Discriminating between Power Normal and Normal Distributions

Ojasvi Rajput[1,**] Debasis Kundu[2†] Sharmishtha Mitra[3‡]

[1,2,3] *Department of Mathematics & Statistics, Indian Institute of Technology Kanpur, Kanpur-208016, Uttar Pradesh, India*

Abstract

In 2008, Gupta and Gupta (2008) introduced the Power Normal distribution, a skewed distribution defined over the entire real line. The Normal distribution is a special case of the Power Normal, which means that for much of the parameter space, these distributions share similarities in terms of their cumulative distribution function (CDF) and probability density function (PDF). Although both models may provide similar fits for the central region of the data, differences often emerge in the tail regions, underscoring the importance of selecting the appropriate model. This paper focuses on the discrimination between these two distributions. Most prior work has focused on discriminating between distributions with the same number of parameters; however, to the best of our knowledge, no work has specifically tackled the challenge of discriminating between distributions with different parameter counts. This distinction adds complexity, as conventional methods often fail in such cases. In this study, we compare several decision criteria based on their probability of correct selection (PCS), including the ratio of maximized log-likelihoods, the KS-distance-based method and Bayesian Information Criterion (BIC). In addition, we develop asymptotic distributions for some of these criteria. The impact of model misspecification is also explored and a dataset is analyzed for illustration.

Keywords: Power normal distribution, normal distributions, Likelihood ratio test, asymptotic distribution, Bayesian information criterion, probability of correct selection, Monte Carlo simulations.

*Corresponding author. Email: ojrj20@iitk.ac.in

†Email: kundu@iitk.ac.in

‡Email: smitra@iitk.ac.in

1 Introduction

Discriminating between two probabilistic models is a classical statistical problem, with applications ranging from reliability analysis to risk modelling and environmental sciences.The foundations of this topic were laid by Cox (1961), who also highlighted the consequences of incorrect model choice, with subsequent developments by Atkinson (1969, 1970), Dyer (1973), and Chen (1980). A substantial body of literature has since emerged (see, for example, Dey and Kundu (2009); Ahmad et al. (2017); Raqab et al. (2018) and the references therein). More recent investigations include discrimination between Poisson–Tweedie and Poisson–exponential–Tweedie count models (Abid and Kokonendji, 2023), between Birnbaum–Saunders and log-normal distributions with attention to misspecification effects (Basu and Kundu, 2023), and among inverse Weibull, log-normal, and inverse Gaussian models (Diyali et al., 2024). Divergence-based approaches have also been proposed, such as the method of Basu and Ng (2025). Related discussions on model misspecification can be found in Wiens (1999); Khakifirooz et al. (2020). While much of this work assumes complete samples, several contributions address censored settings: Diyali et al. (2023) studied discrimination between log-normal and log-logistic models under Type-II censoring, Dey and Kundu (2012) considered the log-normal versus Weibull case, and Kuş et al. (2019) extended the problem to progressive Type-II censoring. However, nearly all of these works assume equal parameter counts, leaving the unequal-dimension case largely unexplored. The present paper addresses this gap.

While much of the existing work focuses on competing distributions with the same number of parameters, situations where the models differ in dimensionality are far more concerning. In such cases, the usual rule of selecting the model with the larger maximized log-likelihood is inherently biased toward the model with more number of parameters. This challenge of discriminating between distributions with different parameter counts has received surprisingly little formal attention in the literature. Developing discrimination tools in such unequal parameter settings is therefore both methodologically important and practically relevant.

Many probability models have been proposed to capture skewness and tail behaviour in data. Among them, the skew-normal distribution, introduced by Azzalini (1985), has received considerable attention due to its flexibility in modeling skewed data on the real line. Despite its appealing properties, practical implementation of the skew-normal model presents certain challenges. Although the maximum likelihood estimators (MLEs) are theoretically consistent and asymptotically normal, they may fail to exist, particularly in small samples. Simulation studies by Gupta and Gupta (2008) further indicate that even in the presence of moderate

skewness, reliably estimating the skewness parameter often requires large sample sizes.

These limitations motivate the search for alternative models that can capture skewness while remaining stable and interpretable. One such alternative is the Power Normal distribution proposed by Gupta and Gupta (2008). This model retains the Normal distribution as a special case ($\alpha = 1$), preserves support over the entire real line, and introduces a flexible mechanism to accommodate skewness.

Given that the Normal distribution continues to be one of the most widely used models across scientific disciplines, it is natural to consider extensions that maintain its interpretability while improving its ability to capture asymmetry. The Power Normal distribution serves this purpose effectively. Interestingly, the Normal and Power Normal distributions can exhibit very similar behaviour in the central region of their probability density and cumulative distribution functions, while differing noticeably in the tails.

This similarity in the bulk of the distribution, combined with differences in tail behaviour, makes it both practically important and theoretically challenging to distinguish between the two models. Consequently, developing methods to discriminate between the Normal model and such skewed alternatives provides a meaningful framework, particularly in settings involving models with unequal parameter dimensions.

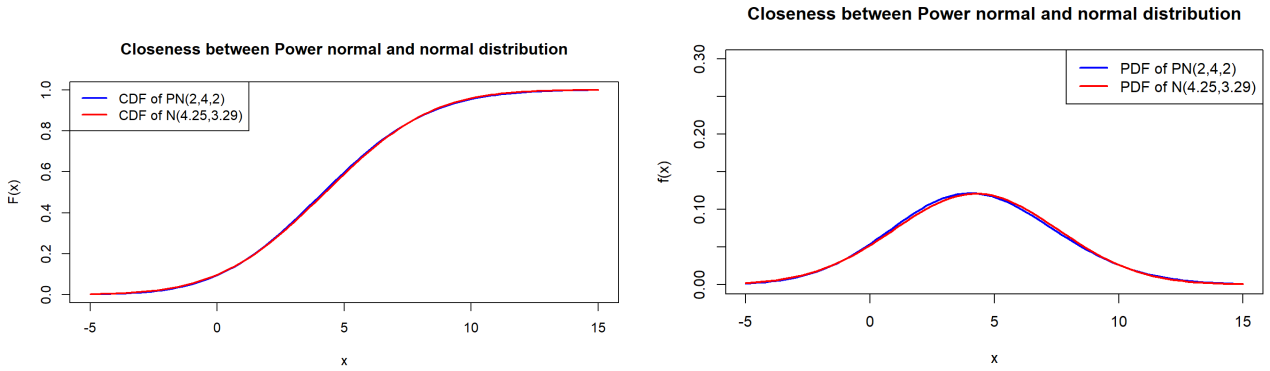
Both distributions share support over the entire real line. Although many real-world reliability datasets are supported on $(0, \infty)$, this is not a limitation: a logarithmic transformation maps $(0, \infty)$ to $\mathbb{R}$, and the same discrimination procedure immediately extends to any two models (with equal or unequal parameter counts) after transformation. Thus, the PN–Normal comparison serves as a representative framework for discrimination problems involving unequal dimensions.

Formally, a random variable X has a Power Normal distribution $PN(\alpha, \mu, \sigma)$ if its density is

$$f_{PN}(x; \alpha, \mu, \sigma) = \frac{\alpha}{\sigma}\left[\Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}\phi\left(\frac{x-\mu}{\sigma}\right), \qquad -\infty < x < \infty, \tag{1}$$

where $\alpha > 0$, $-\infty < \mu < \infty$, $\sigma > 0$, and ϕ, Φ denote the standard normal PDF and CDF respectively. When $\alpha = 1$, this reduces exactly to the Normal distribution.

To illustrate the visual similarity between PN and Normal densities, we identify the Normal distribution closest (in Hellinger distance) to $PN(2, 4, 2)$. The corresponding CDFs and PDFs are shown in Fig. 1. While the two models are nearly indistinguishable in the body of the distribution, the hazard functions (Fig. 2) reveal clear differences, emphasizing the importance of tail behavior when performing model discrimination.



(a) CDF of PN(2,4,2) and N(4.25,3.29). (b) PDF of PN(2,4,2) and N(4.25,3.29).

Figure 1: Similarity in terms of shape of CDF and PDF between PN(2,4,2) and N(4.25,3.29).

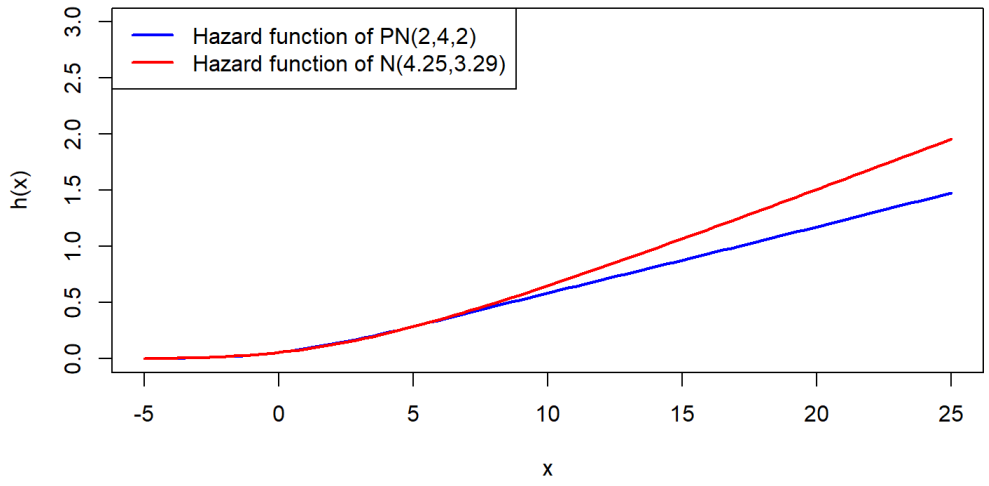


Figure 2: Hazard rate of PN(2,4,2) and N(4.25,3.29).

The theoretical challenges above motivate the need for model selection methods that are robust to differences in parameter dimension. A major contribution of this paper is a detailed investigation of the Bayesian Information Criterion (BIC) in this unequal-parameter setting. We provide the asymptotic development of BIC decision criterion, compare its performance with classical likelihood-based rules, and establish conditions under which BIC yields consistent decisions. Our results show that BIC behaves systematically differently when competing models differ in dimensionality, and we quantify these effects both theoretically and through simulation.

These developments provide a principled foundation for discrimination between PN and Normal models and extend directly to other unequal-parameter comparisons.

The remainder of the paper is structured as follows. Section 2 presents the motivation for the problem using a real-life dataset. Section 3 introduces the discrimination criteria employed in this study and derives their asymptotic distributions. Section 4 outlines the procedure for determining the minimum required sample size. Section 5 reports the results of extensive simulation studies. Section 6 provides a brief discussion on the performance of the BIC criterion under censoring, and Section 7 examines the effect of model misspecification. Section 8 analyzes a real dataset to demonstrate the practical utility of the proposed methods. Finally, Section 9 concludes the paper.

2 Motivation

To motivate the discrimination problem in a real-data context, we discuss the stiffness dataset reported in Johnson et al. (2002), consisting of two measurements—"Shock" (X_1) and "Vibration" (X_2)—taken on each of 30 wooden boards (Table 1).

No.	**Shock**	**Vibration**	**No.**	**Shock**	**Vibration**	**No.**	**Shock**	**Vibration**
1	1889	1651	11	2403	2048	21	2119	1700
2	1645	1627	12	1976	1916	22	1712	1713
3	1943	1685	13	2104	1820	23	2983	2794
4	1745	1600	14	1710	1591	24	2046	1907
5	1840	1841	15	1867	1685	25	1859	1649
6	1954	2149	16	1325	1170	26	1419	1371
7	1828	1634	17	1725	1594	27	2276	2189
8	1899	1614	18	1633	1513	28	2061	1867
9	1856	1493	19	1727	1412	29	2168	1896
10	1655	1675	20	2326	2301	30	1490	1382

Table 1: Two different stiffness measurements of 30 boards.

For all model fitting procedures presented below, the Shock and Vibration datasets were standardized. We analyzed both the marginal data and for the Shock data, we estimated the maximum likelihood estimators (MLEs) under the Power Normal distribution. The MLEs for

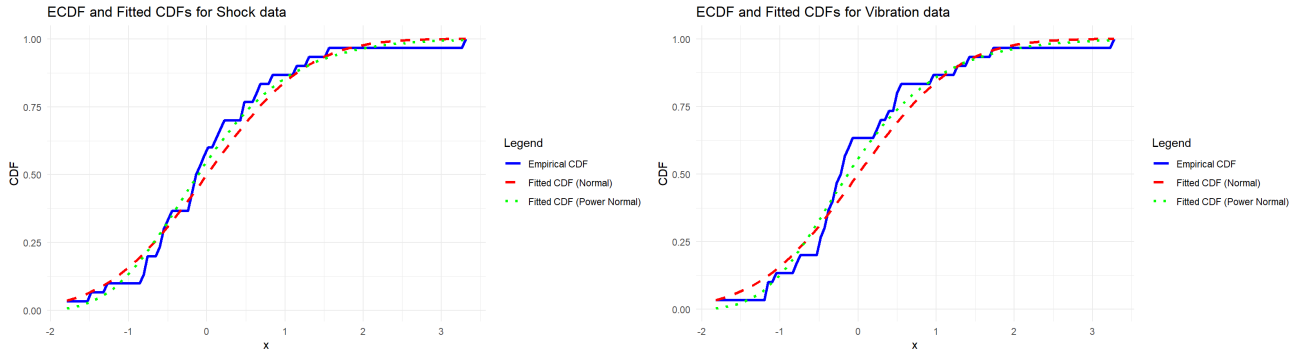
μ, σ, and α were -8.06, 2.62, and 558.20, respectively. Under the Normal distribution for the same data, the MLEs of μ_* and σ_* were $2.671474e-16$ and 0.99, respectively. A similar analysis was performed for the Vibration data. The MLEs of μ, σ, and α under the Power Normal distribution were -14.12, 3.38, and $39,362.34$, respectively. For the Vibration data under the Normal distribution, the MLEs of μ_* and σ_* were $1.915279e-16$ and 0.98, respectively.

Both models provided satisfactory goodness-of-fit in terms of the Kolmogorov–Smirnov statistic:

Model	Shock (X_1)	Vibration (X_2)
Power normal	0.0879 (0.9587)	0.1306 (0.6849)
Normal	0.1148 (0.7819)	0.1790 (0.2915)

Table 2: Kolmogorov–Smirnov distances and associated p-values.

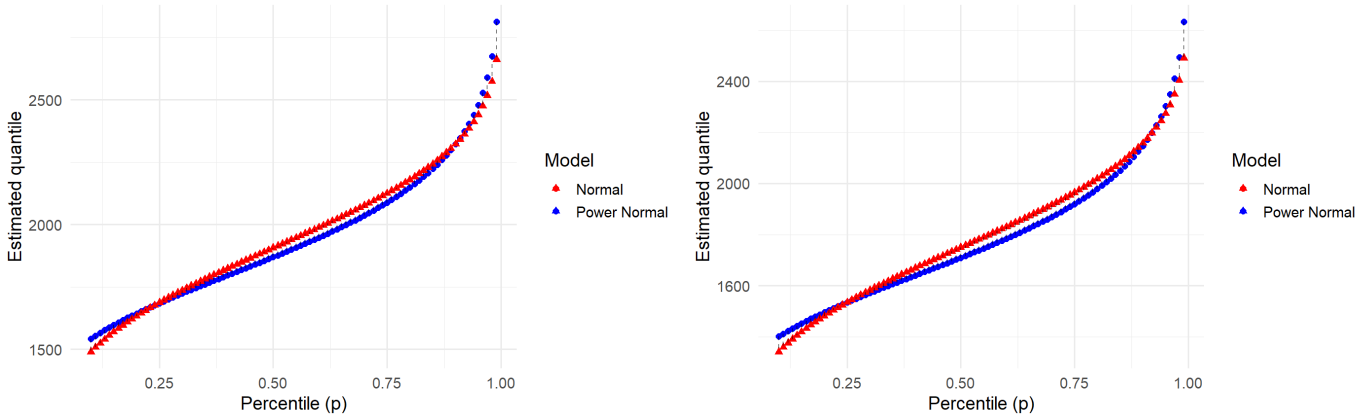
The empirical CDFs and fitted CDFs in Fig. 3 show that both models provide nearly identical fits in the central region.



(a) ECDF and fitted CDFs for Shock data. (b) ECDF and fitted CDFs for Vibration data.

Figure 3: Closeness between empirical CDF and fitted CDFs for the marginals.

However, meaningful differences arise when examining tail quantities. Fig. 4 shows the estimated pth quantiles (for $p \in (0.1, 0.99)$) under each fitted model. Although the curves overlap near the median, substantial divergence appears in the extremes.



(a) Shock data.

(b) Vibration data.

Figure 4: Estimated quantiles under Power Normal (blue) and Normal (red).

Quantitative comparisons for selected percentiles are reported in Table 3.

Marginals	Model	0.60	0.70	0.80
Shock(X_1)	Power normal	0.1263	0.4006	0.7477
	Normal	0.2533	0.5244	0.8416
Vibration(X_2)	Power normal	0.1051	0.3774	0.7261
	Normal	0.2647	0.5236	0.8534

Table 3: Estimates of p^{th} quantiles under each model.

Thus, even when both models fit the observed data well overall, their inferential implications, particularly for tail-based summaries, differ markedly. This underscores the practical need for a principled discrimination framework.

Finally, Table 4 reports the log-likelihood, AIC, and BIC values for both fits. While PN achieves slightly larger likelihood and smaller AIC in each case, the BIC—which more heavily penalizes models with additional parameters, provides a different ranking. Understanding this behavior theoretically and quantifying its practical impact forms a central goal of this paper.

3 Discrimination criterion and asymptotic distribution

In this section, we present the discrimination criterion adopted in this study and develop its asymptotic distribution, which will subsequently be used for determining the minimum required sample size. The most commonly used method for model discrimination in the literature is the

Dataset	**Model**	$\ell(\hat{\theta})$	**AIC**	**BIC**
Shock	Normal	-42.068	88.14	90.93
	Power Normal	-40.103	86.21	90.40
Vibration	Normal	-41.904	87.81	90.60
	Power Normal	-40.210	86.42	90.61

Table 4: Log-likelihood, AIC, and BIC for Shock and Vibration.

ratio of maximized log-likelihoods. This idea was originally introduced by Cox (1961, 1962) for discriminating between overlapping models. Since then, it has been applied in various contexts: for instance, Gupta et al. (2002) used it to differentiate between overlapping distribution families, while Dumonceaux et al. (1973) and Bain and Engelhardt (1980) applied it to compare specific distributions such as the Weibull and gamma. Due to its simplicity and effectiveness, this technique has become a standard choice in many model comparison studies.

However, this method is not suitable for our case. As previously mentioned, it tends to favor the Power Normal model in every instance, simply because it has more parameters. This inherent bias makes it unreliable for our purposes. To overcome this issue, we instead use the Bayesian Information Criterion (BIC), which introduces a penalty for model complexity.

Before presenting the decision criterion, we define the notations used throughout this paper. Let $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\alpha}$ denote the maximum likelihood estimates (MLEs) of the parameters μ, σ, and α under the assumption that the data follows a Power Normal distribution. Similarly, let $\hat{\mu}_*$ and $\hat{\sigma}_*$ denote the MLEs under the Normal distribution.

The BIC for a model is computed using:

$$\mathrm{BIC} = k \ln(n) - 2 \ln(\hat{L}), \tag{2}$$

where k is the number of model parameters, n is the sample size, and $\hat{L}$ is the maximized likelihood.

For the Power Normal model, the BIC becomes:

$$\mathrm{BIC}_{\mathrm{PN}} = 3 \ln(n) - 2L_{\mathrm{PN}}(\hat{\mu}, \hat{\sigma}, \hat{\alpha}),$$

and for the Normal model:

$$\mathrm{BIC}_{\mathrm{N}} = 2 \ln(n) - 2L_{\mathrm{N}}(\hat{\mu}_*, \hat{\sigma}_*).$$

where

$$L_{\mathrm{PN}}(\hat{\mu}, \hat{\sigma}, \hat{\alpha}) = \sum_{i=1}^{n} \ln f_{\mathrm{PN}}(x_i; \hat{\mu}, \hat{\sigma}, \hat{\alpha}),$$
$$L_{\mathrm{N}}(\hat{\mu}_*, \hat{\sigma}_*) = \sum_{i=1}^{n} \ln f_{\mathrm{N}}(x_i; \hat{\mu}_*, \hat{\sigma}_*).$$

Taking the difference between these two BIC values, we define our decision criterion B_n as:

$$B_n = \mathrm{BIC}_{\mathrm{PN}} - \mathrm{BIC}_{\mathrm{N}} = -2\big(L_{\mathrm{PN}}(\hat{\mu}, \hat{\sigma}, \hat{\alpha}) - L_{\mathrm{N}}(\hat{\mu}_*, \hat{\sigma}_*)\big) + \ln(n). \tag{3}$$

We can further define the difference in maximized log-likelihoods, denoted by T_n, given by:

$$T_n = L_{\mathrm{PN}}(\hat{\mu}, \hat{\sigma}, \hat{\alpha}) - L_{\mathrm{N}}(\hat{\mu}_*, \hat{\sigma}_*), \tag{4}$$

and therefore, we have,

$$B_n = -2T_n + \ln(n). \tag{5}$$

We prefer the Power Normal model over the Normal model if:

$$B_n < 0, \tag{6}$$

and choose the Normal model otherwise.

3.1 Asymptotic distribution

In this section, we develop the asymptotic distribution of both the decision criterion given in 4 and 6. Establishing the asymptotic distribution is crucial because it allows us to draw the correspondence between probability of correct selection (PCS) from Monte Carlo (MC) simulations and the PCS derived from the asymptotic distribution. Apart from that, these results help us to find minimum sample size required to discriminate between two distributions for different values of parameters. This connection simplifies the process for practitioners, enabling them to use the asymptotic distribution to calculate PCS for various scenarios more easily.

We use the following notation. Almost sure convergence will be denoted by a.s. For any functions, $f_1(.)$, $E_{PN}(f_1(U))$ and $V_{PN}(f_1(U))$ will denote the mean and variance of U under

the assumption that U follows $PN(\alpha, \mu, \sigma)$. Similarly, we define, $E_N(f_1(U))$ and $V_N(f_1(U))$ as mean and variance of U under the assumption that U follows $N(\mu_*, \sigma_*^2)$. Moreover, if $f_2(.)$ and $f_1(.)$ are two functions, we define $Cov_{PN}(f_2(U), f_1(U))$ and $Cov_N(f_2(U), f_1(U))$, as covariances where U follows $PN(\alpha, \mu, \sigma)$ and $N(\mu_*, \sigma_*^2)$, respectively.

3.2 Asymptotic distribution of T_n

3.2.1 When data is coming from Power normal

Recalling equation (4), we can write,

$$T_n = \sum_{i=1}^{n} \ln f_{\mathrm{PN}}(X_i; \mu, \sigma, \alpha) - \sum_{i=1}^{n} \ln f_{\mathrm{N}}(X_i; \mu_*, \sigma_*) \tag{7}$$

Therefore, if data is coming from $PN(\alpha^0, \mu^0, \sigma^0)$, we have,

$$E_{PN}(T_n|\alpha^0, \mu^0, \sigma^0) = E_{PN}\left[\sum_{i=1}^{n} \ln f_{\mathrm{PN}}(X_i; \mu, \sigma, \alpha) - \sum_{i=1}^{n} \ln f_{\mathrm{N}}(X_i; \mu_*, \sigma_*)\right]. \tag{8}$$

Similarly, we have,

$$V_{PN}(T_n|\alpha^0, \mu^0, \sigma^0) = V_{PN}\left[\sum_{i=1}^{n} \ln f_{\mathrm{PN}}(X_i; \mu, \sigma, \alpha) - \sum_{i=1}^{n} \ln f_{\mathrm{N}}(X_i; \mu_*, \sigma_*)\right]. \tag{9}$$

Now we present the following theorem which helps us in establishment of asymptotic property of our decision criterion T_n. Note that all assumptions required for the application of the relevant theorems, including the Central Limit Theorem and related probabilistic results, have been verified in the Appendix.

Theorem 3.1 *If Data is coming from PN($\alpha_0, \mu_0, \sigma_0$) then T_n is asymptotically normally distributed with mean $E_{PN}(T_n|\alpha_0, \mu_0, \sigma_0)$ and variance $V_{PN}(T_n|\alpha_0, \mu_0, \sigma_0)$*

To prove the above theorem we need the following lemma :

Lemma 3.1 *Under the assumption that data is coming from PN($\alpha_0, \mu_0, \sigma_0$), we have the following results as $n \to \infty$:*
a)$\hat{\alpha} \to \alpha_0$ a.s., $\hat{\mu} \to \mu_0$ a.s.,$\hat{\sigma} \to \sigma_0$ a.s.

b)$\hat{\mu_*} \to \tilde{\mu_*}$ *a.s.* , $\hat{\sigma_*} \to \tilde{\sigma_*}$ *a.s. where*

$$E_{PN}(\ln(f_N(X;\tilde{\mu_*},\tilde{\sigma_*}))) = \max_{\mu_*,\sigma_*} E_{PN}(\ln(f_N(X;\mu_*,\sigma_*)))$$

c)Let us define

$$T_n^\star = L_{PN}(\alpha_0,\mu_0,\sigma_0) - L_N(\tilde{\mu_*},\tilde{\sigma_*})$$

then $\frac{1}{\sqrt{n}}(T_n - E_{PN}(T_n))$ *is asymptotically equivalent to* $\frac{1}{\sqrt{n}}(T_n^\star - E_{PN}(T_n^\star))$

Proof of lemma 3.1: See in appendix.

Proof of Theorem 3.1: Using Central limit theorem and part (b) of Lemma 3.1, it follows that $\frac{1}{\sqrt{n}}(T^\star - E_{PN}(T_n^\star))$ is asymptotically normally distributed with mean zero and variance $V_{PN}(T_n^\star)$. Therefore, using part (c) of Lemma 3.1, the result follows.

It could be seen that values of $\tilde{\mu_*}$ and $\tilde{\sigma_*}$ will depend on $(\alpha_0,\mu_0,\sigma_0)$. We will be calling them as misspecified parameter. Now we provide expressions for asymptotic mean and asymptotic variance as following. Let

$$AM_{PN}(T_n) = \lim_{n\to\infty} \frac{E_{PN}(T_n)}{n} \qquad \text{and} \quad AV_{PN}(T_n) = \lim_{n\to\infty} \frac{V_{PN}(T_n)}{n} \tag{10}$$

where

$$AM_{PN}(T_n) = E_{PN}(\ln f_{PN}(X;\alpha_0,\mu_0,\sigma_0) - \ln f_N(X;\tilde{\mu_*},\tilde{\sigma_*}))$$
$$AV_{PN}(T_n) = V_{PN}(\ln f_{PN}(X;\alpha_0,\mu_0,\sigma_0) - \ln f_N(X;\tilde{\mu_*},\tilde{\sigma_*}))$$

3.2.2 When data is coming from Normal distribution

As discussed in previous subsection, if data is coming from $N(\mu_*^0,\sigma_*^0)$, we have,

$$E_N(T_n|\mu_*^0,\sigma_*^0) = E_N\left[\sum_{i=1}^n \ln f_{\mathrm{PN}}(X_i;\mu,\sigma,\alpha) - \sum_{i=1}^n \ln f_{\mathrm{N}}(X_i;\mu_*,\sigma_*)\right]. \tag{11}$$

Similarly, we have,

$$V_N(T_n|\mu_*^0,\sigma_*^0) = V_N\left[\sum_{i=1}^n \ln f_{\mathrm{PN}}(X_i;\mu,\sigma,\alpha) - \sum_{i=1}^n \ln f_{\mathrm{N}}(X_i;\mu_*,\sigma_*)\right]. \tag{12}$$

Theorem 3.2 *If Data is coming from* $N(\mu_{*0},\sigma_{*0})$ *then* T_n *is asymptotically normally distributed with mean* $E_N(T_n|\mu_{*0},\sigma_{*0})$ *and variance* $V_N(T_n|\mu_{*0},\sigma_{*0})$

To prove the above theorem we need the following lemma :

Lemma 3.2 *Under the assumption that data is coming from N(μ_{*0}, σ_{*0}), we have the following results as $n \to \infty$:*
a)$\hat{\mu_*} \to \mu_{*0}$ a.s. , $\hat{\sigma_*} \to \sigma_{*0}$ a.s.
b)$\hat{\alpha} \to \tilde{\alpha}$ a.s., $\hat{\mu} \to \tilde{\mu}$ a.s.,$\hat{\sigma} \to \tilde{\sigma}$ a.s.

where

$$E_N(\ln(f_{PN}(X;\tilde{\alpha},\tilde{\mu},\tilde{\sigma}))) = \max_{\alpha,\mu,\sigma} E_N(\ln(f_{PN}(X;\alpha,\mu,\sigma)))$$

c)Let us define

$$T_{n\star} = L_{PN}(\tilde{\alpha},\tilde{\mu},\tilde{\sigma}) - L_N(\mu_{*0},\sigma_{*0})$$

then $\frac{1}{\sqrt{n}}(T_n - E_N(T_n))$ is asymptotically equivalent to $\frac{1}{\sqrt{n}}(T_{n\star} - E_N(T_{n\star}))$

Proof of lemma 3.2: See in appendix.
Proof of Theorem 3.2: Using Central limit theorem and part (b) of Lemma 3.2, it follows that $\frac{1}{\sqrt{n}}(T_{n\star} - E_N(T_{n\star}))$ is asymptotically normally distributed with mean zero and variance $V_N(T_{n\star})$. Therefore, using part (c) of Lemma 3.2, the result follows.
It could be seen that values of $\tilde{\alpha},\tilde{\mu},\tilde{\sigma}$ will depend on μ_{*0},σ_{*0}. We will be calling them as mis-specified parameter. Now we provide expressions for asymptotic mean and asymptotic variance as following. Let

$$AM_N(T_n) = \lim_{n\to\infty} \frac{E_N(T_n)}{n} \qquad \text{and} \quad AV_N(T_n) = \lim_{n\to\infty} \frac{V_N(T_n)}{n}$$

where

$$AM_N(T_n) = E_N(\ln f_{PN}(X;\tilde{\alpha},\tilde{\mu},\tilde{\sigma}) - \ln f_N(X;\mu_{*0},\sigma_{*0}))$$
$$AV_N(T_n) = V_N(\ln f_{PN}(X;\tilde{\alpha},\tilde{\mu},\tilde{\sigma}) - \ln f_N(X;\mu_{*0},\sigma_{*0}))$$

3.3 Asymptotic distribution of B_n

3.3.1 When data is coming from Power normal distribution

Recalling equation (3), we can express B_n as:

$$B_n = -2\Bigg[\sum_{i=1}^{n} \ln f_{\mathrm{PN}}(X_i;\mu,\sigma,\alpha) - \sum_{i=1}^{n} \ln f_{\mathrm{N}}(X_i;\mu_*,\sigma_*)\Bigg] + \log(n). \tag{13}$$

If the data is generated from $PN(\alpha^0,\mu^0,\sigma^0)$, the expected value of B_n can be written as:

$$E_{PN}(B_n \mid \alpha^0,\mu^0,\sigma^0) = E_{PN}\Bigg[-2\Bigg(\sum_{i=1}^{n} \ln f_{\mathrm{PN}}(X_i;\mu,\sigma,\alpha) - \sum_{i=1}^{n} \ln f_{\mathrm{N}}(X_i;\mu_*,\sigma_*)\Bigg) + \log(n)\Bigg]. \tag{14}$$

Similarly, the variance of B_n is given by:

$$V_{PN}(B_n \mid \alpha^0,\mu^0,\sigma^0) = V_{PN}\Bigg[-2\Bigg(\sum_{i=1}^{n} \ln f_{\mathrm{PN}}(X_i;\mu,\sigma,\alpha) - \sum_{i=1}^{n} \ln f_{\mathrm{N}}(X_i;\mu_*,\sigma_*)\Bigg) + \log(n)\Bigg]. \tag{15}$$

Now we present the following theorem which helps us in establishment of asymptotic property of our decision criterion B_n.

Theorem 3.3 *If Data is coming from PN(α_0,μ_0,σ_0) then B_n is asymptotically normally distributed with mean $E_{PN}(B_n|\alpha_0,\mu_0,\sigma_0)$ and variance $V_{PN}(B_n|\alpha_0,\mu_0,\sigma_0)$*

Proof of Theorem 3.3: Using Theorem 3.1, we have, T_n follows an asymptotic normal distribution with mean μ_{T_n} and variance $\sigma^2_{T_n}$:

$$T_n \sim AN(\mu_{T_n},\sigma^2_{T_n}) \tag{16}$$

where

$$AM_{PN} = \lim_{n\to\infty} \frac{\mu_{T_n}}{n} \text{ and } AV_{PN} = \lim_{n\to\infty} \frac{\sigma^2_{T_n}}{n} \tag{17}$$

Here AM_{PN} and AV_{PN} are asymptotic mean and asymptotic variance of T_n when data is coming

from Power Normal. From equation (17), for large n ,

$$\mu_{T_n} = O(n). \tag{18}$$

Using equation (16), we start with:

$$\begin{aligned} & T_n \sim AN(\mu_{T_n}, \sigma^2_{T_n}) \\ \implies & -2T_n + \log(n) \sim AN(-2\mu_{T_n} + \log(n), 4\sigma^2_{T_n}) \end{aligned} \tag{19}$$

where the transformation $-2T_n + \log(n)$ applies a linear scaling.
Thus, we conclude that:

$$B_n \sim AN(-2\mu_{T_n} + \log(n), 4\sigma^2_{T_n})$$

or equivalently,

$$B_n \sim AN(\mu_{B_n}, \sigma^2_{B_n})$$

where $\mu_{B_n} = -2\mu_{T_n} + \log(n)$ and $\sigma^2_{B_n} = 4\sigma^2_{T_n}$.

Using equation (19), we observe that $-2\mu_{T_n} = O(n)$. Given that $O(n)$ grows faster than $O(\log(n))$ as $n \to \infty$, it follows that $\mu_{B_n} = -2\mu_{T_n} + \log(n) < 0$ for large n.

3.3.2 When data is coming from Normal distribution

As discussed in previous subsection, if data is generated from $N(\mu^0_*, \sigma^0_*)$, we have,

$$E_N(B_n \mid \mu^0_*, \sigma^0_*) = E_N\bigg[-2\bigg(\sum_{i=1}^{n} \ln f_{\text{PN}}(X_i; \mu, \sigma, \alpha) - \sum_{i=1}^{n} \ln f_{\text{N}}(X_i; \mu_*, \sigma_*) \bigg) + \log(n) \bigg]. \tag{20}$$

Similarly, the variance of B_n is given by:

$$V_N(B_n \mid \mu^0_*, \sigma^0_*) = V_N\bigg[-2\bigg(\sum_{i=1}^{n} \ln f_{\text{PN}}(X_i; \mu, \sigma, \alpha) - \sum_{i=1}^{n} \ln f_{\text{N}}(X_i; \mu_*, \sigma_*) \bigg) + \log(n) \bigg]. \tag{21}$$

Theorem 3.4 *If Data is coming from $N(\mu_{*0}, \sigma_{*0})$ then B_n is asymptotically normally distributed with mean $E_N(B_n|\mu_{*0}, \sigma_{*0})$ and variance $V_N(B_n|\mu_{*0}, \sigma_{*0})$*

Proof of Theorem 3.4: The proof is similar to that of Theorem 3.3 and follows the same

steps.

4 Minimum sample size determination

In this section, we describe a method to determine the minimum sample size needed to tell apart a normal distribution from a Power Normal distribution. This is done for a user-specified probability of correct selection (PCS). The user must either provide an exact PCS value or a range they are comfortable with. Our method is similar to the one proposed by Gupta and Kundu (2003).

There are various ways to measure how different two distributions are — for example, using the Kolmogorov–Smirnov (K–S) distance or the Hellinger distance. When two distributions are very similar, we usually need a larger sample to tell them apart with a certain level of confidence. But when the distributions are quite different, even a smaller sample may be enough. In many real-world cases, if the difference between two distributions is very small, it may not be necessary to discriminate between them.

So, it is important that the user specifies both the desired PCS and a tolerance limit. The tolerance limit defines the smallest difference (in terms of distance between distributions) that is meaningful for the user. If the actual difference is smaller than this limit, the two distributions are considered practically the same, and we do not attempt to separate them.

Once the PCS and tolerance limit are set, we can calculate the minimum sample size needed. In this work, we use the K–S distance to measure the difference between distributions. However, the same approach could also be used with the Hellinger distance, which we do not explore here. Given the PCS, we use the asymptotic behavior of the test statistic B_n to find the required sample size.

From Theorem 3.3, the statistic B_n is asymptotically normally distributed under the Power Normal model, with mean $E_{\mathrm{PN}}(B_n)$ and variance $V_{\mathrm{PN}}(B_n)$. Therefore, the PCS can be approximated as

$$\mathrm{PCS}(\alpha^0, \mu^0, \sigma^0) = \mathbb{P}(B_n < 0 \mid \alpha^0, \mu^0, \sigma^0) \approx \Phi\left(\frac{-E_{\mathrm{PN}}(B_n)}{\sqrt{V_{\mathrm{PN}}(B_n)}}\right) = \Phi\left(\frac{-\sqrt{n} \cdot \mathrm{AM}_{\mathrm{PN}}(\alpha^0, \mu^0, \sigma^0)}{\sqrt{\mathrm{AV}_{\mathrm{PN}}(\alpha^0, \mu^0, \sigma^0)}}\right), \tag{22}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

To compute the minimum sample size n required to achieve a target protection level p^*, we

solve

$$\Phi\left(\frac{-\sqrt{n}\cdot \mathrm{AM_{PN}}(\alpha^0,\mu^0,\sigma^0)}{\sqrt{\mathrm{AV_{PN}}(\alpha^0,\mu^0,\sigma^0)}}\right)=p^*, \tag{23}$$

which yields

$$n_1=\frac{z_{p^*}^2\cdot \mathrm{AV_{PN}}(\alpha^0,\mu^0,\sigma^0)}{\left(\mathrm{AM_{PN}}(\alpha^0,\mu^0,\sigma^0)\right)^2}, \tag{24}$$

where z_{p^*} is the $100p^*$-th percentile of the standard normal distribution.

Similarly, if the data follow a normal distribution with true parameters μ_*^0 and σ_*^0, the minimum sample size required to attain the same PCS level p^* is

$$n_2=\frac{z_{p^*}^2\cdot \mathrm{AV_N}(\mu_*^0,\sigma_*^0)}{\left(\mathrm{AM_N}(\mu_*^0,\sigma_*^0)\right)^2}, \tag{25}$$

where $\mathrm{AM_N}(\mu_*^0,\sigma_*^0)$ and $\mathrm{AV_N}(\mu_*^0,\sigma_*^0)$ denote the asymptotic mean and variance of the test statistic B_n under the Normal model. We now present the asymptotic means and variances for B_n under both choices of the parent distribution, based on the equations derived in Section 3.

α^0	$\mathrm{AM_{PN}}(\alpha^0,\mu^0,\sigma^0)$	$\mathrm{AV_{PN}}(\alpha^0,\mu^0,\sigma^0)$
3	-0.0073	0.0342
4	-0.0111	0.0480
5	-0.0145	0.0639

Table 5: Asymptotic Mean and Variance of B_n when data is coming from $\mathrm{PN}(\alpha^0,1,1)$

σ_*^0	$\mathrm{AM_N}(\mu_*^0,\sigma_*^0)$	$\mathrm{AV_N}(\mu_*^0,\sigma_*^0)$
0.5	0.0004	8.665204e-05
1	0.0004	0.0001
2	0.0004	0.0001

Table 6: Asymptotic Mean and Variance of B_n when data is coming from $\mathrm{N}(0,\sigma_*^0)$

We now present the Kolmogorov–Smirnov (K–S) distance between the Power Normal distribution $\mathrm{PN}(\alpha^0,\mu^0,\sigma^0)$ and the Normal distribution $\mathrm{N}(\tilde{\mu}_*,\tilde{\sigma}_*)$, where $\tilde{\mu}_*$ and $\tilde{\sigma}_*$ denote misspecified parameters for the Normal model obtained using Lemma 3.1. We emphasize that these K–S values are obtained analytically and do not involve any simulation procedure.

The Kolmogorov–Smirnov distance between two distributions with cumulative distribution functions $F(x)$ and $G(x)$ is defined as

$$D = \sup_x \big|F(x) - G(x)\big|.$$

Accordingly, the K–S distance between the Power Normal and Normal distributions is given by

$$D_{\text{PN,N}} = \sup_x \Big|F_{\text{PN}}(x; \alpha^0, \mu^0, \sigma^0) - \Phi\left(\frac{x - \tilde{\mu}_*}{\tilde{\sigma}_*}\right)\Big|,$$

where F_{PN} denotes the cumulative distribution function of the Power Normal distribution and $\Phi(\cdot)$ is the standard Normal cumulative distribution function.

Likewise, we compute the Kolmogorov–Smirnov (K–S) distance between the Normal distribution $\text{N}(\mu_0^*, \sigma_0^*)$ and the Power Normal distribution $\text{PN}(\tilde{\alpha}, \tilde{\mu}, \tilde{\sigma})$, where $\tilde{\alpha}, \tilde{\mu}, \tilde{\sigma}$ denote misspecified parameters for the Power Normal distribution obtained using Lemma 3.2.

In addition to these distances, we provide the corresponding sample sizes required to achieve a protection level of $p^* = 0.9$, calculated using equations (24) and (25), along with the asymptotic quantities from Table 5 and 6. Here, n_{B_n} denotes the required sample size using the test statistic B_n. Based on the asymptotic distribution of B_n and the KS distance between the

α^0	3	4	5
KS(PN($\alpha^0, 1, 1$), N($\tilde{\mu_*}, \tilde{\sigma_*}$))	0.014	0.017	0.017
n_{B_n}	1054.02	639.83	499.15

Table 7: KS distances between $\text{PN}(\alpha^0, 1, 1)$ and $\text{N}(\tilde{\mu_*}, \tilde{\sigma_*})$

σ_*^0	0.5	1	2
KS(PN($\tilde{\alpha}, \tilde{\mu}, \tilde{\sigma}$), N($0, \sigma_*^0$))	5.755796e-05	4.003611e-05	9.101631e-05
n_{B_n}	704.63	920.61	858.38

Table 8: KS distances between $\text{PN}(\tilde{\alpha}, \tilde{\mu}, \tilde{\sigma})$ and $\text{N}(0, \sigma_*^0)$

two distribution functions, the minimum required sample size can be determined such that the probability of correct selection (PCS) reaches a specified protection level p^*, for a given tolerance level K^*. We illustrate the procedure assuming the data are drawn from a Power Normal distribution; however, the methodology for the normal distribution follows analogously. To determine the minimum sample size required to discriminate between Power Normal and normal distribution functions, we consider a user-specified protection level and a given tolerance

level. Suppose the protection level is set at $p^* = 0.9$, and the tolerance level is defined in terms of the Kolmogorov–Smirnov (K–S) distance as $K^* = 0.014$. This tolerance level implies that the practitioner aims to distinguish between the Power Normal and normal distributions only when the K–S distance between them exceeds 0.014.

From Table 7, it is observed that the K–S distance exceeds 0.014 when $\alpha^0 \geq 3$. Therefore, if the data are generated from a Power Normal distribution, a sample size of at most $n = 1056$ is required to achieve the protection level $p^* = 0.9$ under the specified tolerance level.

It is important to note how the usefulness of the Power Normal distribution depends on the degree of skewness in the data. When the parameter α is close to 1, the Power Normal distribution becomes nearly indistinguishable from the Normal distribution, since $\alpha = 1$ corresponds exactly to the Normal case. In such low-skewness situations, the two fitted models produce almost identical CDFs, PDFs, and quantile estimates, and there is little practical benefit in attempting to discriminate between them. Although our minimum sample size analysis focuses on moderately to highly skewed cases ($\alpha = 3, 4, 5$), the observed pattern is consistent with intuition: as $|\alpha - 1|$ increases, the separation between the two models becomes larger and the required sample size for discrimination decreases. Conversely, for values of α very close to 1, the distributions are too similar, and extremely large samples would be needed to distinguish them. Therefore, in practice, when the estimated $\hat{\alpha}$ is close to 1, the Normal model is adequate, whereas for moderate to large skewness, the Power Normal model provides a better representation, particularly in the tails. The proposed BIC-based rule offers a systematic way for practitioners to make this decision.

5 Simulation Results

This section presents a series of simulation experiments to evaluate the performance of the proposed discrimination procedures. For each setting, data are generated from one of the two competing models, and the probability of correct selection (PCS) is computed across 10,000 Monte Carlo replications. When data are generated from the Power Normal distribution $PN(\alpha, 2, 1)$, the value of α is varied. Conversely, when the data come from the Normal model $N(0, \sigma_*)$, the parameter σ_* is varied. We also compare the Monte Carlo based PCS with the PCS obtained from the corresponding asymptotic distribution for the BIC criterion to examine how well asymptotic approximations holds.

5.1 BIC criterion

In this subsection, we compute the probability of correct selection (PCS) using the BIC criterion based on equation (3). We begin by generating data from the Power Normal distribution $PN(1, 1, \alpha)$. For each generated sample, we fit both the Power Normal model and the Normal model, and compute the corresponding value of B_n. Repeating this procedure over 10,000 replications, the PCS is estimated as the proportion of times $B_n < 0$, indicating correct selection of the true model. We then repeat the same procedure by taking the Normal distribution as the parent distribution, in order to assess the performance of the BIC criterion under this setting. Figure 5 illustrates the PCS behavior for different sample sizes and parameter settings. When the parent distribution is Power Normal, the BIC criterion yields relatively low PCS for small samples, but the PCS steadily increases with sample size.

The Hellinger distance between two probability density functions f and g is defined as

$$H(f, g) = \left(\frac{1}{2} \int \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 dx \right)^{1/2}.$$

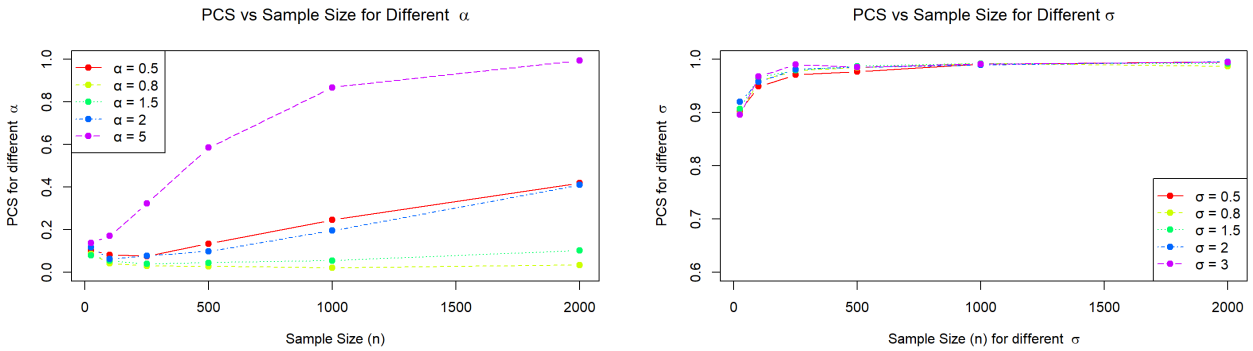
The improvement in PCS is more pronounced as $|\alpha - 1|$ increases. This behavior is consistent with the fact that the Hellinger distance between $PN(\alpha, 2, 1)$ and its closest Normal counterpart also increases (see Table 9). A larger Hellinger distance indicates greater separation between the two models, making them easier to distinguish and thereby leading to higher PCS.

α	2	3	4	5
Closest Normal	$N(2.56, 0.82)$	$N(2.84, 0.74)$	$N(3.02, 0.69)$	$N(3.16, 0.66)$
Hellinger's Distance	0.00038	0.00090	0.00138	0.00179

Table 9: Closest Normal distributions and Hellinger's distances corresponding to $PN(2, 1, \alpha)$.

Figure 6 further confirms that, for sufficiently large sample sizes, the PCS approaches one. This is expected from the consistency of the BIC criterion, which guarantees correct model identification with probability tending to one as the sample size increases.

When the data come from the Normal distribution, the PCS values follow a similar pattern: they are already quite high for moderate sample sizes and continue to improve as n increases. This again demonstrates the strong discriminative power of the BIC-based approach.



(a) PCS based on BIC criterion when data is coming from $PN(1, 1, \alpha)$.

(b) PCS based on BIC criterion when data is coming from $N(0, \sigma)$.

Figure 5: PCS based on BIC criterion for respective choice of parent distribution.

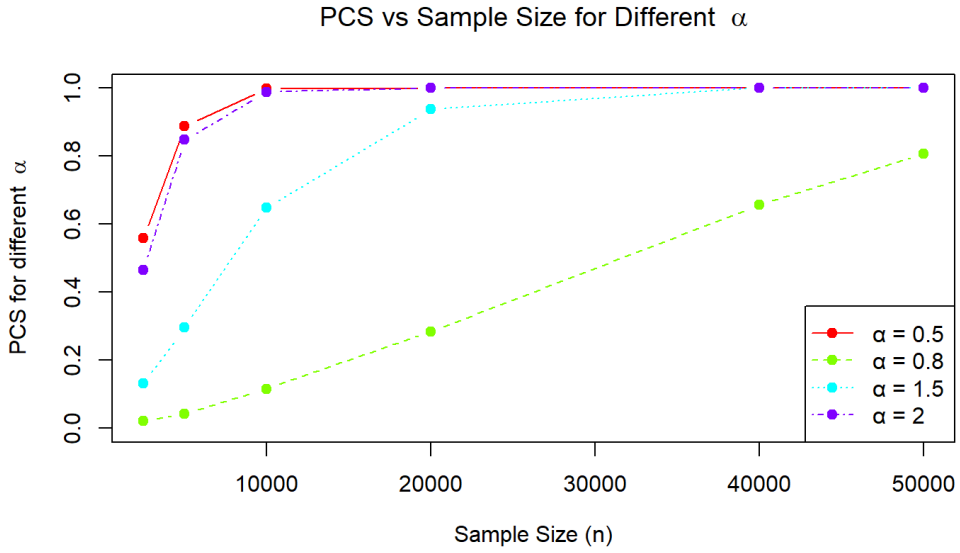


Figure 6: PCS based on BIC criterion criterion when data is coming from $PN(1, 1, \alpha)$.

To examine the agreement between finite-sample and asymptotic behavior, we compare the Monte Carlo estimates of PCS with their asymptotic counterparts derived in Section 3. From Theorem 3.3, the statistic B_n is asymptotically normally distributed under the Power Normal model, with mean $E_{\text{PN}}(B_n)$ and variance $V_{\text{PN}}(B_n)$. Accordingly, the PCS can be approximated using equation (22).

Similarly, when the data are generated from a Normal distribution with true parameters μ_*^0

and σ_*^0, the corresponding PCS can be obtained using the asymptotic distribution of B_n. The asymptotic means and variances of B_n under both choices of the parent distribution are reported in Tables 5 and 6, respectively. These results are presented in Table 10. For the Power Normal parent model, the asymptotic PCS aligns closely with the Monte Carlo PCS once n becomes moderately large.

$\alpha \downarrow n \rightarrow$	500	750	1000	1500	2000
3	0.259	0.412	0.539	0.735	0.843
	(0.276)	(0.423)	(0.542)	(0.710)	(0.826)
4	0.468	0.634	0.768	0.893	0.966
	(0.458)	(0.626)	(0.741)	(0.875)	(0.949)
5	0.606	0.761	0.872	0.977	0.995
	(0.582)	(0.740)	(0.847)	(0.945)	(0.983)

Table 10: Values of PCS using MC simulation and those from asymptotic results (in parentheses) when data is coming from $PN(\alpha, 2, 1)$.

In contrast, when the data are generated from a Normal distribution, the asymptotic PCS is essentially equal to one even for small sample sizes, due to the very small values of AM_N and AV_N (see Table 6). This is fully consistent with the Monte Carlo results reported in Table 11.

$\sigma \downarrow n \rightarrow$	25	50	75	100	200
0.5	0.901	0.937	0.961	0.974	0.979
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
1	0.916	0.938	0.963	0.974	0.982
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
2	0.919	0.939	0.966	0.976	0.984
	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)

Table 11: Values of PCS using MC simulation and those from asymptotic results (in parentheses) when data is coming from $N(0, \sigma)$.

5.2 KS-distance based approach

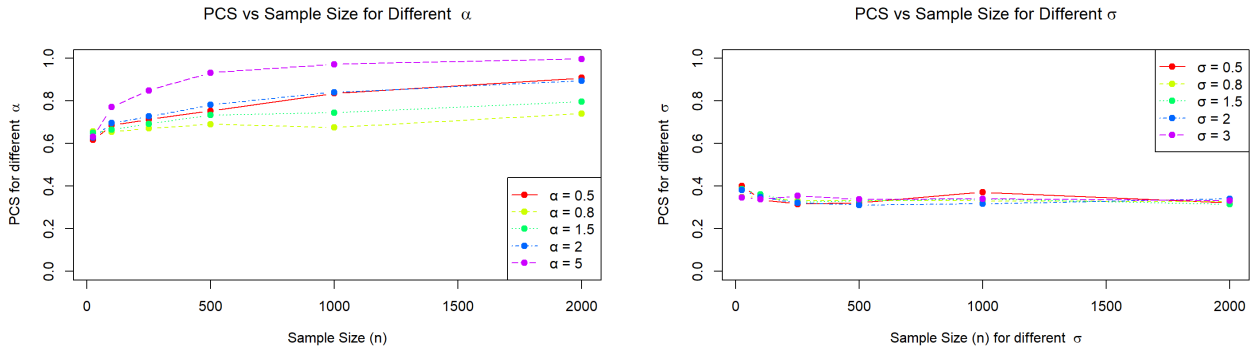
We also consider a Kolmogorov–Smirnov (KS) distance based decision rule for comparison. This method evaluates which fitted distribution lies closer to the empirical cumulative distribution function (ECDF). The KS distance for the Power Normal model is defined as

$$D_{\mathrm{PN}} = \sup_x \big| F_n(x) - F_{\mathrm{PN}}(x; \hat{\alpha}, \hat{\mu}, \hat{\sigma}) \big|,$$

with D_{N} defined analogously for the Normal model. We then compute

$$D_n = D_{\mathrm{PN}} - D_{\mathrm{N}},$$

and select the Power Normal model whenever $D_n < 0$.



(a) PCS based on KS distance criterion when data is coming from $PN(1, 1, \alpha)$.

(b) PCS based on KS distance criterion when data is coming from $N(0, \sigma)$.

Figure 7: PCS based on KS distance criterion for respective choice of parent distribution.

When the data are generated from the Power Normal model, the KS-based rule performs adequately, and the PCS improves with sample size and with increasing $|\alpha - 1|$. However, when the data come from the Normal distribution, the KS-based approach performs poorly: the PCS falls below 0.5 and even slightly decreases with increasing sample size. These results indicate that the KS distance is not a reliable discrimination tool for this problem.

It is worth commenting on the general performance of the discrimination strategies considered in this work. For discrimination problems where the competing models have the same number of parameters, the existing literature consistently reports that the ratio of maximized log-likelihoods (RML) performs very well and often outperforms alternative methods. However, this guidance does not extend to the present setting, because the Power Normal and Normal distributions differ in their parameter dimensions. In such cases, the RML method is no longer applicable, as it systematically favours the model with more parameters. Since the problem of discriminating between distributions with unequal parameter counts has not been explored previously, no generally accepted optimal rule exists. Based on the results obtained in this study, the BIC criterion appears to perform the most reliably across a wide range of scenarios, yielding consistent and stable probabilities of correct selection. Nevertheless, a theoretically optimal discrimination rule for unequal-dimensional model families may exist, and identifying

such a method remains an open direction for future research.

6 Discrimination Performance for Incomplete and Censored Data

To further investigate the applicability of the proposed discrimination strategies in the presence of incomplete observations, we analyze a publicly available insurance dataset included in the `copula` package in R. The data were originally compiled by the U.S. Insurance Services Office (ISO) and consist of general liability claims selected from late settlement periods. Each record corresponds to one of the $n = 1500$ randomly sampled claims and contains three key pieces of information: the indemnity payment (referred to as the "loss"), the corresponding Allocated Loss Adjustment Expense (ALAE), and the applicable policy limit.

The ALAE variable represents the portion of the insurer's expenses that can be directly attributed to settling a specific claim, such as attorney fees and costs associated with claim investigation. Let Y_i denote the recorded loss amount and X_i the associated ALAE for the ith claim.

An important feature of this dataset is the presence of right-censoring due to policy limits. For each claim, the policy limit ℓ_i specifies the maximum indemnity amount payable under the contract. If the actual loss exceeds this limit, only the policy limit is recorded, resulting in a censored observation. Formally, we define the censoring indicator

$$\delta_i = \begin{cases} 1, & \text{if } Y_i \leq \ell_i \quad \text{(loss fully observed)}, \\ 0, & \text{if } Y_i > \ell_i \quad \text{(loss censored)}, \end{cases} \qquad i = 1, \ldots, n.$$

Using this definition, the dataset contains a total of 34 censored loss observations and 1466 fully observed ones. This structure makes the dataset particularly useful for evaluating the performance of model discrimination techniques in the presence of censoring or incomplete information. A scatterplot of (loss, ALAE) on the log scales is depicted in Figure 8. In this dataset, 34 out of the 1500 claims are right–censored due to policy limits, and we inherited the same censoring structure for (loss,ALAE). Both the Normal and Power Normal models were fitted via maximum likelihood under this censoring scheme.

For both the LOSS and ALAE datasets, we fitted the Normal and Power Normal models after standardizing the observations. Under the Normal model, the MLEs for the LOSS data were

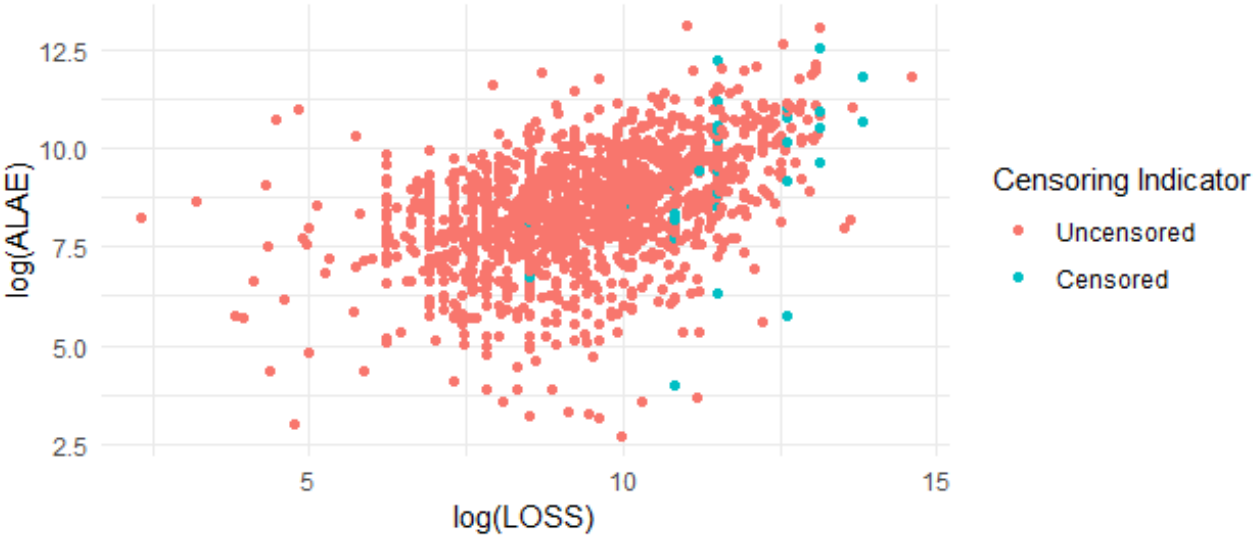


Figure 8: Scatterplot for loss and ALAE (on the log-scale).

$\widehat{\mu_*} = 7.05$ and $\widehat{\sigma_*} = 3.14$, whereas for the ALAE data the corresponding estimates were $\widehat{\mu_*} = 7.82$ and $\widehat{\sigma_*} = 3.59$. In contrast, under the Power Normal model, the LOSS dataset produced the estimates $\widehat{\mu} = -8.28$, $\widehat{\sigma} = 6.37$, and $\widehat{\alpha} = 45.24$, while the ALAE dataset yielded $\widehat{\mu} = -9.13$, $\widehat{\sigma} = 7.25$, and $\widehat{\alpha} = 39.19$. The detailed log-likelihood, AIC, and BIC values corresponding to these fitted models are reported in Table 12, which clearly illustrates the improved fit of the Power Normal model for both datasets even in the presence of censoring.

Dataset	**Model**	$\ell(\hat{\theta})$	**AIC**	**BIC**
LOSS (standardized)	Normal	-164.881	333.762	344.388
	Power Normal	-152.004	310.008	325.948
ALAE (standardized)	Normal	-180.289	364.578	375.204
	Power Normal	-170.373	346.746	362.685

Table 12: Log-likelihood, AIC, and BIC for standardized LOSS and ALAE datasets under Normal and Power Normal models.

This empirical study illustrates that the proposed discrimination methodology remains effective for truncated or incomplete datasets, and the Power Normal distribution continues to offer a markedly superior fit in the presence of skewness and heavy tails.

7 Effect of model misspecification

In this section, we explore the effect of model misspecification on the average estimate of the p^{th} quantile. We are particularly interested in examining the consequences of incorrect model selection when estimating the p^{th} quantile. Specifically, we aim to assess the severity of the impact that such model misspecification has on the relative bias of the estimated p^{th} quantile. Suppose, data is coming from PN$(\alpha^0, \mu^0, \sigma^0)$ and we find the estimates of p^{th} quantile (let p_0 denote true value of the p^{th} quantile) assuming data is indeed coming from PN(say $p_{\hat{PN}}$) and then assuming wrong model i.e. Normal(say $p_{\hat{N}}$). Then, Relative bias under true model is given as:

$$RB_1 = \frac{p_0 - p_{\hat{PN}}}{p_0},$$

where as that of under wrong model is given as:

$$RB_2 = \frac{p_0 - p_{\hat{N}}}{p_0}.$$

First, we generate data from the Power Normal distribution $PN(3, 2, 1)$. Under the correct model assumption (Power Normal), we estimate the p^{th} quantile for different values of p $(0.90, 0.95, \text{and } 0.99)$ across varying sample sizes $n = 25, 50, 75,$ and 100 and then we find the value of relative bias (RB_1). Next, we perform the same estimation under the wrong model (Normal distribution) and compute RB_2. The results are obtained in Table 13.

$p \downarrow$ n $\rightarrow$	25	50	75	100
0.90	0.017	0.008	0.005	0.003
	(0.019)	(0.010)	(0.007)	(0.006)
0.95	0.021	0.009	0.006	0.003
	(0.023)	(0.017)	(0.018)	(0.016)
0.99	0.044	0.019	0.008	0.004
	(0.045)	(0.044)	(0.043)	(0.041)

Table 13: Value above the parenthesis denotes average value of Relative bias under true model where as that value in the parenthesis denotes that of under wrong model (Normal in this case).

We then repeat this entire procedure by altering the parent distribution from Power Normal to Normal $(0, 2)$

$p \downarrow$ n $\rightarrow$	25	50	75	100
0.90	0.009	0.004	0.003	0.002
	(0.040)	(0.018)	(0.017)	(0.008)
0.95	0.014	0.007	0.006	0.001
	(0.052)	(0.022)	(0.019)	(0.009)
0.99	0.008	0.005	0.004	0.003
	(0.064)	(0.024)	(0.018)	(0.017)

Table 14: Value above the parenthesis denotes average value of Relative bias under true model where as that value in the parenthesis denotes that of under wrong model (Power Normal in this case).

From Tables 13 and 14, it is evident that the relative bias under the true model is smaller compared to that under the wrong model assumption. Furthermore, the difference between the two relative biases grows as we progress from the 0.90-th quantile to the 0.99-th quantile. This indicates that as we move toward the tail region, the impact of model misspecification on the p-th quantile becomes more serious.

8 Data analysis

Considering the dataset introduced in Section 1 (see Table 1), we now apply the BIC criterion to determine the most appropriate model for the respective marginal datasets, namely 'Shock' and 'Vibration'.

We begin by analyzing the 'Shock' data. Using the MLEs provided in Section 1, the maximized log-likelihood under the Power Normal model is $L_{\text{PN}}(\hat{\mu}, \hat{\sigma}, \hat{\alpha}) = -40.103$, while for the Normal model, it is $L_{\text{N}}(\hat{\mu_*}, \hat{\sigma_*}) = -42.068$. Based on these log-likelihood values, the corresponding BIC values are $BIC_{\text{PN}} = 90.40$ and $BIC_{\text{N}} = 90.93$. Thus, the difference in BIC values is

$$B_n = BIC_{\text{PN}} - BIC_{\text{N}} < 0,$$

indicating that the Power Normal model is preferred over the Normal model for this dataset.

Similarly, for the 'Vibration' data, the maximized log-likelihood under the Power Normal model is $L_{\text{PN}}(\hat{\mu}, \hat{\sigma}, \hat{\alpha}) = -39.466$, and under the Normal model, it is $L_{\text{N}}(\hat{\mu_*}, \hat{\sigma_*}) = -42.065$. Using these values, the BIC values are $BIC_{\text{PN}} = 89.13$ and $BIC_{\text{N}} = 90.93$. Here again, the difference in BIC values is

$$B_n = BIC_{\text{PN}} - BIC_{\text{N}} < 0,$$

confirming that the Power Normal model is the better choice for this dataset as well.

9 Conclusion

In this study, we have investigated the problem of discriminating between Power Normal and Normal distributions, addressing a challenge that has not received attention in the literature. Unlike most prior work, which focuses on discrimination between distributions with the same number of parameters, we have specifically tackled the complexities arising when the parameter counts differ.

Our findings reveal that conventional methods, such as the ratio of maximized log-likelihoods and the Kolmogorov-Smirnov (KS) distance-based approach, are inadequate for this task. This underscores the need for alternative criteria. The Bayesian Information Criterion (BIC) emerges as an effective solution, providing better results in terms of the probability of correct selection (PCS). Additionally, we derived the asymptotic distribution of the BIC criterion.

To validate our approach, we conducted Monte Carlo (MC) simulations, calculating PCS values both through simulation and using asymptotic results. Our analysis also explored the impact of model misspecification on p^{th} quantile , highlighting the importance of selecting the appropriate model.

Extending the problem of discrimination to higher dimensions, particularly in light of the bivariate extension of the Power Normal distribution proposed by Kundu and Gupta (2013), presents another promising direction for further research. More work is needed in this direction.

Appendix:

Verification of assumptions of White (1982):

Assumption A1:

Assumption A1 requires that the observations $U_1, \ldots, U_n$ are independent and identically distributed random vectors with common distribution G defined on a measurable Euclidean space, and that G admits a Radon–Nikodym density with respect to a dominating measure.

In our setup, let $(X_i)_{i=1}^n$ be a sequence of independent and identically distributed real-valued random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We identify $U_i = X_i$ for all i.

We assume that $X_i \sim PN(\alpha_0, \mu_0, \sigma_0)$, so that the common distribution G is defined on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra. The distribution G is absolutely continuous with respect to Lebesgue measure λ, and hence admits a Radon–Nikodym density $g = dG/d\lambda$ given by

$$g(x) = \frac{\alpha_0}{\sigma_0}\left[\Phi\left(\frac{x-\mu_0}{\sigma_0}\right)\right]^{\alpha_0-1}\phi\left(\frac{x-\mu_0}{\sigma_0}\right), \qquad x \in \mathbb{R}.$$

Since ϕ and Φ are measurable functions, it follows that g is measurable. Therefore, G admits a measurable Radon–Nikodym density with respect to Lebesgue measure.

Hence, Assumption A1 is satisfied.

Assumption A2:

Assumption A2 requires that the family of distribution functions $F(u, \theta)$ admits Radon–Nikodym densities $f(u, \theta)$ that are measurable in u for each $\theta \in \Theta$, continuous in θ for each u, and that the parameter space Θ is a compact subset of $\mathbb{R}^p$.

In our setting, the parametric family is given by the Normal distribution with parameter vector $\theta = (\mu, \sigma) \in \mathbb{R} \times (0, \infty)$. Let $F(x, \theta)$ denote the corresponding distribution function, which admits the density

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}.$$

Measurability: For each fixed θ, the function $f(x; \mu, \sigma)$ is measurable in x, since it is composed of continuous (hence measurable) functions of x.

Continuity in θ: For each fixed $x \in \mathbb{R}$, the function $f(x; \mu, \sigma)$ is continuous in (μ, σ) on $\mathbb{R} \times (0, \infty)$, and hence continuous on any subset $\Theta \subset \mathbb{R} \times (0, \infty)$.

Compactness: To satisfy the compactness requirement, we restrict the parameter space to

$$\Theta = \{(\mu, \sigma) : |\mu| \leq M,\ \epsilon \leq \sigma \leq M\},$$

where $M > 0$ and $\epsilon > 0$ are fixed constants. This set is closed and bounded, and hence compact in $\mathbb{R}^2$.

We assume that M is chosen sufficiently large and ϵ sufficiently small so that the pseudo-true parameter θ^* lies in the interior of Θ, ensuring that this restriction is without loss of generality.

Therefore, all conditions of Assumption A2 are satisfied.

Assumption A3:

Assumption A3 requires that: (a) $E[\log g(X)]$ exists and there exists an integrable function $m(x)$ such that $\sup_{\theta\in\Theta} |\log f_N(x;\theta)| \le m(x)$; and (b) the Kullback–Leibler divergence

$$I(g : f, \theta) = E\left[\log \frac{g(X)}{f_N(X;\theta)}\right]$$

has a unique minimum at some $\theta^* \in \Theta$.

(a) Integrability:

Let $X \sim PN(\alpha_0, \mu_0, \sigma_0)$ with density g. Define $Z = (X - \mu_0)/\sigma_0$. Then

$$\log g(X) = \log \alpha_0 - \log \sigma_0 + (\alpha_0 - 1) \log \Phi(Z) + \log \phi(Z).$$

Using

$$\log \phi(z) = -\tfrac{1}{2} \log(2\pi) - \tfrac{z^2}{2},$$

we have

$$|\log \phi(z)| \le C + \tfrac{1}{2} z^2,$$

so that $E|\log \phi(Z)| < \infty$ whenever $E(Z^2) < \infty$. Since the Power Normal distribution admits finite second moments (see Gupta and Gupta (2008)), this condition holds.

Next, note that $U = \Phi(Z)$ follows a Beta$(\alpha_0, 1)$ distribution under the Power Normal model. Hence,

$$E|\log \Phi(Z)| = -E[\log U] = \frac{1}{\alpha_0} < \infty.$$

Therefore, $E[\log g(X)]$ exists.

We now construct a dominating function for $\log f_N(x;\theta)$. For $(\mu, \sigma) \in \Theta$,

$$\log f_N(x; \mu, \sigma) = -\log \sigma - \frac{1}{2} \log(2\pi) - \frac{(x-\mu)^2}{2\sigma^2}.$$

Since Θ is compact, there exist constants $C_1, C_2 > 0$ such that

$$\sup_{\theta\in\Theta} |\log f_N(x;\theta)| \le C_1 + C_2 x^2.$$

Let $m(x) = C_1 + C_2 x^2$. Since $E[X^2] < \infty$ under the Power Normal distribution, $E[m(X)] < \infty$, establishing the required domination.

(b) Uniqueness of the KL minimizer.

Since $E[\log g(X)]$ does not depend on θ, minimizing $I(g : f, \theta)$ is equivalent to maximizing

$$Q(\mu, \sigma) = E[\log f_N(X; \mu, \sigma)].$$

A direct calculation yields

$$Q(\mu, \sigma) = -\log \sigma - \frac{Var(X) + (E[X] - \mu)^2}{2\sigma^2} + C,$$

where C is constant in (μ, σ).

For fixed σ, $Q(\mu, \sigma)$ is strictly concave in μ, with unique maximizer $\mu^* = E[X]$. Substituting this, we obtain

$$Q(\sigma) = -\log \sigma - \frac{Var(X)}{2\sigma^2} + C,$$

which has a unique maximizer at $\sigma^{*2} = Var(X)$.

Thus, the Kullback–Leibler divergence is uniquely minimized at

$$\theta^* = (E[X], \sqrt{Var(X)}).$$

By choosing M sufficiently large and ϵ sufficiently small, we ensure that θ^* lies in the interior of Θ.

Therefore, Assumption A3 is satisfied.

Assumption A4:

Assumption A4 requires that the first derivatives of the log-density,

$$\frac{\partial}{\partial \theta_i} \log f(u, \theta), \quad i = 1, \ldots, p,$$

exist, are measurable in u for each $\theta \in \Theta$, and are continuously differentiable functions of θ for each u.

In our setting, the parametric model is the Normal distribution with parameter vector $\theta = (\mu, \sigma)$ and density

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}.$$

The log-density is

$$\log f(x;\mu,\sigma) = -\log\sigma - \frac{1}{2}\log(2\pi) - \frac{(x-\mu)^2}{2\sigma^2}.$$

Existence and measurability:

The first derivatives are

$$\frac{\partial}{\partial\mu}\log f(x;\mu,\sigma) = \frac{x-\mu}{\sigma^2}, \qquad \frac{\partial}{\partial\sigma}\log f(x;\mu,\sigma) = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}.$$

For each fixed $\theta \in \Theta$, these are compositions of continuous functions of x with denominators bounded away from zero (since $\sigma \geq \epsilon > 0$). Hence, they are measurable functions of x.

Continuous differentiability in θ:

For each fixed $x \in \mathbb{R}$, the second derivatives are

$$\frac{\partial^2}{\partial\mu^2}\log f(x;\mu,\sigma) = -\frac{1}{\sigma^2}, \qquad \frac{\partial^2}{\partial\mu\partial\sigma}\log f(x;\mu,\sigma) = -\frac{2(x-\mu)}{\sigma^3},$$

$$\frac{\partial^2}{\partial\sigma^2}\log f(x;\mu,\sigma) = \frac{1}{\sigma^2} - \frac{3(x-\mu)^2}{\sigma^4}.$$

These functions are continuous in (μ,σ) on Θ, since σ is bounded away from zero. Therefore, the first derivatives of $\log f(x,\theta)$ are continuously differentiable in θ for each x.

Hence, Assumption A4 of White (1982) is satisfied.

Assumption A5:

Assumption A5 requires that the second derivatives of the log-density and the products of first derivatives,

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(u,\theta), \quad \frac{\partial}{\partial\theta_i}\log f(u,\theta)\frac{\partial}{\partial\theta_j}\log f(u,\theta),$$

are dominated by functions integrable with respect to the true distribution G, uniformly over $\theta \in \Theta$.

In our setting, the model is Normal with $\theta = (\mu,\sigma)$. From previous calculations,

$$\frac{\partial}{\partial\mu}\log f(x;\mu,\sigma) = \frac{x-\mu}{\sigma^2}, \qquad \frac{\partial}{\partial\sigma}\log f(x;\mu,\sigma) = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3},$$

and

$$\frac{\partial^2}{\partial \mu^2} \log f(x;\mu,\sigma) = -\frac{1}{\sigma^2}, \quad \frac{\partial^2}{\partial \mu \partial \sigma} \log f(x;\mu,\sigma) = -\frac{2(x-\mu)}{\sigma^3},$$

$$\frac{\partial^2}{\partial \sigma^2} \log f(x;\mu,\sigma) = \frac{1}{\sigma^2} - \frac{3(x-\mu)^2}{\sigma^4}.$$

Uniform bounds for second derivatives: Since $|\mu| \leq M$ and $\epsilon \leq \sigma \leq M$, we have $|x-\mu| \leq |x|+M$ and $\sigma^{-k} \leq \epsilon^{-k}$. Hence,

$$\sup_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \mu^2} \log f(x;\theta) \right| \leq C_1,$$

$$\sup_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \mu \partial \sigma} \log f(x;\theta) \right| \leq C_2(1+|x|),$$

$$\sup_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \sigma^2} \log f(x;\theta) \right| \leq C_2(1+x^2),$$

for constants $C_1, C_2 > 0$.

Uniform bounds for score functions:

Similarly,

$$\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \mu} \log f(x;\theta) \right| \leq C_3(1+|x|), \qquad \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \sigma} \log f(x;\theta) \right| \leq C_3(1+x^2).$$

Thus, for all i, j,

$$\sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_i} \log f(x;\theta) \frac{\partial}{\partial \theta_j} \log f(x;\theta) \right| \leq C_4(1+x^4).$$

Let $m(x) = C(1+x^4)$ for sufficiently large $C > 0$. Then all second derivatives and score products are bounded by $m(x)$ uniformly over $\theta \in \Theta$.

Since the true distribution is $PN(\alpha_0, \mu_0, \sigma_0)$ and admits a finite fourth moment (see Gupta and Gupta (2008)), it follows that

$$E[m(X)] < \infty.$$

Hence, both the second derivatives and the products of score functions are dominated by an integrable function, and Assumption A5 is satisfied.

Assumption A6:

Assumption A6 requires that: (a) θ^* is an interior point of Θ, (b) $B(\theta^*) = E[s(X,\theta^*)s(X,\theta^*)']$ is nonsingular, and (c) θ^* is a regular point of $A(\theta)$.

(a) Interior point:

From Assumption A3, the pseudo-true parameter is

$$\theta^* = (\mu^*, \sigma^*) = (E[X], \sqrt{Var(X)}).$$

Since the Power Normal distribution is non-degenerate, $Var(X) > 0$, so $\sigma^* > 0$. By choosing M sufficiently large and $\epsilon > 0$ sufficiently small, we ensure that

$$|\mu^*| < M, \quad \epsilon < \sigma^* < M,$$

so that θ^* lies in the interior of Θ.

(b) Nonsingularity of $B(\theta^*)$:

Let $s(X, \theta^*) = (S_1, S_2)'$. Suppose there exist constants a_1, a_2 such that

$$a_1 S_1 + a_2 S_2 = 0 \quad \text{almost surely.}$$

Substituting,

$$a_1 \frac{X - \mu^*}{\sigma^{*2}} + a_2 \left(-\frac{1}{\sigma^*} + \frac{(X - \mu^*)^2}{\sigma^{*3}} \right) = 0.$$

Multiplying through by σ^{*3},

$$a_1 \sigma^* (X - \mu^*) + a_2 \big(-(\sigma^{*2}) + (X - \mu^*)^2 \big) = 0 \quad \text{a.s.}$$

Above condition isn't true unless X is degenerate. Since $Var(X) > 0$, this cannot occur, and thus $a_1 = a_2 = 0$.

Hence, $B(\theta^*)$ is positive definite and therefore nonsingular.

(c) Regularity of $A(\theta)$:

Recall that

$$A(\theta) = E\left[\frac{\partial^2}{\partial \theta \partial \theta'} \log f(X, \theta) \right].$$

In the present case $\theta = (\mu, \sigma)$, so $A(\theta)$ is the 2×2 matrix

$$A(\theta) = \begin{pmatrix} E\left[\dfrac{\partial^2}{\partial \mu^2} \log f(X, \theta) \right] & E\left[\dfrac{\partial^2}{\partial \mu \partial \sigma} \log f(X, \theta) \right] \\ E\left[\dfrac{\partial^2}{\partial \sigma \partial \mu} \log f(X, \theta) \right] & E\left[\dfrac{\partial^2}{\partial \sigma^2} \log f(X, \theta) \right] \end{pmatrix}.$$

Using the expressions derived earlier,

$$\frac{\partial^2}{\partial \mu^2} \log f(X, \theta) = -\frac{1}{\sigma^2}, \qquad \frac{\partial^2}{\partial \mu \partial \sigma} \log f(X, \theta) = -\frac{2(X-\mu)}{\sigma^3},$$

$$\frac{\partial^2}{\partial \sigma^2} \log f(X, \theta) = \frac{1}{\sigma^2} - \frac{3(X-\mu)^2}{\sigma^4}.$$

Taking expectations, we obtain

$$A(\theta) = \begin{pmatrix} -\frac{1}{\sigma^2} & -\frac{2E[X-\mu]}{\sigma^3} \\ -\frac{2E[X-\mu]}{\sigma^3} & \frac{1}{\sigma^2} - \frac{3E[(X-\mu)^2]}{\sigma^4} \end{pmatrix}.$$

At $\theta = \theta^* = (\mu^*, \sigma^*)$, we have $E[X - \mu^*] = 0$, so

$$A(\theta^*) = \begin{pmatrix} -\frac{1}{\sigma^{*2}} & 0 \\ 0 & \frac{1}{\sigma^{*2}} - \frac{3Var(X)}{\sigma^{*4}} \end{pmatrix}.$$

Since $\sigma^{*2} = Var(X)$, this simplifies to

$$A(\theta^*) = \begin{pmatrix} -\frac{1}{\sigma^{*2}} & 0 \\ 0 & -\frac{2}{\sigma^{*2}} \end{pmatrix}.$$

This matrix is diagonal with strictly negative entries, and hence is nonsingular.

Moreover, from Assumptions A4 and A5, the entries of $A(\theta)$ are continuous in θ. Therefore, by continuity of the determinant, $A(\theta)$ remains nonsingular in a neighborhood of θ^*, so its rank is constant in that neighborhood.

Hence, θ^* is a regular point of $A(\theta)$. Hence, all parts of Assumption A6 are satisfied.

Justification of the application of the Central Limit Theorem

To justify the application of the Central Limit Theorem, we verify that

$$E_{PN}\left[(\log g(X))^2\right] < \infty,$$

where g denotes the density of the Power Normal distribution.

Let $X \sim PN(\alpha_0, \mu_0, \sigma_0)$ and define $Z = (X - \mu_0)/\sigma_0$. Then

$$\log g(X) = \log \alpha_0 - \log \sigma_0 + (\alpha_0 - 1) \log \Phi(Z) + \log \phi(Z).$$

Using $(a + b + c)^2 \leq C(a^2 + b^2 + c^2)$ for some constant $C > 0$, it suffices to show that

$$E\big[(\log \Phi(Z))^2\big] < \infty \quad \text{and} \quad E\big[(\log \phi(Z))^2\big] < \infty.$$

Since

$$\log \phi(z) = -\tfrac{1}{2} \log(2\pi) - \tfrac{z^2}{2},$$

we have

$$(\log \phi(z))^2 \leq C(1 + z^4)$$

for some constant $C > 0$. Hence,

$$E[(\log \phi(Z))^2] \leq C(1 + E[Z^4]) < \infty,$$

since the Power Normal distribution admits finite moments of fourth order (see Gupta and Gupta (2008)).

Let $U = \Phi(Z)$. Under the Power Normal model, $U \sim \text{Beta}(\alpha_0, 1)$. Therefore,

$$E[(\log \Phi(Z))^2] = E[(\log U)^2].$$

Using the Beta density,

$$E[(\log U)^2] = \alpha_0 \int_0^1 (\log u)^2 u^{\alpha_0 - 1} \, du.$$

This integral is finite for all $\alpha_0 > 0$, and evaluates to

$$E[(\log U)^2] = \frac{2}{\alpha_0^2} < \infty.$$

Combining the above results, we obtain

$$E_{PN}\big[(\log g(X))^2\big] < \infty.$$

Hence, the required moment condition is satisfied.

Proof of Lemma 3.1(a)

The proof relies on an application of the Strong Law of Large Numbers (SLLN). To apply the SLLN, it suffices to verify that

$$E_{PN}\big[\log f_{PN}(X;\alpha,\mu,\sigma)\big] < \infty.$$

From the expression of the power-normal density,

$$\log f_{PN}(X;\alpha,\mu,\sigma) = \log\alpha - \log\sigma + (\alpha-1)\log\Phi(Z) + \log\phi(Z), \quad Z = \frac{X-\mu}{\sigma}.$$

The required integrability of $\log f_{PN}(X;\alpha,\mu,\sigma)$ follows directly from Assumption A3(a), where it has already been shown that

$$E\big|\log\Phi(Z)\big| < \infty \quad \text{and} \quad E\big|\log\phi(Z)\big| < \infty.$$

Hence,

$$E_{PN}\big[\log f_{PN}(X;\alpha,\mu,\sigma)\big] < \infty.$$

From Strong law of large number, we have:

$$-\frac{1}{n}\sum_{i=1}^{n} log f_{PN}(X_i;\bar{\alpha},\bar{\mu},\bar{\sigma}) \to -E_{PN}(log f_{PN}(X;\bar{\alpha},\bar{\mu},\bar{\sigma})) \text{ a.s.}$$

where a.s. means almost sure convergence. Also, $-E_{PN}(log f_{PN}(X;\bar{\alpha},\bar{\mu},\bar{\sigma}))$ has a minimum at true value ie: $(\alpha^0,\mu^0,\sigma^0)$ which are identifiably unique (as established earlier).
Note that Lemma 2.2 of White (1980) is stated for a general class of objective functions $Q_n(\theta)$ and does not depend on the specific least squares structure. In the present setting, the negative log-likelihood

$$Q_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\log f_{PN}(X_i;\theta)$$

satisfies the same conditions, since (i) $Q_n(\theta)$ converges uniformly almost surely to its expectation by the SLLN and the integrability conditions established earlier, and (ii) the limiting function has a unique minimizer due to identifiability. Therefore, Lemma 2.2 applies in this context.

Moreover,

$$-\frac{1}{n}\sum_{i=1}^{n} logf_{PN}(X_i;\hat{\alpha},\hat{\mu},\hat{\sigma}) = \inf_{\bar{\alpha},\bar{\mu},\bar{\sigma}\in\Theta} -\frac{1}{n}\sum_{i=1}^{n} logf_{PN}(X_i;\bar{\alpha},\bar{\mu},\bar{\sigma})$$

Using Lemma 2.2 of White (1980), we have, $(\hat{\alpha},\hat{\mu},\hat{\sigma}) \rightarrow (\alpha^0,\mu^0,\sigma^0)$ a.s.

Proof of Lemma 3.1(b):

Using similar flow of logic as used in Lemma 3.1(a), proof of Lemma 3.1(b) can be established.

Proof of Lemma 3.1(c):

Now for proving statement of Lemma 3.1(c), we need to show that :

$$\frac{1}{\sqrt{n}}[T_n - E_{PN}(T_n|\alpha^0,\mu^0,\sigma^0) - T_n^* + E_{PN}(T_n^*)] \xrightarrow{P} 0$$

In order to do so let us focus on the establishment of the statement:

$$\frac{1}{\sqrt{n}}[T_n - T_n^*] \xrightarrow{P} 0$$

followed by proving :

$$E_{PN}\left(\frac{1}{\sqrt{n}}[T_n - T_n^*]\right) \longrightarrow 0$$

Before that let:

$$h(X_i;\chi) = logf_{PN}(X_i;\alpha,\mu,\sigma) - logf_N(X_i;\mu_*,\sigma_*)$$

where $\chi = (\alpha,\mu,\sigma,\mu_*,\sigma_*) = (\chi_1,\chi_2,\chi_3,\chi_4,\chi_5)$. Now consider,

$$\begin{aligned}\frac{1}{\sqrt{n}}[T_n - T_n^*] &= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}[h(X_i;\hat{\chi}) - h(X_i;\tilde{\chi})]\right] \\ &= \sum_{j=1}^{5}\left[\frac{1}{n}\sum_{i=1}^{n} h_j'(X_i;\tilde{\chi_j})\right]\sqrt{n}(\hat{\chi_j} - \tilde{\chi_j}) + o_p(1)\end{aligned}$$

where $h_j^{'}(.) = \frac{\partial h(.)}{\partial \chi_j}$. Here $\hat{\chi} = (\hat{\alpha}, \hat{\mu}, \hat{\sigma}, \hat{\mu_*}, \hat{\sigma_*})$ and $\tilde{\chi} = (\alpha^0, \mu^0, \sigma^0, \tilde{\mu_*}, \tilde{\sigma_*})$.$\hat{\chi_j}$ is the j-th component of $\hat{\chi}$ and $\tilde{\chi_j}$ is the j-th component of $\tilde{\chi}$ for $j \in \{1, 2, \ldots, 5\}$. Thus, using weak law of large numbers (WLLN), we have:

$$\frac{1}{n}\sum_{i=1}^{n} h_j^{'}(X_i; \tilde{\chi_j}) \xrightarrow{P} E_{PN}(h_j^{'}(X; \chi))\Big|_{\chi=\tilde{\chi}} \tag{26}$$

$$E_{PN}(h_j^{'}(X; \chi))\Big|_{\chi=\tilde{\chi}} = \Bigg[\int_{-\infty}^{\infty} \frac{\partial}{\partial \chi_j} log\Bigg(\frac{f_{PN}(x; \alpha, \mu, \sigma)}{f_N(x; \mu_*, \sigma_*)}\Bigg) f_{PN}(x; \alpha^0, \mu^0, \sigma^0)\, dx\Bigg]\Bigg|_{\chi=\tilde{\chi}} \tag{27}$$

Using dominated convergence theorem (DCT), we have:

$$E_{PN}(h_j^{'}(X; \chi))\Big|_{\chi=\tilde{\chi}} = 0 \tag{28}$$

Using above results and the fact that $\sqrt{n}(\hat{\chi_j} - \tilde{\chi_j})$ converges to a normal distribution with zero mean and some finite variance(using Theorem 3.2 of White (1982)), we have

$$\frac{1}{\sqrt{n}}[T_n - T_n^*] \xrightarrow{P} 0$$

Now,

$$\begin{aligned}
E_{PN}\Bigg(\frac{1}{\sqrt{n}}[T_n - T_n^*]\Bigg) &= E_{PN}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[h(X_i; \hat{\chi}) - h(X_i; \tilde{\chi})] \\
&\leq \sum_{j=1}^{5} E_{PN}\Bigg|\Bigg[\frac{1}{n}\sum_{i=1}^{n} h_j^{'}(X_i; \tilde{\chi_j})\Bigg]\sqrt{n}(\hat{\chi_j} - \tilde{\chi_j})\Bigg| + o(1) \\
&\leq \sum_{j=1}^{5} \Bigg[E_{PN}[\sqrt{n}(\hat{\chi_j} - \tilde{\chi_j})]^2 E_{PN}[\frac{1}{n}\sum_{i=1}^{n} h^{'}(X_i; \tilde{\chi_j})]^2\Bigg]^{\frac{1}{2}} + o(1)
\end{aligned}$$

Here $o(1)$ refers to a sequence that tends towards 0. Last step is written using Cauchy-Schwarz inequality. Since $E_{PN}[\sqrt{n}(\hat{\chi_j} - \tilde{\chi_j})]^2 < \infty$ (using theorem 3.2 of White (1982)) and using equation (28), we have,

$$E_{PN}\Bigg[\frac{1}{n}\sum_{i=1}^{n} h^{'}(X_i; \tilde{\chi_j})\Bigg]^2 \longrightarrow 0$$

Thus, $E_{PN}\left(\frac{1}{\sqrt{n}}[T_n - T_n^*]\right) \longrightarrow 0$
Hence result follows.

Proof of Lemma 3.2:

Along the same direction as done in case of Lemma 3.1

Statements and Declarations:

- Funding:No funding was received for conducting this study
- Conflict of interest:The authors have no competing interests to declare that are relevant to the content of this article.

References

Abid, R. and Kokonendji, C. C. (2023). Choice between and within the classes of poisson-tweedie and poisson-exponential-tweedie count models. *Communications in Statistics - Simulation and Computation*, 52(5):2115–2129.

Ahmad, M. A., Raqab, M. Z., and Kundu, D. (2017). Discriminating between the generalized Rayleigh and Weibull distributions: Some comparative studies. *Communications in Statistics-Simulation and Computation*, 46(6):4880–4895.

Atkinson, A. C. (1969). A test for discriminating between models. *Biometrika*, 56(2):337–347.

Atkinson, A. C. (1970). A method for discriminating between models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 32(3):323–345.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178.

Bain, L. J. and Engelhardt, M. (1980). Probability of correct selection of Weibull versus Gamma based on likelihood ratio. *Communications in statistics-theory and methods*, 9(4):375–381.

Basu, S. and Kundu, D. (2023). Model misspecification of log-normal and Birnbaum-Saunders distributions. *Communications in Statistics-Simulation and Computation*, pages 1–21.

Basu, S. and Ng, H. K. T. (2025). A density power divergence measure to discriminate between generalized exponential and Weibull distributions. *Statistical Papers*, 66(1):1–25.

Chen, W. (1980). On the tests of separate families of hypotheses with small sample size. *Journal of Statistical Computation and Simulation*, 11(3-4):183–187.

Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 105–123.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society: Series B (Methodological)*, 24(2):406–424.

Dey, A. K. and Kundu, D. (2009). Discriminating among the log-normal, Weibull, and generalized exponential distributions. *IEEE Transactions on Reliability*, 58(3):416–424.

Dey, A. K. and Kundu, D. (2012). Discriminating between the Weibull and log-normal distributions for Type-II censored data. *Statistics*, 46(2):197–214.

Diyali, B., Kumar, D., and Singh, S. (2023). Discriminating between log-normal and log-logistic distributions in the presence of Type-II censoring. *Computational Statistics*, pages 1–25.

Diyali, B., Kumar, D., and Singh, S. (2024). Discriminating among inverse Weibull, lognormal, and inverse Gaussian distributions. *Quality and Reliability Engineering International*, 40(4):1698–1718.

Dumonceaux, R., Antle, C. E., and Haas, G. (1973). Likelihood ratio test for discriminating between two models with unknown location and scale parameters. *Technometrics*, 15(1):19–27.

Dyer, A. R. (1973). Discrimination procedures for separate families of hypotheses. *Journal of the American Statistical Association*, 68(344):970–974.

Gupta, R. D. and Gupta, R. C. (2008). Analyzing skewed data by power normal model. *Test*, 17(1):197–210.

Gupta, R. D. and Kundu, D. (2003). Discriminating between Weibull and generalized exponential distributions. *Computational statistics & data analysis*, 43(2):179–196.

Gupta, R. D., Kundu, D., and Manglick, A. (2002). Probability of correct selection of gamma versus ge or Weibull versus ge based on likelihood ratio statistic. In *Recent advances in statistical methods*, pages 147–156. World Scientific.

Johnson, R. A., Wichern, D. W., et al. (2002). Applied multivariate statistical analysis.

Khakifirooz, M., Tseng, S. T., and Fathi, M. (2020). Model misspecification of generalized gamma distribution for accelerated lifetime-censored data. *Technometrics*, 62(3):357–370.

Kundu, D. and Gupta, R. D. (2013). Power-normal distribution. *Statistics*, 47(1):110–125.

Kuş, C., Pekgör, A., and Kınacı, İ. (2019). Discriminating between the lognormal and Weibull distributions under progressive censoring. *Cumhuriyet Science Journal*, 40(2):493–504.

Raqab, M. Z., Al-Awadhi, S. A., and Kundu, D. (2018). Discriminating among Weibull, log-normal, and log-logistic distributions. *Communications in Statistics-Simulation and Computation*, 47(5):1397–1419.

White, H. (1980). Nonlinear regression on cross-section data. *Econometrica: Journal of the Econometric Society*, pages 721–746.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25.

Wiens, B. L. (1999). When log-normal and gamma models give different results: a case study. *The American Statistician*, 53(2):89–93.