# Hypothesis Testing for Coefficient of Variation in an Inverse Gaussian Population

Debaraj Sen* Yogendra P. Chaubey† Krishna K. Saha‡

**Abstract**
The inverse Gaussian distribution provides an attractive family of probability densities in modeling the coefficient of variation (CV) as it may conveniently be parameterized in terms of the mean and CV. Tests for mean and dispersion parameters have been investigated for this family in the literature, however, the coefficient of variation has not received much attention in this respect. Noting that the coefficient of variation plays a very important role in many practical data analysis situations, this article considers the uniformly most powerful invariant test for the problem. Some approximations to the distribution of the resulting test statistic have been investigated.

**Key Words:** Maximal invariant, invariance, Neyman-Pearson lemma

## 1. Introduction

The inverse Gaussian distribution has received considerable attention as a model for describing positively skewed data after the pioneering work of Tweedie (1957a, 1957b) and the subsequent review paper by Folks and Chhikara(1978). The probability density function *(pdf)* of the inverse Gaussian random variable $X$ is given by,

$$f(x|\mu,\lambda) = \left\{\frac{\lambda}{2\pi x^3}\right\}^{1/2} \exp\left\{-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right\}; \quad x>0, \quad \mu>0, \quad \lambda>0 \tag{1}$$

The above density will be denoted by $IG(\mu,\lambda)$. Its cumulative distribution function *(cdf)* can be written as

$$F(x\,|\mu,\lambda) = \Phi\left\{\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right\} + \exp\left\{\frac{2\lambda}{\mu}\right\}\ \Phi\left\{-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right)\right\} \tag{2}$$

where $\Phi(.)$ denotes the *c.d.f.* of a standard normal variable. It is also interesting to note that

$$\bar{X} \sim IG(\mu, n\lambda) \quad \text{and} \quad \lambda \sum_{i=1}^{n}\left(\frac{1}{X_i}-\frac{1}{\bar{X}}\right) \sim \chi^2_{n-1} \tag{3}$$

For a broad review and application of the IG family and other related results, the reader may refer to the text by Chhikara and Folks (1989) and Seshadri (1998).

We will find it convenient to parametrize the above IG density in terms of $(\mu, \upsilon)$, where $\upsilon$ denotes the square of the coefficient of variation given by

$$\upsilon = \frac{\mu}{\lambda}.$$

*Department of Mathematics and Statistics, Concordia University, Montréal, QC H3G 1M8, Canada
†Department of Mathematics and Statistics, Concordia University, Montréal, QC H3G 1M8, Canada
‡Department of Mathematics and Statistics, Central Connecticut State University, CT 06050, USA

Noting that the coefficient of variation plays a very important role in many practical data analysis situations (see Searles (1964), Srivastava (1974)), this article considers the uniformly most powerful invariant test for the problem that is outlined in Section 2. Section 3 considers the computational aspect of the distribution function of the test statistic. A very accurate approximation is developed in Section 4 that has gone through an extensive numerical analysis. A selection of this analysis is presented in Section 5.

## 2. Uniformly Most Powerful Invariant Test

Our interest lies in testing hypotheses about the parameter $\upsilon$, based on a random sample $\mathbf{X} = (X_1, X_2, ..., X_n)$ from $IG(\mu, \mu\upsilon^{-1})$. Using the sufficiency reduction of the data, we find that the likelihood function $\ell(\mu, \upsilon|\mathbf{X})$ factors as follows;

$$\begin{aligned} \ell(\mu, \upsilon|\mathbf{x}) &= \mathrm{e}^{n\upsilon} \prod_{i=1}^{n} \left\{ \left( \frac{\mu}{2\upsilon\pi x_i^3} \right)^{1/2} \right\} \exp\left\{ -\frac{1}{2\mu\upsilon} \sum x_i - \frac{\mu}{2\upsilon} \sum \frac{1}{x_i} \right\} \\ &= g_\theta(S(x))\ h(x) \end{aligned} \tag{4}$$

where

$$g_\theta(S(x)) = \mathrm{e}^{n\upsilon} \prod_{i=1}^{n} \left\{ \left( \frac{\mu}{2\upsilon\pi} \right)^{1/2} \right\} \exp\left\{ -\frac{1}{2\mu\upsilon} \sum x_i - \frac{\mu}{2\upsilon} \sum \frac{1}{x_i} \right\} \tag{5}$$

is a function of parameters $(\mu, \upsilon)$, only through

$$S(x) = \left( \sum x_i, \sum \frac{1}{x_i} \right),$$

and

$$h(x) = \prod_{i=1}^{n} \frac{1}{x_i^{3/2}}, \tag{6}$$

is a function of the data only. Thus, by the Fisher's factorization theorem, it follows that the bivariate statistic $(X_{\cdot} = \sum X_i, X_{-} = \sum \frac{1}{X_i})$ is sufficient for $(\mu, \upsilon)$. Note also that $IG(\mu, \lambda)$ is a convex exponential family (see Seshadri (1993), Prop. 2.6), hence $S(X)$ is complete for $(\mu, \upsilon)$. Looking at the likelihood ratio under $H_0 : \upsilon = \upsilon_0$ against $H_1 : \upsilon = \upsilon_1$, we find that an UMP test as given by Neyman-Pearson lemma is not available. This will be made more precise later by considering the group of scale changes: $G_c = \{g_c\}$, where $g_c(y) = cy, c > 0$ under which the problem of testing remains invariant: $\mathbf{X} \to \mathbf{Y} = (Y_1, Y_2, ..., Y_n)$ where

$$\begin{aligned} Y_i &= X_i / X_{i+1},\ i = 1, ..., n-1, \\ Y_n &= X_n. \end{aligned}$$

The distribution for $Y$ has been worked out in Jorgenson (1982) (see Eq. 3.21) for the generalized inverse Gaussian distribution. Specializing it to the inverse Gaussian case, the joint *pdf* of $Y$ is given by

$$\frac{n^{n/2}}{2^n K_{-1/2}^n(1/\upsilon)} \prod_{i=1}^{n} y_i^{-\frac{i}{2}-1} \exp\left\{ -\frac{1}{2} \left( \frac{\mu}{\upsilon} \sum_{k=1}^{n} \prod_{i=k}^{n} y_i^{-1} + \frac{1}{\mu\upsilon} \sum_{k=1}^{n} \prod_{i=k}^{n} y_i \right) \right\}. \tag{7}$$

Integrating out $y_n$ from the above expression, the joint density of the maximal invariant $(Y_1, Y_2, ..., Y_{n-1})$ is given by (see Eq. 3.23 of Jorgenson (1982));

$$\frac{K_{-n/2}((T_1(y))^{\frac{1}{2}}(T_2(y))^{\frac{1}{2}}/\upsilon)}{2^{n-1}K^n_{-1/2}(1/\upsilon)} \prod_{i=1}^{n-1} y_i^{-\frac{i}{2}-1} \left(\frac{T_1(y)}{T_2(y)}\right)^{-\frac{n}{4}} \tag{8}$$

where

$$T_1(y) = \sum_{k=1}^{n} \prod_{i=k}^{n-1} y_i^{-1}, \; T_2(y) = \sum_{k=1}^{n} \prod_{i=k}^{n-1} y_i \tag{9}$$

and $K_a(.)$ denotes the modified Bessel function of the third kind that may be defined by the integral representation

$$K_a(u) = \frac{1}{2}\int_0^{\infty} t^{a-1}\exp\left\{-\frac{u}{2}(t+t^{-1})\right\} dt. \tag{10}$$

The above shows that the distribution of $Y$ depends on the maximal invariant in the parameter space, namely $\upsilon$. Hence, the most powerful invariant test about $\upsilon$ may be based on $(T_1(y), T_2(y))$. Jorgenson (1982) shows that the statistic $T^* = \sqrt{T_1(y)T_2(y)}$ is a maximal invariant under the scale transformations, hence its distribution depends on $\upsilon$ only. Therefore the most powerful invariant test of $H_0 : \upsilon = \upsilon_0$ against a simple alternative may be based on the distribution of the test statistic $T^*$. He further showed that in fact $T^* = \sqrt{X_. X_{-1}}$, and that (see his Eq. 5.6), the ratio of the non-null *pdf* to that of the null *pdf* of $T^*$ is given by

$$U(t^*; \upsilon_0, \upsilon) = \frac{K_{-n/2}(t^*/\upsilon)K^n_{-1/2}(1/\upsilon_0)}{K_{-n/2}(t^*/\upsilon_0)K^n_{-1/2}(1/\upsilon)}.$$

This shows that there is no uniformly most powerful test for a composite alternative.

In what follows we will consider the modified test statistic $T$ given by

$$T = \bar{X} \sum \left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right),$$

that can be explicitly shown to have distribution depending on the parameter $\upsilon$ only. The uniformly most powerful invariant test for $H_0 : \upsilon = \upsilon_0$ vs. $H_1 : \upsilon = \upsilon_1$ is based on the test statistic $T^*$,

$$CR : \{X : T(X) \geq t_\alpha\},$$

where $t_\alpha$ is obtained from

$$\Pr_{\upsilon=\upsilon_0}[T \geq t_\alpha] = \alpha.$$

In the following section we discuss computation of the distribution of $T^*$ and some approximations. Different approximations are compared with exact percentage points also in the next section.

## 3. Distribution of the Test Statistic

We can write

$$T = \bar{X} \sum_{i=1}^{n} \left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right) = ZY, \tag{11}$$

where

$$Z \sim IG(\upsilon, n) \;\; \text{and} \;\; Y \sim \chi^2_{n-1}.$$

To demonstrate this we use the properties of $\bar{X}$ and $\lambda \sum_{i=1}^{n} \left(\frac{1}{X_i} - \frac{1}{\bar{X}}\right)$ from Eq. (3) and the fact that for any constant $c$, $cX \sim IG(c\mu, c\lambda)$, given that $X \sim IG(\mu, \lambda)$. The exact Distribution of T then can be expressed as

$$P(T \leq t) = \int_0^\infty P\left(Z \leq \frac{t}{y}\right) f_Y(y) dy \tag{12}$$

or

$$P(T \leq t) = \int_0^\infty P\left(Y \leq \frac{t}{z}\right) f_Z(z) dz. \tag{13}$$

Note that $P\left(Z \leq \frac{t}{y}\right)$ may be calculated from cumulative distribution of the standard normal distribution as

$$P(Z \leq z) \;\; = \;\; \Phi\left\{\sqrt{\frac{n}{z}}\left(\frac{z}{\upsilon} - 1\right)\right\} + \exp\{\frac{2n}{\upsilon}\} \; \Phi\left\{-\sqrt{\frac{n}{z}}\left(\frac{z}{\upsilon} + 1\right)\right\}$$

Computationally the latter representation may be better because $exp(\frac{2n}{\upsilon})$ can blow up for large $n$ or small $\upsilon$.

## 4. Approximations

The approximations for the test statistic $T$ are moment based approximations based on the limiting behaviour of the components $Z$ and $V$. Using the moment properties of IG and Chi-square distributions, we give below the first three cumulants of the statistic $T$,

$$\begin{aligned} E(T) &= \upsilon\nu \\ V(T) &= \frac{\upsilon^3}{\nu}(2\nu + nu^2) + \upsilon^2 + 2\nu \\ k_3(T) &= (\nu^3 + 6\nu^2 + 8\nu)\left(\upsilon^3 + 3\frac{\upsilon^4}{n} + 3\frac{\upsilon^5}{n^2}\right) - 3\upsilon\nu\left(\upsilon^2 + \frac{\upsilon^3}{n}\right)(2\nu + \nu^2) + 2\upsilon^3\nu^3. \end{aligned}$$

where $\nu = n - 1$.

### 4.1 First Approximation

Noting that for large $n$, $Y/\nu$ converges to 1 with probability 1, we may approximate $T$ by an IG random variable. Note that $E(T/(\upsilon\nu)) = 1$, we approximate the distribution of $T/(\upsilon\nu)$ by that of $IG(1, \lambda_1)$, where $\lambda_1$ may be obtained by equation the second moment. Since $Var(\frac{T}{\upsilon\nu}) = \frac{1}{\lambda_1}$, thus $\lambda_1 = \frac{n\nu}{(2+\nu)\upsilon+2n}$. Note that for large $n$, $\lambda_1 \approx \frac{n}{\upsilon}$, therefore for large $n$

$$T/\nu \approx IG(\upsilon, n).$$

### 4.2 Second Approximation

Using a similar approach to the distribution of $Z$, instead of that of $Y$, we can approximate $T$ by a multiple of Chi-square random variable. So we have approximately $\frac{T}{\upsilon\nu} \sim Gamma(\alpha, \beta)$, where $\alpha\beta = 1$ and $\alpha\beta^2 = \text{Var}(T/(\upsilon\nu)$. Thus we get

$$\beta = \frac{(2+\nu)\upsilon + 2n}{n\nu} = 1/\lambda_1 \;\; \text{and} \;\; \alpha = \lambda_1. \tag{14}$$

### 4.3 Third Approximation

This approximation combines the strengths of the above approximations through a mixture approach where

$$\frac{T}{\upsilon\nu} = w.IG(1, \lambda_1) \oplus (1 - w).Gamma\left(\lambda_1, \frac{1}{\lambda_1}\right),$$

with $w$ being the mixing coefficient to be determined appropriately, and $\oplus$ denotes that the distribution (and consequently the pdf) is the mixture of the constituent random variables. Therefore

$$P\left(\frac{T}{\upsilon\nu} \leq x\right) = w.\text{cdf of } IG(1, \lambda) + (1 - w).\text{cdf of } Gamma\left(\lambda_1, \frac{1}{\lambda_1}\right).$$

To obtain the appropriate value of $w$ we equate the third central moments:

$$E\left(\frac{T}{\upsilon\nu}\right)^3 \quad = \quad E[IG(1, \lambda_1)]^3 + (1 - w)E\left[Gamma\left(\lambda_1, \frac{1}{\lambda_1}\right)\right]^3.$$

This gives

$$k_3(T/(\upsilon\nu)) = w.k_3[IG(1, \lambda_1)] + (1 - w).k_3\left[Gamma\left(\lambda_1, \frac{1}{\lambda_1}\right)\right],$$

Noting that $k_3[IG(1, \lambda)] = \frac{3}{\lambda^2}$ and $k_3[Gamma\left(\lambda, \frac{1}{\lambda}\right)] = \frac{2}{\lambda^2}$, we get

$$w = \frac{\lambda_1^2}{\upsilon^3\nu^3}k_3(T) - 2. \tag{15}$$

The simplicity of the mixture approximation is in approximation of percentiles in the fact that they can be also obtained by simple mixtures (see Chaubey, 1989)). Thus

$$t_\alpha = (\upsilon\nu)[wX_\alpha^{(1)} + (1 - w)X_\alpha^{(2)}],$$

where $X_\alpha^{(1)}$ denotes the percentile of $IG(1, \lambda_1)$ and $X_\alpha^{(2)}$ denotes that for the random variable $Gamma(\lambda_1, 1/\lambda_1)$.

## 5. Comparison of the Approximations

For a qualitative judgement of the three approximations, the exact distribution function and corresponding approximations were computed for a selection of values of the sample size $n$ and the exact value of the CV as given in terms of $\upsilon$. These are accompanied by a `boxplot` of the corresponding errors. Figures 1-4 represent these for sample sizes 10, 20, 30 and 50 respectively. From the graph of the distribution functions, it is difficult to discriminate between different approximations; but the error plots seem quite useful. We also compare the exact values of the percentiles for different values of $\upsilon_0$ and $\alpha = .01$ and .05. These values are displayed in the table below. The computations are performed using numerical integration routines from `R` (see R Development Core Team, 2005).

**Table 1:** A comparison of the mixture approximation for the percentiles of the test statistic $T$

| $\upsilon = 0.1$, $\alpha = 0.05$ | | | | |
|---|---|---|---|---|
| | $n = 10$ | $n = 20$ | $n = 30$ | $n = 50$ |
| Exact | 1.907226 | 1.604352 | 1.481422 | 1.364071 |
| IG Approx. | 1.928325 | 1.620498 | 1.493704 | 1.372344 |
| Gamma Approx. | 1.906014 | 1.602906 | 1.480233 | 1.363272 |
| Mixture Approx. | 1.908313 | 1.604630 | 1.481531 | 1.364133 |
| $\upsilon = 0.3$, $\alpha = 0.05$ | | | | |
| Exact | 1.959070 | 1.638569 | 1.508182 | 1.383821 |
| IG Approx | 1.978577 | 1.653179 | 1.519290 | 1.391281 |
| Gamma Approx | 1.956529 | 1.634673 | 1.504930 | 1.381517 |
| Mixture Approx | 1.962573 | 1.639464 | 1.508575 | 1.383957 |
| $\upsilon = 0.1$, $\alpha = 0.01$ | | | | |
| Exact | 2.470374 | 1.940409 | 1.736048 | 1.547014 |
| IG Approx | 2.602469 | 2.009526 | 1.782771 | 1.575322 |
| Gamma Approx | 2.453586 | 1.932056 | 1.730547 | 1.543780 |
| Mixture Approx | 2.468925 | 1.939647 | 1.735577 | 1.546775 |
| $\upsilon = 0.3$, $\alpha = 0.01$ | | | | |
| Exact | 2.590431 | 2.008769 | 1.786265 | 1.581772 |
| IG Approx | 2.705068 | 2.069618 | 1.827709 | 1.607070 |
| Gamma Approx | 2.543698 | 1.985152 | 1.770659 | 1.572561 |
| Mixture Approx | 2.587937 | 2.007021 | 1.785142 | 1.581184 |

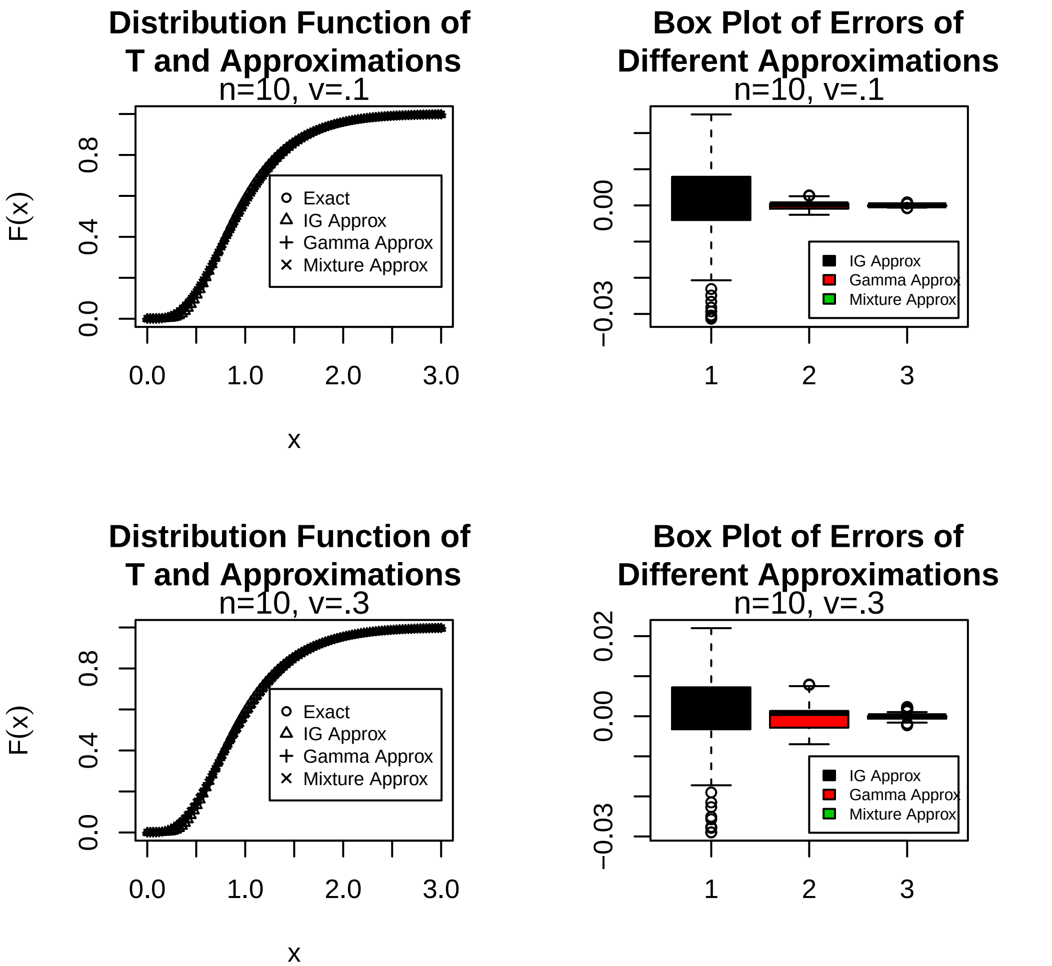


**Figure 1**: Exact and Approximate Distributions of $T$ and Their Error Analysis for $n = 10$

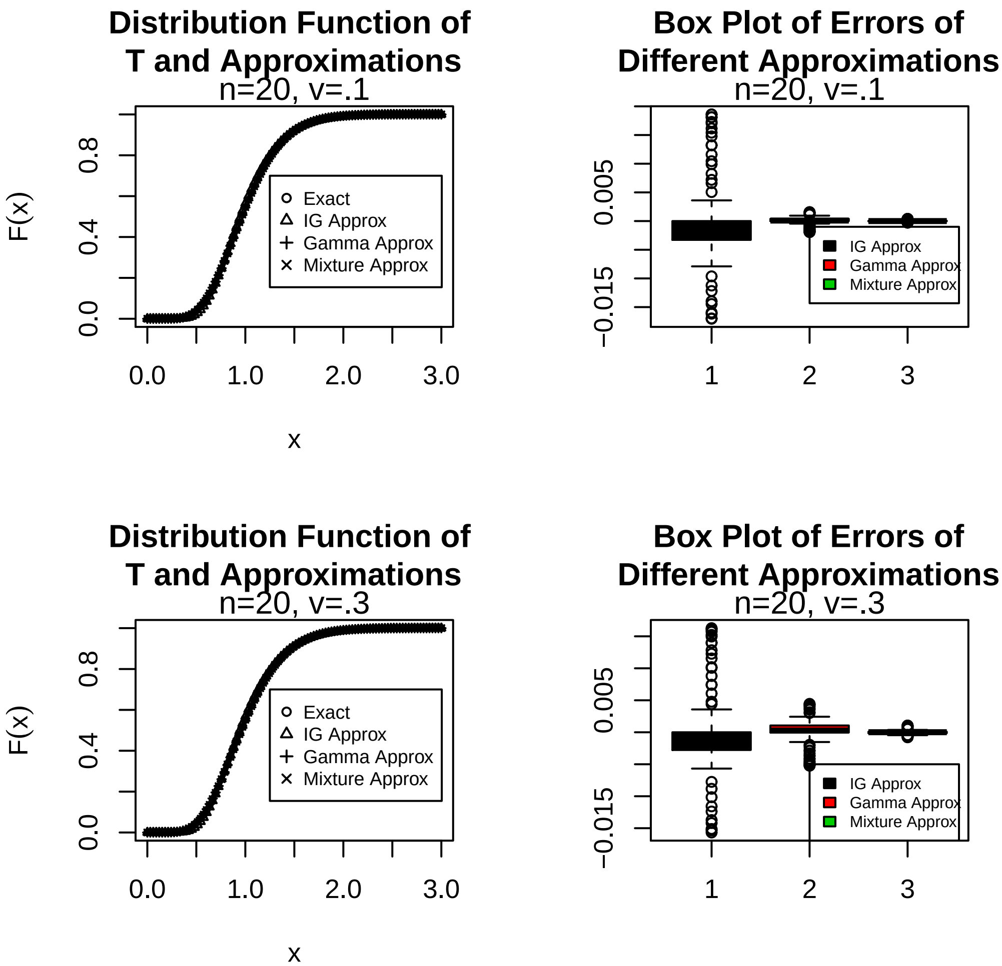


**Figure 2**: Exact and Approximate Distributions of $T$ and Their Error Analysis for $n = 20$

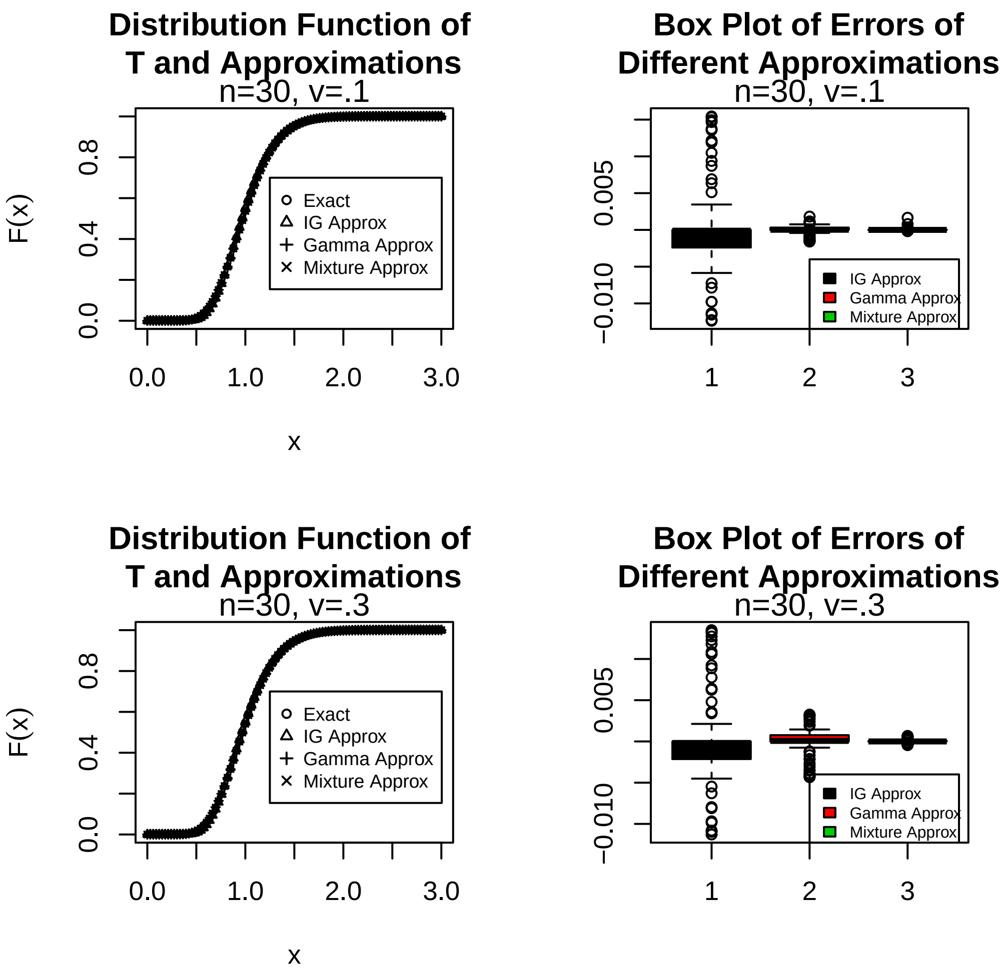


**Figure 3**: Exact and Approximate Distributions of $T$ and Their Error Analysis for $n = 30$

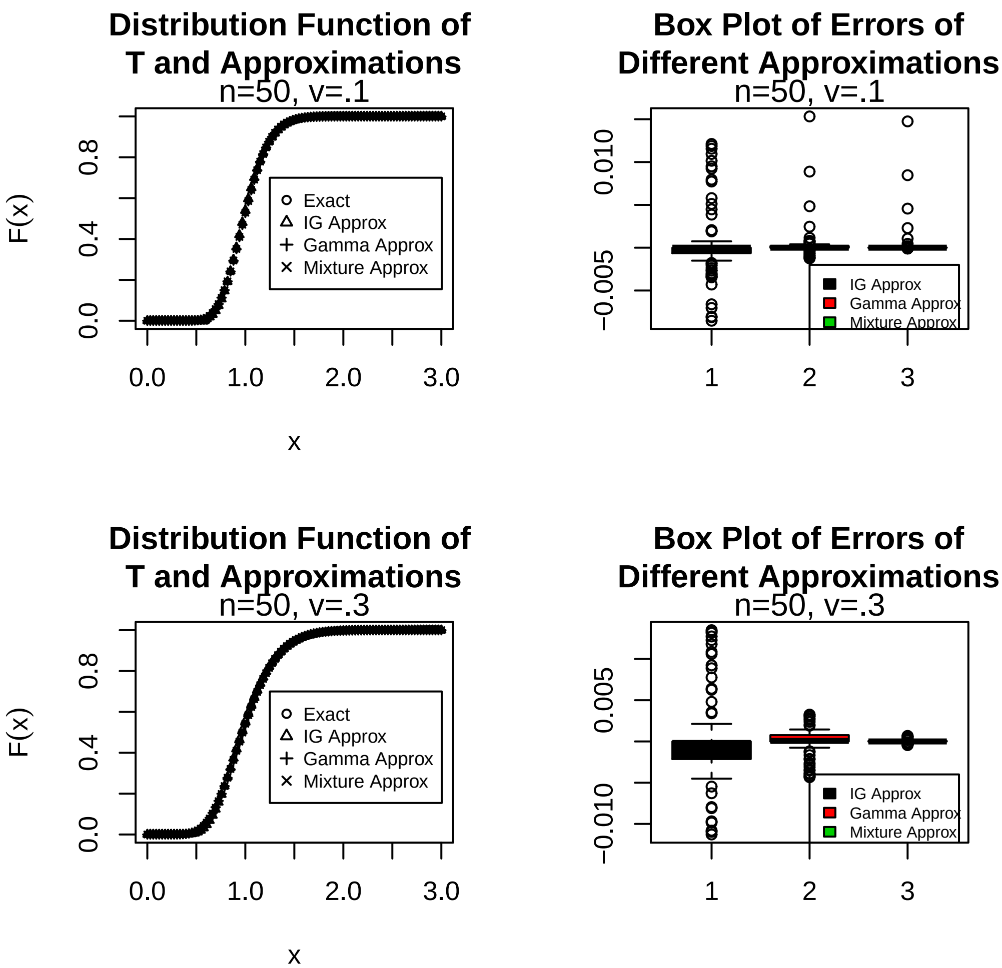


**Figure 4**: Exact and Approximate Distributions of $T$ and Their Error Analysis for $n = 50$

From the graphs of the distribution functions in Figures 1-4, it seems that all the three approximations provide similar qualitative performance. However, the accompanying error plots show that the IG approximation is not as good as the $\chi^2$ approximation for small $\upsilon$. For larger values of $\upsilon$, IG approximation is quite good, however, mixture approximation may provide substantial improvement over both of them providing a very accurate approximation. The mixture approximation may also provide a fairly accurate (up to two decimals) approximation to the percentiles.

## Acknowledgements

The first author thankfully acknowledges the financial support from Concordia University, Canada to present this manuscript at the 2013 Joint Statistical Meetings held in Montreal, Quebec, Canada. The second author would like to acknowledge the partial support of this research through a discovery grant from NSERC, Canada.

## REFERENCES

Chaubey, Y. P. (1989), "Edgeworth Expansions with Mixtures," *Metron* XLVII, 53–64.

Chhikara, R. S., and Folks, J.L. (1989), *The Inverse Gaussian Distribution: Theory, Methodology and applications*, New York: Marcel Dekker.

Folks, J. L., and Chhikara, R. S. (1978), "The Inverse Gaussian Distribution and Its Statistical Application - A Review", *J. Roy. Statist. Soc.*, B40, 263-289.

Jorgenson, B. (1982), *Statistical Properties of the Generalized Inverse Gaussian Distribution,* Lecture Notes in Statistics #9, New York: Springer-Verlag.

R Development Core Team (2005), *R: A Language and Environment for Statistical Computing,* R Foundation for Statistical Computing, Vienna, Austria.

Searles, D.T. (1964), "The Utilization of a Known Coefficient of Variation in the Estimation Procedure," *J. of American Statistical Association*, 59, 1225-1226.

Seshadri, V. (1993), *The Inverse Gaussian Distribution: A Case Study in Exponential Families,* Oxford: Clarenden Press.

Seshadri, V. (1998), *The Inverse Gaussian Distribution: Statistical Theory and Applications,* New York: Springer Verlag.

Srivastava, V. K. (1974), "On The Use of Coefficient Of Variation in Estimating Normal Mean," *J. Indian Society of Agricultural Statistics*, 26, 33-36.

Tweedie, M. C. K. (1957a), "Statistical Properties of Inverse Gaussian Distributions-I," *Ann. Math. Statist.*, 28, 362-377.

Tweedie, M. C. K. (1957b), "Statistical Properties of Inverse Gaussian Distributions-II," *Ann. Math. Statist.*, 28, 696-705.