2008

APPLICATIONS OF THE BIVARIATE GAMMA DISTRIBUTION IN NUTRITIONAL EPIDEMIOLOGY AND MEDICAL PHYSICS

Jolene Barker
Virginia Commonwealth University

Virginia Commonwealth University

School of Medicine

This is to certify that the thesis prepared by Jolene Kristen Barker entitled APPLICATIONS OF THE BIVARIATE GAMMA DISTRIBUTION IN NUTRITIONAL EPIDEMIOLOGY AND MEDICAL PHYSICS has been approved by his or her committee as satisfactory completion of the thesis or dissertation requirement for the degree of Master of Science

---

Viswanathan Ramakrishnan, Ph.D., Director of Thesis

Jessica McKinney Ketchum, Ph.D., School of Medicine

Elisabeth Weiss, M.D., School of Medicine

Shumei S. Sun, Ph.D., F.A.H.A, Chair, Department of Biostatistics

---

Jerome F. Strauss, III, M.D., Ph.D., Dean, School of Medicine

---

Dr. F. Douglas Boudinot, Dean of the School of Graduate Studies

December 12, 2008

APPLICATIONS OF THE BIVARIATE GAMMA DISTRIBUTION IN NUTRIONAL EPIDEMIOLOGY AND MEDICAL PHYSICS

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

by

JOLENE KRISTEN BARKER
B.S., Physics, Radford University, 2006

Director:
VISWANATHAN RAMAKRISHNAN, PH.D.
ASSOCIATE PROFESSOR
DEPARTMENT OF BIOSTATISTICS

Virginia Commonwealth University
Richmond, Virginia
September 2008

Acknowledgement

This work would not have been possible without the support and encouragement of my thesis advisor, Dr. Ramesh. I am eternally grateful that he agreed to advise me throughout this endeavor even though I had not had a formal class with him. At first, I was concerned that our lack of previous interaction would have him "headed for the hills" after a couple of weeks as my advisor, but we developed a wonderful working relationship.

Dr. Jessica McKinney Ketchum, one of my committee members, has also been abundantly helpful, and has assisted me in numerous ways throughout my time at VCU. I would also like to thank Russ Boyle for wrestling the "black box" of funding on my behalf. I am very grateful for the generous stipend support that the Department of Biostatistics gave me.

Many thanks are due to Dr. Robert Johnson who has been especially inspiring because of his constant commitment to student success which does not go unnoticed by those who have been under his instruction. Whenever a first-year student inquires about his reputation with students, he is always well spoken of by his previous and current students.

I am also grateful for the close friends I have made at VCU: Stephanie, Andre, Tina and Rhonda. While I thoroughly enjoyed spending many long hours with Stephanie studying for the qualifiers, I most definitely wish that neither she nor I ever have to spend a month and a half holed up in a basement studying, blasting music and eating food from McDonalds at 4am. You all are such an encouragement.

I am very grateful for the spiritual and emotional support from the congregation at Saint Giles. I am very fortunate to have met such great people there and they will be very much missed.

I cannot end without thanking my family, on whose constant encouragement and love I have relied on throughout my higher education. Special thanks are due to my mother for constantly reminding me that the will of God will never lead me to a place where His grace is not sufficient.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Abstract

APPLICATIONS OF THE BIVARIATE GAMMA DISTRIBUTION IN

NUTRITIONAL EPIDEMIOLOGY AND MEDICAL PHYSICS

By Jolene Kristen Barker, M.S.

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2008

Major Director: Viswanathan Ramakrishnan, Ph.D.
Associate Professor, Department of Biostatistics

In this thesis the utility of a bivariate gamma distribution is explored. In the field of nutritional epidemiology a nutrition density transformation is used to reduce collinearity. This phenomenon will be shown to result due to the independent variables following a bivariate gamma model. In the field of radiation oncology paired comparison of variances is often performed. The bivariate gamma model is also appropriate for fitting correlated variances.

A method for simulating bivariate gamma random variables is presented. This method is used to generate data from several bivariate gamma models and the asymptotic properties of a test statistic, suggested for the radiation oncology application, is studied.

1. Introduction

In the theory of statistics we are introduced to many univariate and multivariate models. For example, the Gaussian (or normal) model, used in hypothesis testing and regression of continuous outcomes, the chi-squared model, used in testing goodness-of-fit, the binomial model, used in the analyses of dichotomous data, the Poisson model, used in the analyses of count data, and so on. In this thesis we consider a seldom used model, the bivariate gamma model and discuss its properties and its applications. The need for the bivariate gamma model is motivated by introducing two situations where it might be appropriate.

1.1 Nutrition Application

In nutritional epidemiology, it is often of interest to examine the relationships between diet and certain health abnormalities. These studies require a model that contains a response variable, Y, a specific nutrient intake variable, N, which may be associated with the response variable, and the total nutrient intake variable, T, which also may be associated with the response variable. For example, the response variable may be a dichotomous variable, such as the presence or absence of heart disease and a specific nutrient intake variable of interest may be fat. Statistical models are built to examine the effects of nutrient intake on the response. In these models it is suggested that the total intake, which has great potential for confounding, be included as a covariate (Willett 1990, Palmgren 1993). Hence, a regression model takes the suggested following form:

$$f(Y) = \beta_0 + \beta_1 N + \beta_2 T + \varepsilon, \tag{1.1}$$

where $f(Y)$ is a function of the response Y.

In applications, when one attempts to fit the model (1.1), the collinearity between the specific nutrient variable and the total nutrient intake variable could lead to unstable estimates for the regression parameters.

Eliminating or reducing the collinearity has been a problem of interest in the literature. This may be achieved in several ways. Some of the common methods are using centered data for the predictor variables, principle components regression which uses the method of least squares to allow biased estimators on a set of artificial variables of the correlation matrix that are then possibly eliminated to reduce the variance, and ridge regression which modifies the method of least squares to allow biased estimators of the regression coefficients (Myers 1990). This thesis will look at an alternate method that is specific to the nutrition application. Here we will consider the "nutrient density transformation" method. This method builds the "nutrient density" model, which substitutes the nutrient density $U = N/T$ in place of N, along with T included as a covariate in the model (Jain, Cook, Davis, Grace, Howe and Miller 1980, Hegsted 1985, Smith, Slettery and French 1991, Willett and Sampfer 1894). This method is both biologically and statistically sound.

Nutrient density is a measure of dietary composition. There are two analogous approaches to calculating nutrient density. One is defined as the ratio of a specific nutrient such as carbohydrate, protein or fat to the total caloric intake. The other is used for macronutrients. This approach expresses intake of a specific nutrient as a percentage of total caloric intake (Willett 1998). So one can either express the nutrient as a ratio of the total energy or as a percentage of energy (Willett 1990). Nutrient density, interpreted

as the vitamin or mineral content of food or diet per unit energy, has long been a useful measure in the nutritional sciences (Backstrand 2003). The concept of nutrient density recognizes the close relationship between energy intake and consumption of other nutrients. Vitamins and minerals are almost always consumed together with significant amounts of energy; therefore, intakes of energy and micronutrients are often strongly correlated (Backstrand 2003).

From a statistical point of view, in some nutritional data, nutrient density (U) in addition to being biologically meaningful as discussed above, seems to be less correlated with the total nutrients (T) than the nutrient intake (N) thereby eliminating the problem of collinearity.

In this thesis a theoretical rationale for this phenomenon is considered. Conditions under which the transformation will or will not eliminate collinearity will be provided. Specifically, we will provide conditions under which, ρ_{UT}, the correlation between U and T, will be smaller in absolute value compared to ρ_{NT}, the correlation between N and T. A collection of bounds for the mean of the nutrient density will be derived, within which ρ_{UT} can be expected to be smaller than ρ_{NT}. It will also be proven that the correlation ρ_{UT} is zero if the joint distribution of N and T is bivariate gamma.

1.2 Radiation Oncology Application

In radiation oncology, it is often of interest to compare the geometric targeting error of two treatments. Consider the simple case of evaluating the error in targeting the tumor centroid (i.e., the center of mass) for two treatment arms, say A and B. Let x_{ij} be the difference between the anticipated location of the tumor centroid relative to the isocenter and its actual location for the *j*th patient at the *i*th fraction for one of the treatments. The isocenter is defined as the point in space through which the central beam of radiation passes.

The systematic error (error for patient j averaged over N fractions), namely $\bar{x}_j = \sum_{i=1}^{N} x_{ij} \Big/ N$ is the measurement of interest. The distribution of $\bar{x}_j$ in the patient population is usually assumed to be normal with mean 0 and variance Σ^2 (scalar valued variance of all systematic errors.) We denote this by $N(0, \Sigma^2)$. The study hypothesis for the geometric endpoint is formulated in terms of Σ. Although all the measurements, x_{ij}, are available, the primary statistical analysis comparing the two treatments based on this endpoint is usually based on $\bar{x}_j$. In addition, in most experiments the two treatment arms are applied in succession on the same patients.

One example of this application is the study comparing Brachytherapy and a new proposed Image-guided adaptive radiation therapy (IGART). These are applied to the same patients in sequence and the systematic errors are observed. Brachytherapy (from the Greek brachy, meaning “short”), is a form of localized radiotherapy (radiation therapy given at a short distance). In this, radioactive seeds or sources are placed inside

of, or next to, the area requiring treatment and the radiation is targeted around these seeds. Brachythearpy has minimal side effects as the area near the tumor or the tumor itself is given a high radiation dose while reducing the radiation exposure in the surrounding healthy tissues. Brachytherapy is an option for patients with localized (organ-confined) cancer and it provides a good alternative to surgical removal of the cancerous tissue. In the treatment of prostate cancer, it is minimally invasive as the radioactive seeds are about the size of a grain of rice. They are implanted through very thin needles. Depending on different variables, between 50 and 100 seeds are used (American Brachytherapy Society).

Image-guided adaptive radiation therapy (IGART) is the development of image-guided target localization and patient-specific adaptation techniques (Song 2005). IGART is a closed-loop treatment process that is designed to include the individual treatment information, such as patient-specific anatomic variation and delivered dose assessed during the therapy course in treatment evaluation and planning optimization (Yan 2008). In an IGART process, treatment planning and modification decisions are typically made on the basis of patient-specific anatomic and biological variations determined from multiple image measurements of patient anatomy obtained at various times during the course of treatment delivery (Yan 2008).

In off-line IGART sessions, onboard cone-beam computed tomography (CBCT) allows one to acquire volumetric information of a patient prior to treatment on a routine basis. Cone-beam CT is a scanner that uses a cone shaped x-ray beam rather than a conventional linear fan beam to provide images of the bony structures of the skull.

Cone-beam CT scanners use a square 2 dimensional array of detectors to capture the cone shaped beam thus providing a volume of data. Subsequently reconstruction software is applied on the cone-beam CT volumetric data to produce a stack of 2D gray scale level images of the anatomy (www.conebeam.com). This makes it possible to adaptively modify the patient treatment plan with consideration of organ deformation as well as previously delivered doses.

Each procedure is applied to each patient and the distances $\bar{x}_{j,A}$ and $\bar{x}_{j,B}$ are measured. As a result the measurements are correlated. They are assumed to be jointly bivariate normal with means (0,0) variances, (Σ_A^2, Σ_B^2), and some covariance Σ_{AB}. The null hypothesis of interest is $H_0: \Sigma_A/\Sigma_B \leq 1$ against the alternative $H_a: \Sigma_A/\Sigma_B > 1$. Equivalently, the alternative hypothesis is that IGART reduces the systematic error significantly.

When two independent samples are available, the tests for this null hypothesis would be a straight forward *F* test based on the sample variances. However, here, since the procedures are administered to the same individuals, the independence assumption is violated and hence the test can not be preformed using the standard *F* test. We propose a test statistic similar to the *F* statistic; however, we do not assume the distribution of this statistic to be an *F* distribution. Suppose the estimate of the variances of the two treatments for the *j*th individual be denoted by $\bar{x}_{j,A}$ and $\bar{x}_{j,B}$. We derive the distribution of the statistic $(\bar{x}_{j,A}^2 / \bar{x}_{j,B}^2)(\Sigma_B / \Sigma_A)$ under the null hypothesis under appropriate assumptions.

1.3 Bivariate Gamma Distribution

In both examples discussed above, we will argue in chapter 3 that a bivariate gamma model is appropriate. In this section we will introduce this model by defining its probability density function (pdf). Two random variables X and Y are jointly distributed as a bivariate gamma when the pdf is

$$f(x,y \mid \alpha,\beta,\gamma) = \frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)} x^{\alpha-1}(y-x)^{\gamma-1}e^{-\beta t}, \tag{1.2}$$

where $0 < X < Y$ and α, β, and γ are all greater than 0 (Rohatgi 1976, page 118). Bivariate gamma distributions have found useful applications in many areas. For example, in the modeling of rainfall at two nearby rain gauges, data obtained from rainmaking experiments, the dependence between annual streamflow and aerial precipitation, wind gust data, and the dependence between rainfall and runoff (Nadarajah 2006). They have also found applications in reliability theory, renewal processes, and stochastic routing problems (Nadarajah 2006).

In chapter 2 we will discuss the properties of the bivariate gamma distribution. We will discuss the marginal and conditional distributions of X and Y. Consequently, the marginal expectations and variances of X and Y and the correlation between X and Y will be calculated.

In chapter 3 we will argue why the bivariate gamma distribution may be useful in explaining the reduction of multicollinearity in the nutrition application and why it is appropriate in radiation oncology applications. Specifically, in the nutrition case, we will show how the expected correlation between the nutrient density and the total intake is in

fact exactly zero under this model. In the radiation oncology case we will provide a test statistic for testing the hypothesis of interest discussed in section 1.2.

In chapter 4 we will provide a method for simulating bivariate gamma data and utilize this to study the asymptotic properties of a test statistic suggested for the radiation oncology application.

2. Properties of the Bivariate Gamma Model

In this chapter first we will revisit the bivariate gamma probability density function (pdf) and study its properties. The marginal distribution of the two random variables will be derived. Consequently, the marginal expectations and variances of X and Y will be derived. The marginal distributions will be used to determine the conditional distributions. Finally, the correlation between X and Y will be derived.

The joint pdf for two random variables X and Y following a bivariate gamma distribution could be considered

$$f(x,y \mid \alpha,\beta,\gamma) = \frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)} x^{\alpha-1}(y-x)^{\gamma-1}e^{-\beta y}, \tag{2.1}$$

where $0 < X < Y$ and α, β and γ are all constants greater than 0 (Rhotaghi 1976, page 118). Notice that based on the condition $0 < x < y$ the term $y - x$ guarantees that the probability density function is greater than 0. The marginal distribution of X could be calculated by eliminating Y by integrating the joint pdf over the range of Y. That is, the marginal pdf of X is

$$f(x) = \frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)} \int_x^\infty x^{\alpha-1}(y-x)^{\gamma-1}e^{-\beta y}dy. \tag{2.2}$$

To compute the integral substitute $u = y - x$. Notice that if y equals x then u is 0 and the integral in terms of u has the form

$$f(x) = \frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)} x^{\alpha-1} \int_0^\infty u^{\gamma-1}e^{-\beta(x+u)}du. \tag{2.3}$$

After rearranging terms as below (2.4) we notice that the integral is that of a univariate gamma function where the constants are γ and β. That is, the integral equals $\Gamma(\gamma)/\beta^{\gamma}$.

$$f(x)=\frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)}x^{\alpha-1}e^{-\beta x}\int_0^{\infty}u^{\gamma-1}e^{-\beta u}du \tag{2.4}$$

Thus the equation (2.4) reduces to the pdf

$$f(x)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}. \tag{2.5}$$

where $0<x<\infty$ and α and β are constants greater than 0. Notice that the marginal distribution of X only depends on α and β but not on γ. Also, we recognize, the distribution of X is a gamma with parameters (α,β). In a similar fashion the marginal distribution of Y could be calculated by eliminating X by integrating the joint pdf over the range of X. That is, the marginal pdf of Y is

$$f(y)=\frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)}\int_0^{\infty}x^{\alpha-1}y^{\gamma-1}(1-\frac{x}{y})^{\gamma-1}e^{-\beta y}dx \tag{2.6}$$

To compute the integral substitute $u=x/y$. In this case if y equals x then u is 1 and the integral in terms of u has the form

$$f(y)=\frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)}\int_0^{1}(yu)^{\alpha-1}y^{\gamma-1}(1-u)^{\gamma-1}e^{-\beta y}ydu \tag{2.7}$$

After rearranging terms we notice that the integral takes the form of a beta function where the constants are α and γ.

$$f(y)=\frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)}y^{\alpha+\gamma-1}e^{-\beta y}\int_0^{1}u^{\alpha-1}(1-u)^{\gamma-1}du \tag{2.8}$$

Thus the integral is $B(\alpha,\gamma)$ or equivalently $\Gamma(\alpha)\Gamma(\gamma)/\Gamma(\alpha+\gamma)$. Substituting this in (2.8) yields the marginal pdf,

$$f(y)=\frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha+\gamma)}y^{\alpha+\gamma-1}e^{-\beta y} \tag{2.9}$$

where $0<y<\infty$ and $\alpha,\beta,$ and γ are constants greater than 0. Thus the marginal distribution of Y is also a gamma with parameters ($\alpha+\gamma,\beta$). The marginal expectations and variances of X and Y are therefore (Casella and Berger 2002)

$$\mu_X = E(X)=\frac{\alpha}{\beta},$$
$$\mu_Y = E(Y)=\frac{\alpha+\gamma}{\beta},$$
$$V(X)=\frac{\alpha}{\beta^2},$$
$$V(Y)=\frac{\alpha+\gamma}{\beta^2}$$

The conditional distribution of X given Y is the ratio of the joint density in (2.1) to the marginal distribution of Y in (2.9). After substituting and simplifying, the conditional distribution of X given Y can be shown to have the pdf

$$f(x\mid y)=\frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\Gamma(\gamma)}x^{\alpha-1}(y-x)^{\gamma-1}y^{-\alpha-\gamma+1}. \tag{2.10}$$

Similarly, the conditional distribution of Y given X is the ratio of the joint density in (2.1) to the marginal distribution of X in (2.5). After cancelling the appropriate terms the conditional distribution of Y given X has the pdf

$$f(y\mid x)=\frac{\beta^{\gamma}}{\Gamma(\gamma)}(y-x)^{\gamma-1}e^{-\beta(y-x)}. \tag{2.11}$$

Next we consider the correlation between X and Y. The correlation indicates the strength and direction of a linear relationship between two random variables. We will calculate the Pearson correlation coefficient which is the covariance of the two variables divided by the product of their standard deviations. That is, the correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is defined as:

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{2.12}$$

For any random variables X and Y, the covariance is defined as

$$Cov(X,Y) = E(XY) - \mu_X \mu_Y \tag{2.13}$$

To calculate the correlation between X and Y first calculate the expected value of the bivariate random vector (X,Y). To do this we will solve the integral

$$E(X,Y) = \frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)} \int_0^\infty y e^{-\beta y} \int_0^y x^\alpha (y-x)^{\gamma-1} dxdy \tag{2.14}$$

After rearranging terms we obtain

$$E(X,Y) = \frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)} \int_0^\infty y^\gamma e^{-\beta y} \int_0^y x^\alpha (1-\frac{x}{y})^{\gamma-1} dxdy \tag{2.15}$$

To compute the integral substitute $u = x/y$ as before and solve. That is,

$$E(X,Y) = \frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)} \int_0^\infty y^{\alpha+\gamma+1} e^{-\beta y} \int_0^1 u^\alpha (1-u)^{\gamma-1} dudy \tag{2.16}$$

We now notice that the last integral is in the form of a beta function with parameters $(\alpha+1,\gamma)$. Substituting this the integral reduces to,

$$E(X,Y)=\frac{\beta^{\alpha+\gamma}\alpha}{(\alpha+\gamma)\Gamma(\alpha+\gamma)}\int_0^\infty y^{(\alpha+\gamma+2)-1}e^{-\beta y}dy \tag{2.17}$$

The integral is in the form of a gamma function so it reduces to $\Gamma(\alpha+\gamma+2)/\beta^{\alpha+\gamma+2}$. After cancelling the appropriate terms using the relationships of a gamma function, we get $E(X,Y)=\alpha(\alpha+\gamma+1)/\beta^2$. Since we already know μ_x and μ_Y, we now can calculate $COV(X,Y)$ which turns out to be α/β^2. Dividing this fraction by $\sigma_X\sigma_Y$,which is $\sqrt{Var(X)}\sqrt{Var(Y)}$, yields

$$\rho_{XY}=\frac{\sqrt{\alpha}}{\sqrt{\alpha+\gamma}} \tag{2.18}$$

Next to apply this to the nutrient density and the radiation oncology applications we consider the joint distribution of $U=X/Y$ and $V=T$. Using the change of variable technique to the joint pdf of X and Y, it can be shown that the joint distribution of U and V is,

$$f(u,v;\alpha,\beta,\gamma)=\frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)}u^{\alpha-1}(1-u)^{\gamma-1}v^{\alpha+\gamma-1}e^{-\beta v}, \tag{2.19}$$

where $0<u<1,\ 0<v<\infty$ and α,β and γ are positive. The Jacobian of this transformation is 1. The above density clearly factors into a gamma $(\alpha+\gamma,\beta)$ density for V and a beta (α,γ) density for U. Therefore, U and T are independent and hence $\rho_{UT}=0$. Also, $E(U)=\alpha/\alpha+\gamma$, as U follows a beta distribution. The pdf of U is below,

$$f(u;\alpha,\gamma)=\frac{1}{B(\alpha,\gamma)}u^{\alpha-1}(1-u)^{\gamma-1}, \tag{2.20}$$

where $0 < u < 1$ and α and γ are positive.

3. Application of the Bivariate Gamma Model

In this chapter we will show why the bivariate gamma model is appropriate in the nutrition and radiation oncology examples. For the nutrition example, we will show how the collinearity between the nutrient intake and the total intake translates to zero correlation under the bivariate gamma model. Then we will also provide an interval for the correlation within which one could expect the correlation between the nutrient density and the total intake is smaller than the correlation between the nutrient intake and the total intake.

3.1. Nutrition

As introduced in chapter 1, in nutrition the usual regression model employed to find associations between nutrients and health abnormalities has the following form:

$$f(Y) = \beta_0 + \beta_1 N + \beta_2 T + \varepsilon, \tag{3.1}$$

where Y is the outcome, for example it is the absence or presence of a certain health abnormality, N, is the measure of a specific nutrient intake, and, T, is the measure of total nutrient intake which is treated as a covariate. Although the model in (3.1) seems to adequately explore these relationships, the collinearity between N and T leads to unstable estimates for the regression parameters. To eliminate or reduce the collinearity one suggestion is to consider the nutrient density model which substitutes the nutrient density $U = N/T$ in place of N, with T included as a covariate in the model.

To determine if the nutrient density model will be superior over the usual regression model in providing stable estimates for the regression parameters one must

consider the conditions under which, ρ_{UT}, the correlation between U and T, will be smaller in absolute value compared to ρ_{NT}, the correlation between N and T. That is, we want to find conditions under which

$$|\rho_{UT}| < |\rho_{NT}|. \tag{3.2}$$

Let R represent the intake other than the nutrients. Then N and R are positive random variables such that $T = N + R$ with means μ_N and μ_R so that $\mu_T = \mu_N + \mu_R$. Let the variances of N and R be σ_N^2 and σ_R^2, respectively. Result 1 provides an inequality for the correlation ρ_{UT} as a function of the correlation ρ_{NT}.

Result 1: The correlation between the nutrient density, U, and the total intake, T, given by

$$\rho_{UT} = \frac{\mu_N - E(U)\mu_T}{\sqrt{V(U)V(T)}} \tag{3.3}$$

where $E(.)$ and $V(.)$ denote the expectation and the variance, respectively, can be written as a function of ρ_{NT}.

Proof : The numerator in (3.3) follows since $E(U) = E[(N/T)T] = E(N) = \mu_N$. To show the right hand side of (3.3) is a function of ρ_{NT} it is shown the that $V(T)$ is a function of ρ_{NT}. We have, (denoting covariance by Cov)

$$V(T) = V(N) + V(R) + 2Cov(N,R) = \sigma_N^2 + \sigma_R^2 + 2Cov(N,R)$$

and the correlation between N and T is

$$\rho_{NT} = \frac{Cov(N,T)}{\sqrt{V(N)V(T)}} = \frac{\sigma_N^2 + Cov(N,R)}{\sqrt{\sigma_N^2\left(\sigma_N^2 + \sigma_R^2 + Cov(N,R)\right)}}.$$

Now multiplying and dividing the right hand side by 2 and then adding and subtracting σ_R^2 to the numerator and rewriting terms in terms of the variance of T yields,

$$\rho_{NT} = \sqrt{\frac{\sigma_N^2 + \sigma_R^2 + 2Cov(N,R)}{4\sigma_N^2}} - \frac{\sigma_R^2 - \sigma_N^2}{\sqrt{4\sigma_N^2[\sigma_N^2 + \sigma_R^2 + 2Cov(N,R)}}$$

$$= \sqrt{\frac{V(T)}{4\sigma_N^2}} - \frac{\sigma_R^2 - \sigma_N^2}{\sqrt{4\sigma_N^2 V(T)}} \tag{3.4}$$

Squaring both sides of (3.4) and solving for $V(T)$ and simplifying gives the following expression for $V(T)$:

$$V(T) = \sigma_N^2\left[\left(\frac{\sigma_R^2}{\sigma_N^2} - 1\right) + 2\rho_{NT}^2 + \sqrt{\left[\left(\frac{\sigma_R^2}{\sigma_N^2} - 1\right) + 2\rho_{NT}^2\right]^2 + \left(\frac{\sigma_R^2}{\sigma_N^2} - 1\right)^2}\right]. \tag{3.5}$$

The expression (3.3) will be zero if and only if

$$E(U) = E\left(\frac{N}{T}\right) = \frac{E(N)}{E(T)} = \frac{\mu_N}{\mu_T}. \tag{3.6}$$

That is, if and only if the mean of the nutrient density is the ratio of the mean of the macronutrient intake to the mean of the total intake. This equality seems to hold, at least approximately, for a variety of macronutrients. The reason for this, might be that the nutrient variable N and the remainder variable R approximately follow a gamma distribution and also that N and $T = N + R$ jointly follow a bivariate gamma distribution.

As shown below in Result 2, under the bivariate gamma model, equation (3.6) will exactly hold. Consequently, the expected correlation between the nutrient density and the total intake will be exactly zero.

Result 2: Suppose N and T are jointly distributed as a bivariate gamma distribution with probability density function

$$f(n,t;\alpha,\beta,\gamma)=\frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)}n^{\alpha-1}(t-n)^{\gamma-1}e^{-\beta t} \tag{3.7}$$

where $0<n<t$ and α,β and γ are all constants greater than 0 (Rhotaghi 1976, page 118). (The condition $0<n<t$ is appropriate in the context of nutrient data analysis since $N>0,\ R>0$ and $N<T=N+R$. That is, the intake of calories from the specific nutrients must be smaller than the intake of total calories. Then U and T are independent and

$$E\left(\frac{N}{T}\right)=\frac{E(N)}{E(T)}=\frac{\alpha}{\alpha+\gamma}. \tag{3.8}$$

Proof: Integrating (3.7) with respect to n and t shows the marginal distributions of N and T are respectively gamma with parameters (α,β) and $(\alpha+\gamma,\beta)$. Therefore the marginal expectations and variances of N and T are given in chapter 2 with N being X and T being Y. The correlation between N and T is

$$\rho_{NT}=\sqrt{\frac{\alpha}{\alpha+\gamma}} \tag{3.9}$$

the square root of the ratio of the respective means. Now transforming the variables N and T to $U = N/T$ and $V = T$ and using the change of variable technique it can be shown that the joint distribution of U and T is,

$$f(u,v;\alpha,\beta,\gamma) = \frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)} u^{\alpha-1}(1-u)^{\gamma-1} v^{\alpha+\gamma-1} e^{-\beta v}, \tag{3.10}$$

where $0 < u < 1,\ 0 < v < \infty$ and α, β and γ are positive. The above density clearly factors into a gamma $(\alpha+\gamma, \beta)$ density for T and for a beta (α, γ) density for U. Therefore, U and T are independent and hence $\rho_{UT} = 0$. Also, $E(U) = \alpha/\alpha + \gamma$, as U follows beta distribution.

When employing this transformation method, one may find that the data may not exactly follow a gamma distribution and hence the equality in (3.9) may not hold exactly. For these situations it would be useful to know how close the left-hand side and the right-hand side of (3.9) should be in order to achieve a reduction in the correlation ρ_{UT}. This leads to the following inequality that follows directly from equations (3.1) and (3.2).

Result 3: For random variables N and T with correlation, ρ_{NT}, where $0 < \rho_{NT} < 1$, and for random variable $U = N/T$ as defined, the inequality in (3.2) will hold if and only if

$$\left| E(U) - \frac{\mu_N}{\mu_T} \right| \le \frac{\rho_{NT}\sqrt{V(T)V(U)}}{\mu_T} \tag{3.11}$$

Proof: Squaring both sides of the inequality (3.2) and substituting (3.3) will yield (3.11) after some algebra.

The rationale for the stipulation of strict inequality on ρ_{NT} is as follows. In equation (3.11) if $\rho_{NT} = 0,$ the variables N, T and U should satisfy equation (3.8). The consequence of this under the bivariate gamma distribution is that $\alpha = 0,$ which is a violation under the gamma model. On the other hand, if $\rho_{NT} = 1,$ the two variables N and T become linearly dependent and hence the variance of the ratio, which is a constant, would be 0. Now, using the three results, the following theorem could be stated.

Theorem: Given any two random variables N and $T,$ for the transformation $U = N/T$, ρ_{UT}^2 will be less than $\rho_{NT}^2,$ if and only if, either a) the random variables N and T jointly follow a bivariate gamma distribution or b) the expected value of the transformed variable U satisfies the inequality in (3.11).

It is obvious from the equation (3.3) that, for given values of $\mu_N, \mu_T, E(U),$ and $V(U),$ the correlation ρ_{UT} will be small for large values of $\rho_{NT}.$ The representation in (3.11) suggests, when the specific nutrient variable is strongly correlated with the total intake there is a better chance of reducing the collinearity by transforming to nutrient density. This is because the width of the interval in (3.11) will be relatively small and hence the chance that the inequality will hold is reduced. Under this situation transforming to the nutrient density could in fact introduce collinearity unless N and T jointly follow a bivariate gamma distribution.

Separately from the correlation ρ_{NT} the interval in (3.11) depends on, σ_N^2, the variance of the specific nutrients (which increases the width for larger values), $\mu_T,$ the

mean of the total intake (which decreases the width for larger values) and the ratio, σ_R^2/σ_N^2. The three-dimensional plots of the correlation ρ_{UT} as a function of ρ_{NT} and $E(U)$ are presented in Figures 1(a)-1(d) to provide a visual representation of the interval. Plotting the square of the equation in (3.3) produced these figures. In all of these figures the mean of the total intake was set at 2500 and the variance of the nutrient density was set at 1. The variance of the specific nutrients was set at two levels, 10 (micronutrients; figures 1a) and 1c)) and 250 (macronutrients; figure 1b) and 1d)). Similarly, the ratio between the variances of R and N was set at two levels, namely, 10 (figures 1a) and 1b)) and 20 (figures 1c) and 1d)).

The x-axes of the figures represent the square of the correlation between the specific nutrients and the total intake. The y-axes represent the square of the difference between the expected value of the nutrient density and the ratio of the means. The z-axes represent the square of the correlation between the nutrient density and the total intake.

The figures demonstrate the following:

1) The variance of the nutrient density doesn't appear to cause the shape of the function to change; however, it does seem to change the scale of the difference between $E(U)$ and the ratio of the means.
2) The range of values of the mean $E(U)$ for which the correlation ρ_{UT}^2 is close to 0 or smaller than the correlation ρ_{NT}^2 is considerably narrower when the ratio between the variance is smaller. (Compare figures 1a and 1b and figures 1c and 1d.)

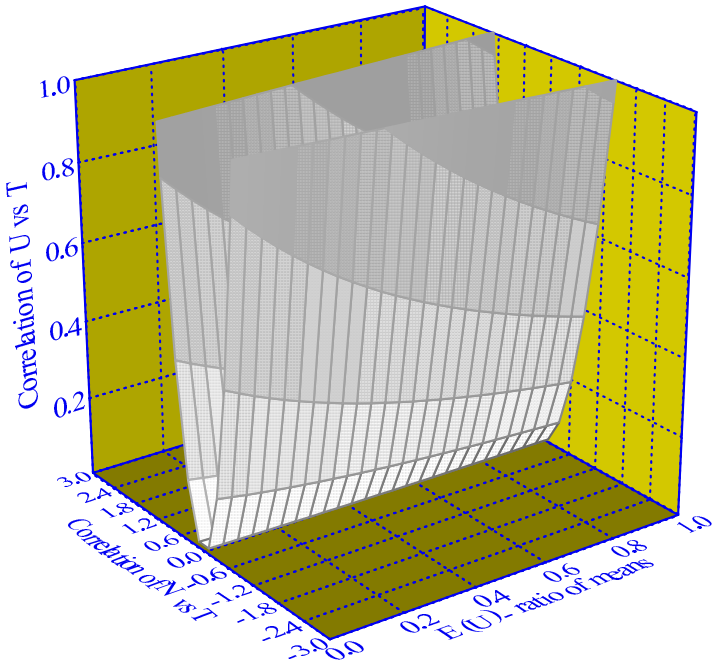
3) There is a quadratic drop in the magnitude of reduction in ρ_{UT}^2 as ρ_{NT}^2 increases (figures 1a-1d). This pattern becomes more obvious when the difference between the expected value of U and the ratio of the means of N and T is farther from 0. (Also see figures 2a-2d.)

4) The reduction in ρ_{UT}^2 compared to ρ_{NT}^2 is accomplished only for larger values of ρ_{NT}^2 (figures 1a-1d). As the difference between the expected value of U and the ratio of the means of N and T is farther from 0 the range of values of ρ_{NT}^2 for which the reduction is accomplished becomes smaller.

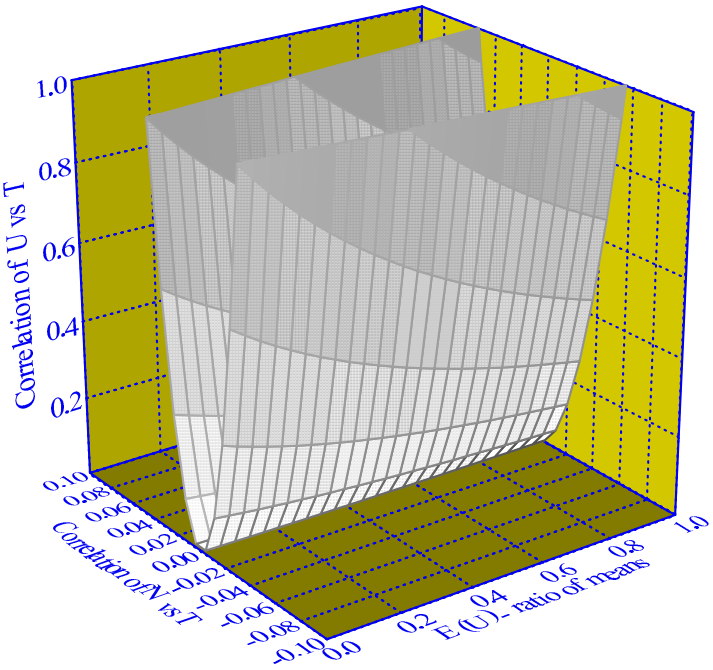
To further illustrate the points in 3 and 4, ρ_{UT}^2 was plotted against ρ_{NT}^2 for specific values of the difference between the expected value of U and the ratio of the means of N and T. These plots are shown in Figures 2a) and 2d). Since the variance of the specific nutrient wasn't found to alter shape (other than changing the scale) these plots were produced only for $\sigma_N^2 = 100$. These plots clearly indicate that ρ_{UT}^2 is closer to 0 for a broad range of values of ρ_{NT}^2 when the difference between the nutrient density and the ratio of the means is smaller (Figures 2b) and 2d)) than when this distance is larger (Figures 2a) and 2c)). Similarly, the reduction is better achieved for macronutrients (Figures 2a) and 2b)) compared to micronutrients (Figures 2c) and 2d)).

Figure 1. Region where the reduction in ρ_{UT}^2 is achieved

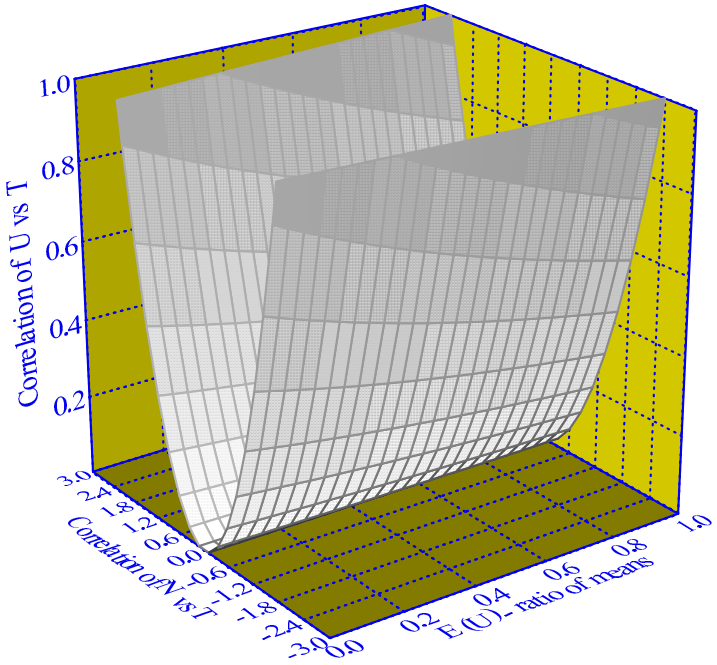
a) $\sigma_R^2 / \sigma_N^2 = 10,\ \sigma_N = 10$



b) $\sigma_R^2 / \sigma_N^2 = 10,\ \sigma_N = 250$



c) $\sigma_R^2 / \sigma_N^2 = 20,\ \sigma_N = 10$



d) $\sigma_R^2 / \sigma_N^2 = 20,\ \sigma_N = 250$

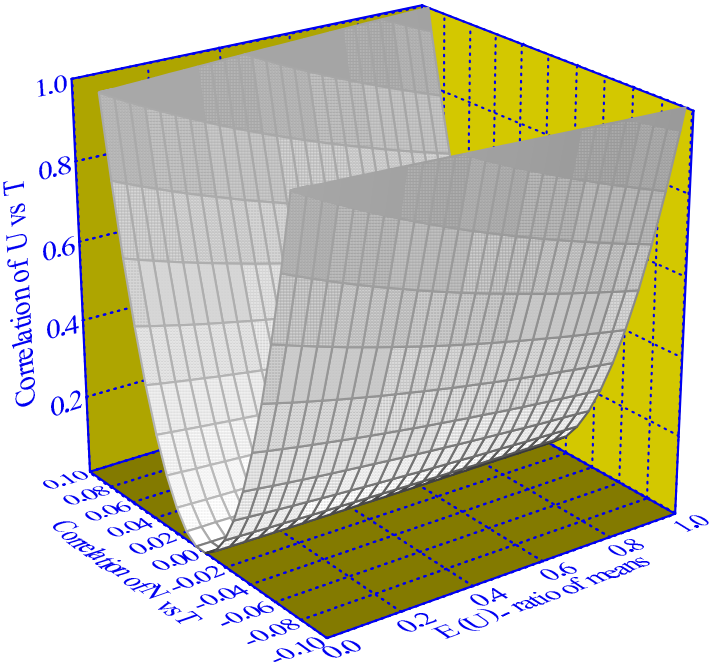
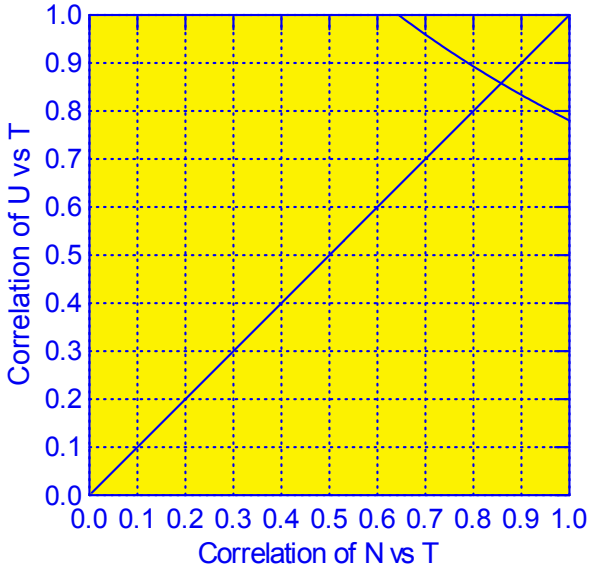
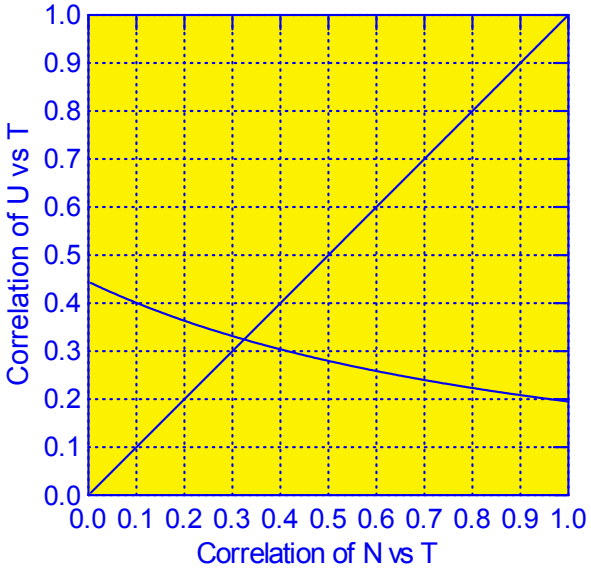


Figure 2. Reduction in ρ_{UT}^2 for specific values of the difference $[E(U) - \mu_N / \mu_T]$.

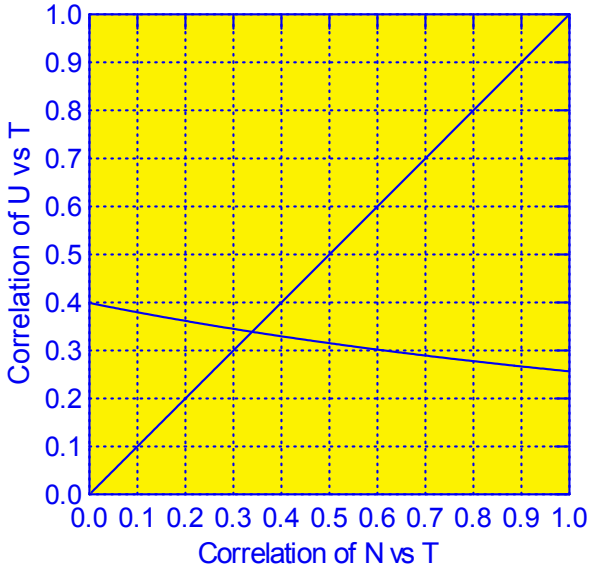
a) $E(U) - \mu_N / \mu_T = 1.2,\ \sigma_R^2 / \sigma_N^2 = 10$



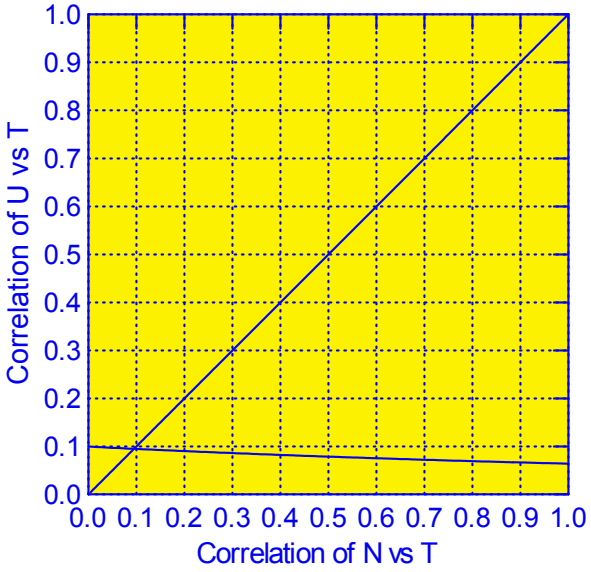
b) $E(U) - \mu_N / \mu_T = 0.6,\ \sigma_R^2 / \sigma_N^2 = 10$



c) $E(U) - \mu_N / \mu_T = 1.2,\ \sigma_R^2 / \sigma_N^2 = 20$



d) $E(U) - \mu_N / \mu_T = 0.6,\ \sigma_R^2 / \sigma_N^2 = 20$



3.2 Radiation Oncology

As introduced in chapter 1, in radiation oncology it may be of interest to study the error in targeting a tumor's centroid of two treatments, Brachytherapy and IGART. The geometric endpoint for the study hypothesis is formulated in terms of Σ. Hence, the variance in measurement is of interest. Also each treatment is applied to the same patient so the sample is paired which means that the variances are dependent. The issue is the method of analysis for the ratio Σ_A/Σ_B. Since the measurements, $\bar{x}_{j,A}^2$ and $\bar{x}_{j,B}^2$ are based on normal random variables these squared transformations would be gamma distributed (or some function of a chi-squared) (Casella and Berger 2002). In addition, since the two treatments are paired, they are correlated. We also can assume without loss of generality that the variance under treatment B is smaller than the variance under treatment A. Then, under the null hypothesis that $\Sigma_A / \Sigma_B = 1$, the joint distribution of $\bar{x}_{j,A}^2$ and $\bar{x}_{j,B}^2$ could be assumed to follow a bivariate gamma distribution with parameters $(\alpha_1, \alpha_2, 1)$. Consider now the transformation $w_j = \left(\bar{x}_{j,B}/\bar{x}_{j,A}\right)^2$. Then the variable w_j is equivalent in distribution to the variable $U = N/T$ in chapter 2. That is w_j will follow a beta distribution with parameters (α_1, α_2).

Here we assume that scale parameter $\beta = 1$ and that it is the same for the two random variables. This is true only under the null hypothesis. We will discuss the alternative hypothesis in Chapter 4. The mean and the variance of w_j are given by,

$E(w_j) = \alpha_1/\alpha_1 + \alpha_2$ and $V(w_j) = \alpha_1\alpha_2/(\alpha_1 + \alpha_2)^2(\alpha_1 + \alpha_2 + 1)$. Let $\bar{w} = \sum_{j=1}^{M} w_j \Big/ M$

where M is the number of subjects. Then, for large M by the central limit theorem,

$$Z = \sqrt{M}\left(\bar{w} - E(w_j)\right)\Big/\left(\sqrt{V(w_j)}\right)$$

is distributed as $N(0,1)$. This can be used to test the null hypothesis stated earlier.

4. Simulating a Bivariate Gamma Model

In this chapter we will provide a method for simulating bivariate gamma data and utilize this to study the asymptotic properties of a test statistic suggested for the radiation oncology application.

4.1 Simulating Bivariate Gamma Data

Generating bivariate Gamma random variables in the general case, where the $x < y$ constraint is not imposed, is complex. Several articles have appeared in the literature discussing the simulation of this general case (Arwini 2005). However in the special case where $x < y$ there seems to be an easier approach. We discuss this in this section. Using the gamma random number generator in SAS, RANGAM, two gamma random variables, X and Y, were generated simultaneously. Say, the distributions of X and W are respectively gamma with parameters (α, β) and (γ, β). Suppose for given $X = x$, the random variable, $Y = x + W$, is defined. Then, we claim the joint distribution of X and Y is a bivariate gamma. Consider the conditional density of Y given $X = x$. That is, the density of $Y = x + W$. Since W is a gamma (γ, β) and x is a constant by change of variable technique, the distribution of Y is

$$f_Y(y) = J \frac{\beta^\gamma}{\Gamma(\gamma)} (y - x)^{\gamma - 1} e^{-(y-x)\beta}, \tag{4.1}$$

where the Jacobian, J, is 1. Now the joint density of Y and X is given by,

$$f_{Y,X}(y, x) = f_{Y/X=x}(y) f_X(x).$$

We know, X is a gamma(α,β) random variable. Therefore, multiplying (4.1) by the probability density function of a gamma(α,β) yields,

$$f_{Y,X}(y,x)=\frac{\beta^{\gamma}}{\Gamma(\gamma)}(y-x)^{\gamma-1}e^{-\beta(y-x)}\frac{\Gamma(\alpha)}{\beta^{\alpha}}x^{\alpha-1}e^{-x\beta} \tag{4.2}$$

Simplifying yields,

$$f_{Y,X}(y,x)=\frac{\beta^{\alpha+\gamma}}{\Gamma(\alpha)\Gamma(\gamma)}x^{\alpha-1}(y-x)^{\gamma-1}e^{-\beta y}.$$

Notice that this is the bivariate gamma density as defined in equation (2.1). Therefore generating x and W and for each $X=x$ computing $Y=x+W$ gives a sample from a bivariate gamma distribution. We will use this method to simulate bivariate gamma data. Then we will test if this method produces the expected bivariate gamma random variables. We will compute the means, variances and covariance of the two random variables and compare it with the expected means, variances and covariance in terms of Bias and Mean Squared Error (MSE).

4.2 Simulation Parameters

The two shape parameters, α and γ, were set to (5,5) and (5,10). The location parameter, β, was set to 1/20 and 1/2.The sample size was fixed to be a 100 and simulations were repeated 1000 times. The expected means and variances for X and Y, the gamma variables, are as defined in equation (2.4) and the correlation between X and Y is defined in (3.18).

Once the samples were generated the following means were computed over the 1000 simulations: the mean of the means of X, the mean of the means of Y, the mean of the variances of X, the mean of the variances of Y, the mean of the correlations between X and Y, the mean of the variances of correlations between X and Y.

Using these, the mean squared error (MSE) and bias were calculated. The MSE measures the precision of an estimate W for a parameter θ. That is, the MSE of an estimator quantifies the amount by which an estimator differs from the true value of the quantity being estimated. The MSE is a function of θ defined by

$$E_\theta(W-\theta)^2 = Var_\theta W + (E_\theta W - \theta)^2 = Var_\theta W + (Bias_\theta W)^2,$$

where $Bias_\theta W$ is the bias of the estimator W and is defined as the difference between the expected value of W and θ. That is, $Bias_\theta W = E_\theta W - \theta$. If the bias for an estimator is 0 then that estimator is unbiased and satisfies $E_\theta W = \theta$ for all θ.

The known means and variances, μ_X, μ_Y, σ_X^2, and σ_Y^2, of X and Y were calculated for different values of the parameters $\{\alpha, \gamma \text{ and } \beta\}$. Table 1 displays these means and variances.

Table 1: Known Means and Variances

	μ_X	μ_Y	σ_X^2	σ_Y^2
$\alpha = 5, \gamma = 10, \beta = 0.05$	0.25	0.75	0.0125	0.0375
$\alpha = 5, \gamma = 5, \beta = 0.05$	0.25	0.5	0.0125	0.025
$\alpha = 5, \gamma = 10, \beta = 0.5$	2.5	7.5	1.25	3.75
$\alpha = 5, \gamma = 5, \beta = 0.5$	2.5	5	1.25	2.5

Once the statistics mentioned above were generated the known means and variances were used to calculate the Bias and MSE for each estimator. Plots were constructed comparing the $Bias^2$ and MSE for each of the statistics produced.
For plotting purposes, the negative log of the Bias and MSE were used. We will discuss them in relative terms.

The parameters $(\gamma = 5, \beta = 0.5, 0.05)$ and $(\gamma = 10, \beta = 0.5, 0.05)$ are symbolized by a blue diamond, a pink box, a yellow triangle, and a teal "x". The results are presented in table 2 and figures 3,4,5,6 and 7. We will discuss the results in the next section.
To study the asymptotic properties of Z based on the ratio of the gammas we performed a separate simulation study as described earlier. The results are presented in table 3 and figures 8, 9, 10, 11, 12, 13 and 14 . We will discuss the results in the next section.

Table 2: Bias^2 and Mean Squared Error of Sample Statistics

		β=.5		β=.05	
		$Bias^2$	*MSE*	$Bias^2$	*MSE*
α=5 γ=5	$\overline{x}$	0.01129	128.1	0.00011288	1.28099
	s_X^2	0.19547	241.66	0.000019547	2.4166
	$\overline{y}$	0.05798	520.68	0.00057979	0.05207
	s_Y^2	4.06632	1707.3	0.00040663	0.17073
	r_{XY}	0.05321	30.83471	0.05321	30.83471
α=5 γ=10	$\overline{x}$	0.000028162	127.62	2.8162E-07	1.2762
	s_X^2	0.35333	359.54	0.000035333	3.59537
	$\overline{y}$	0.0577	518.8	0.00057703	0.05188
	s_Y^2	1.4297	3846.4	0.00014297	0.38464
	r_{XY}	0.01385	51.29254	0.01385	51.29254
*for clarity of presentation all numbers have been multiplied by 10000					

Figure 3: Plot of the Bias^2 and Mean Squared Error for the Variance of *Y*

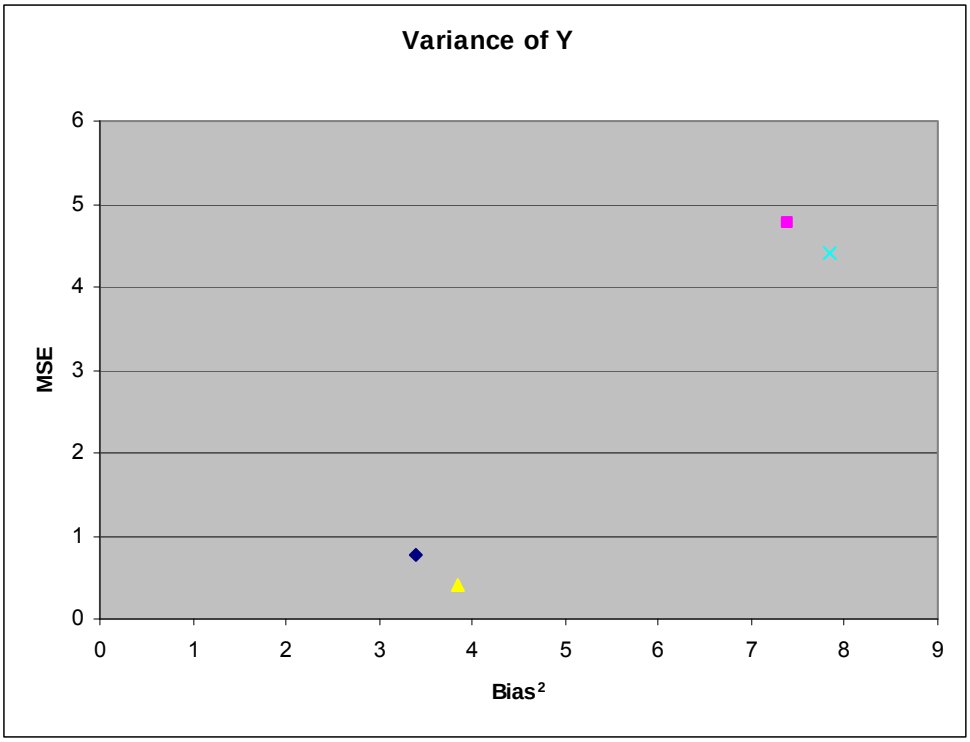


Figure 4: Plot of the Bias^2 and Mean Squared Error for the Variance of *X*

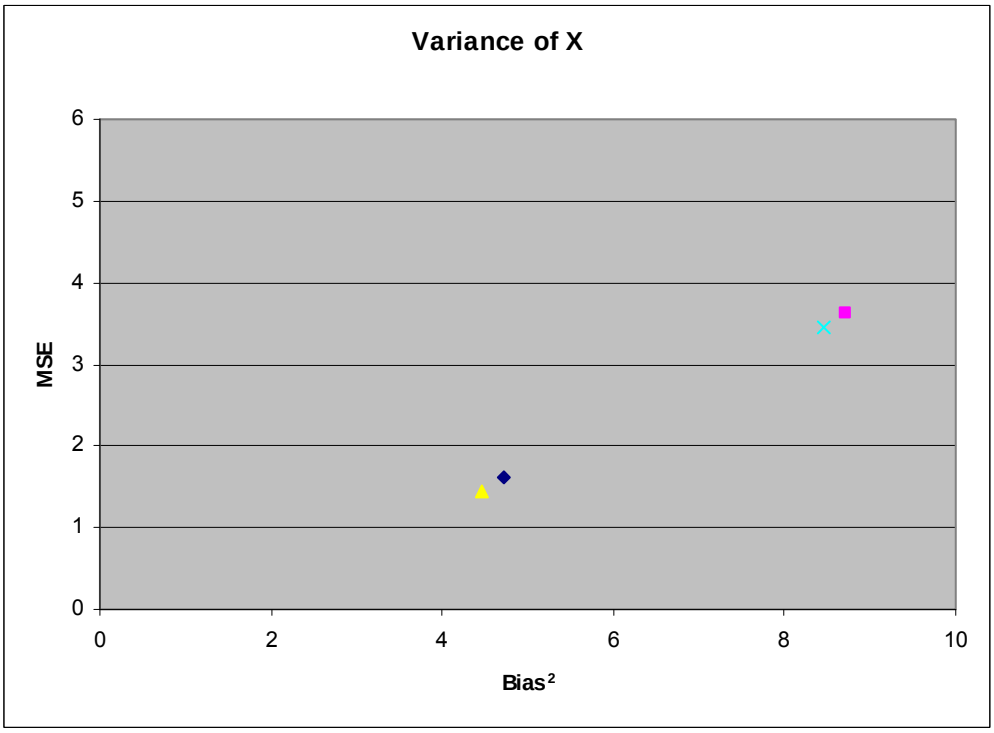


Figure 5: Plot of the Bias^2 and Mean Squared Error for the Mean of *X*

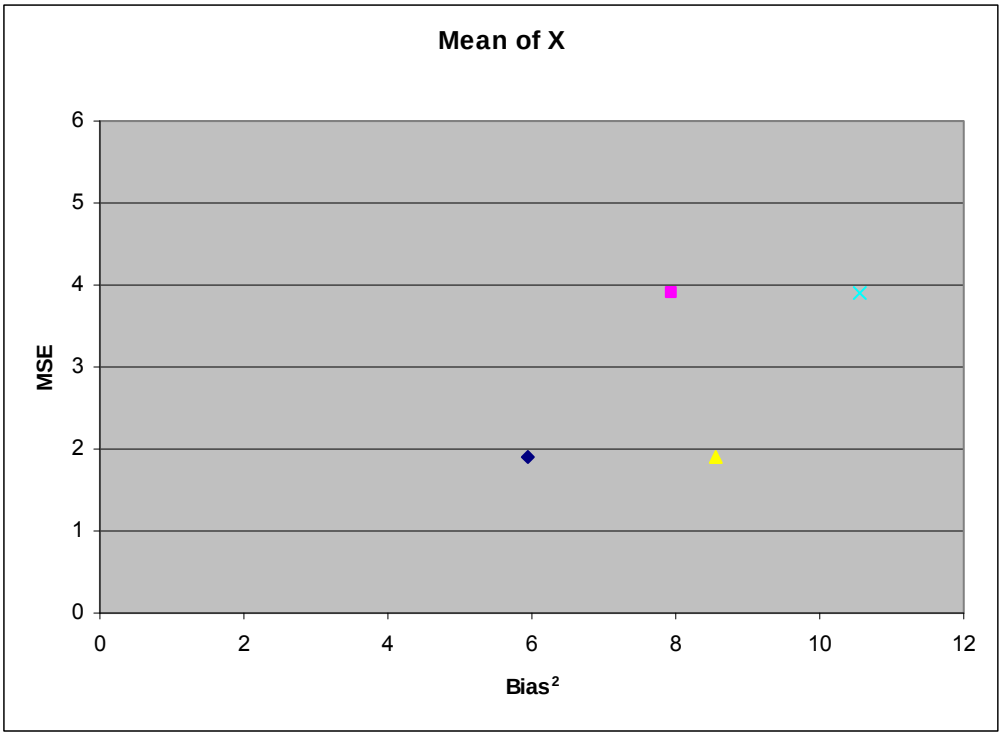


Figure 6: Plot of the Bias^2 and Mean Squared Error for the Mean of *Y*

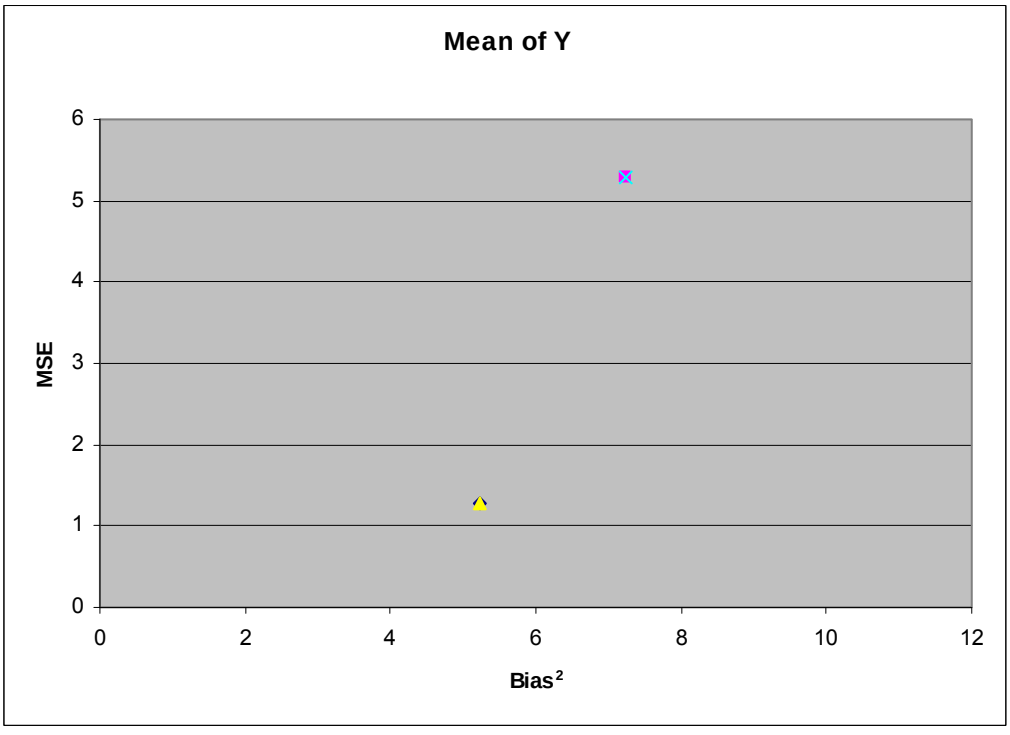


Figure 7: Plot of the Bias^2 and Mean Squared Error for the Correlation between *X* and *Y*

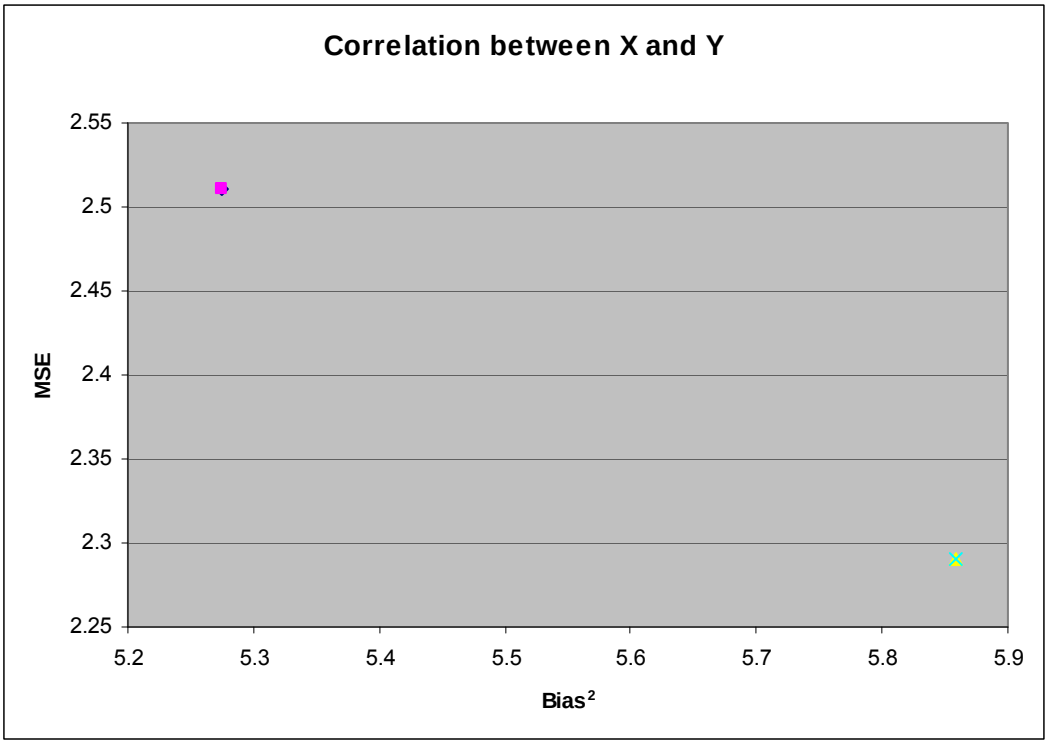


Table 3: Significance Level of Simulated Data in Percent

sample size	γ=5 $\beta = 0.05$	γ=10 $\beta = 0.05$
5	60	40
20	25	20
50	8	14
100	8	7
500	1.4	2
1000	0.7	0.7

Figure 8: Histogram of the Z values for a sample size of 5 with α=5 γ=5 β=0.05

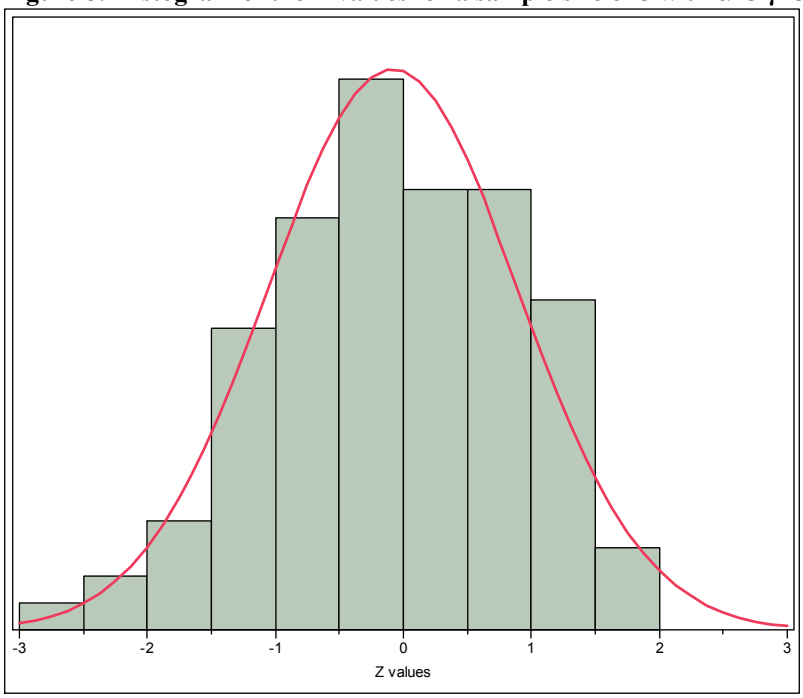


Figure 9: Histogram of the Z values for a sample size of 50 with α=5 γ=5 β=0.05

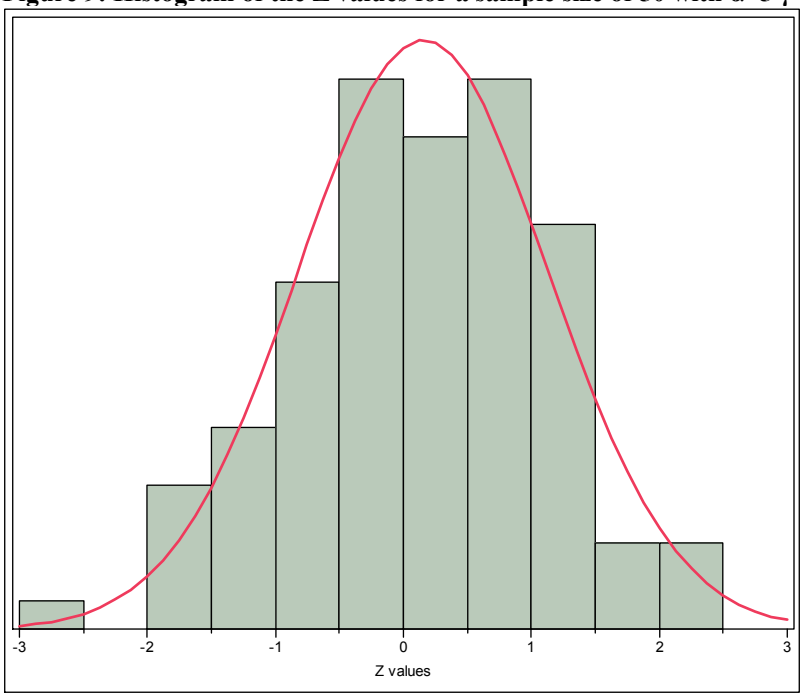


Figure 10: Histogram of the Z values for a sample size of 500 with α=5 γ=5 β=0.05

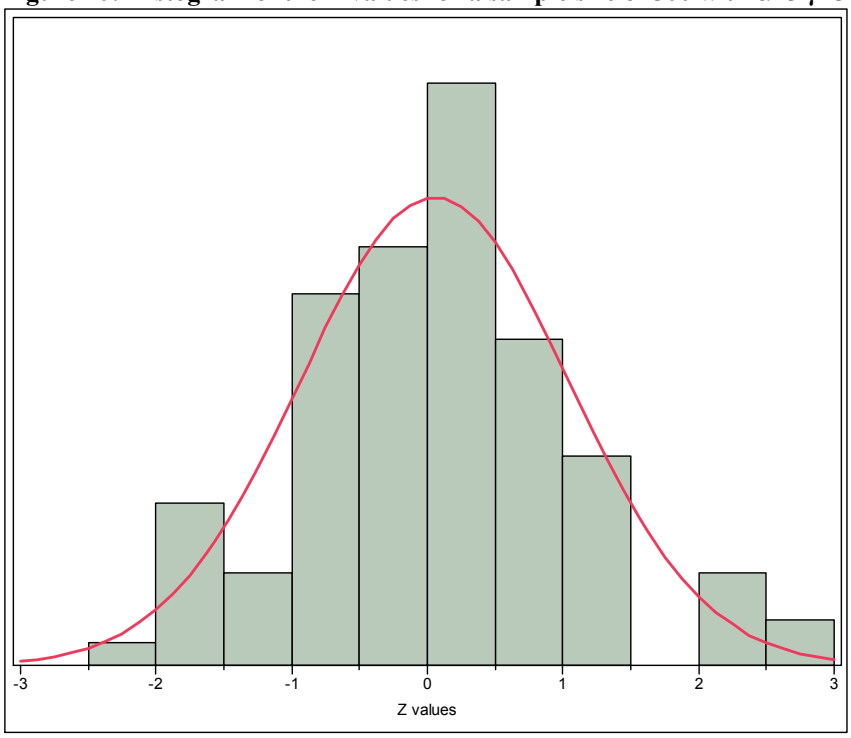


Figure 11: Histogram of the Z values for a sample size of 1000 with α=5 γ=5 β=0.05

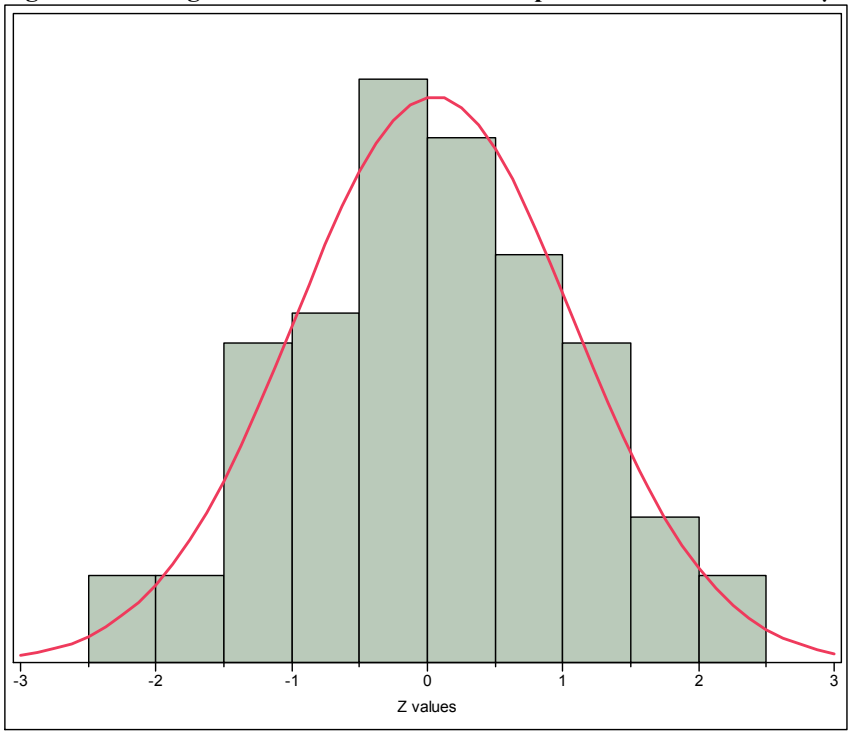


Figure 12: Histogram of the Z values for a sample size of 5 with α=5 γ=10 β=0.05

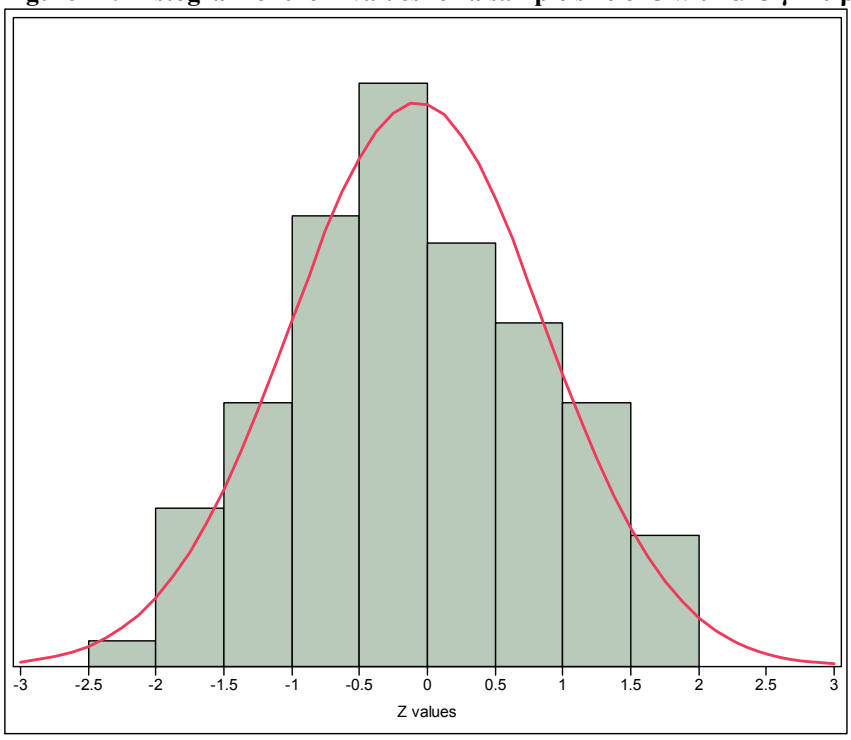


Figure 13: Histogram of the Z values for a sample size of 50 with α=5 γ=10 β=0.05

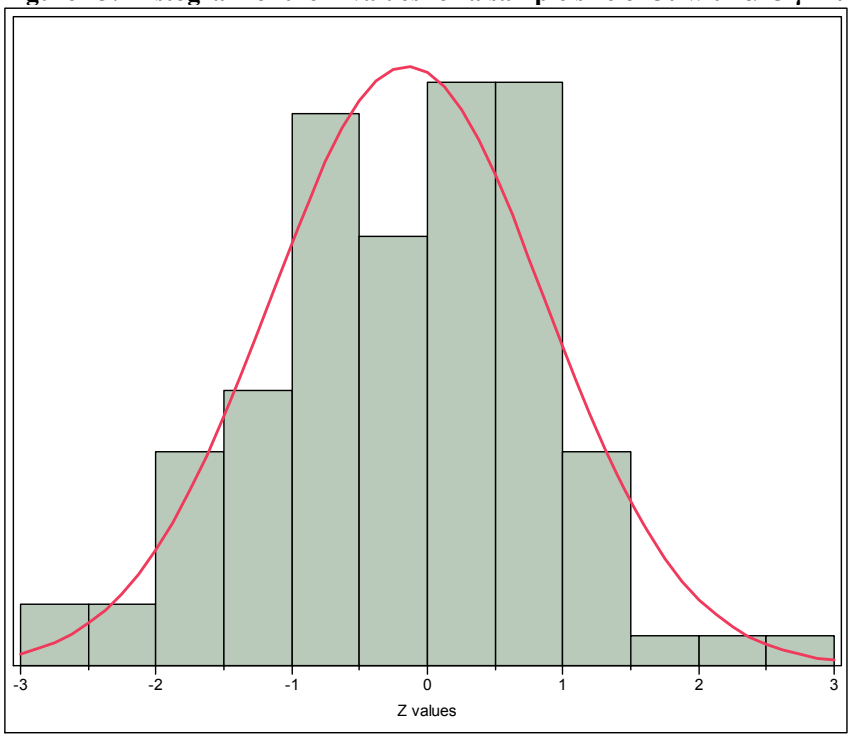
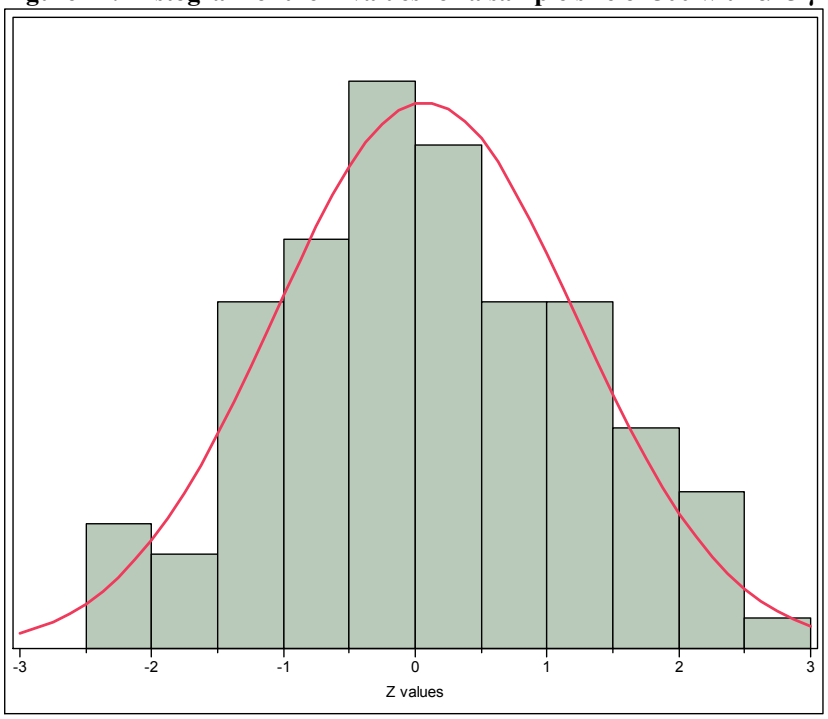


Figure 14: Histogram of the Z values for a sample size of 500 with α=5 γ=10 β=0.05



4.3 Discussion of the Simulation Results

From the table, in general, we could conclude that this method of generating gamma is appropriate. Notice that, all of the MSE's and Biases are quite small. The figures and the table are of the order 10^{-4}. We will discuss the graphs in relative terms. Examining figure 3 we see that for the variance of Y the Bias and MSE are sensitive to changes in β, the location parameter. For $\beta = 0.5$ the points $\gamma = 5$ and $\gamma = 10$ are similar and for $\beta = 0.05$ the points for $\gamma = 5$ and $\gamma = 10$ are similar. Increasing the location parameter by ten fold also increases the MSE by ten fold. The Bias for $(\gamma = 5,10$ and $\beta = 0.05)$ is twice that of $(\gamma = 5,10$ and $\beta = 0.5)$ and the MSE for $(\gamma = 5,10$ and $\beta = 0.05)$ is ten times that of $(\gamma = 5,10$ and $\beta = 0.5)$. However, for the same location parameter doubling the shape parameter γ increases the Bias by 0.45 while the MSE is decreased by 0.35.

From figure 5 we see that for the mean of X, Bias and MSE are equally sensitive to changes in β and γ. For $\beta = 0.5$ the points $\gamma = 5$ and $\gamma = 10$ have similar MSE values but differ in Bias by 2.6. For $\beta = 0.05$ the points $\gamma = 5$ and $\gamma = 10$ have similar MSE values but differ in Bias by 2.6. For the same shape parameter increasing the location parameter by 10 fold decreases the Bias and MSE by 2.

From figure 4 we see that with the same shape parameter for the variance of X the Bias is twice as sensitive as the MSE. For the same location parameter the Bias and MSE have similar values with the larger shape parameter having smaller values for each.

Looking at figure 6, the mean of Y, we that for the same location parameter doubling the shape parameter does not change the Bias or MSE by more than 2 thousandths. For the same shape parameter increasing the location parameter by 10 fold decreases the Bias by 2 and the MSE by 4.

Figure 7, the correlation between X and Y, shows us that for the same shape parameter changing the location parameter has no effect on the Bias and MSE. For the same location parameter doubling the shape parameter increases the Bias by 0.58, but decreases the MSE by 0.22.

Regarding the asymptotic properties of Z, changes in the location parameter should not affect the significance level for either shape parameter as the null distribution for the ratio (3.8) proposed to test these hypotheses $w_j(\Sigma_A/\Sigma_B)^2$ does not depend on β. So we fixed $\beta = 0.05$ without loss of generality. For both values of γ as sample size increased the significance level decreased. The samples of 500 and 1000 were under 0.025. Comparing figures 8 and 12 we can see that an increase in the shape parameter creates a distribution with thicker tails. Comparing figures 10 and 15 we see that increasing the shape parameter produces a distribution with thicker tails and a wider peak.

4.4 Limitations and Future problems

To compute the sample size needed and/or to compute power the non-null distribution of beta is needed. No conceptual motivation from a bivariate gamma immediately follows.

Literature Cited

Literature Cited

Arwini, Khadiga, Dobson C.T.J, Felipussi S Scharcanski . Comparison of distance measures between bivariate gamma processes. Preprint (2005).

Backstrand Jeffrey K. (2003), Quantitative Approaches to Nutrient Density for Public Health Nutrition. *Public Health Nutrition,* 6(8)

Casella, G., and Berger, R. L., (2002), *Statistical Inference*, Pacific Grove, California, Duxbury

Hegsted, D.M. (1985), Dietary Standard: Dietary Planning and Nutrition Education, *Clinical Nutrition,* 26, 757-768.

Jain, M., Cook, G. M. Davis, F. G., Grace, M. G., Howe, G. R. and Miller, A. B. (1980), A Case-Control Study of Diet and Colorectal Cancer, *International Journal of Cancer,* 26, 757-768.

Myers, R. H., (1990), *Classical and Modern Regression with Applications,* Pacific Grove, California, Duxbury

Nadarajah, S., and Gupta, A. K. (2006), Some Bivariate Gamma Distributions. *Applied Mathematics Letters,* 19, pgs 767-774

Palmgren J. (1993), Controlling for Total Energy Intake in Regression Models for Assessing Macronutrient Effects on Disease. *European Journal of Clinical Nutrition,* 47, S46-S50.

Rohatgi, V. K., (1976), *An Introduction to Probability Theory and Mathematical Statistics,* New York, John Wiley

Song, W., Schaly, B., Bauman, G., Battista, J., Van Dyk, J., (2005), Image-Guided Adaptive Radiation Therapy (IGART): Radiobiological and Dose Escalation Considerations for Localized Carcinoma of the Prostate, *Medical Physics,* 32(7), 2193-2203

Smith, K. R., Slattery, M. L., and French, T. K. (1991), Collinear Nutrients and the Risk of Colon Cancer, *Journal of Clinical Epidemiology,* 44, 715-723.

Willett, W. (1990), Total Energy Intake and Nutrient Composition: Dietary Recommendations for Epidemiologist, *International Journal of Cancer,* 46, 770-771.

Willett, W., and Stampfer, M. J. (1986), Total Energy Intake: Implications for Epidemiologic Analyses, *American Journal of Epidemiology,* 124, 17-27.

Willett, W., (1998), *Nutritional Epidemiology,* New York, New York, Oxford University Press

Yan, Di., (2008), Developing Quality Assurance Processes for Inage-Guided Adaptive Radiation Therapy, International Journal of Radiation Oncology Biological Physics, 78, S28-S32.

Appendix A

```
libname simulate 'C:\Documents and Settings\barkerjk\Desktop\jolene';

/* Simulation One */
data simulate.sample; *generating samples;

  Do i=1 to 1000;
     do j=1 to 100;
         x=.05*RANGAM(123456789, 5);   /* gamma with shape 5 & scale 20
*/
         w=.05*RANGAM(987654321, 5);  /* gamma with shape 10 & scale 20
*/
             y=x+w;
         OUTPUT;
     end;
  END;

run;

data sample1; *tagging samples 1-100;
merge simulate.sample;
if (i=j or 1<=j<=100) then set=i;
else set=0;
run;

PROC MEANS MEAN VAR noprint; *output dataset of means and variances;
     VAR x y;
     BY set;
       output out= means MEAN(x y)= meanx meany VAR(x y)=varx vary;
run;

proc means data=means Mean VAR noprint;*find the mean of the means and
the mean of the variances;
var meanx meany varx vary;
output out=mmeans MEAN (meanx meany varx vary)=mmeanx mmeany mvarx
mvary
Var(meanx meany varx vary)=varmeanx varmeany vvarx vvary;
run;
proc print data=mmeans;run;

* BIAS of estimate=estimate-actual value;
* MSE=Var(estimate)-(Bias*Bias);

data est;
set mmeans;
alpha=5;gamma=5;beta=.05;
```

```
biasx=mmeanx-alpha*beta;
biasx2=biasx*biasx;
biasvarx=mvarx-alpha*(beta*beta);
biasvarx2=biasvarx*biasvarx;
biasy=mmeany-(alpha+gamma)*beta;
biasy2=biasy*biasy;
biasvary=mvary-((alpha+gamma)*(beta*beta));
biasvary2=biasvary*biasvary;
msex=varmeanx+biasx2;
msey=varmeany+biasy2;
msevarx=vvarx+biasvarx2;
msevary=vvary+biasvary2;
run;
proc print data=est;run;

PROC CORR data=sample1 outp=b noprint; *finding the correlations and
creating a dateset for them;
     VAR x;
       with y;
     BY set;
run;

data new; *cleaning the dataset;
set b;
if _type_='CORR';
drop _type_;
run;

proc means data=new Mean VAR noprint; *calculating the mean of the
correlations;
var x;
output out=mcorr mean (x)=mmeancorrxy var (x)=meanvarcorr;
run;

data est2;
set mcorr;
alpha=5;gamma=5;beta=.05;
biascorr=mmeancorrxy-sqrt(alpha/(alpha+gamma));
biascorr2=biascorr*biascorr;
msecorr=meanvarcorr+biascorr2;
run;
proc print data=est2;run;

/* Simulation Two */

data  simulate.sample2; *generating samples;
m1= 5;
m2= 10;
meanb=m1/(m1+m2);
varb=m1*m2/(((m1+m2)** 2 )*(m1+m2+ 1 ));
flag= 0;
Do i=1 to 100;
  Do j=1 to 500; *sample size;
```

```
        x= .05*RANGAM( 123456789 , 5 );   /* gamma with shape 5 & scale
20 */
        w= .05*RANGAM( 987654321 , 10 );  /* gamma with shape 10 &
scale 20 */
        y=x+w;
        v=x/y;
        z=SQRT( 1/ 500 )*(v-meanb)/sqrt(varb);
        if z>= 1.96 OR z <=- 1.96 then flag= 1;
        else flag= 0;
        OUTPUT;
  END ;
END ;
Run ;
run ;


data sample3; *tagging samples where 1<= j <= n ;
merge simulate.sample2;
if (i=j or 1 <=j<= 500 ) then set=i;
else set= 0;
run ;
proc sort data=sample3 out =sample3; by set; run;
PROC MEANS MEAN VAR NOPRINT data=sample3; *output dataset with a mean
and variance for each of the 100 sets;
    BY set;
    output out = sum1 sum (z)= sumz VAR (z)= varz mean(z)= meanz;
run ;
proc univariate data=sum1 plots; var sumz; run;
data simulate.flaga;
set sum1;
if sumz> 1.96 or sumz<- 1.96 then flag= 1;
else flag= 0;
run ;
proc freq data=simulate.flaga; tables flag; run;

proc univariate data = simulate.flaga noprint;
*title "Histogram for the Summation of Z with n=1000 gamma=10
Beta=0.05";
histogram sumz /cfill=ligr normal cframe=liy barwidth=8 cv=black;
inset mean std max min;
run;
*title;
```

VITA

Jolene Barker was born in Williamsburg, Virginia. She was raised between Williamsburg and Southwest Virginia. After high school she attended Radford University where she majored in Physics and minored in Mathematics. While at Radford University she was a research intern for the Department of Nuclear Research and Engineering at Georgia Tech and for the Applied Biosciences Center at Virginia Tech. Jolene is currently a research biostatistician for the Dickson Institute at Carolinas Medical Center in Charlotte, North Carolina. She is the biostatistician for the Department of Obstetrics and Gynecology at CMC and she is the data analyst for LiveWELL, Carolinas!, an employee program that partners with the community to promote preventive health and encourage healthy choices.