Original Research Article

RELATIONSHIP AMONG CONTINUOUS PROBABILITY DISTRIBUTIONS AND INTERPOLATION

ABSTRACT

Aims: The Study seeks to determine the relationship that exists among Continuous Probability Distributions and the use of Interpolation Techniques to estimate unavailable but desired value of a given probability distribution.

Study design: Statistical Probability tables for Normal, Student t, Chi-squared, F and Gamma distributions were used to compare interpolated values with statistical tabulated values. Charts and Tables were used to represent the relationships among the five probability distributions.

Methodology: Linear Interpolation Technique was employed to interpolate unavailable but desired values so as to obtain approximate values from the statistical tables. The data were analyzed for interpolation of unavailable but desired values at 95% α-level from the five continuous probability distribution.

Results: Interpolated values are as close as possible to the exact values and the difference between the exact value and the interpolated value is not pronounced. The table and chart established showed that relationships do exist among the Normal, Student-t, Chi-squared, F and Gamma distributions.

Conclusion: Interpolation techniques can be applied to obtain unavailable but desired information in a data set. Thus, uncertainty found in a data set can be discovered, then analyzed and interpreted to produce desired results. However, understanding of how these probability distributions are related to each other can inform how best these distributions can be used interchangeably by Statisticians and other Researchers who apply statistical methods employed in practical applications.

Keywords: Distribution, Interpolation, Normal, Chi-squared, Student-t, F, Gamma

1. INTRODUCTION

Statistical Probability Distribution is used to provide estimates of the likelihood of occurrence of future events. It defines or describes the likelihoods for a range of possible future outcomes of an experiment and the probability associated with each outcome. Statistical distributions provide the foundation for the analysis of empirical data and for many statistical procedures. Empirical results can be sensitive to the degree to which distributional characteristics such as the mean, variance, skewness, and kurtosis of the data can be modeled by the assumed statistical distribution. (Ayienda, C. 2013) [1].

In this Study, we compare the relationships that exists among five continuous probability distributions namely: Normal, Chi-squared, Student-t, F and Gamma distributions. These relationships that exist imply

UNDER PEER REVIEW

that some tabulated probability distributions can be used interchangeably by Statisticians and other Researchers who apply statistical methods. Some of such distribution relationships are the Normal and Chi-squared distributions, Normal and Gamma distributions, Chi-squared and F distributions, Chi-squared and Student-t distributions, the Student-t and F distributions, and the Gamma and Chi-squared distributions. Etc.

The normal and student-t distributions are closely related to the sampling distribution of means; while the chi-squared and F distributions are closely related to the sampling distribution of variances.

1.1 Continuous Random Variable

A random variable X, is said to be Continuous Random Variable if the number of possible values of outcomes that the variable can take is infinite.

The distribution function (x) of a continuous random variable is given by,

$\int_{-\infty}^{\infty} f(x)\, \mathrm{d}x = 1$

1.2 Normal Distribution

The normal distribution is probably the most important distribution in all of probability and statistics. Many populations have distributions that can be fit very closely by an appropriate normal (or bell) curve.

A random variable X is said to be normally distributed if its distribution X ~ (μ, δ). The distribution has a belled-shape and the mean, median and mode are equal and located in the center of the distribution. Normal distribution is given by,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} \quad , \qquad -\infty < x < \infty \; ; \; -\infty < \mu < \infty \; ; \; \sigma^2 > 0$$

where x is the value of a continuous random variable.

The mean, $E(x) = \mu$

variance, $var(x) = \sigma^2$

and skewness = 0

Proof,

$E(x) = \int_{-\infty}^{\infty} x\, f(x)\; dx$

$= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad \Rightarrow \quad \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-\frac{(x-\mu)}{\sigma\sqrt{2}}} dx$

Let $z = \frac{(x-\mu)}{\sigma\sqrt{2}} \quad \Rightarrow \quad x = \mu + \sigma\sqrt{2}\, z \qquad dx = \sigma\sqrt{2}\, dz$

$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu + \sigma\sqrt{2}\, z\; e^{-z^2}\, \sigma\sqrt{2}\, dz \quad \Rightarrow \quad \frac{1}{\sqrt{\pi}} \left\{ \int_{-\infty}^{\infty} \mu\, e^{-z^2} dz + \int_{-\infty}^{\infty} \sigma\sqrt{2}\, z\, e^{-z^2} dz \right\}$

$= \frac{\mu}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2}\, dz + \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} z e^{-z^2}\, dz$

$= \frac{2\mu}{\sqrt{\pi}} \int_{0}^{\infty} e^{-z^2}\, dz + 0$ even function = 2, odd function = 0

$= \frac{2\mu}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2}$ from gamma function $\int_{-\infty}^{\infty} e^{-z^2}\, dz = \frac{\Gamma\left(\frac{1}{2}\right)}{2} = \frac{\sqrt{\pi}}{2}$

UNDER PEER REVIEW

= μ

$Var(x) = E(x^2) - [E(x)]^2 = E(x-\mu)^2$

$= \int_{-\infty}^{\infty}(x-\mu)^2 f(x)\, dx$

$= \int_{-\infty}^{\infty}(x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \Rightarrow \int_{-\infty}^{\infty}(x-\mu)^2 \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$

Let $z = \frac{x-\mu}{\sigma} \Rightarrow x - \mu = \sigma z \quad dx = \sigma dz$

$= \int_{-\infty}^{\infty}(\sigma z)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\sigma dz \Rightarrow \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2\, e^{-\frac{z^2}{2}}\, dz$

$= \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^{\infty} z^2\, e^{-\frac{z^2}{2}}\, dz$ even function = 2

Let $u = \frac{z^2}{2} \Rightarrow z^2 = 2u \Rightarrow z = \sqrt{2u}$

$2z\, dz = 2du \Rightarrow dz = \frac{du}{z} = \frac{du}{\sqrt{2u}}$

$= \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^{\infty} z^2\, e^{-u}\, \frac{du}{z} \Rightarrow \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^{\infty} \frac{z^2}{z}\, e^{-u}\, du \Rightarrow \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^{\infty} z\, e^{-u}\, du$

$= \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^{\infty} \sqrt{2u}\, e^{-u}\, du \Rightarrow \frac{2\sigma^2\sqrt{2}}{\sqrt{2\pi}}\int_0^{\infty} u^{\frac{1}{2}}\, e^{-u}\, du$

rewriting in gamma form, $\frac{2\sigma^2}{\sqrt{\pi}}\int_0^{\infty} u^{\frac{3}{2}-1}\, e^{-u}\, du$

$= \frac{2\sigma^2}{\sqrt{\pi}} \frac{\Gamma\left(\frac{3}{2}\right)}{1^{\frac{3}{2}}}$ from gamma function $\int_0^{\infty} x^{\alpha-1}\, e^{-x} = \frac{\Gamma(\alpha)}{\beta^{\alpha}}$

$= \frac{2\sigma^2}{\sqrt{\pi}} \frac{\sqrt{\pi}}{2}$ from gamma function $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$

$= \sigma^2$

Suppose Z is a corresponding random variable to X, where $Z = \frac{x-\mu}{\sigma}$ and if the mean, μ = 0 and standard deviation, σ =1 then the distribution becomes a **Standard Normal** distribution. Thus, the

Standardization $Z \sim N(0,1)$ is given as,

$f(z) = Pr(Z \leq x) = \frac{1}{\sqrt{2\pi}} e^{-\left[\frac{x^2}{2}\right]}$

1.3 Chi-squared Distribution

The chi-squared distribution is a special case of the normal distribution. When the standard normal variable Z is squared, with 1 degree of freedom where $Z^2 = \left(\frac{x-\mu}{\sigma}\right)^2$, then the random variable X is said to have a **Chi-squared** distribution with v degrees of freedom.

As the degree of freedom gets large, the curve is less skewed, more normal.

Thus, the pdf of a chi-square distribution is given by,

$f(x; v) = \frac{1}{2^{v/2}\Gamma(v/2)} x^{\left(\frac{v}{2}\right)-1} e^{-\frac{x}{2}}$ $x \geq 0$, $v > 0$

The mean, $E(x) = v$

variance, $var(x) = 2v$

and skewness = $\sqrt{\frac{8}{v}}$

Proof,

$E(x) = \int_0^\infty x\, f(x)\; dx$

$= \int_0^\infty x \;\frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} x^{\left(\frac{v}{2}\right)-1} e^{-\frac{x}{2}}\, dx \quad \Rightarrow \quad \frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} \int_0^\infty x\; x^{\left(\frac{v}{2}\right)-1} e^{-\frac{x}{2}}\, dx$

$= \frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} \int_0^\infty x^{\left(\frac{v}{2}+1\right)-1}\; e^{-\frac{x}{2}}\, dx$

$= \frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)}\; \frac{\Gamma\left(\frac{v}{2}+1\right)}{\left(\frac{1}{2}\right)^{\frac{v}{2}+1}}$ from gamma function $\int_0^\infty x^{\alpha-1}\, e^{-x} = \frac{\Gamma(\alpha)}{\beta^\alpha}$

$= \frac{\left(\frac{1}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)}\; \frac{\Gamma\left(\frac{v}{2}+1\right)}{\left(\frac{1}{2}\right)^{\frac{v}{2}}\left(\frac{1}{2}\right)^{1}} \quad \Rightarrow \quad \frac{\Gamma\left(\frac{v}{2}+1\right)}{\Gamma\left(\frac{v}{2}\right)\left(\frac{1}{2}\right)}$

$= \frac{\left(\frac{v}{2}\right)!}{\left(\frac{v}{2}-1\right)!\left(\frac{1}{2}\right)}$ from gamma function $\Gamma(\alpha) = (\alpha + 1 - 1) = \alpha!$

$= \frac{\left(\frac{v}{2}\right)}{\frac{1}{2}} \quad \Rightarrow \quad \frac{v}{2} \times \frac{2}{1} = v$

$Var(x) = E(x^2) - [E(x)]^2$

$E(x^2) = \int_0^\infty x^2\, f(x)\; dx$

$= \int_0^\infty x^2 \;\frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} x^{\left(\frac{v}{2}\right)-1} e^{-\frac{x}{2}}\, dx \quad \Rightarrow \quad \frac{1}{2^{\frac{v}{2}}\Gamma\left(\frac{v}{2}\right)} \int_0^\infty x^2\; x^{\left(\frac{v}{2}\right)-1} e^{-\frac{x}{2}}\, dx$

$= \frac{\left(\frac{1}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} \int_0^\infty x^{\left(\frac{v}{2}+2\right)-1}\; e^{-\frac{x}{2}}\, dx$

$= \frac{\left(\frac{1}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)}\; \frac{\Gamma\left(\frac{v}{2}+2\right)}{\left(\frac{1}{2}\right)^{\frac{v}{2}+2}}$ from gamma function $\int_0^\infty x^{\alpha-1}\, e^{-x} = \frac{\Gamma(\alpha)}{\beta^\alpha}$

$= \frac{\left(\frac{1}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)}\; \frac{\Gamma\left(\frac{v}{2}+2\right)}{\left(\frac{1}{2}\right)^{\frac{v}{2}}\left(\frac{1}{2}\right)^{2}} \quad \Rightarrow \quad \frac{\Gamma\left(\frac{v}{2}+2\right)}{\Gamma\left(\frac{v}{2}\right)\left(\frac{1}{2}\right)^2}$

$= \frac{\left(\frac{v}{2}\right)\left(\frac{v}{2}+1\right)!}{\left(\frac{v}{2}-1\right)!\left(\frac{1}{2}\right)^2}$ from gamma function $\Gamma\left(\frac{v}{2}+2\right) = \left(\frac{v}{2}\right)\left(\frac{v}{2}+2-1\right)!$

$= \frac{\frac{v^2}{4}+\frac{v}{2}}{\left(\frac{1}{4}\right)} \quad \Rightarrow \quad \frac{v^2+2v}{4} \times \frac{4}{1} = v^2 + 2v$

But $Var(x) = E(x^2) - [E(x)]^2$

$= v^2 + 2v - (v)^2 = 2v$

UNDER PEER REVIEW

For small degree of freedom the chi-square distribution is skewed to the right, but as the degree of freedom increases, the distribution becomes symmetrical and approaches the normal distribution.

This distribution arises in many areas of statistics, for example, assessing the goodness-of-fit of models, particularly those fitted to contingency tables.

1.4 Student-t Distribution

In probability and statistics, Student's t-distribution is a continuous probability distribution that arises when estimating the mean of a normally distributed population in situations where the sample size is small. The t-distribution is commonly used to estimate the mean μ of a normal distribution when the variance σ^2 is not known.

The Student's t-distribution is given by,

$$f(t;\, v) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

where the parameter v is a positive integer and the variable x is a real number. **Γ** is the gamma function defined by,

$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}\,dy$ $\quad\quad \alpha > 0$

The mean, $E(x) = 0$ $\quad\quad v > 1$

variance, $var(x) = \frac{v}{v-2}$ $\quad v > 2$

and skewness = 0 $\quad\quad v > 3$ (Wikipedia, 2010) [2]

Proof,

Suppose Z and Y are independent. If $Z \sim N(0,1)$ and Y $\sim \chi^2(v)$, from definition, a random variable,

$$X = \frac{Z}{\sqrt{\frac{Y}{v}}} \sim T(v)$$

$$E(x) = E\left(\frac{Z}{\sqrt{\frac{Y}{v}}}\right) \Rightarrow \quad E\left(\frac{Z\sqrt{v}}{\sqrt{Y}}\right) \Rightarrow \quad \sqrt{v}\,E(Z)E\left(\sqrt{\frac{1}{Y}}\right)$$

But $Z \sim N(0,1)$ $\Rightarrow$ $\sqrt{v}$ x 0 x $E\left(\sqrt{\frac{1}{Y}}\right) = 0$

$$Var(x) = E(x^2) - [E(x)]^2$$

$$E(x^2) = \int_0^\infty x^2 f(x)\,dx$$

$= 2\int_0^\infty x^2 f(x)\,dx$ $\quad\quad$ even function = 2

$$= 2\int_0^\infty x^2\,\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}dx$$

Let $c = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}$ $\Rightarrow$ $\quad 2c\int_0^\infty x^2\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}dx$

UNDER PEER REVIEW

Let $t = \frac{x^2}{v}$ ⇒ $x^2 = tv$ ⇒ $x = \sqrt{tv} = (tv)^{\frac{1}{2}} = t^{\frac{1}{2}} v^{\frac{1}{2}}$

$dx = \frac{1}{2} (v)^{\frac{1}{2}} t^{-\frac{1}{2}} dt$ ⇒ $dx = \frac{\sqrt{v}}{2} \frac{1}{\sqrt{t}} dt$

$= 2c \int_0^\infty tv\,(1+t)^{-\frac{v+1}{2}} \frac{\sqrt{v}}{2} \frac{1}{\sqrt{t}} dt$ ⇒ $\frac{2c}{2} \int_0^\infty tv\,(1+t)^{-\frac{v+1}{2}} v^{\frac{1}{2}} \; t^{-\frac{1}{2}} dt$

$= cv^{\frac{3}{2}} \int_0^\infty t^{\frac{1}{2}} (1+t)^{-\frac{v}{2}-\frac{1}{2}} dt$ ⇒ $cv^{\frac{3}{2}} \int_0^\infty t^{\frac{3}{2}-1} (1+t)^{-\frac{3}{2}-\left(\frac{v}{2}-1\right)} dt$

$= cv^{\frac{3}{2}} \int_0^\infty t^{\frac{3}{2}-1} \frac{1}{(1+t)^{\frac{3}{2}+\left(\frac{v}{2}-1\right)}} dt$

$= cv^{\frac{3}{2}} B\left[\frac{3}{2}, \frac{v}{2}-1\right]$ from beta function $\int_0^\infty \frac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}} = B(\alpha,\beta)$

Recall $c = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}$ ⇒ $\frac{1}{\sqrt{v}\, B\left(\frac{V}{2}, \frac{1}{2}\right)}$

$= \frac{1}{\sqrt{v}} \frac{1}{B\left(\frac{V}{2}, \frac{1}{2}\right)} v^{\frac{3}{2}} B\left[\frac{1}{2}+1, \frac{v}{2}-1\right]$

$= v^{-\frac{1}{2}+\frac{3}{2}} \frac{\Gamma\left(\frac{v}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\Gamma\left(\frac{1}{2}\right)} \frac{\Gamma\left(\frac{1}{2}+1\right)\Gamma\left(\frac{v}{2}-1\right)}{\Gamma\left(\frac{1}{2}+1+\frac{v}{2}-1\right)}$ from beta function $B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$= v \frac{\Gamma\left(\frac{v}{2}+\frac{1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\Gamma\left(\frac{1}{2}\right)} \frac{\Gamma\left(\frac{1}{2}+1\right)\Gamma\left(\frac{v}{2}-1\right)}{\Gamma\left(\frac{1}{2}+\frac{v}{2}\right)}$ ⇒ $v \frac{\Gamma\left(\frac{1}{2}+1\right)\Gamma\left(\frac{v}{2}-1\right)}{\Gamma\left(\frac{v}{2}\right)\Gamma\left(\frac{1}{2}\right)}$

$= v \frac{\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)\left(\frac{v}{2}-1\right)^{-1}\left(\frac{v}{2}-1\right)!}{\left(\frac{v}{2}-1\right)!\,\sqrt{\pi}}$ from gamma function $\mathbf{\Gamma}(\alpha) = \alpha\,\mathbf{\Gamma}(\alpha)$

$= v \left[\frac{\frac{1}{2}\sqrt{\pi}}{\left(\frac{v}{2}-1\right)\sqrt{\pi}}\right]$ ⇒ $\frac{\frac{v}{2}}{\frac{v-2}{2}}$ $= \frac{v}{2} \times \frac{2}{v-2}$

$= \frac{v}{v-2}$

$Var(x) = E(x^2) - [E(x)]^2$

$= \frac{v}{v-2} - 0 = \frac{v}{v-2}$

This distribution plays a role in a number of widely-used statistical analyses, including the Student's *t*-test for assessing the statistical significance of the difference between two sample means, the construction of confidence intervals for the difference between two population means, and in linear regression analysis. It is often used in Bayesian inference because it provides a multivariate distribution with ‘thicker’ tails than the multivariate normal.

1.5 F Distribution

F distribution is the distribution of the ratio of two independent Chi-squared random variables divided by their degrees of freedom. Suppose X_1 and X_2 are two independent random variables such that $X_1 \sim X^2{}_n$ and $X_2 \sim X^2{}_r$ respectively,

we define a random variable $X = \frac{\frac{X_1}{n}}{\frac{X_2}{r}} \sim F_{n,r}$ where n and r are positive numbers and degrees of freedom for X_1 and X_2. The F distribution for a random variable X is given by,

UNDER PEER REVIEW

$$f_{n,r}(x) = \frac{\left(\frac{n}{r}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1}}{B\left(\frac{n}{2},\frac{r}{2}\right)\left[1+\left(\frac{n}{r}\right)^{x}\right]^{\frac{n+r}{2}}}$$

where $B\left(\frac{n}{2},\ \frac{r}{2}\right) = \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{r}{2}\right)}{\Gamma\left(\frac{n}{2}+\frac{r}{2}\right)}$ and $\mathbf{\Gamma}$ is the gamma function.

The mean, $E(x) = \frac{r}{r-2}$ $\quad r > 2$

variance, $var(x) = \frac{2r^2(n+r-2)}{n(r-2)^2(r-4)}$ $\quad r > 4$

and skewness = $\frac{(2n+r-2)\sqrt{8(r-4)}}{(r-6)\sqrt{r(n+r-2)}}$ $\quad r > 6$ (Wikipedia, 2010) [2]

Proof,

$$E(x) = \int_0^\infty x\, f(x)\ dx$$

$$= \int_0^\infty x\, \frac{x^{\frac{n}{2}-1}\left(\frac{n}{r}\right)^{\frac{n}{2}}}{B\left(\frac{n}{2},\ \frac{r}{2}\right)\left[1+\left(\frac{n}{r}\right)^{x}\right]^{\frac{n+r}{2}}}\ dx \quad \Rightarrow \quad \int_0^\infty \frac{x^{\left(\frac{n}{2}+1\right)-1}\left(\frac{n}{r}\right)^{\frac{n}{2}}}{B\left(\frac{n}{2},\ \frac{r}{2}\right)\left[1+\left(\frac{n}{r}\right)^{x}\right]^{\frac{n+r}{2}}}\ dx$$

Let y = $\left(\frac{n}{r}\right)^{x}$ $\quad \frac{dy}{dx} = \frac{n}{r} \quad \Rightarrow \quad dx = \frac{r\,dy}{n}$ and $x = \frac{yr}{n}$

$$E(x) = \int_0^\infty \frac{\left(\frac{yr}{n}\right)^{\left(\frac{n}{2}+1\right)-1}\left(\frac{n}{r}\right)^{\frac{n}{2}}\left(\frac{r}{n}\right)dy}{B\left(\frac{n}{2},\ \frac{r}{2}\right)[1+y]^{\frac{n+r}{2}}} \quad \Rightarrow \quad \int_0^\infty \frac{\left(\frac{r}{n}\right)^{\left(\frac{n}{2}+1\right)-1} y^{\left(\frac{n}{2}+1\right)-1}\left(\frac{n}{r}\right)^{\frac{n}{2}}\left(\frac{r}{n}\right)dy}{B\left(\frac{n}{2},\ \frac{r}{2}\right)[1+y]^{\frac{n+r}{2}}}$$

$$= \int_0^\infty \frac{\left(\frac{r}{n}\right)^{\frac{n}{2}}\left(\frac{n}{r}\right)^{\frac{n}{2}} y^{\left(\frac{n}{2}+1\right)-1}\left(\frac{r}{n}\right)dy}{B\left(\frac{n}{2},\ \frac{r}{2}\right)[1+y]^{\frac{n+r}{2}}} \quad \Rightarrow \quad \int_0^\infty \frac{y^{\left(\frac{n}{2}+1\right)-1}\left(\frac{r}{n}\right)dy}{B\left(\frac{n}{2},\ \frac{r}{2}\right)[1+y]^{\frac{n+r}{2}}}$$

$$= \frac{\left(\frac{r}{n}\right)}{B\left(\frac{n}{r},\ \frac{r}{2}\right)} \int_0^\infty \frac{y^{\left(\frac{n}{2}+1\right)-1}}{[1+y]^{\frac{n+r}{2}}} \quad \Rightarrow \quad \frac{\left(\frac{r}{n}\right)}{B\left(\frac{n}{r},\ \frac{r}{2}\right)} \int_0^\infty \frac{y^{\left(\frac{n}{2}+1\right)-1}}{[1+y]^{\frac{n}{2}+\frac{r}{2}}}$$

$$= \frac{\left(\frac{r}{n}\right)}{B\left(\frac{n}{r},\ \frac{r}{2}\right)} \int_0^\infty \frac{y^{\left(\frac{n}{2}+1\right)-1}}{[1+y]^{\left(\frac{n}{2}+1+\frac{r}{2}-1\right)}}$$

Let $\alpha = \frac{n}{2} + 1$ and $\beta = \frac{r}{2} - 1$

$$= \frac{\left(\frac{r}{n}\right)}{B\left(\frac{n}{r},\ \frac{r}{2}\right)}\ B\left(\frac{n}{2}+1,\ \ \frac{r}{2}-1\right)$$

from beta function $\int_0^\infty \frac{y^{\alpha-1}}{(1+y)^{\alpha+\beta}} = B(\alpha,\beta)$

$$= \frac{\left(\frac{r}{n}\right) B\left(\frac{n}{2}+1,\ \frac{r}{2}-1\right)}{B\left(\frac{n}{r},\ \frac{r}{2}\right)}$$

equation 1

$$= \frac{\left(\frac{r}{n}\right)\Gamma\left(\frac{n}{2}+1\right)\Gamma\left(\frac{r}{2}-1\right)}{\Gamma\left(\frac{n}{2}+1+\frac{r}{2}-1\right)} \times \frac{\Gamma\left(\frac{n}{2}+\frac{r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{r}{2}\right)}$$

from beta function $B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$$= \frac{\left(\frac{r}{n}\right)\Gamma\left(\frac{n}{2}+1\right)\Gamma\left(\frac{r}{2}-1\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{r}{2}\right)}$$

$$= \frac{\left(\frac{r}{n}\right)\left(\frac{n}{2}\right)!\left(\frac{r}{2}-1\right)^{-1}\left(\frac{r}{2}-1\right)!}{\left(\frac{n}{2}-1\right)!\left(\frac{r}{2}-1\right)!}$$

From gamma function $\mathbf{\Gamma}(\alpha) = (\alpha-1)\ \mathbf{\Gamma}(\alpha-1)$

UNDER PEER REVIEW

$= \left(\frac{r}{n}\right)\left(\frac{n}{2}\right)! \text{ X } \frac{1}{\frac{r}{2}-1} \qquad \Rightarrow \qquad \frac{\left(\frac{r}{2}\right)}{\frac{r}{2}-1} = \frac{r}{2} \text{ X } \frac{2}{r-2}$

$= \frac{r}{r-2}$

In general, $E(x^m) = \frac{\left(\frac{r}{n}\right)^m B\left(\frac{n}{2}+m,\ \frac{r}{2}-m\right)}{B\left(\frac{n}{r},\ \frac{r}{2}\right)}$ from equation 1

Hence, $Var(x) = E(x^2) - [E(x)]^2$

$E(x^2) = \frac{\left(\frac{r}{n}\right)^2 B\left(\frac{n}{2}+2,\ \frac{r}{2}-2\right)}{B\left(\frac{n}{r},\ \frac{r}{2}\right)}$

$= \frac{\left(\frac{r}{n}\right)^2 \Gamma\left(\frac{n}{2}+\frac{r}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{r}{2}\right)} \text{ x } \frac{\left(\frac{r}{n}\right)\Gamma\left(\frac{n}{2}+2\right)\Gamma\left(\frac{r}{2}-2\right)}{\Gamma\left(\frac{n}{2}+2+\frac{r}{2}-2\right)}$ from beta function $B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$= \frac{\left(\frac{r}{n}\right)^2 \Gamma\left(\frac{n}{2}+2\right)\Gamma\left(\frac{r}{2}-2\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{r}{2}\right)}$

$= \frac{\left(\frac{r}{n}\right)^2 \left(\frac{n}{2}\right)\left(\frac{n}{2}+2-1\right)!\left(\frac{r}{2}-1\right)^{-1}\left(\frac{r-2}{2}\right)^{-1}\left(\frac{r}{2}-1\right)!}{\left(\frac{n}{2}-1\right)!\left(\frac{r}{2}-1\right)!}$

from gamma function $\mathbf{\Gamma}(\alpha) = (\alpha-1)(\alpha-2)\mathbf{\Gamma}(\alpha-1)$

$\left(\frac{r}{n}\right)^2\left(\frac{n}{2}\right)\left(\frac{n}{2}+1\right) \text{ X } \frac{1}{\left(\frac{r}{2}-1\right)\left(\frac{r}{2}-2\right)} \qquad \Rightarrow \qquad \left(\frac{r}{n}\right)^2 \frac{\left(\frac{n^2}{4}+\frac{n}{2}\right)}{\left(\frac{r}{2}-1\right)\left(\frac{r}{2}-2\right)}$

$= \frac{\frac{r^2}{n^2}\left(\frac{n^2}{4}+\frac{n}{2}\right)}{\left(\frac{r}{2}-1\right)\left(\frac{r}{2}-2\right)} \qquad \Rightarrow \qquad \frac{\frac{r^2}{n^2}\left(\frac{n^2+2n}{4}\right)}{\left(\frac{r-2}{2}\right)\left(\frac{r-4}{2}\right)} \qquad = \frac{\frac{r^2}{n^2}\,\frac{n\,(n+2)}{4}}{\frac{(r-2)\,(r-4)}{4}}$

$= \frac{r^2\,(n+2)}{4n} \text{ X } \frac{4}{(r-2)\,(r-4)} \qquad = \frac{r^2\,(n+2)}{n\,(r-2)\,(r-4)}$

$Var(x) = E(x^2) - [E(x)]^2$

$= \frac{r^2\,(n+2)}{n\,(r-2)\,(r-4)} - \left(\frac{r}{r-2}\right)^2 \qquad \Rightarrow \qquad \frac{r^2\,(n+2)}{n\,(r-2)\,(r-4)} - \frac{r^2}{(r-2)^2}$

$= \frac{r^2\,(n+2)\,(r-2) - n\,(r-4)\,r^2}{n\,(r-2)^2\,(r-4)} \qquad \Rightarrow \qquad \frac{r^2\,(nr-2n+2r-4-nr+4n)}{n\,(r-2)^2\,(r-4)}$

$= \frac{r^2\,(2n+2r-4)}{n\,(r-2)^2\,(r-4)} \qquad = \frac{2r^2\,(n+r-2)}{n\,(r-2)^2\,(r-4)}$

For small degrees of freedom, the F distribution is skewed to the right, but for large values of degree of freedom for v, the distribution becomes approximately normal.

The *F* distribution is used in many statistical tests e.g. test for equality of variances, tests for differences in means in ANOVA and tests for regression models. Widely used to assign P-values to mean square ratios in the analysis of variance.

1.6 Gamma Distribution

A random variable X is said to be a gamma distribution with parameters α and β if its distribution $X \sim \Gamma(\alpha,\beta)$. Gamma distribution is used to describe waiting times between successive occurrences of a random event. Gamma distribution is given by,

UNDER PEER REVIEW

$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}$ $\qquad 0 < x < \infty \qquad \alpha > 0, \quad \beta > 0$

where β is a scale parameter, α is a shape parameter. The function Γ is defined by,

$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy \qquad \alpha > 0 ; \qquad y > 0$

The function Γ is recursive satisfying the relationship:

- $\Gamma(1) = \int_0^{\infty} e^{-y}\, dy = 1$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- $\Gamma(\alpha + 1) = (\alpha + 1 - 1)! = \alpha!$
  So that, for an integer $n \geq 2$; $\Gamma(n) = (n-1)!$
- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for $\alpha > 1$
  Thus, $\Gamma(\alpha + 1) = (\alpha + 1 - 1)\Gamma(\alpha + 1 - 1) = \alpha\Gamma(\alpha)$ for any positive real α

When $\beta = 1$ then the gamma distribution converts to a **Standard Gamma** Distribution given as,

$f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \qquad x > 0 ; \quad \alpha > 0$

The gamma function Γ can be expressed in terms of Beta function $B(\alpha, \beta)$ as,

$$\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\, dy$$

The mean $E(x) = \alpha\beta$

Variance, $Var(x) = \alpha\beta^2$

Skewness = $2\alpha^{-\frac{1}{2}}$

Proof

$E(x) = \int_0^{\infty} x\, f(x)\, dx$

$= \int_0^{\infty} x \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx \quad \Rightarrow \quad \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x\; x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx$

$= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{(\alpha+1)-1} e^{-\frac{x}{\beta}}\, dx$

$= \frac{\Gamma(\alpha+1)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^{\alpha}}$ From gamma function $\int_0^{\infty} x^{(\alpha+1)-1} e^{-\frac{x}{\beta}}\, dx = \Gamma(\alpha+1)\beta^{\alpha+1}$

$= \frac{\Gamma(\alpha+1)\beta^{\alpha+1}\beta^{-\alpha}}{\Gamma(\alpha)} \quad \Rightarrow \quad \frac{\Gamma(\alpha+1)\beta^{1}}{\Gamma(\alpha)}$

$= \frac{\alpha!\beta^{1}}{(\alpha-1)!}$ From gamma function $\Gamma(\alpha+1) = \Gamma(\alpha+1-1) = \alpha!$

$= \alpha\beta$

$Var(x) = E(x^2) - [E(x)]^2$

UNDER PEER REVIEW

$E(x^2) = \int_0^\infty x^2 \, f(x) \; dx$

$= \int_0^\infty x^2 \; \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} \; dx \qquad \Rightarrow \qquad \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^2 \;\; x^{\alpha-1} e^{-\frac{x}{\beta}} \; dx$

$\frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \;\; x^{(\alpha+2)-1} e^{-\frac{x}{\beta}} \; dx$

$= \frac{\Gamma(\alpha+2)\beta^{\alpha+2}}{\Gamma(\alpha)\beta^\alpha}$ From gamma function $\int_0^\infty x^{(\alpha+2)-1} e^{-\frac{x}{\beta}} \; dx \; = \boldsymbol{\Gamma}(\alpha+2)\beta^{\alpha+2}$

$= \frac{\Gamma(\alpha+2)\beta^{\alpha+2}\beta^{-\alpha}}{\Gamma(\alpha)} \qquad \Rightarrow \qquad \frac{\Gamma(\alpha+2)\beta^2}{\Gamma(\alpha)}$

$= \frac{(\alpha+1)\Gamma(\alpha+2-1)!\beta^2}{\Gamma(\alpha-1)!}$ From gamma function $\Gamma(\alpha+2) = (\alpha+1)\Gamma(\alpha+2-1)!$

$= (\alpha + 1)\alpha!\,\beta^2 \;\; = \; \alpha(\alpha + 1)\beta^2$

$Var(x) = E(x^2) - [E(x)]^2$

$= \; \alpha(\alpha + 1)\beta^2 - (\alpha\beta)^2 \;\; \Rightarrow \qquad (\alpha^2 + \alpha)\beta^2 - \alpha^2\beta^2$

$= \; \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2$

$= \; \alpha\beta^2$

1.7 Interpolation

Most data set among probability distributions are unknown, and hence unavailable. Thus most statistical tables are incomplete due to those unavailable but desired values. Therefore this study introduces us to Interpolation Methods. These Interpolation techniques are applied, notably in mathematical tables, for estimating the unknown value of a function, say $y_x = f(x)$, from two (or more) known values that embrace it, i.e. values of y_a and y_b such that $a < x < b$.

Interpolation is the technique of estimating the value of a function for any intermediate value of the independent variable within a given range. If the function *f*(*x*) is known explicitly, then the value of *y* corresponding to any value of *x* can easily be found. (Louis and Dupuis, 2002) [3].

Interpolation seeks to fill in missing information in some small region of the whole dataset. We use interpolation to interpolate values from computations, develop numerical integration and differentiation schemes, and to develop finite element methods.

Interpolation function is a function that passes exactly through a set of data points. An interpolation function is also called 'Interpolant'.

For most practical purposes it is sufficient to quote the significance of results within limits. For example probability that α greater than 10 %, or α lies between 5% and 10%, and so on. Sometimes degree of freedom or tabulated data numbers may have gaps in between successive tabulated ones whose value entries are needed for exact results. This is why we make rough interpretations between entries in order to quote the actual values needed for our results. Note that interpolations is a sort of approximation to the actual values.

There are different forms of interpolation techniques which includes:

1. **Linear Interpolation** - is a method of curve fitting using linear polynomials to construct new data points within the range of a discrete set of known data points. It is used to approximate a value of

UNDER PEER REVIEW

some function *f*, using two known values of that function at other points. Basic operations in linear interpolation between two values is commonly used in computer graphics.

2. **Spline Interpolation** - is a form of interpolation where the interpolant is a special type of piecewise polynomial called a 'spline'. Instead of fitting a single, high-degree polynomial to all of the values at once, it fits low degree polynomials to small subsets of the values. It avoids the problem of oscillation at the edges of an interval when using polynomial interpolation with polynomials of high degree over a set of equispaced interpolation points. Splines are used to model automobile bodies.
3. **Polynomial Interpolation** - is the interpolation of a given data set by the polynomial of lowest possible degree that passes through the points of the datasets. Polynomial interpolation can be used to approximate complicated curves. An application is the evaluation of the natural logarithm.
4. **Trigonometric Interpolation** - is interpolation with trigonometric polynomials. Such as sum of sines and cosines of given periods. This form of interpolation is especially suited for interpolation of periodic functions. They are also used to form discrete Fourier transform.
5. **Multivariate Interpolation** - is interpolation on functions of more than one variable. The function to be interpolated is known at given points (x_i, y_i, z_i, . . .) and the interpolation problem consists of yielding values at arbitrary points (*x, y, z, . . .*). Multivariate interpolation is used in geostatistics, where it is used to create a digital elevation model from a set of points on the earth surface, for example spotting heights in topographic survey or depths in hydrographic survey. (Wikipedia, 2010) [2].

This study implies the general method of linear interpolation on data pairs (x, y).

1.8 Statement of the Problem

Statistical distributions provide the foundations for the analysis of empirical data and for many statistical procedures. In experimental work, one often encounters problems where a standard statistical probability density function is applicable. Also empirical results can be sensitive to the degree to which distributional characteristics such as the mean, variance, and skewness of the data can be modelled by the assumed statistical distribution. It was found necessary to compare the relationship that exists among these practical applications.

Almost all data suffer from some kind of uncertainty/biasness and it is always important to know whether those uncertainties/biasness come from a sort of unavailable but desired values or uncertain behaviors. If such certain data is given on fixed positions, then the question of interpolation arises. For the purpose of this research, it is been discovered that some data set contain unavailable but desired values which are unknown and this makes it difficult to interpret and analyze due to those unavailable (or unknown) values.

Consequently, an important open problem of this study is how probability distributions and interpolation interact.

1.9 Theoretical Review

Evans and Rosenthal (2009) [4], carried out a research work on continuous probability distributions.

Suppose $Xi \sim N(\mu, \sigma^2)$ for *i* = 1, 2, . . . , *n* and that they are independent random variables.

If $\bar{X} = \frac{(X_1 + \ldots + X_n)}{n}$ then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

The chi-squared distribution with *n* degrees of freedom is the distribution of the random variable

$$Z = {X_1}^2 + {X_2}^2 + \ldots + {X_n}^2,$$

where $X_1, \ldots, X_n$ are independent and identically distributed, each with the standard normal distribution *N*(0, 1)

UNDER PEER REVIEW

The Student-t distribution with *n* degrees of freedom is the distribution of the random variable

$$Z = \frac{X}{\sqrt{\frac{X_1{}^2 + X_2{}^2 + ...+ X_n{}^2}{n}}}$$

where *X*, X_1, . . . , X_n are independent and identically distributed, each with the standard normal distribution *N*(0, 1).

The F distribution with *m* and *n* degrees of freedom is the distribution of the random variable

$$Z = \frac{(X_1{}^2 + X_2{}^2 + ...+ X_n{}^2)/m}{(Y_1{}^2 + Y_2{}^2 + ...+ Y_n{}^2)/n},$$

where X_1, . . . , X_m, Y_1, . . . , Y_n are independent and identically distributed, each with the standard normal distribution.

Thus $X \sim \chi^2(m)$ and $Y \sim \chi^2(n)$

Rathore (2003) [5] considered four specific probability distributions – normal, the student-t, the chi-squared, and the F distributions, and the special features of each distributions, in particular, the situations in which these distributions can be useful. These four distributions play a very pivotal role in econometric theory and practice.

The work outlined that these probability distribution are important because they help us to find out the probability distribution of estimators (or statistics), such as the sample mean and sample variance. Equipped with that knowledge, we will be able to draw inferences about their true population values.

The work also stated the properties of these distributions. For the normal distribution,

- The normal distribution curve is symmetrical around its mean value μ_x.
- The probability of obtaining a value of a normally distributed random variable far away from its mean value of becomes progressively smaller.
- Approximately 68% of the area under the normal curve lies between the values of ($\mu_x \pm \sigma_x$), approximately 95% of the area lies between ($\mu_x \pm 2\sigma_x$), and approximately 99.7% of the area lies between ($\mu_x \pm 3\sigma_x$). The total area under the curve is 1 or 100%.
- A linear combination of two (or more) normally distributed random variable is itself normally distributed.

For the student-t distribution,

- The t distribution like the normal distribution is symmetric.
- The mean of the t distribution, like the standard normal distribution, is zero, but its variance is defined for degree of freedom greater than 2.

For the chi-squared distribution,

- The chi-squared distribution takes only positive values and ranges from 0 to infinity.
- Unlike the normal distribution, the chi-squared distribution is a skewed distribution, the degree of skewness depends on the degree of freedom. For small degree of freedom the distribution is skewed to the right, but as the degree of freedom increases, the distribution becomes symmetrical and approaches the normal distribution.
- The mean value of a chi-squared random variable is v and its variance is 2v, where v is the degree of freedom.

For the F distribution,

- Like the chi-squared distribution, the f distribution is also skewed to the right and also ranges between 0 and infinity.
- Like t and chi-squared distributions, the F distribution approaches the normal distribution as v_1 and v_2 the degree of freedoms becomes large.

UNDER PEER REVIEW

- The square of a t-distribution random variable with v degree of freedom has an F distribution with 1 and v degree of freedoms respectively. i.e. ${t_v}^2 = F_{1,v}$
- Just as there is a relationship between the F and t- distributions, there is a relationship between the F and chi-square distributions, which is $F_{(m,n)} = \frac{\chi^2}{m}$ as $n \to \infty$

Another research work carried out by Geyer (2003), shows the relationships among these continuous distributions.

For **normal distribution**, if *Z* is normally distributed i.e. *N*(0, 1), then Z^2 is Gamma $\left(\frac{1}{2}, \frac{1}{2}\right)$ distributed.

For **Chi-squared distribution**, if v is large, then $\chi^2(v) \approx N(v, 2v)$

Chi-squared (*v*) = Gamma $\left(\frac{v}{2}, \frac{1}{2}\right)$

If Z and Y are independent, X is normally distributed, *N*(0, 1) and Y is chi-Squared distributed, $\chi^2(v)$, then $\frac{X}{\sqrt{\frac{Y}{v}}}$ is Student-t distributed, t(v).

If X and Y are independent and are Chi-squared (*μ*) and Chi-squared (*v*) distributed, respectively, then $\frac{(X/\mu)}{(Y/v)}$ is F distributed, F(μ, v).

For **Student-t Distribution**, if v is large, then t(v) $\approx N(0,1)$

If Z and Y are independent, X is normally distributed, *N*(0, 1) and Y is chi-Squared distributed, $\chi^2(v)$, then $\frac{X}{\sqrt{\frac{Y}{v}}}$ is Student-t distributed, t(v).

If X is t(v) distributed, then X^2 is F($1, v$) distributed.

If t(1) = Cauchy(0, 1).

For **F Distribution**, If X and Y are independent and are Chi-squared (*μ*) and Chi-squared (*v*) distributed, respectively, then $\frac{(X/\mu)}{(Y/v)}$ is F distributed, F(μ, v).

If X is t(v) distributed, then X^2 is F($1, v$) distributed.

In another research by Walck (2007) [6], he indicated in F distribution that, for large values of degree of freedoms i.e. n and r, the F-distribution tends to a normal distribution. He also indicated in the student-t distribution, that as the parameter v → ∞, the Student's t-distribution approaches the standard normal distribution.

Schlegel et al. (2012) [7] also carried out a research work on interpolation of data. The study analyze the effects of the usual linear interpolation schemes for visualization of the normal distributed data. In addition, the study demonstrated that, the methods known in geostatistics and machine learning have favorable properties for visualization purposes, because in many fields of science or engineering, we are confronted with uncertain data. For that reason, the visualization of uncertainty received a lot of attention, especially in recent years. In the majority of cases, normal distributions are used to describe uncertain behavior, because they are able to model many phenomena encountered in science. Therefore, in most applications, uncertain data is (or is assumed to be) normally distributed.

Bursal (1996) [8], stated in his work ''On Interpolating between Probability distributions'', that uncertainty in simulations of physical systems is common and may be due to parameter variations or noise. He found out that, computing these distributions is often a time-consuming task and needs to be repeated when some systems parameters are changed. His work presented formulas for interpolating between

UNDER PEER REVIEW

probability density and mass functions in spaces of arbitrary dimensionality. It is found that these interpolating formulas give accurate results even when the functions one is interpolating between are not that close. Moreover, the work concluded that interpolation is a lucrative alternative to Monte Carlo simulation and even to the Generalized Cell Mapping method when complete probability distributions, as opposed to only the low-order statistics, are needed. It is expected that this interpolation technique will relieve much of the burden of repeated, time consuming simulations as certain relevant parameters are varied.

The research work by USIAD (2014) [9], on ''Spatial Interpolation with Demographic and Health Survey Data: Key Considerations'', had a review that the most important criteria for selecting an interpolation method for use with Demographic and Health Survey data are an accurate and statistically rigorous map, and inclusion of a corresponding map surface with estimates of the uncertainty or potential error associated with the spatial interpolation. One of the conclusion of the study was that maps can be interpolated, using spatial interpolation and geospatial analysis which provides demographic information about the population of interest. However, the maps produced should use publicly available data, be standardized across countries, be easily reproducible, and be created as comparable maps to facilitate policy and program decision-making. These spatial interpolation surface maps should be produced for a few indicators in each survey, along with comprehensive descriptions of methods and guidelines on data use and interpretation.

1.10 Scope of the Study

This Study is limited only to five (5) Continuous Probability Distributions which are the Normal, Chi-squared, Student-t, F and Gamma distributions. The study is concerned about the Relationships that exists among these Five Continuous Distributions and Interpolation Method. For the purpose of this study, Linear Interpolation technique was employed for estimating unavailable but desired values.

2. METHODOLOGY

Statistical Probability tables for Normal, Student t, Chi-squared, F and Gamma distributions were used to compare interpolated values with statistical tabulated values.

Linear Interpolation Technique was employed to interpolate unavailable but desired values so as to obtain approximate values from the statistical tables. The data were analyzed for interpolation of unavailable but desired values at 95% α-level from the four continuous probability distribution.

Charts and Tables were used to represent the relationships among the five probability distributions.

2.1 Linear Interpolation

Linear interpolation is a method of curve fitting using linear polynomials to construct new data points within the range of a discrete set of known data points. Linear interpolation is obtained by passing a straight line between two data points.

The estimation formula for linear interpolation is given by,

$$f(x) = f(x_0)\,\frac{(x_1 - x)}{(x_1 - x_0)} + f(x_1)\,\frac{(x - x_0)}{(x_1 - x_0)}$$

Or $$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0)$$

Where $f(x_i)$ are the exact function for which values are known only at a discrete set of data points, x_0, x_1 are the data points (also referred to as interpolation points or nodes). (Cuneyt, S. 2010) [10].

UNDER PEER REVIEW

Given below is a procedure for linear interpolation, which is approximately the same as the actual values. Consider the following example for each probability distribution;

(1) Find the 95% significance level of $N(0,1)$ for z = 0.6

Table 2.1 Standard normal values with one missing value denoted by '*'.

z	f(x) = *N*(0,1)
0.4	0.6736
0.5	0.7088
0.6	*
0.7	0.7734

Using linear interpolation, we can interpolate the missing value for z = 0.6 using the points x_0 = 0.5 and x_1 = 0.7. By calculation,

$$f(x) = 0.7088\,\frac{(0.7-0.6)}{(0.7-0.5)} + 0.7734\,\frac{(0.6-0.5)}{(0.7-0.5)}$$

$$= \frac{0.7088(0.1)}{0.2} + \frac{0.7734(0.1)}{0.2}$$

= 0.7088(0.5) + 0.7734(0.5)

⇒ 0.3544 + 0.3867 = 0.7411

Note that the exact value for z = 0.6 is 0.7422

The difference from interpolated value = 0.0011

Thus, we can see that the interpolated value is approximately exact to 2 decimal places and by using interpolation method, the missing information about z = 0.6 is not entirely lost.

(2) Find the value of χ^2 for 75 *df* at 95% significance level.

Table 2.2 Chi-squared values.

df	f(x) = χ^2 (v)
60	43.188
70	51.739
80	60.391
90	69.126

Using linear interpolation, we can interpolate the missing value of χ^2 for 75 *df,* using the points x_0 = 70 and x_1 = 80. By calculation,

UNDER PEER REVIEW

$f(x)$ = 51.739 $\frac{(80-75)}{(80-70)}$ + 60.391 $\frac{(75-70)}{(80-70)}$

= 51.739(0.5) + 60.391(0.5)

⇒ 25.8695 + 30.1955 = 56.065

Note that the exact value for 75 df at 0.05 α level = 56.054

The difference from interpolated value = 0.011

(3) Using T distribution, obtain $T(49)$. For 1 - α = 0.05,

Table 2.3 $T(v)$ values with one missing value denoted by '*'.

V	f(x) = $T(v)$
47	1.678
48	1.677
49	*
50	1.676

Computing the missing value for v = 49 using linear interpolation, let the points x_0 = 48 and x_1 = 50. By calculation,

$f(x)$ = 1.678 $\frac{(50-49)}{(50-48)}$ + 1.676 $\frac{(49-48)}{(50-48)}$

= 1.678(0.5) + 1.676(0.5)

⇒ 0.839 + 0.838 = 1.677

Note that the exact value for v = 49 is 1.677

The difference from interpolated value = 0.000

Here, we can see that the interpolated value is the same as the actual value, thus the missing information about v is discovered.

(4) An analysis for *F* values given in the table below, shows incomplete values at v_1 = 12 for α = 0.05. Obtain the missing value of F$(12, 75)$

UNDER PEER REVIEW

Table 2.4 Incomplete *F* values.

v_2	v_1 = 12
30	2.09
40	2.00
60	1.92
120	1.83

Using linear interpolation, we can interpolate the missing value; for v_2 = 75, using the points x_0 = 60 and x_1 = 120. Thus computing, we have

$$f(x) = 1.92\,\frac{(120-75)}{(120-60)} + 1.83\,\frac{(75-60)}{(120-60)}$$

$$= 1.92(0.75) + 1.83(0.25)$$

$$\Rightarrow 1.44 + 0.4575 = 1.90$$

Note that the exact value for F$(12, 75)$ = 1.88

The difference from interpolated value = 0.02

(5) Find the value of the gamma function for x = 9

Table 2.5 Table showing an incomplete *Gamma* values.

x	α = 5
4	0.371
6	0.715
8	0.900
10	0.971

Computing the missing value for x = 9 using linear interpolation, let the points x_0 = 8 and x_1 = 10. By calculation,

$$f(x) = 0.900\,\frac{(10-9)}{(10-8)} + 0.971\,\frac{(9-8)}{(10-8)}$$

$$= 0.900(0.5) + 0.971(0.5)$$

UNDER PEER REVIEW

$\Rightarrow 0.45 + 0.4855 = 0.936$

Note that the exact value for x = 9 is 0.945

The difference from interpolated value = 0.009

Thus, we can see that the interpolated value is approximately close to the actual value, and for large values of degree of freedom the interpolated value is approximatively the same as the actual values. Thus, by using interpolation method, the unavailable but desired information about the distribution is not entirely lost.

2.2 Chart

Charts are visual display or communication of information in form of graphs, flow charts, tree diagrams etc. It is also a graphical representation of the flow of data about an information system.

3. RESULTS AND DISCUSSION

From {table 3.1}, we see that interpolated values are as close as possible to the exact values and the difference between the exact value and the interpolated value is not pronounced implying that the use of interpolation is good enough.

The results of interpolation is presented in table 3.1 below

Table 3.1 Interpolated Results.

Distribution	Exact Value	Interpolated Value	Difference
Normal z of 0.6	0.7422	0.7411	0.0011
Chi-squared χ^2 (75)	56.054	56.065	0.011
Student-t $T(49)$	1.677	1.677	0.000
F $F(12, 75)$	1.88	1.90	0.02
Gamma $\Gamma(5, 9)$	0.945	0.936	0.009

UNDER PEER REVIEW

From {table 3.2}, we see that relationship among these continuous probability distribution exist, that they are related to each other and can be used interchangeably among themselves.

Table 3.2 Relationship among Normal, Chi-squared, Student-t, F and Gamma Distributions.

Distribution	Relationship with Other Distribution	Expression
Normal	When the Standard normal distribution is squared with 1 degree of freedom, then the distribution is said to have a **Chi-square** distribution with v degrees of freedom. i.e. the sum of squared independent standard normal follows a chi-square distribution. where v corresponds to the number of z variable added together. The value of z at 95% i.e. 0.05 = 1.96 $(1.96)^2$ = 3.841 = $\chi^2(1)$ at 0.05 α - level The square of the standard normal distribution divided by 2 is equal to a **Gamma** distribution with one parameter $\alpha = \frac{1}{2}$	$Z^2 = \left(\frac{x-\mu}{\sigma}\right)^2 \sim \chi^2(v)$ i.e. $Z_1{}^2 + Z_2{}^2 + \ldots + Z_n{}^2 \sim \chi^2(n)$ $\sum_{i=1}^{n} Z_i{}^2 \sim \chi^2(n)$ $Z_1{}^2 \sim \chi^2(1)$ $Z_1{}^2 + Z_2{}^2 \sim \chi^2(2)$ If Z ~ $N(0,1)$, then $\frac{Z^2}{2} \sim \boldsymbol{\Gamma}(\frac{1}{2})$
Chi-squared	When the degrees of freedom increase, the mean value of the chi-squared distribution is approximately equal to that of the **Normal** distribution.	For large v, $\chi^2(v) \approx N(v, 2v)$

UNDER PEER REVIEW

	The ratio of an independent normal distribution random variable and the square root of a chi-squared distribution (itself divided by its degree of freedom) is equal to a **Student-t** distribution. The ratio of two independent chi-squared random variables divided by their degrees of freedom is the same as an **F** distribution.	If X ~ N (0, 1) and Y ~ χ^2(v), then $\frac{X}{\sqrt{\frac{Y}{v}}} \sim T(v)$ i.e. Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma)$ i.e. $X_i \sim N(\mu, \sigma)$ $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ Thus, $\bar{x} = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ and if y = $\frac{nS^2}{\sigma^2} \sim \chi^2$(n -1) Then $t = \frac{Z}{\sqrt{\frac{y}{n-1}}} \sim t(n-1)$ $= \frac{N(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}}$ If X ~ χ^2(r) and Y ~ χ^2(v) then, $\frac{(X/r)}{(Y/v)} \sim F(r, v)$ i.e. recall that $\frac{nS^2}{\sigma^2} \sim \chi^2$(n -1) thus, $\frac{\frac{n_1 S_1{}^2}{\sigma^2}}{n_1 - 1} \div \frac{\frac{n_2 S_2{}^2}{\sigma^2}}{n_2 - 1}$ $= \frac{\frac{\chi^2(n-1)}{n-1}}{\frac{\chi^2(n_2-1)}{n_2-1}} \sim F(n_1 - 1,\ n_2 - 1)$
Student-t	When the variation parameter for a student-t distribution becomes large, the student-t distribution is approximately a Standard **Normal** distribution.	For large v, (i.e. $v \to \infty$) T(v) ≈ N (0, 1)

UNDER PEER REVIEW

	When the students-t distribution is squared, it is equal to an **F** distribution with 1 and v degrees of freedom.	T^2(v) ~ F$(1, v)$ i.e. $t^2 = Z^2 \Big/ \frac{y}{n-1}$ $= \frac{(N(0,1))^2}{\frac{\chi^2(n-1)}{n-1}}$ Recall that $\left(\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim \chi^2(1)$ $= \frac{\frac{\chi^2(1)}{1}}{\frac{\chi^2(\mathrm{n}-1)}{n-1}} \sim$ F$(1,\ n-1)$
F	When the degrees of freedom for an F distribution increases, the mean value of an F distribution approaches that of the **Normal** distribution.	For large r, v; F(r, v) ~ N $(0,\ 1)$
Gamma	If a random variable X follows a gamma distribution with parameters $\alpha = \frac{n}{2}$ and β = 2 then it is identical to a **Chi-squared** distribution with n degrees of freedom.	$\Gamma\left(\frac{n}{2}, 2\right) \sim \chi^2$(n)

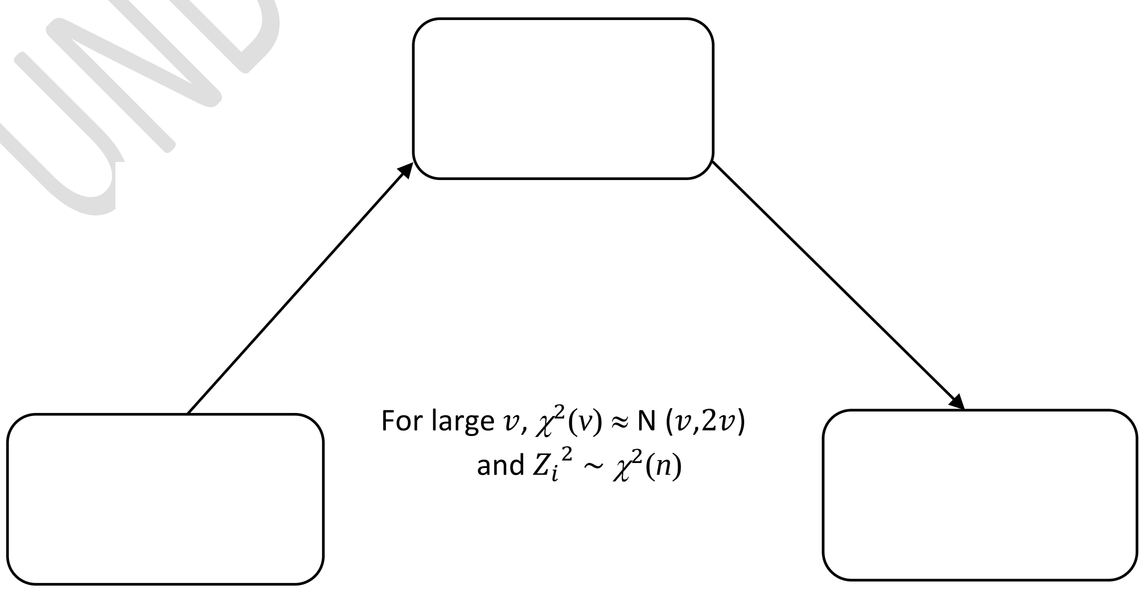
UNDER PEER REVIEW

From {figure 3.1}, we see the interaction that exist among the Normal, Chi-squared, Student-t, F and Gamma distributions. It indicates the relationship that links these distributions. That for large degrees of freedom, the chi-square distribution is approximately normal with mean v and variance $2v$.

Also for large degrees of freedom, the student-t and F distribution are approximately normally distributed. i.e. N $(0,\ 1)$.

Moreover, the ratio of two independent chi-square random variables follows an F distribution. We can also see from the chart that the square of a student-t distribution follow an F distribution. Lastly, chi-squared distribution relates to Student-t distribution, t(v) with $\frac{X}{\sqrt{\frac{Y}{v}}}$ if X is normally distributed, *N*(0, 1) and Y is chi-Square distributed, χ^2(*v*).

UNDER PEER REVIEW



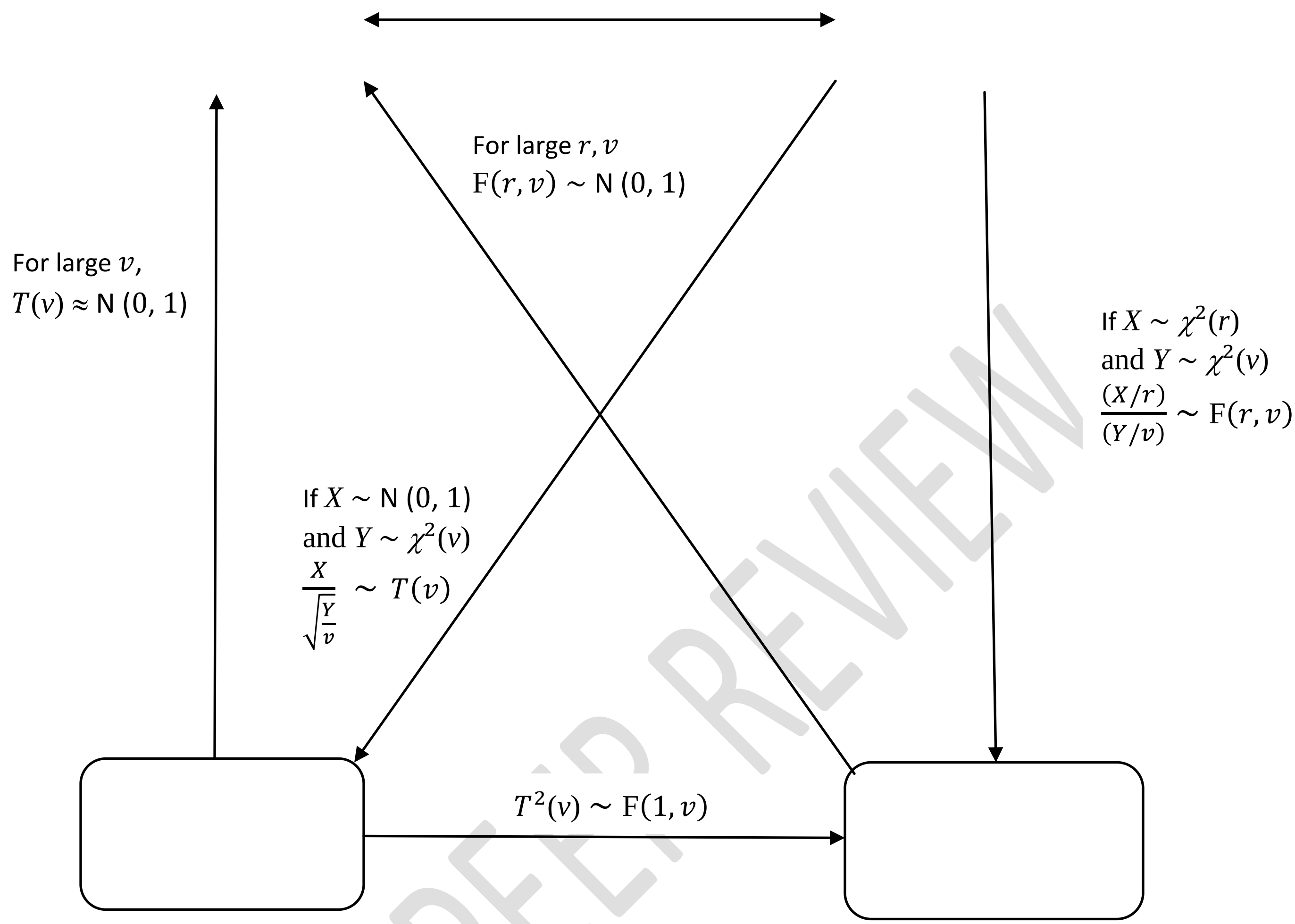


Figure 3.1 Relationships among Continuous Probability distributions.

UNDER PEER REVIEW

4. CONCLUSION

From the study, we have been able to show that Interpolation techniques can be applied to obtain unavailable but desired information in a data set. Thus, uncertainty found in a data set can be discovered, then analyzed and interpreted to produce desired results. However, understanding of how these probability distributions are related to each other can inform how best these distributions can be used interchangeably by Statisticians and other Researchers who apply statistical methods employed in practical applications.

REFERENCES

1. Ayienda K. Gamma and Related Distributions. University of Nairobi; 2013.

2. Wikipedia. Continuous Distributions. 2010.
Accessed 21 October 2021.
Available: https://en.m.wikipedia.org/continuousdistributions

3. Louis L, Frangois D. Statistical Tables, Explained and Applied. World Scientific Publishing Co.; 2002.

4. Michael J, Jeffrey R. Probability and Statistics: The Science of Uncertainty. 2nd ed. University of Toronto; 2009.

5. Siddharth R. Basic Econometrics/Statistics, Statistical Distributions: Normal, T, Chi-Sq, & F. Delhi University; 2003.

6 Christian W. Hand-Book on Statistical Distributions for Experimentalists. University of Stockholm; 2007.

7. Steven S, Nico K, Gerik S. On the Interpolation of Data with Normally Distributed Uncertainty for Visualization. University of Leipzig; 2012.

8. Bursal F. On Interpolating between probability distributions. Elsevier Inc., 1996.

9. United States Agency for International Development. Spatial Interpolation Demographic and Health Survey Data: Key Considerations. ICF International. 2014;

10. Cuneyt S. Numerical Methods: Interpolation. Middle East Technical University; 2010.

UNDER PEER REVIEW