acknowledgements

This work was supported by the High-Performance Computing Cluster at the University of Kent.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neuri.2025.100195.

References

[1] J.R. Crawford, P.H. Garthwaite, Comparison of a single case to a control or normative sample in neuropsychology: development of a Bayesian approach, Cogn. Neuropsychol. 24 (2007) 343–372.
[2] S. Das, A. Adhikary, A.A. Laghari, S. Mitra, Eldo-care: EEG with kinect sensor based telehealthcare for the disabled and the elderly, Neurosci. Inf. 3 (2023) 100130.
[3] J.D. Higbee, J.B. McDonald, A comparison of the GB2 and skewed generalized log-t distributions with an application in finance, J. Econom. 240 (2024) 105064.
[4] S. Kim, S. Song, Cyber risk measurement via loss distribution approach and GARCH model, Commun. Stat. Appl. Methods 30 (2023) 75–94.
[5] I. Ktena, O. Wiles, I. Albuquerque, S.-A. Rebuffi, R. Tanno, A.G. Roy, S. Azizi, D. Belgrave, P. Kohli, T. Cemgil, A. Karthikesalingam, S. Gowal, Generative models improve fairness of medical classifiers under distribution shifts, Nat. Med. 30 (2024) 1166–1173.
[6] S. Yin, H. Li, A.A. Laghari, L. Teng, T.R. Gadekallu, A. Almadhor, FLSN-MVO: edge computing and privacy protection based on federated learning Siamese network with multi-verse optimization algorithm for industry 5.0, IEEE Open J. Commun. Soc. (2024) 1–17.
[7] C.A. Taylor, J.M. Bell, M.J. Breiding, L. Xu, Traumatic brain injury-related emergency department visits, hospitalizations, and deaths - United States, 2007 and 2013, Morb. Mortal. Wkly. Rep., CDC Surveill. Summ. 66 (2017) 1–18.
[8] M. Zimmermann, K. Schultz-Nielsen, G. Dumas, I. Konvalinka, Arbitrary methodological decisions skew inter-brain synchronization estimates in hyperscanning-EEG studies, Imaging Neurosci. 2 (2024) 1–19.
[9] J.R. Crawford, D.C. Howell, Comparing an individual's test score against norms derived from small samples, Clin. Neuropsychol. 12 (1998) 482–486.
[10] J.R. Crawford, P.H. Garthwaite, A. Azzalini, D.C. Howell, K.R. Laws, Testing for a deficit in single-case studies: effects of departures from normality, Neuropsychologia 44 (2006) 666–677.
[11] J.R. Crawford, P.H. Garthwaite, Methods of testing for a deficit in single-case studies: evaluation of statistical power by Monte Carlo simulation, Cogn. Neuropsychol. 23 (2006) 877–904.
[12] A. Azzalini, A. Capitanio, The Skew-Normal and Related Families, Institute of Mathematical Statistics Monographs, vol. 3, Cambridge University Press, Cambridge, 2014.
[13] A. Azzalini, A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, J. R. Stat. Soc., Ser. B, Stat. Methodol. 65 (2003) 367–389.
[14] G. Chen, N. Balakrishnan, A general purpose approximate goodness-of-fit test, J. Qual. Technol. 27 (1995) 154–161.
[15] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.
[16] H. Jeffreys, An invariant form for the prior probability in estimation problems, Proc. R. Soc. Lond. Ser. A, Math. Phys. Sci. 186 (1946) 453–461.
[17] D. Sun, J.O. Berger, Reference priors with partial information, Biometrika 85 (1998) 55–71.
[18] L. Cohen, S. Sharifi-Malvajerdi, K. Stangl, A. Vakilian, J. Ziani, Bayesian strategic classification, Adv. Neural Inf. Process. Syst. 37 (2025) 111649–111678.
[19] F.J. Rubio, M.F.J. Steel, Inference in two-piece location-scale models with Jeffreys priors, Bayesian Anal. 9 (2014) 1–22.
[20] V. Fortuin, Priors in Bayesian deep learning: a review, Int. Stat. Rev. 90 (2022) 563–591.
[21] M.D. Branco, M.G. Genton, B. Liseo, Objective Bayesian analysis of skew-t distributions, Scand. J. Stat. 40 (2013) 63–85.
[22] H. Dette, C. Ley, F. Rubio, Natural (non-)informative priors for skew-symmetric distributions, Scand. J. Stat. 45 (2018) 405–420.
[23] T.C.O. Fonseca, M.A.R. Ferreira, H.S. Migon, Objective Bayesian analysis for the student-t regression model, Biometrika 95 (2008) 325–333.
[24] F.W. Scholz, M.A. Stephens, K-sample Anderson-Darling tests, J. Am. Stat. Assoc. 82 (1987) 918–924.
[25] A. Gramfort, M. Luessi, E. Larson, D.A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, M.S. Hämäläinen, MEG and EEG data analysis with MNE-python, Front. Neurosci. 7 (2013) 1–13.
[26] X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany, P. Maguire, D. Rosas, N. Makris, A. Dale, B. Dickerson, B. Fischl, Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer, NeuroImage 32 (2006) 180–194.
[27] D. Machin, M.J. Campbell, S.B. Tan, S.H. Tan, Sample Sizes for Clinical, Laboratory and Epidemiology Studies, John Wiley & Sons, 2018.
[28] M.A. Shafto, L.K. Tyler, M. Dixon, J.R. Taylor, J.B. Rowe, R. Cusack, A.J. Calder, W.D. Marslen-Wilson, J. Duncan, T. Dalgleish, R.N. Henson, C. Brayne, F.E. Matthews, The Cambridge centre for ageing and neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing, BMC Neurol. 14 (2014) 204.