國立交通大學

統計學研究所

碩 士 論 文

研 究 生：李忠庭

指導教授：彭南夫 博士

中 華 民 國 九 十 三 年 六

一個機率函數之估計法

AN APPROXIMATION METHOD OF A PROBABILITY FUNCTION

研 究 生：李忠庭　　Student：Chung -Ting Lee

指導教授：彭南夫 博士　　Advisor：Dr. Nan-Fu Peng

國 立 交 通 大 學

統計學研究所

碩 士 論 文

A Thesis

Submitted to Institute Statistics

College of Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Statistics

June 2004

Hsin-chu, Taiwan, Republic of China

中華民國九十三年六月

一個機率函數之估計法

研究生：李忠庭　　　　　　　　　指導教授：彭南夫 博士

國立交通大學統計學研究所

摘　　　要

本論文是探討估計有關分配相同且相互獨立隨機變數之和的分佈函數，對於沒有已知或公式解的分佈函數，我們將提供一種簡單的三次多項式估計法，使得對正的未知隨機變數和有準確的估計。並有多次模擬比較的結果來做驗証。

誌謝

在統計所兩年的時間裡，除了得到課業上的專業知識外，也學到許多電腦方面的技能，有這些收穫，最要感謝我的指導教授 彭南夫老師認真負責的指導，帶領著我順利完成此篇論文。

再來，我要感謝我的父母，由於他們的支持，讓我能毫無憂慮的追求我的學業，並且當我遇到困難時，也會不斷的鼓勵我，給予我解決問題的勇氣。我也要感謝我的女朋友婉玥，有了她的陪伴與細心照料，使我信心百倍，讓我渡過了兩年愉快的學習生涯。

此外，我還要特別感謝我的研究所同學文祥對我在學業上的幫助，還有其他同學欣妤、怡均、政輝、志浩、超毅、慶富、巧慧、淑珍、宇青、崢珮及好友彥廷、宗憲帶給我許多歡樂，豐富了我的生活。最後，將此論文獻給我的師長、家人、朋友、及所有有緣相聚的人。

目錄

第一章、簡介

1.1 研究動機

在此篇論文裡，主要要探討的是估計有關分配相同且相互獨立（*identically distributed and independent*）隨機變數（*random variable*）之和的分佈函數。假設我們在更新過程（*renewal process*）中的到達時距（*inter arrival*）是來自指數分佈（*Exponential distribution*）時，就可利用分配相同且相互獨立的指數分佈相加形成珈瑪分佈（*Gamma distribution*）來得到我們想要的結果。但是有許多分配之和並無像指數分佈相加時的良好性質，例如:韋伯分配(*Weibull distribution*)、貝他分配(*Bata distribution)*……等。以韋伯分配為例: 當到達時距 $X_1, X_2, \ldots, X_n$ 為 n 個分配相同且獨立來自 $Weibull(c,\beta)$ 分佈的隨機變數，(其機率密度函數為 $f(x)=\frac{c}{\beta}x^{c-1}e^{-\frac{x^c}{\beta}}$)，則我們有興趣的機率值就是

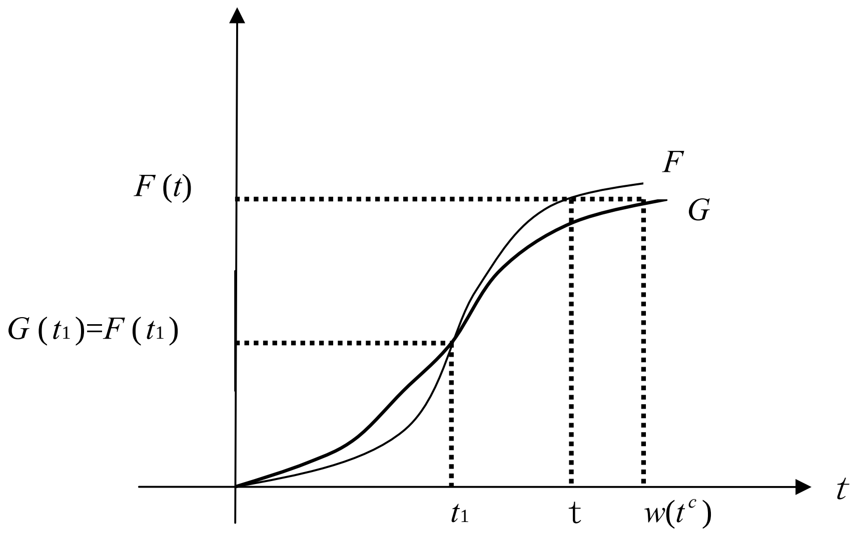
$$P(S_n \le t) \text{，其中 } S_n = X_1 + \ldots + X_n \text{。}$$

而此機率值並無已知的公式或解法。在本篇論文中我們將提供一種較簡單以及擁有不錯準確度的逼近方法並配合附錄的參數表來估計我們感興趣的機率值，並期待這樣的方法能讓工業界和各個領域在使用時，能夠以更快的計算速度和較低的困難度中，得到精確的結果。

第二章、三次多項式的近似方法

在這一章中，我們探討$F_n(t)=P(S_n \le t)$的估計問題，根據盧信銘*(2003)*提出了一個方法，只利用三個點（三個分位點（*quantile*））就可以得到非常近似的答案。我們是以函數方式（*functionwisely*）求得，而非以一個點一個點方式。我們的想法與作法如下：

如果X_i為來自某一獨立同分配的隨機變數，$i=1,\ldots,n$，且為正的隨機變數（其值皆為正數），Y_i為來自另一獨立同分配的隨機變數，$i=1,\ldots,n$，亦為正的隨機變數，且X_i、Y_i經由變數變換具有 $Y_i=X_i^c$ 的關係，在此，我們希望用 $Y_1+\ldots+Y_n$ 的分佈函數來近似 $X_1+\ldots+X_n$ 的分佈函數。假設$X_1+\ldots+X_n$服從F(F為一個累積機率密度函數(*cdf*))，$Y_1+\ldots+Y_n$服從G(G為另一個累積機率密度函數)，因為X_i、Y_i皆為正的隨機變數，所以$F(0)=0$、$G(0)=0$，其圖形畫出如下:



註:

$t_1=w(t_1^c)$

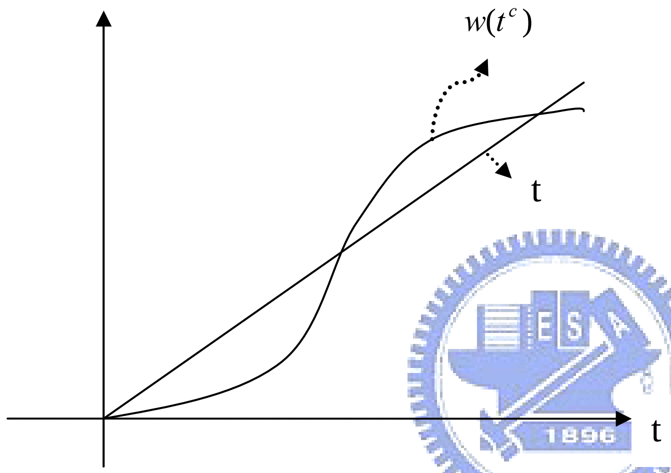
$t<w(t^c)$　，$t>t_1$

$t>w(t^c)$　，$t<t_1$

由圖可看出此為一單峰之隨機變數，因此我們以下只對具單峰的隨機變數做推論；因為這兩個分佈函數都是連續可微分（ *continuously differentiable* ）的函數，所以對於每一個 $t \geq 0$，存在 $w(t)$，使得

$$P(X_1 + \ldots + X_n \leq t) = P(X_1{}^c + \ldots + X_n{}^c \leq w(t^c)) = P(Y_1 + \ldots + Y_n \leq w(t^c))$$ 。

同時 $w(t^c)$ 是連續可微分，以及 $w(0) = 0$ 。我們大致畫出 t 與 $w(t^c)$ 的關係圖:



也因為 $w(0) = 0$ ，在本文中我們所考慮的 $w(t^c)$ 之形式為 $\alpha t^{3c} + \gamma t^{2c} + \tau t^c$（註：也可考慮 $\alpha t^{2c} + \gamma t^c$ 之形式，但其結果並不好。)。若前面圖中 F 與 G 之角色互換，則結果相反，但不論如何，我們猜想 $w(t)$ 可用三次方的形式來做估計；其中三個未知的參數，我們利用 $X_1 + \ldots + X_n$ 及 $Y_1 + \ldots + Y_n$ 的 0.1、0.5 以及 0.9 這三個分位點來決定。首先，我們令 t_p 滿足

$$P(X_1 + \ldots + X_n \leq t_p) = p$$ ，

並且 θ_p 滿足

$$P(Y_1 + \ldots + Y_n \leq \theta_p) = p$$ ，

因此，t_p與θ_p分別為$X_1+\ldots+X_n$與$Y_1+\ldots+Y_n$的第p分位點。當$p=0.1,\ 0.5,\ 0.9$時，我們利用$\alpha\, t_p^{3c}+\gamma\, t_p^{2c}+\tau\, t_p^{c}=\theta_p$得到三個線性方程式，而這三個方程式可求得此三個未知參數（α, γ, τ）。求的此三個位知數後，我們即可得到$\hat{w}(t^c)=\alpha\, t^{3c}+\gamma\, t^{2c}+\tau\, t^{c}$，亦即$P(X_1+\ldots+X_n\le t)\cong P(Y_1+\ldots+Y_n\le \hat{w}(t^c))$。通常，我們要求$Y_1+\ldots+Y_n$的累積機率密度函數為已知，這個的好處是：如果得知（α, γ, τ），則$X_1+\ldots+X_n$的累積機率密度函數之近似函數，可以很容易求得。我們發現，這種優良近似的現象在各式正的隨機變數中都會發生。

第三章 各種分配之實際模擬

根據第二章所述之理論，我們利用多種不同的分配，加以實際模擬如下:

3.1 估計韋伯分配隨機變數和的研究方法

3.1.1 估計方法

假設 X_i 為來自 $Weibull(c,1)$ 分佈的隨機變數，則經由變數變換 $Y_i = X_i^c$ 可知 Y_i 為來自 $Exponential(1)$ 分佈的隨機變數，$i=1,...,n$ 。因此，我們希望用 $Y_1+...+Y_n$ 的分佈函數來近似 $X_1+...+X_n$ 的分佈函數，且 $Y_1+...+Y_n$ 為 $gamma(n,1)$分配。對於每一個 $t \geq 0$，我們希望可以找到 $w(t^c)$ ，使得

$$P(X_1+...+X_n \leq t) = P(Y_1+...+Y_n \leq w(t^c))$$ 。

同時 $w(t^c)$ 是連續可微分。因為 $w(0)=0$，在本文中我們所考慮的 $w(t^c)$ 之形式為 $\alpha t^{3c} + \gamma t^{2c} + \tau t^c$。其中三個未知參數，我們利用 $X_1+...+X_n$ 及 $Y_1+...+Y_n$ 的 0.1、0.5 以及 0.9 這三個分位點來決定。

3.1.2 模擬步驟:

我們把以上說明化為下列步驟：

（一）決定模擬樣本數 $N=10^7$ ，令 $T_1,T_2,...,T_N$ 為 N 個分配相同且獨立來自 $F_n(x) = P(X_1+...+X_n \leq x)$ 的隨機變數，再將其排序得到 $T_{(1)},T_{(2)},...,T_{(N)}$ ，並且令 $\hat{t}_p = T_{(Np)}$ 為 t_p 的估計值，其中 $p=$0.1, 0.5, 0.9。因為模擬樣本數 $N=10^7$ ，因此當 $p=$0.1, 0.5, 0.9 時，Np 都是正整數，所以 $T_{(Np)}$ 是

第 p 樣本分位點（ $p-th\ sample\ quantile$ ）

亦即 $\hat{t}_{0.1}=T_{(10^6)}$，$\hat{t}_{0.5}=T_{(5\times10^6)}$，$\hat{t}_{0.9}=T_{(9\times10^6)}$ 分別為 $t_{0.1}$，$t_{0.5}$，$t_{0.9}$ 的估計值。）

（二）以相同的方式估算出 θ_p，模擬樣本數 $N=10^7$，令 $Q_1,Q_2,...,Q_N$ 為 N 個分配相同且獨立來自 $F_n(y)=P(Y_1+...+Y_n\le y)$ 的隨機變數，再將其排序得到 $Q_{(1)},Q_{(2)},...,Q_{(N)}$，並且令 $\hat{\theta}_p=Q_{(Np)}$ 為 θ_p 的估計值，其中 p=0.1, 0.5, 0.9。經過排序後，找出 $\hat{\theta}_{0.1}=Q_{(10^6)}$，$\hat{\theta}_{0.5}=Q_{(5\times10^6)}$，$\hat{\theta}_{0.9}=Q_{(9\times10^6)}$ 分別為 $\theta_{0.1}$，$\theta_{0.5}$，$\theta_{0.9}$ 的估計值。

（三）我們由

$$\alpha\,\hat{t}_{0.1}^{3c}+\gamma\,\hat{t}_{0.1}^{2c}+\tau\,\hat{t}_{0.1}^{c}=\theta_{0.1}，$$

$$\alpha\,\hat{t}_{0.5}^{3c}+\gamma\,\hat{t}_{0.5}^{2c}+\tau\,\hat{t}_{0.5}^{c}=\theta_{0.5}，$$

$$\alpha\,\hat{t}_{0.9}^{3c}+\gamma\,\hat{t}_{0.9}^{2c}+\tau\,\hat{t}_{0.9}^{c}=\theta_{0.9}，$$

解得 α,γ,τ。

（四）因此在固定某個 n 及形態參數 c 下，我們得到 $w(t^c)$ 的估計形式為

$\hat{w}(t^c)=\alpha\,t^{3c}+\gamma\,t^{2c}+\tau\,t^{c}$。

（五）所以我們有興趣的機率值，可寫成下面的近似方程式:

$P(X_1+...+X_n\le t)\cong P(Y_1+...+Y_n\le\hat{w}(t^c))=p$。

3.1.3 模擬結果

在本實驗中，我們做 $n=2,3,...,10$， 以及 c=1, 2, …9 的例子，而在附錄表(一)中只列出 c=1，n=5 及 3 的結果。附錄表一列出了相對誤差值，(定義為

$\dfrac{|P_1 - \hat{P}_1|}{P_1}$，其中 $P(X_1+...+X_n \le t) = P_1$， $P(Y_1+...+Y_n \le \hat{w}(t^c) = \hat{P}_1$)。而我們從表一中可發現，由最大相對誤差的觀點，當 n=3，其相對誤差最大在 t=1 時的 0.0010105，當 n=5 時， t=2 時的 0.003474，其餘的表現都很好，且在真實機率值 P_1 越大時，誤差有越來越小的跡象。而未列出的 c 值結果亦同。

3.2 估計指數分配隨機變數和的研究方法

在前一節中，我們用指數分配來近似韋伯分配得到不錯的結果，在此節中，我們反過來以韋伯分配近似指數分配。

3.2.1 估計方法

如果 X_i 為來自參數為 1 的指數分配($Exponential(1)$)隨機變數，Y_i 為來自具有參數(c 、1)的韋伯分佈($Weibull(c,1)$)隨機變數，則經由變數變換 $Y_i = X_i^{1/c}$ ，$i = 1,...,n$ 。因此對於每一個 $t \ge 0$ ，我們希望找到 $w(t)$ ，使得

$$P(X_1+...+X_n \le t) = P(Y_1+...+Y_n \le w(t^{1/c}))\text{。}$$

因為 $w(0)=0$ ，在此我們假設 $w(t^{1/c})$ 之形式為 $\alpha t^{3/c} + \gamma t^{2/c} + \tau t^{1/c}$ 。其中三個未知參數，我們利用 $X_1+...+X_n$ 及 $Y_1+...+Y_n$ 的 0.1、0.5 以及 0.9 這三個分位點來決定。

3.2.2 模擬步驟:

我們一樣地做下列步驟：

（一）決定模擬樣本數 $N=10^7$，令 $T_1,T_2,...,T_N$ 為 N 個分配相同且獨立來自 $F_n(x)=P(X_1+...+X_n \le x)$ 的隨機變數，再將其排序得到 $T_{(1)},T_{(2)},...,T_{(N)}$，並且令 $\hat{t}_p=T_{(Np)}$ 為 t_p 的估計值，其中 p=0.1, 0.5, 0.9，$T_{(Np)}$ 是第 p 樣本分位點（$p-th$ $sample$ $quantile$）

亦即 $\hat{t}_{0.1}=T_{(10^6)}$，$\hat{t}_{0.5}=T_{(5\times10^6)}$，$\hat{t}_{0.9}=T_{(9\times10^6)}$ 分別為 $t_{0.1}$，$t_{0.5}$，$t_{0.9}$ 的估計值。

（二）以相同的方式估算出 θ_p，模擬樣本數 $N=10^7$，令 $Q_1,Q_2,...,Q_N$ 為 N 個分配相同且獨立來自 $F_n(y)=P(Y_1+...+Y_n \le y)$ 的隨機變數，再將其排序得到 $Q_{(1)},Q_{(2)},...,Q_{(N)}$，並且令 $\hat{\theta}_p=Q_{(Np)}$ 為 θ_p 的估計值，其中 p=0.1, 0.5, 0.9。經過排序後，找出 $\hat{\theta}_{0.1}=Q_{(10^6)}$，$\hat{\theta}_{0.5}=Q_{(5\times10^6)}$，$\hat{\theta}_{0.9}=Q_{(9\times10^6)}$ 分別為 $\theta_{0.1}$，$\theta_{0.5}$，$\theta_{0.9}$ 的估計值。）

（三）我們由

$$\alpha\,\hat{t}_{0.1}^{3/c}+\gamma\,\hat{t}_{0.1}^{2/c}+\tau\,\hat{t}_{0.1}^{1/c}=\hat{\theta}_{0.1}\text{，}$$

$$\alpha\,\hat{t}_{0.5}^{3/c}+\gamma\,\hat{t}_{0.5}^{2/c}+\tau\,\hat{t}_{0.5}^{1/c}=\hat{\theta}_{0.5}\text{，}$$

$$\alpha\,\hat{t}_{0.9}^{3/c}+\gamma\,\hat{t}_{0.9}^{2/c}+\tau\,\hat{t}_{0.9}^{1/c}=\hat{\theta}_{0.9}\text{，}$$

解得 α,γ,τ。

（四）因此在固定個數 n 及參數 a、b 下，我們得到 $w(t^{1/c})$ 的估計形式為 $\hat{w}(t^{1/c})=\alpha\,t^{3/c}+\gamma\,t^{2/c}+\tau\,t^{1/c}$。

（五）所以我們有興趣的機率值，可寫成下面的近似方程式:

$P(X_1+...+X_n \le t)\cong P(Y_1+...+Y_n \le \hat{w}(t^{1/c}))=p$。

3.2.3 模擬結果

在本實驗中，我們做 $n = 2, 3, ..., 10$， 以及 c=1, 2, …9 的例子，而在附錄表(二)中只列出 c=1 及 5，n=5 的結果。附錄表一列出了相對誤差值，(定義為 $\frac{|P_1 - \hat{P}_1|}{P_1}$，其中 $P(X_1 + ... + X_n \le t) = P_1$，$P(Y_1 + ... + Y_n \le \hat{w}(t^{1/c}) = \hat{P}_2$)。而從表中可發現，由最大相對誤差的觀點，當 c=5，其相對誤差最大在 t=3 時的 0.0048237，當 c=1 時， t=2 時的 0.0040875，其餘的表現都很好，未列出的 c 值結果亦同。

3.3 非正隨機變數不適合此方法之探討

經過幾次試驗此三次多項式的估計方法皆得到不錯的結果之後，我們想要進一步的試驗不同性質的分配，3.1 節和 3.2 節中所用的分配皆為正的值，於是此節中，我們想要拿具有正值和負值的分配，而選擇的是具有良好性質的標準常態分配*(Normal(0,1))*。

3.3.1 估計方法

我們一樣探討 $F_n(t) = P(S_n \le t)$ 的估計問題，其中 $S_n = X_1 + ... + X_n$。$X_1, X_2, ..., X_n$ 為分配相同且相互獨立的標準常態分配*(Normal(0,1))*。 Y_i 為自具有自由度 1 的卡方分配*(chi-square distribution)*，此種組合剛好 $Y_i = X_i^2$，

$i=1,...,n$。因此，我們用 $Y_1+...+Y_n$ 的分佈函數來近似 $X_1+...+X_n$ 的分佈函數。這兩個分佈函數也是連續且可微分的函數，所以對於每一個 $t\geq 0$，我們希望找到 $w(t^2)$，使得

$$P(X_1+...+X_n\leq t)=P(Y_1+...+Y_n\leq w(t^2))$$ 。

同時 $w(t^2)$ 是連續且可微的函數，不同的是 $w(0)\neq 0$，所以我們所考慮的 $w(t^2)$ 形式不能為 $\alpha t^6+\gamma t^4+\tau t^2$，必須加上常數項 ρ 這樣的話即有四個未知參數，須取四個點來得到四個方程式，為了避免四個點不好取得，決定增為四次方程式，所以 $w(t^2)$ 形式為 $\alpha t^8+\gamma t^6+\tau t^4+\delta t^2+\rho$。其中有五個未知的參數 α、r、τ、δ、ρ，我們利用 $X_1+...+X_n$ 及 $Y_1+...+Y_n$ 的 0.1、0.3、0.5、0.7 以及 0.9 這五個分位點來決定。我們令 t_p 滿足:

$$P(X_1+...+X_n\leq t_p)=p$$ ，

並且 θ_p 滿足

$$P(Y_1+...+Y_n\leq \theta_p)=p$$

當 0.1、0.3、0.5、0.7、 0.9 時，我們利用 $\alpha t_p^8+\gamma t_p^6+\tau t_p^4+\delta t^2+\rho=\theta_p$ 得到五個線性方程式，再利用此五個方程式解五個未知參數: α、r、τ、δ、ρ。

3.3.2 模擬步驟:

再以相同的方式做模擬:

（一）模擬樣本數 $N=10^7$，令 $T_1,T_2,...,T_N$ 為 N 個分配相同且獨立來自

$F_n(x)=P(X_1+...+X_n\le x)$ 的隨機變數，再將其排序得到 $T_{(1)},T_{(2)},...,T_{(N)}$，

並且令 $\hat{t}_p=T_{(Np)}$ 為 t_p 的估計值。因為 $N=10^7$，當 0.1、0.3、0.5、0.7、

0.9 時，Np 皆為正整數，$T_{(Np)}$ 是第 p 樣本分位點。

（二）再以相同的方式估算出 θ_p。

（三）我們由

$$\alpha\,\hat{t}_{0.1}^8+\gamma\,\hat{t}_{0.1}^6+\tau\,\hat{t}_{0.1}^4+\delta t_{0.1}^2+\rho=\hat{\theta}_{0.1}$$

$$\alpha\,\hat{t}_{0.3}^8+\gamma\,\hat{t}_{0.3}^6+\tau\,\hat{t}_{0.3}^4+\delta t_{0.3}^2+\rho=\hat{\theta}_{0.3}$$

$$\alpha\,\hat{t}_{0.5}^8+\gamma\,\hat{t}_{0.5}^6+\tau\,\hat{t}_{0.5}^4+\delta t_{0.5}^2+\rho=\hat{\theta}_{0.5}$$

$$\alpha\,\hat{t}_{0.7}^8+\gamma\,\hat{t}_{0.7}^6+\tau\,\hat{t}_{0.7}^4+\delta t_{0.7}^2+\rho=\hat{\theta}_{0.7}$$

$$\alpha\,\hat{t}_{0.9}^8+\gamma\,\hat{t}_{0.9}^6+\tau\,\hat{t}_{0.9}^4+\delta t_{0.9}^2+\rho=\hat{\theta}_{0.9}$$

解得 α、r、τ、δ、ρ。

（四）因此在固定個數 n 之下，我們得到 $w(t^2)$ 的估計形式為 $\hat{w}(t^2)=\alpha\,t^8+\gamma\,t^6+\tau\,t^4+\delta t^2+\rho$。

（五）所以我們有興趣的機率值，可寫成下面的近似方程式:

$P(X_1+...+X_n\le t)\cong P(Y_1+...+Y_n\le\hat{w}(t^2))=p$。

3.3.3 模擬結果

在本實驗中，我們做 n=2，3，…，10， 的例子，而在我們的附錄表(三)中列出 $n=5$ 的結果。附錄表五中包含了五個參數 α、r、τ、δ、ρ 及 $w(t^2)$，另外也列出了相對誤差值，(定義為 $\dfrac{|\,P_1-\hat{P}_1\,|}{P_1}$，其中

$P(X_1+...+X_n \le t) = P_1$，$P(Y_1+...+Y_n \le \hat{w}(t^2) = \hat{P}_1$)。我們從表三中可發現，其誤差是相當大的，因為算出來之$w(t^2)$有可能為負值或是很大，在此次試驗中t=-2, -1, 0, 1 時，算出來之$w(t^2)$為負值，而卡方分配之值皆為正數，所以造成$\hat{P}_1$=0，得到的誤差為 1.0000；當 t=-3, 2, 3 時，因為$w(t^2)$太大，所以皆得到$\hat{P}_1$=1，因此誤差亦很大，所以可知此方法是失敗的，而失敗的原因是因為*Normal*(0, 1)非正的隨機變數，我們將在下一節中探討。

3.4 估計常態隨機變數絕對值之和

由 3.4 的試驗結果得知，把三次方估計法使用在常態分配上面並不適當，難道是有負值的關係嗎？為了了解此說法的正確性，在本節裡，我們特別把所有常態分配之值，皆加上絕對值，使它也是一個正的分配(卡方分配)對正的分配(常態分配加絕對值)做估算。

3.4.1 估計方法

在此$X_1, X_2, ..., X_n$為分配相同、相互獨立並加上絕對值的標準常態分配$X_i \overset{d}{=} |Z|$ i=1, 2⋯, n ，其中 Z 為服從 *Normal*(0, 1)分配的隨機變數。 Y_i為自具有自由度 1 的卡方分配*(chi-square distribution)*， 亦保有$Y_i = X_i^{\ 2}$，$i = 1,...,n$。因此，我們用 $Y_1+...+Y_n$ 的分佈函數來近似 $X_1+...+X_n$ 的分佈函數。這兩個分佈函數也是連續且可微分的函數， 同樣的，

$P(X_1+...+X_n \le t)=P(Y_1+...+Y_n \le w(t^2))$。我們想找出 $w(t^2)$，同時 $w(t^2)$ 是連續且可微的函數，為了和 3.3 節的一樣，我們設 $w(t^2)$ 形式為 $\alpha t^8+\gamma t^6+\tau t^4+\delta t^2+\rho$。其中有五個未知的參數 α、r、τ、δ、ρ，我們利用 $X_1+...+X_n$ 及 $Y_1+...+Y_n$ 的 0.1、0.3、0.5、0.7 以及 0.9 這五個分位點來決定。

我們令 t_p 滿足:

$$P(X_1+...+X_n \le t_p)=p，$$

並且 θ_p 滿足

$$P(Y_1+...+Y_n \le \theta_p)=p$$

當 0.1、0.3、0.5、0.7、 0.9 時，我們利用 $\alpha t_p^8+\gamma t_p^6+\tau t_p^4+\delta t^2+\rho=\theta_p$ 得到五個線性方程式，再利用此五個方程式解五個未知參數: α、r、τ、δ、ρ。因為 $w(0)=0$ 的關係，所以理論上得到的 ρ 值應該等於零。

3.4.2 模擬步驟:

對照 3.3.2 之方法。

3.4.3 模擬結果

在本實驗中，我們做 $n=2,3,...,10$，例子，而在附錄表四中列出包含了五個參數 α、r、τ、δ、ρ 及 $w(t^2)$，並列出了相對誤差值，(定義為 $\frac{|P_1-\hat{P}_1|}{P_1}$，其中 $P(X_1+...+X_n \le t)=P_1$，$P(Y_1+...+Y_n \le \hat{w}(t^2)=\hat{P}_1$)。由最大相對誤差的觀點，當 n=5，其相對誤差最大在 t=2 時的 0.013763，其餘的表現都很好。而未

列出的值結果亦同。

3.5 以眾數取代中位數的結果

前面所用的三次多項式估計法，皆是取三個分位點(0.1、0.5、0.9)，但是這三個點不一定就是最適當的，在此節中我們將把中位數(0.5)這個分位點捨掉，改以眾數($mode$)這點來取代。

3.5.1 估計方法

若 X_i 為來自 $Weibull(c,1)$ 分佈的隨機變數，則經由變數變換 $Y_i = X_i^{\,c}$ 可知 Y_i 為來自 $Exponential(1)$ 分佈的隨機變數，$i=1,\ldots,n$。因此，我們用 $Y_1+\ldots+Y_n$ 的分佈函數來近似 $X_1+\ldots+X_n$ 的分佈函數。對於每一個 $t\geq 0$，我們希望可以找到 $w(t)$，使得

$$P(X_1+\ldots+X_n \leq t) = P(Y_1+\ldots+Y_n \leq w(t)) \text{。}$$

同時 $w(t)$ 是連續可微分，以及 $w(0)=0$。又因 $w(0)=0$，我們假設 $w(t)$ 之形式為 $\alpha t^{3c}+\gamma t^{2c}+\tau t^{c}$。其中三個未知的參數 $\alpha,\ \gamma,\ \tau$，我們利用 $X_1+\ldots+X_n$ 及 $Y_1+\ldots+Y_n$ 的 0.1、$mode$ 以及 0.9 這三個分位點來決定。首先，我們令 t_p 滿足

$$P(X_1+\ldots+X_n \leq t_p) = p \text{，}$$

並且 θ_p 滿足

$$P(Y_1+\ldots+Y_n \leq \theta_p) = p \text{，}$$

因此，t_p 與 θ_p 分別為為 $X_1+\ldots+X_n$ 與 $Y_1+\ldots+Y_n$ 的第 p 分位點。當 $p=0.1$, *mode*, 0.9 時，我們利用 $\alpha t_p^{3c}+\gamma t_p^{2c}+\tau t_p^{c}=\theta_p$ 得到三個線性方程式，而這三個方程式可求得此三個未知參數（α, γ, τ）。

3.5.2 模擬步驟:

（一）模擬樣本數 $N=10^7$，令 $T_1,T_2,\ldots,T_N$ 為 N 個分配相同且獨立來自 $F_n(x)=P(X_1+\ldots+X_n\le x)$ 的隨機變數，再將其排序得到 $T_{(1)},T_{(2)},\ldots,T_{(N)}$，並且令 $\hat{t}_p=T_{(Np)}$ 為 t_p 的估計值，其中 $p=0.1$, *mode*, 0.9。因為模擬樣本數 $N=10^7$，因此當 $p=0.1$, 0.9 時，Np 都是正整數，所以 $T_{(Np)}$ 是第 p 樣本分位點，亦即 $\hat{t}_{0.1}=T_{(10^6)}$，$\hat{t}_{0.9}=T_{(9\times10^6)}$ 分別為 $t_{0.1}$，$t_{0.9}$ 的估計值。

另外，*mode* 的取法是把所得 10^7 樣本數加以分組，累算出個數最多的組別，做為 *mode* 的估計點 $\hat{t}_{\mathrm{mode}}$。

（二）以相同的方式估算出 θ_p，模擬樣本數 $N=10^7$，令 $Q_1,Q_2,\ldots,Q_N$ 為 N 個分配相同且獨立來自 $F_n(y)=P(Y_1+\ldots+Y_n\le y)$ 的隨機變數，將其排序得到 $Q_{(1)},Q_{(2)},\ldots,Q_{(N)}$，並且令 $\hat{\theta}_p=Q_{(Np)}$ 為 θ_p 的估計值，其中 $p=0.1$, 0.9。經過排序後，找出 $\hat{\theta}_{0.1}=Q_{(10^6)}$，$\hat{\theta}_{0.9}=Q_{(9\times10^6)}$ 分別為 $\theta_{0.1}$，$\theta_{0.9}$ 的估計值。

相同的，*mode* 的取法是把所得 10^7 樣本數加以分組，累算出個數最多的組別，做為 *mode* 的估計點 $\hat{Q}_{\mathrm{mode}}$。

（三）我們由

$$\alpha\,\hat{t}_{0.1}^{3c}+\gamma\,\hat{t}_{0.1}^{2c}+\tau\,\hat{t}_{0.1}^{c}=\theta_{0.1}\text{，}$$

$$\alpha\,\hat{t}_{\mathrm{mode}}^{3c}+\gamma\,\hat{t}_{\mathrm{mode}}^{2c}+\tau\,\hat{t}_{\mathrm{mode}}^{c}=\theta_{\mathrm{mode}}\text{，}$$

$$\alpha\,\hat{t}_{0.9}^{3c}+\gamma\,\hat{t}_{0.9}^{2c}+\tau\,\hat{t}_{0.9}^{c}=\theta_{0.9}\text{，}$$

解得 α, γ, τ 。

（四）因此在固定某個 n 及形態參數 c 下，我們得到 $w(t)$ 的估計形式為

$\hat{w}(t)=\alpha\,t^{3c}+\gamma\,t^{2c}+\tau\,t^{c}$ 。

（五）所以我們有興趣的機率值，可寫成下面的近似方程式:

$P(X_1+\ldots+X_n\le t)\cong P(Y_1+\ldots+Y_n\le\hat{w}(t))=p$ 。

3.5.3 模擬結果

在此試驗中，我們分別以中位數估計法和眾數估計法做 $n=2,3,\ldots,10$ ， 以及 c=0.5,1,2,…9 的例子，而在附錄表(五)中列出 c=0.5、1 及 5，n=5 的結果。並列出中位數估計法的相對誤差值，(定義為 $\frac{|P_1-\hat{P}_1|}{P_1}$ ，其中 $P(X_1+\ldots+X_n\le t)=P_1$，$P(Y_1+\ldots+Y_n\le\hat{w}(t^2)=\hat{P}_1$)。和眾數法的相對誤差值(定義為 $\frac{|P_1-\hat{\hat{P}}_1|}{P_1}$ ，$P(X_1+\ldots+X_n\le t)=P_1$ ，$P(Y_1+\ldots+Y_n\le\hat{w}(t^2)=\hat{\hat{P}}_1$)。再加以比較 $\hat{P}_1$ 和 $\hat{\hat{P}}_1$。從表五中可發現，當 c=1 時，在 t=1,3,5 時中位數法誤差小於眾數法，而在 t=2,4,6,7,8 時誤差是眾數法較小，而全部的最大相對誤差是 t=1 時眾數法的 0.0124476。當 c=0.5 時，在 t=7,9,15 中位數法誤差小於眾數法，而在 t=1,2,3,4,5,12,20 時誤差是眾數法較小，而全部的最大相對誤差是 t=1 時中位數

法的 0.00305。當 c=5 時，在 t=4.7 和 5 時中位數法誤差小於眾數法，而在 t=4,4.3,4.5,6 時誤差是眾數法較小，而全部的最大相對誤差是 t=4.7 時眾數法的 0.004923。其餘的都表現很好，而未列出的值結果亦同，由此可知中位數法與眾數法各有優劣。

第四章、結論

最後，我們雖然不能得到真正分配相同且獨立隨機變數之和的分佈函數。但我們透過其他與其有次方關係的隨機變數和之分佈，找出它們之間的關係，只要二者都是正的隨機變數，初步可以找到一個三次多項式的關係存在，配合著算出的參數表，即得到一個簡單的方程式來估計我們感興趣的隨機變數之和的分佈。未來我們研究的方向可再加入更多次方的式子，看看是否能讓我們的估計更準確、更適合，也可以再針對其他負的隨機變數做研究，使得此方法的運用能更廣泛。期待我們的研究能夠讓更多人受惠，在遇到相關問題時能夠有多一種估計方法可以做選擇。

參考文獻

1. Bickel, P. J. & Doksum, K. A. (2001). *Mathematical Statistics, Basic Ideas and Selected Topics*. 2nd ed.. Volume1. New Jersey: Prentice-Hall.
2. Elsayed, A. (1996). *Reliability Engineering*. MA: Addison-Wesley.
3. Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
4. Ross, S. M. (1996). *Stochastic Processes*. 2nd ed.. New York: Wiley.
5. Ross, S. M. (1997). *Simulation*. 2nd ed.. New York: Academic Press.
6. Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: J. Wiley & sons.
7. Stuart, A. & Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics*. 5th ed., Volume 1, Distribution Theory. London: Charles Griffin & Co..
8. Trivedi, K. S. (2002). *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. 2nd ed.. New York: Wiley.

9.盧信銘(2003)

附錄

[表一、相對誤差表]

在固定 n=3 與 c=1 的情況下，任意輸入 t 值，

$P(X_1+...+X_n \le t)=P_1, P(Y_1+...+Y_n \le \hat{w}(t^c))=\hat{P}_1$, 相對誤差為 $\dfrac{|P_1-\hat{P}_1|}{P_1}$ 。

t	真實機率 P_1	估計機率 $\hat{P}_1$	相對誤差
1	0.080152	0.080071	0.001010
2	0.322935	0.323186	0.000778
3	0.576254	0.5762903	0.000062
4	0.761504	0.761917	0.000617
5	0.87470	0.874601	0.0001115

固定 n=5 與 c=1 的情況下，任意輸入 t 值

t	真實機率 P_1	估計機率 $\hat{P}_1$	相對誤差
2	0.00528445	0.0526609	0.003474
3	0.1844376	0.184465	0.000150
4	0.370659	0.3705504	0.000294
5	0.55909	0.558861	0.000399
6	0.714325	0.713987	0.0004734

7	0.826149	0.826284	0.0001629
8	0.899427	0.8994128	0.0000166

[表二、相對誤差表]

在固定 n=5 與 c=1 的情況下，任意輸入 t 值

$P(X_1+\ldots+X_n \le t)=P_1, P(Y_1+\ldots+Y_n \le \hat{w}(t^{1/c}))=\hat{P}_1$, 相對誤差為 $\dfrac{|P_1-\hat{P}_1|}{P_1}$。

t	真實機率 P_1	估計機率 $\hat{P}_1$	相對誤差
2	0.052645	0.0528371	0.0036281
3	0.1839756	0.1847272	0.0040875
4	0.3704939	0.3710592	0.0015249
5	0.5587293	0.5588994	0.0003060
6	0.7145179	0.7138352	0.0009558
7	0.8262037	0.8257325	0.0005700
8	0.9002639	0.9002379	0.0000288

在固定 n=5 與 c=5 的情況下，任意輸入 t 值

$$P(X_1+\ldots+X_n\le t)=P_1, P(Y_1+\ldots+Y_n\le \hat{w}(t^2))=\hat{P}_1,$$ 相對誤差為 $$\frac{|P_1-\hat{P}_1|}{P_1}$$ 。

t	真實機率 P_1	估計機率 $\hat{P}_1$	相對誤差
2	0.052656	0.052909	0.0048237
3	0.184478	0.185365	0.0048082
4	0.369974	0.370136	0.0004405
5	0.559211	0.558969	0.0004327
6	0.713825	0.713481	0.0004819
7	0.825949	0.825699	0.0003014
8	0.8995499	0.899527	0.00002445

[表三、相對誤差表]

在固定 n=5 的情況下，解得 α、r、τ、δ、ρ，

$\hat{w}(t)=\alpha\, t^{8}+\gamma\, t^{6}+\tau\, t^{4}+\delta t^{2}+\rho$，任意輸入 t 值

$P(X_1+\ldots+X_n\le t)=P_1, P(Y_1+\ldots+Y_n\le \hat{w}(t^2))=\hat{P}_1$, 相對誤差為 $\dfrac{|P_1-\hat{P}_1|}{P_1}$。

t	α	r	τ	δ	ρ	$\hat{w}(t^2)$
-3	0.73101	-11.64981	52.776103	-52.528326	4.37677	110.4484
-2	-7.303371	135.77078	-714.0532	750.8696	4.327455	-1597.3788
-1	0.542801	-9.192380	44.395359	-45.410422	4.38503	-5.279360
0	-10469.604	184019.06	-918239.66	928243.246	-72.10521	-72.10521
1	4.690211	-79.84025	387.533749	-397.5495	4.343049	-80.82274
2	-3.523513	12.478964	164.395781	239.35476	4.361745	1573.90966
3	5.526529	-96.197736	476.369582	-485.23570	4.379513	354.599462

t	真實機率 P_1	估計機率 $\hat{P}_1$	相對誤差
-3	0.0898	1.000000	10.135857
-2	0.18783	0.000000	1.0000000
-1	0.32542	0.000000	1.0000000
0	0.49920	0.000000	1.0000000
1	0.67194	0.000000	1.0000000
2	0.81479	1.000000	0.2273101
3	0.91047	1.000000	0.0983217

[表四、相對誤差表]

在固定 n=5 的情況下，解得 α、r、τ、δ、ρ，

$\hat{w}(t) = \alpha\, t^8 + \gamma\, t^6 + \tau\, t^4 + \delta t^2 + \rho$，任意輸入 t 值，

$P(X_1 + ... + X_n \le t) = P_1, P(Y_1 + ... + Y_n \le \hat{w}(t^2)) = \hat{P}_1$, 相對誤差為 $\dfrac{|P_1 - \hat{P}_1|}{P_1}$ 。

t	α	r	τ	δ	ρ	$\hat{w}(t^2)$
2	0.000001	-0.000065	0.001175	0.275137	0.086715	1.202144
3	0.000000	0.000047	0.002060	0.309354	-0.005889	2.642737
4	-0.000002	0.00013	-0.004063	0.330349	0.083509	4.587979
5	0.000001	0.000001	-0.000479	0.293919	0.038389	7.110343
6	-0.000002	0.000158	-0.004189	0.325572	-0.052056	9.910262

t	真實機率 P_1	估計機率 $\hat{P}_1$	相對誤差
2	0.054489	0.055239	0.0137639
3	0.24675	0.24695	0.0008105
4	0.53141	0.53132	0.0001693
5	0.78424	0.78512	0.0011211
6	0.92339	0.921089	0.0024908

[表五、相對誤差表]

在固定 n=5 與 c=1 的情況下，以中位數估計法，任意輸入 t 值，

$$P(X_1+...+X_n \le t) = P_1, P(Y_1+...+Y_n \le \hat{w}(t^2)) = \hat{P}_1, \text{相對誤差為} \frac{|P_1-\hat{P}_1|}{P_1}。$$

在固定 n=5 與 c=1 的情況下，以眾數估計法，任意輸入 t 值

$$P(X_1+...+X_n \le t) = P_1, P(Y_1+...+Y_n \le \hat{w}(t^2)) = \hat{\hat{P}}_1, \text{相對誤差為} \frac{|P_1-\hat{\hat{P}}_1|}{P_1}$$

t	真實機率 P_1	$\hat{P}_1$	相對誤差	$\hat{\hat{P}}_1$	相對誤差	較好
1	0.0036794	0.0036397	0.0107626	0.00363359	0.0124476	$\hat{P}_1$
2	0.0052597	0.052651	0.0010304	0.052647	0.000958	$\hat{\hat{P}}_1$
3	0.184516	0.184502	0.0000791	0.184489	0.0001484	$\hat{P}_1$
4	0.3704107	0.3705167	0.0002861	0.3703119	0.0002667	$\hat{\hat{P}}_1$
5	0.558836	0.5590467	0.0003761	0.5591152	0.0004985	$\hat{P}_1$
6	0.7142273	0.7141577	0.0000974	0.7142779	0.0000708	$\hat{\hat{P}}_1$
7	0.82645	0.826513	0.0000757	0.826408	0.0000505	$\hat{\hat{P}}_1$
8	0.899452	0.899463	0.0000115	0.8994624	0.0000104	$\hat{\hat{P}}_1$

在固定 n=5 與 c=0.5 的情況下，以中位數估計法，任意輸入 t 值

$P(X_1+...+X_n \le t) = P_1, P(Y_1+...+Y_n \le \hat{w}(t^2)) = \hat{P}_1$, 相對誤差為 $\dfrac{|P_1-\hat{P}_1|}{P_1}$ 。

在固定 n=5 與 c=0.5 的情況下，以眾數估計法，任意輸入 t 值

$P(X_1+...+X_n \le t) = P_1, P(Y_1+...+Y_n \le \hat{w}(t^2)) = \hat{\hat{P}}_1$, 相對誤差為 $\dfrac{|P_1-\hat{\hat{P}}_1|}{P_1}$

t	真實機率 P_1	$\hat{P}_1$	相對誤差	$\hat{\hat{P}}_1$	相對誤差	較好
1	0.036384	0.036273	0.00305	0.036294	0.00248	$\hat{\hat{P}}_1$
2	0.112779	0.112769	0.0000851	0.112778	0.0000053	$\hat{\hat{P}}_1$
3	0.199667	0.199515	0.0007632	0.199526	0.0007081	$\hat{\hat{P}}_1$
4	0.285127	0.284791	0.001176	0.284805	0.001128	$\hat{\hat{P}}_1$
5	0.364084	0.363949	0.0003702	0.363951	0.0003658	$\hat{\hat{P}}_1$
7	0.499024	0.4990303	0.0000124	0.4994306	0.0008139	$\hat{P}_1$
9	0.6051265	0.605291	0.0002716	0.6049286	0.0003272	$\hat{P}_1$
12	0.721359	0.7217506	0.0005425	0.7216734	0.0004355	$\hat{\hat{P}}_1$
15	0.800804	0.800706	0.0001223	0.8006457	0.0001987	$\hat{P}_1$
20	0.882105	0.882305	0.0002269	0.882294	0.0002151	$\hat{\hat{P}}_1$

在固定 n=5 與 c=5 的情況下，以中位數估計法，任意輸入 t 值

$P(X_1+\ldots+X_n \le t) = P_1, P(Y_1+\ldots+Y_n \le \hat{w}(t^2)) = \hat{P}_1$, 相對誤差為 $\dfrac{|P_1-\hat{P}_1|}{P_1}$ 。

在固定 n=5 與 c=5 的情況下，以眾數估計法，任意輸入 t 值

$P(X_1+\ldots+X_n \le t) = P_1, P(Y_1+\ldots+Y_n \le \hat{w}(t^2)) = \hat{\hat{P}}_1$, 相對誤差為 $\dfrac{|P_1-\hat{\hat{P}}_1|}{P_1}$

t	真實機率 P_1	$\hat{P}_1$	相對誤差	$\hat{\hat{P}}_1$	相對誤差	較好
4	0.106553	0.106657	0.0009666	0.106634	0.000747	$\hat{\hat{P}}_1$
4.3	0.264746	0.266221	0.005571	0.265861	0.004212	$\hat{\hat{P}}_1$
4.5	0.415983	0.417053	0.002571	0.415906	0.0001855	$\hat{\hat{P}}_1$
4.7	0.583067	0.581803	0.002166	0.580196	0.004923	$\hat{P}_1$
5	0.80505	0.80294	0.002623	0.801732	0.004126	$\hat{P}_1$
6	0.999133	0.998713	0.0004199	0.998732	0.0004017	$\hat{\hat{P}}_1$