International Journal of
Science and Engineering Investigations
vol. 6, issue 65, June 2017
ISSN: 2251-8843

Sampling Problem from Elliptically Contoured Distribution and Its Application in Psychiatric Screening Instruments

Olosunde A. A.[1], Soyinka A. T.[2]
[1]Department of Statistics, Obafemi Awolowo University, Ile-Ife
[2]Neuro-Psychiatry Hospital Aro, Abeokuta
([1]akinolosunde@gmail.com, [2]soyinkataiwo@yahoo.ie)

Abstract- In this paper, we employed the series expansion of incomplete gamma function via the Taylor series to obtain the function for pivotal quantity from which the confidence intervals for each parameter of a special family of elliptically contoured distribution called the exponential power distribution (EPD) were derived. The results obtained were further used to determine the estimated sample size when sampling from EPD as well as the pivotal test statistic for the test of mean differences between two independent EPD samples. Psychiatric screening instruments were used to demonstrate the applicability of the results.

Keywords- *Incomplete Gamma, Pivotal Quantity, Exponential Power Distribution, Test Statistics*

I. INTRODUCTION

By definition the density function of exponential power distribution (EPD) is:

$$f(x,\mu,\sigma,\beta)=\frac{1}{\sigma\Gamma\left(1+\frac{1}{2\beta}\right)2^{1+\frac{1}{2\beta}}}\exp^{-\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right]} \quad (1)$$

$$-\infty<x<\infty;-\infty<\mu<\infty$$

where $\beta>0$ is the shape parameter and $\sigma>0$ is the scale parameter.

Note: When $\beta=\frac{1}{2}$ equation (1) becomes a Laplace function. When $\beta=1$ the equation becomes a normal density that approaches a uniform density as β values increases beyond one towards infinity. However for $\beta<1$ the distribution has heavier tails that is useful in providing robustness towards outliers [1] [2]. Density (1) has generalized normal distribution and very useful in modeling real life phenomenon e.g. [1] have used it as an underlining distribution in repeated measurements, have also used it in modeling poultry feeds data.

Likewise the corresponding cumulative distribution function (CDF) for the EPD is

$$F(x)=\frac{1}{2}+\frac{1}{2}\left[\frac{\gamma\left(\frac{1}{2\beta},\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|\right]^{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}\right] \quad (2)$$

where γ is the incomplete gamma function

Next we express equation (2) in terms of the incomplete gamma series expansion.

As given by [3] the incomplete gamma series expansion is given as:

$$F_p[y]=\frac{\gamma(p,y)}{\Gamma(p)}=\frac{y^p e^{-y}}{\Gamma(p+1)}\left[\begin{array}{l}1+\frac{y}{p+1}+\frac{y^2}{(p+1)(p+2)}\\+\frac{y^3}{(p+1)(p+2)(p+3)}+....+\end{array}\right] \quad (3)$$

So we proceed to re-write equation (2) in the form of equation (3) to obtain equation (4):

$$F_{\frac{1}{2\beta}}\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right]=\frac{\gamma\left(\frac{1}{2\beta},\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|\right]^{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}$$

$$=\frac{\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right]^{\frac{1}{2\beta}}}{\Gamma\left(\frac{1}{2\beta}+1\right)}\left[\begin{array}{l}1-\frac{\frac{1}{2\beta}\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right]}{1!\left(\frac{1}{2\beta}+1\right)}+\frac{\frac{1}{2\beta}\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right]^2}{2!\left(\frac{1}{2\beta}+2\right)}\\-\frac{\frac{1}{2\beta}\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right]^3}{3!\left(\frac{1}{2\beta}+3\right)}+\frac{\frac{1}{2\beta}\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right]^4}{4!\left(\frac{1}{2\beta}+4\right)}+....+\end{array}\right] \quad (4)$$

By further simplification we can express equation (4) as:

$$F_{\frac{1}{2\beta}}\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right] = \frac{\left[\frac{1}{2}\right]^{\frac{1}{2\beta}}}{\Gamma\left(\frac{1}{2\beta}+1\right)} \left[\frac{\left|\frac{x-\mu}{\sigma}\right|}{0!} + \frac{\left(-\frac{1}{2}\right)\left|\frac{x-\mu}{\sigma}\right|^{2\beta+1}}{1!(2\beta+1)} + \frac{\left(\frac{1}{2}\right)^2\left|\frac{x-\mu}{\sigma}\right|^{4\beta+1}}{2!(4\beta+1)} + \frac{\left(-\frac{1}{2}\right)^3\left|\frac{x-\mu}{\sigma}\right|^{6\beta+1}}{3!(6\beta+1)} + \frac{\left(\frac{1}{2}\right)^4\left|\frac{x-\mu}{\sigma}\right|^{8\beta+1}}{4!(8\beta+1)} + ... + \right] \quad (5)$$

Equation (5) resembles a separate Taylor series expansion for product of two functions when $\beta \geq 1$ and $0 < \beta < 1$.

Hence Equation (5) approximate to:

$$F_{\frac{1}{2\beta}}\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right] = \frac{\left[\frac{1}{2}\right]^{\frac{1}{2\beta}}}{\Gamma\left(\frac{1}{2\beta}+1\right)}\left(e^{1/2}\right)\ln\left[1+\left|\frac{x-\mu}{\sigma}\right|\right] \quad (6)$$

for every $\beta \geq 1$

Or

$$F_{\frac{1}{2\beta}}\left[\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{2\beta}\right] = \frac{\left[\frac{1}{2}\right]^{\frac{1}{2\beta}}}{\Gamma\left(\frac{1}{2\beta}+1\right)}\left(e^{-1/2}\right)\tan\left|\frac{x-\mu}{\sigma}\right| \quad (7)$$

for $0 < \beta < 1$

Next we review some moment properties of EPD.

II. MOMENTS OF EPD

If a random variable X has the pdf equation (1) then its mth moments [8] can be obtained from the relation:

$$E\left(X^m\right) = \frac{\left[(-1)^m + 1\right]\int_0^\infty \left(\sigma(2Z)^{\frac{1}{2\beta}} - \mu\right)^m Z^{\frac{1}{2\beta}-1} e^{-Z} dZ}{2\Gamma\left(\frac{1}{2\beta}\right)}$$

By evaluating this equation the estimates of the first four moments $m = 1, 2, 3, 4$ are:

$$E(X) = \mu$$

$$E\left(X^2\right) = \mu^2 + \frac{\sigma^2 2^{\frac{2}{2\beta}}\Gamma\left(\frac{3}{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}$$

$$E\left(X^3\right) = \mu^3 + \frac{3\mu\sigma^2 2^{\frac{2}{2\beta}}\Gamma\left(\frac{3}{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}$$

$$E\left(X^4\right) = \mu^4 + \frac{6\mu^2\sigma^2 2^{\frac{2}{2\beta}}\Gamma\left(\frac{3}{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)} + \frac{\sigma^4 2^{\frac{4}{2\beta}}\Gamma\left(\frac{5}{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}$$

In addition, its central moment estimates are:

$$E\left|X - E(X)\right| = \frac{\sigma 2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}; \; Var(X) = \frac{\sigma^2 2^{\frac{2}{2\beta}}\Gamma\left(\frac{3}{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}$$

$$E\left(X - E(X)\right)^3 = 0$$

Skewness=0

$$E\left(X - E(X)\right)^4 = \frac{\sigma^4 2^{\frac{4}{2\beta}}\Gamma\left(\frac{5}{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}; \text{ Kurtosis=}\frac{\Gamma\left(\frac{5}{2\beta}\right)\Gamma\left(\frac{1}{2\beta}\right)}{\Gamma^2\left(\frac{3}{2\beta}\right)}$$

Next we proceed to obtain the function for the pivotal quantity of EPD.

III. PIVOTAL QUANTITY FUNCTION FOR EPD

Definition 3.1:- Let X be a random variable and define $F(a) = P[X \leq a]$; then a random variable $U = -2\ln F(X)$ has the density $\left(\frac{1}{2}\right)e^{-\frac{U}{2}}I_{(0,\infty)}(U)$ which is a χ_2^2 density defined for every $F(X)$ over the uniform distribution interval $(0,1)$. Likewise the random variable $V = -2\ln\left[1 - F(X)\right]$ is χ_2^2 density. So if $U_1, U_2, ..., U_n$ are pivotal points each i.i.d. χ_2^2 density then a sum across the points $PQ(X_i, \theta) = \sum_{i=1}^{n} U_i = -2\sum_{i=1}^{n}\ln F(X_i)$ has a χ_{2n}^2 density and so is a pivotal quantity (PQ) for θ.

Hence $PQ_1(X_i,\theta)=\sum_{i=1}^{n}U_i$ and $PQ_2(X_i,\theta)=\sum_{i=1}^{n}V_i$ $=-2\sum_{i=1}^{n}\ln\left[1-F(X_i)\right]$ are pivotal quantities.

So we proceed to apply Definition 1.1 to equations (6) and (7). Given:

$$PQ(X_i,\theta)=-2\sum_{i=1}^{n}\ln F(X_i) \tag{8}$$

where $F(X_i)=F_{\frac{1}{2\beta}}\left[\frac{1}{2}\left|\frac{x_i-\mu}{\sigma}\right|^{2\beta}\right]$ in equations (6) and (7) then after simplification,

$$\begin{aligned} PQ=2n&\left[\ln\left(2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right)\right)-\frac{1}{2}\right]\\ &-2\sum_{i=1}^{n}\ln\left[\ln\left(1+\left|\frac{x_i-\mu}{\sigma}\right|\right)\right] \end{aligned} \tag{9}$$

for every $\beta\geq 1$

or

$$\begin{aligned} PQ=2n&\left[\ln\left(2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right)\right)-\frac{1}{2}\right]\\ &-2\sum_{i=1}^{n}\ln\left[\tan\left|\frac{x_i-\mu}{\sigma}\right|\right] \end{aligned} \tag{10}$$

for $0\prec\beta\prec 1$

IV. CONFIDENCE INTERVAL FOR EACH PARAMETER OF EPD

A. *Definition 4.1*

If $PQ(X_i,\theta)=q(X_1,X_2,...,X_n;\theta)$ is a pivotal quantity with pdf defined within the fixed interval $0<\gamma<1$; then there will exist q_1 and q_2 depending on γ such that $P(q_1<PQ<q_2)=\gamma$. q_1 and q_2 are the confidence intervals for the parameter θ in PQ and γ is the confidence probability that the true value of θ is between the interval q_1 and q_2 [5], [8].

B. *Confidence interval for the shape parameter, β*

From the stated Definition 4.1 given:

$$P(q_1<PQ<q_2)=\gamma \tag{11}$$

Then we have the following Proposition:

Proposition 1: Let random variables $X_1,X_2,...,X_n$ from distribution (1). Then the confidence interval for the shape parameter β is given as:

$$e^{\left(\frac{q_1}{2n}+\frac{1}{2}\right)}\sum_{i=1}^{n}\left[Z_{i1}\right]^{\frac{1}{n}}<2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right)<e^{\left(\frac{q_2}{2n}+\frac{1}{2}\right)}\sum_{i=1}^{n}\left[Z_{i1}\right]^{\frac{1}{n}}$$

for every $\beta\geq 1$

or

$$e^{\left(\frac{q_1}{2n}-\frac{1}{2}\right)}\sum_{i=1}^{n}\left[Z_{i2}\right]^{\frac{1}{n}}<2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right)<e^{\left(\frac{q_2}{2n}-\frac{1}{2}\right)}\sum_{i=1}^{n}\left[Z_{i2}\right]^{\frac{1}{n}}$$

for $0<\beta<1$

Proof:

We substitute the PQ in equations (9) and (10) into equation (11) to obtain

$$P\left(\begin{aligned} &q_1<2n\left[\ln\left(2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right)\right)-\frac{1}{2}\right]\\ &-2\sum_{i=1}^{n}\ln\left[\ln\left(1+\left|\frac{x_i-\mu}{\sigma}\right|\right)\right]<q_2 \end{aligned}\right)=\gamma \tag{12}$$

or

$$P\left(\begin{aligned} &q_1<2n\left[\ln\left(2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right)\right)-\frac{1}{2}\right]\\ &-2\sum_{i=1}^{n}\ln\left[\tan\left|\frac{x_i-\mu}{\sigma}\right|\right]<q_2 \end{aligned}\right)=\gamma \tag{13}$$

Supposing we re-express a section of equations (12) and (13) such that:

$$Z_{i1}=\ln\left[1+\left|\frac{x_i-\mu}{\sigma}\right|\right] \tag{14}$$

$$Z_{i2}=\tan\left|\frac{x_i-\mu}{\sigma}\right| \tag{15}$$

Then equation (12) and (13) becomes:

$$P\left(\begin{aligned} &q_1<2n\left[\ln\left(2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right)\right)-\frac{1}{2}\right]\\ &-2\sum_{i=1}^{n}\ln Z_{i1}<q_2 \end{aligned}\right)=\gamma \tag{16}$$

$$P\begin{pmatrix} q_1 < 2n\left[\ln\left(2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right)\right)-\frac{1}{2}\right] \\ -2\sum_{i=1}^{n}\ln\left[\tan\left|\frac{x_i-\mu}{\sigma}\right|\right] < q_2 \end{pmatrix} = \gamma \quad (17)$$

Hence from equations (16) and (17), the upper and lower confidence interval for the shape parameter of EPD is:

$$e^{\left(\frac{q_1}{2n}+\frac{1}{2}\right)}\sum_{i=1}^{n}\left[Z_{i1}\right]^{\frac{1}{n}} < 2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right) < e^{\left(\frac{q_2}{2n}+\frac{1}{2}\right)}\sum_{i=1}^{n}\left[Z_{i1}\right]^{\frac{1}{n}} \quad (18)$$

for every $\beta \geq 1$

or

$$e^{\left(\frac{q_1}{2n}-\frac{1}{2}\right)}\sum_{i=1}^{n}\left[Z_{i2}\right]^{\frac{1}{n}} < 2^{\frac{1}{2\beta}}\Gamma\left(\frac{1}{2\beta}+1\right) < e^{\left(\frac{q_2}{2n}-\frac{1}{2}\right)}\sum_{i=1}^{n}\left[Z_{i2}\right]^{\frac{1}{n}} \quad (19)$$

for $0<\beta<1$

Since q_1 and q_2 has density $\chi^2_{2n,\left(\frac{1\pm A}{2}\right)}$ we can obtain the confidence intervals of β to a probability area of A% at specific sample size n; provided the estimate of $\mu\left(\hat{\mu}\right)$ and $\sigma\left(\hat{\sigma}\right)$ are known.

C. *Confidence interval for the location parameter (μ) with known variance*

Considering equations (14) and (15) as pivotal quantity (PQ); we obtain the confidence interval for the location parameter μ.

$$P\left(q_1 \prec \ln\left[1+\left|\frac{x_i-\mu}{\sigma}\right|\right] \prec q_2\right) = \gamma \quad (20)$$

Supposing observed values $x_i's$ are sampling distribution of sample means then the confidence interval for the location parameter μ from the mean of its sample means is:

$$P\left[\overline{X}-\frac{\sigma}{\sqrt{n}}\left(e^{q_1}-1\right) < \mu < \overline{X}+\frac{\sigma}{\sqrt{n}}\left(e^{q_2}-1\right)\right] = \gamma \quad (21)$$

Likewise from equation (15), the confidence interval for μ is:

$$P\left[\overline{X}-\frac{\sigma}{\sqrt{n}}\tan^{-1} q_1 < \mu < \overline{X}+\frac{\sigma}{\sqrt{n}}\tan^{-1} q_2\right] = \gamma \quad (22)$$

Where q_1 and q_2 are standard normal deviates.

So constructing the 95% confidence interval for μ in the two cases for a unit scale we have:

$$\left[\overline{X}+\frac{0.859}{\sqrt{n}} \prec \mu \prec \overline{X}+\frac{6.099}{\sqrt{n}}\right]$$

and

$\left[\overline{X}-\frac{1.099}{\sqrt{n}} \prec \mu \prec \overline{X}+\frac{1.099}{\sqrt{n}}\right]$ respectively.

Note: An estimate of the scale parameter $\left(\hat{\sigma}\right)$ can be obtained from the variance of X.

However, assuming equations (14) and (15) are standard normal, then the equation can be extended to its corresponding t-distribution with $n-1$ degrees of freedom

V. SAMPLE SIZE ESTIMATION FOR SAMPLING FROM EPD

Making *n* the subject of the formula in equation (21) we obtain the sample size estimation formula required to sample from an EPD as:

$$n=\frac{\left(e^{q}-1\right)^2\sigma^2}{\left|\overline{X}-\mu\right|^2} \quad (25)$$

where q is the normal deviate.

Assuming q is the normal deviate at the level of type I (α) and type II (β) error then:

$$q = z_{\alpha/2}+z_{\beta} \quad (26)$$

Hence:

$$n=\frac{\left(e^{z_{\alpha/2}+Z_\beta}-1\right)^2\sigma^2}{\left|\overline{X}-\mu\right|^2}=\frac{\left(e^{z_{\alpha/2}+Z_\beta}-1\right)^2\sigma^2}{E^2} \quad (27)$$

where E is the estimated mean deviation or error to be tolerated.

Finally substituting E as the mean deviation in equation

$$E\left(X-E\left(X\right)\right)^4=\frac{\sigma^4 2^{\frac{4}{2\beta}}\Gamma\left(\frac{5}{2\beta}\right)}{\Gamma\left(\frac{1}{2\beta}\right)}; \text{ Kurtosis}=\frac{\Gamma\left(\frac{5}{2\beta}\right)\Gamma\left(\frac{1}{2\beta}\right)}{\Gamma^2\left(\frac{3}{2\beta}\right)}$$

we have the sample size as:

$$n=\frac{\left(e^{z_{\alpha/2}+Z_\beta}-1\right)^2\Gamma^2\left(\frac{1}{2\hat{\beta}}\right)}{2^{\frac{1}{\hat{\beta}}}\Gamma^2\left(\frac{1}{\hat{\beta}}\right)} \tag{28}$$

for $\hat{\beta}\geq 1$, or:

$$n=\frac{\left[\tan^{-1}\left(z_{\alpha/2}+Z_\beta\right)\right]^2\Gamma^2\left(\frac{1}{2\hat{\beta}}\right)}{2^{\frac{1}{\hat{\beta}}}\Gamma^2\left(\frac{1}{\hat{\beta}}\right)} \tag{29}$$

for $0\prec\hat{\beta}\prec 1$.

From equation (28) when $\beta=1$; the sample size for the full curve $N=2n$ is:

$$N=\sqrt{\pi}\left(e^{Z_{\alpha/2}+Z_\beta}-1\right)^2 \tag{30}$$

VI. PIVOTAL QUANTITY FOR THE TEST OF MEAN DIFFERENCES BETWEEN TWO INDEPENDENT EPD SAMPLES

The pivotal quantity for the test of mean differences between two independent samples of EPD can be derived from equation (21) and (22) provided the scale parameter or its estimated value is known when H_0: $\mu_1=\mu_2$. After simplification, the pivotal quantity is thus given as:

$$Z_1=\ln\left[1+\frac{\bar{y}-\bar{x}}{\frac{\sigma_1}{\sqrt{n_1}}+\frac{\sigma_2}{\sqrt{n_2}}}\right] \tag{31}$$

and

$$Z_2=\tan\left[\frac{\bar{y}-\bar{x}}{\frac{\sigma_1}{\sqrt{n_1}}+\frac{\sigma_2}{\sqrt{n_2}}}\right] \tag{32}$$

Where Z_1 and Z_2 are standard normal deviates. The null hypothesis will be rejected when $|Z_i|>|Z_{\alpha/2}|$ for $i=1,2$.

VII. APPLICATION

Assuming the random samples $X_1,\ldots,X_n$ are scores obtained from scoring a given questionnaire or event or a likert scale; such that the scoring order $X_1<X_2<X_3<\ldots..<X_n$ denotes ranks; then partitioning the random scores at lower and upper rejection region say at points X_a and X_b indicates a truncation of the CDF within the ordered samples $X_1<X_2<X_3<\ldots..<X_n$ at area $\left(\frac{1-A}{2}\right)\%$ each. The truncation at points X_a and X_b may be based on clinical significance or not. However, irrespective of whether a point $X_{med}=\mu$ exist within the acceptance region that divides the entire CDF to two distinct groups of diverse (medical) condition or not; we can determine the sample size of the pivot points between the two partitions that will span an acceptance region with area A%.

A. *Samples size estimation formula for likert scale questionnaire with assumed EPD*

Depending on the scale of questionnaire under study, the above sample size formula can be adjusted to suit all cases of likert scale measurements. If a likert scale has options Yes/Neutral/No then its mean deviation is scaled over a unit standard scale. Also for likert scale Strongly Agree (SA), Agree (A), Neutral (N), Disagree (D), Strongly Disagree (SD) the mean deviation is a multiple of two unit standard scales. So for likert scale measurements, as the options increases on either side of the divides, the multiplying factor of the unit scale increases.

So we have:

$$n=\frac{\left(e^{z_{\alpha/2}+Z_\beta}-1\right)^2\sigma^2}{\left|\bar{X}-\mu\right|^2}=\frac{\left(e^{z_{\alpha/2}+Z_\beta}-1\right)^2\sigma^2}{(k\sigma)^2}=\frac{\left(e^{z_{\alpha/2}+Z_\beta}-1\right)^2}{k^2}$$

for half side of the EPD curve. The total sample size is thus $N=2n=\frac{2\left(e^{z_{\alpha/2}+Z_\beta}-1\right)^2}{k^2}$ where k is the multiplying factor of the standard unit scale for the mean deviation.

TABLE I. NUMBER OF LIKERT SCALED QUESTIONNAIRES WHEN K=1 AND K=2 .

1-α	$Z_{\alpha/2}$	$Z_{\beta=0.2}=0.85$, $k=1$, $k=2$		$Z_{\beta=0.1}=1.282$, $k=1$, $k=2$	
90%	1.645	N=247.386	N=61.846	N=647.566	N=156.1415
95%	1.96	N=487.34	N=121.8348	N=1208.829	N=302.207
99%	2.576	N=1770.527	N=442.63	N=4300.449	N=1075.112

Sample size estimates for likert scaled questionnaire

B. Samples size estimation formula for observed scores assumed to be EPD

Note that in scoring format the smallest possible score is the minimum $X_1 = X_{\min}$, the highest possible score is the maximum $X_n = X_{\max}$ while the median score $X_{med} = \mu$ varies depending on event. For instance, in a likert scale, the $X_{med} = \mu$ is the neutral point, for General Health Questionnaire (GHQ 28-items) and Strength and Difficulty Questionnaire (SDQ 25-items), the $X_{med} = \mu$ are 5 and 15 indicating the points that divides observations to case and non-case condition. So when the information available is restricted to $X_{\min}, X_{med}, X_{\max}$ then we are faced with the problem of determining the precise estimator for the scale parameter and the mean deviation in equation (27).

1- Using equation (27), and assuming that the scale parameter is the range $(\sigma = X_{\max} - X_{\min})$ and the mean deviation is the maximum error $E = \max|X_i - \mu|$ then for GHQ 28-items questionnaire with $X_{\min} = 1$; location parameter $X_{med} = \mu = 5$ and scale parameter $X_{\max} = 28$ the sample size at 95% confidence and 90% desired power is given as

$$n = \frac{\left(e^{1.96+1.282} - 1\right)^2 \sigma^2}{E^2} = \frac{604.4}{23^2} \times 27^2 = 832.926$$

To accommodate for the second half of the GND then we double the result N=1665.85.

2- Another application is estimating the sample size of patient that will be required to ensure that the random scores of patient from Strength and Difficulty Questionnaire (SDQ) is distributed within 90% confidence area with 80% desired power. SDQ 50 items questionnaire has properties $X_{\min} = 0$, $X_{med} = \mu = 15, X_{\max} = \sigma = 40$.

$$n = \frac{\left(e^{1.645+0.85} - 1\right)^2 \sigma^2}{E^2} = \frac{123.6929}{25^2} \times 40^2 = 316.65396$$

Thus when doubled we have N=633.3076.

3- To determine the sample size of patients required to repeat a study on prevalence of ADHD with previous information of about 5.3% - 18% prevalence range and general consensus rate of 8% at 95% confidence and 90% power. (Journal of child and Adolescence mental health 2006, 18(1):1-5). ADHD- Attention Deficit and Hyperactivity Disorder.

$$X_{\min} = 5.3\%, X_{med} = \mu = 8\%, X_{\max} = 18\%,$$

$$n = \frac{\left(e^{1.96+1.282} - 1\right)^2 \sigma^2}{E^2} = \frac{604.4}{10^2} \times 12.7^2 = 974.8599$$

So accommodating for the second half we have N=1949.719.

Note that σ and E can be estimated via MLE or moments provided we have empirical evidence from previous study.

References

[1] Lindsey J.K. (1999). Multivariate Elliptically Contoured Distributions for Repeated Measurements. Biometrics 55, 1277-1280.

[2] Saralees Nadarajah (2005). A Generalized Normal Distribution. Journal of Applied Statistics Vol 32, No 7, 685-694.

[3] Roy Takenaga (1966). The Evaluation of the Incomplete Gamma Function. http:// www.ams.org/journal-terms-of-use.

[4] Paris, R.B. (2010). Incomplete Gamma Function. NIST Handbook of mathematical functions, Cambridge University press. ISBN 978-0521192255.

[5] Suhasini, S.R. (2010). Advanced Statistical Inference. Suhasini subbarao@stat.tamu.edu.

[6] Sterling, M (2011). Manual of the General Health Questionnaire. Measure of emotional distress. ajp.physiotherapy.asn.au/ajp/vol_57/4/Clin.pdf.

[7] Goodman, R (2000). Scoring the Self-Report Strengths and Difficulties Questionnaire. bjp.rcpsych.org/content/177/6/534.long British Journal of Psychiatry.

[8] Mood, Alexander McFar1ane, (1913)- Introduction to the theory of statistics (McGraw-Hi1l series in probability and statistics). ISBN 0-07-042864-6.