# 國 立 交 通 大 學

## 統計學研究所

## 碩 士 論 文

容忍區間

**Tolerance Interval**

研 究 生 ：陳沛君
指導教授 ：洪慧念 博士

中 華 民 國 九 十 五 年 六 月

# 容忍區間

# Tolerance Interval

研 究 生：陳沛君　　Student： Pei-Chun Chen
指導教授：洪慧念　　Advisor： Dr. Hui-Nien Hung

國 立 交 通 大 學
統計學研究所
碩 士 論 文

**A Thesis**
**Submitted to Institute of Statistics**
**College of Science**
**National Chiao Tung University**
**in Partial Fulfillment of the Requirements**
**for the Degree of**
**Master**
**in**
**Statistics**
**June 2006**

**Hsinchu, Taiwan, Republic of China**
中華民國九十五年六月

# 容忍區間

研究生：陳沛君　　　　指導教授：洪慧念 博士

國立交通大學統計學研究所

## 摘要

在統計這個龐大的領域中，「估計」是個最重要的項目，估計又可以分為點估計和區間估計，一般的點估計可以估計參數，一些特殊量的估計如中位數，分位數等等，區間估計則是希望給定一個區間使得此區間覆蓋真值的機率達到一定水準。這篇文章的主題是有關於覆蓋區間的區間估計，我們提出一個系統化的方法，並且探討此方法與傳統參數的區間估計有何關係，試著用不同於以往的計算方法來連接它們。

# Tolerance Interval

Student：Pei-Chun Chen　　　Adviser：Dr. Hui-Nien Hung

Institute of Statistics

National Chiao Tung University

## Abstract

In the statistics history, estimation is one of the most important research topics, we know that estimation have two major parts, one is the point estimation, and the other is the interval estimation. In general, we use point estimation to estimate parameters or some important quantities, for example: medium, quantile….etc. The interval estimation gives us an interval which covers the true parameters with a fixed percentage. In this paper, we will discuss the interval estimation of the coverage interval. We propose a systematic method to find an interval estimation of the coverage interval. Also we discuss the relationship between our method and traditional interval estimation methods.

# 誌 謝

不知不覺在統研所的研究生活即將進入尾聲，在洪慧念教授的指導下，研究論文也將如期完成，心中真是百感交集，除了期待與不捨，當然還有感激與感動。很慶幸自己是交大統研所的一份子，認識了和藹可親的教授們、親切助人的學長姐、互相幫助的朋友們以及熱心親切的所辦助理，使這兩年的研究生活過的愉快又溫馨。

即將畢業的我，首先要感謝我的指導教授_洪慧念教授，謝謝老師這兩年的支持鼓勵和包容，花費很多時間耐心的指導我，慢慢的帶領我進行論文的研究，除了學業方面，也謝謝教授在生活中教導我很多做人道理和處事態度，身為洪慧念教授的學生讓我覺得很驕傲也很開心，希望自己能不負教授期望的迎向人生的另一個旅程。接著感謝陳鄰安教授、徐南蓉教授和黃信誠教授的中肯建議和指導，使得這篇論文能更加完整。

最後我要感謝我的家人:父親 陳淑森、母親 林美英、妹妹 亞妤和珮文，感謝父母從小到大的栽培、支持以及無怨無悔的付出，感謝妹妹們給我打氣、加油，感謝他們給我一個溫馨又快樂的家庭，使我能夠專心的求學及生活。最後還要感謝小宛茹、婉文、秀慧、孟樺、大宛茹、鷰筑和小馬，在這兩年的生活中陪伴我，在我最需要幫忙的時候支持我以及鼓勵我，讓我有勇氣面對挫折與挑戰，我會永遠珍惜你們的友誼。在鳳凰花開、驪歌輕唱之際，將這篇文章和所有的好朋友與親人分享，由衷的感謝所有曾經幫助過我的人。

陳沛君　謹誌于

國立交通大學統計研究所

中華民國九十五年六月

目錄

## 第一章　緒論

在進入這篇文章的主題之前，我們要先簡單介紹幾種區間，其中包含有覆蓋區間(coverage interval)、參考區間(reference interval) 、統計覆蓋區間(statistical coverage interval)、信賴區間(confidence interval)、容忍區間(tolerance interval)和預測區間 (prediction interval)，這些區間有些較常出現在工業統計的領域，有些是醫學統計的領域...等，但是不管它們最常出現在哪種領域，他們的原理都離不開統計。

### 1-1. 覆蓋區間

覆蓋區間是指一個區間能以一個正確的特定比率包含某個具特定分配的隨機變數。在度量衡學中$[a,b]$被定義為一個覆蓋率$100p\%$的覆蓋區間是指$[a,b]$有$100p\%$包含分位數$Y$而這個$Y$具有密度函數$g(y)$，也就是說$\int_a^b g(y)dy = p$，從這個式子我們可以發現覆蓋區間$[a,b]$可以被定義為$(g_1(\theta), g_2(\theta))$，也就是指這個區間是參數的函數，舉例: 一組樣本$X = (x_1, x_2, \cdots x_n)$來自$\mathrm{N}(\mu,1)$則一個覆蓋率$100p\%$的覆蓋區間$[a,b]$為 $(\mu - z_{\frac{1-p}{2}}, \mu + z_{\frac{1-p}{2}})$。

### 1-2. 參考區間

從字面上可以看出是指一個讓人參考的區間，又可以稱為 normal range 或 reference range，可以解釋為一個隨機變數在正常的情形下會落在的區間，最常在醫學研究中被使用，我們可以在意義上認為參考區間就是生物學上的覆蓋區

間。舉例來說，我們可以量一群正常人的血液 pH 值(當然，要如何判斷一個人是否為”正常人”是很困難的，我們不在這裡討論。)，所得到的 pH 值稱為參考值，這些參考值可能會形成某些分配，假設有 95%的人其 pH 值落在(7.36~7.43N/A)間，則這個區間就稱為人類血液 pH 值的參考區間(7.36~7.43N/A)。

## 1-3. 統計覆蓋區間

對於覆蓋區間的估計，統計學家常使用各種不同的估計方法去估計覆蓋區間 $(g_1(\theta), g_2(\theta))$，例如藉由點估計的各種方法分別去估計 $g_1(\theta)$ 和 $g_2(\theta)$，這樣會得到一個 $(g_1(\theta), g_2(\theta))$ 的估計區間，我們也可以求取 $(g_1(\theta), g_2(\theta))$ 的區間估計，不管是點估計或是區間估計，指的都是一種統計覆蓋區間，所以統計覆蓋區間泛指估計覆蓋區間的任何估計，以下的容忍區間和預測區間都屬於統計覆蓋區間。

## 1-4. 信賴區間

信賴區間指的是一個會包含一個我們感興趣的未知參數的估計區間，這個估計的區間是由一組給定的樣本計算出來的範圍，假若我們從同樣的分配不停的抽出獨立的樣本，則每一個樣本都可以計算出一個 $\gamma$ 比率的信賴區間，這時候我們可以發現，這麼多個區間會有大概 $\gamma$ 比率的區間包含未知參數，這個 $\gamma$ 比率我們稱為信賴水準(confidence level)。

## 1-5. 容忍區間

容忍區間比較常被使用在工業統計,在製造業中容忍區間被使用來檢查產品的規格是否落在可容忍的容忍範圍(tolerance limit)內。計算上和信賴區間不同的地方在於,信賴區間是由同屬於某個分配的很多組獨立的樣本所得到,這些區間會有某個比率包含未知參數,而容忍區間則是覆蓋區間的信賴區間估計,也就是說我們要找一個區間使其能在一定信心水準下使得未來一個新的資料落在此區間達到一固定水準。亦即對於一個覆蓋率$\beta$信賴水準$\gamma$的容忍區間$\left(T_1(\underset{\sim}{X}),T_2(\underset{\sim}{X})\right)$,其數學表示式為: $P_{\underset{\sim}{X}|\theta}\left(P_{X_{n+1}|\theta}\left(X_{n+1}\in\left(T_1(\underset{\sim}{X}),T_2(\underset{\sim}{X})\right)\right)>\beta\right)\geq\gamma$ ,其中$(X_1,X_2,\cdots X_n,X_{n+1})$為 iid,我們能用以下三種方式形容:

(I)我們有$\gamma$比率的信心這個區間會包含未來隨機變數的機率大於等於$\beta$。

(II)對於一個覆蓋區間的信賴區間估計,有$\gamma$的機率未來觀測到的值會掉進區間$\left(T_1(\underset{\sim}{X}),T_2(\underset{\sim}{X})\right)$的機率至少會是$\beta$。

(III)對於一個信賴區間估計機率為$\gamma$,則會有$\beta$比率(相對於有$m$個觀測值,$m\neq 1$,$\beta=\dfrac{k}{m}$)的觀測值會掉進區間$\left(T_1(\underset{\sim}{X}),T_2(\underset{\sim}{X})\right)$。

1-6. 預測區間

預測區間是個預測某些未來的值的區間,它和容忍區間有些類似,預測區間是根據現有樣本或過去樣本來估計一個(或一組小數目的)未來觀察值。$\gamma$比率的預測區間是指我們有$\gamma$比率的信心預測未來的一個值會落入這個區間。數學式子表示為$\mathrm{P}_{\underset{\sim}{X},X_{n+1}|\theta}(T_1(\underset{\sim}{X})\leq X_{n+1}\leq T_2(\underset{\sim}{X}))=E_{\underset{\sim}{X}|\theta}\left(P_{X_{n+1}|\theta}\left(X_{n+1}\in\left(T_1(\underset{\sim}{X}),T_2(\underset{\sim}{X})\right)\right)\right)=\gamma$。

在下面的章節中，首先是文獻回顧(第二章)，因為在這篇論文中，使用或引用很多已經被發表的重要文章，由於它們是這麼的重要，所以我們必須讓讀者先了解它們的內容，接著我們會再次介紹信賴區間 (第三章)和容忍區間 (第四章)，雖然前面文章已經簡單的介紹過它們,但是在以下的兩個章節中我們介紹的內容將會有關於它們的數學表示式、計算方法和簡單的例子，在信賴區間部分我們加入有關它的近似區間的一些內容，而在容忍區間方面我們運用文獻回顧(第二章)中所提及的有關貝氏的容忍區間想法，並試著用不同於以往的方法找出容忍區間，最後再跟讀者介紹兩篇容忍區間在非統計方面的運用的論文。以下的內容僅會包含單邊有上界的區間，當然，我們也會提及如何利用單邊的區間來求出雙邊的區間。

## 第二章　文獻回顧

在這一章裡，我們有信賴區間的近似區間、常態分配的容忍區間和貝氏的容忍區間，這幾篇的文獻回顧與本論文有很大的相關性，所以我們特別把這幾篇文章的內容放在文獻回顧這個章節裡。

### 2-1. 信賴區間的近似區間

藉由貝氏的想法，B. L. Welch and H. W. Peers. (1963)找出了信賴區間的近似區間。假如我們有一組隨機樣本 S，其密度函數為 $p(S,\theta)dS$ ，接著假設$\theta$有先驗分配$q(\theta)d\theta$ ，則藉由貝氏的想法我們可以知道$\theta$的後驗分配

$\theta \mid \underset{\sim}{X} \sim \dfrac{p(S,\theta)q(\theta)d\theta}{\int p(S,t)q(t)dt}$，定義 $g(S,\alpha)$ 滿足 $\dfrac{\int^{g(S,\alpha)} p(S,t)q(t)dt}{\int p(S,t)q(t)dt} = \alpha$ 就可以推得 $\Pr\{\theta < g(S,\alpha) \mid S\} = \alpha$，也就是說給定樣本 S 則 $g(S,\alpha)$ 就是 $\theta$ 的一個 $\alpha$ 分位數。但是如果 $\theta$ 沒有先驗分配，也就是 $\theta$ 是一個固定的值，則 $\theta$ 的一個 $\alpha$ 分位數就不可以用上面的方法推出。我們想找一個 $h(S,\alpha)$ 滿足 $\Pr\{\theta < h(S,\alpha) \mid \theta\} = \alpha$，在這個情況下 $\theta$ 是一個固定的值，S 是隨機的，則 $h(S,\alpha)$ 就可以說是參數 $\theta$ 單邊 $\alpha$ 信賴水準下信賴區間的上界。Lindley(1958)發現了我們可以找 $h(S,\alpha)$ 滿足 $\dfrac{\int^{h(S,\alpha)} p(S,t)w(t)dt}{\int p(S,t)w(t)dt} = \alpha$，其中 $w(t)$ 是一個任意的加權函數，且若令 $w(t)$ 為 $\sqrt{\text{Fisher information}}$，則 $\Pr\{\theta < h(S,\alpha) \mid \theta\}$ 和 $\alpha$ 的誤差會最小。

## 2-2. 常態分配的容忍區間

假設我們有一組樣本 $X = (x_1, x_2, \cdots x_n)$ 來自 $\mathrm{N}(\mu,\sigma^2)$。在 $\mu$ 和 $\sigma^2$ 未知的情形下，為了得到單邊有上界的 $\beta$ 覆蓋率 $\gamma$ 信賴水準的容忍區間。所以我們假設 $\beta$ 覆蓋率 $\gamma$ 信賴水準的單邊有上界容忍區間為 $(-\infty, \overline{X} + ks)$，其中 $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$、$s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$、$k$ 滿足 $\mathrm{P}_{\underset{\sim}{X}|\mu,\sigma^2}(P_{X_{n+1}|\mu,\sigma^2}(X_{n+1} < \overline{X} + ks) \geq \beta) = \gamma$ ( $X_{n+1}$ 是未來的觀測值)，這裡的 $k$ 是被決定的值，叫做 tolerance factor，$\gamma$ 就是信賴水準，$\beta$ 是覆蓋率。在推導 $k$ 之前，我們需要先了解 non-centrality $t$ 分配，假設 $Y$ 是標準常態分配，而 $V$ 是 $\chi^2_\upsilon$,且與 $Y$ 獨立則若 $t' = \dfrac{Y+\delta}{\sqrt{V}}\sqrt{\upsilon}$ 稱為 non-centrality t 分

配記做$t'(\upsilon,\delta)$。當$\delta=0$，$t'(\upsilon,\delta)$就是 centrality t 分配$t(\upsilon)$。我們定義$t'(\upsilon,\delta)$的

$1-\alpha$分位數為$t_{1-\alpha}{}'(\upsilon,\delta)$，我們可以發現$f_{\upsilon}(t,\delta)=f_{\upsilon}(-t,-\delta)$和

$t_{\alpha}{}'(\upsilon,\delta)=-t_{1-\alpha}{}'(\upsilon,-\delta)$。

對於右邊的$\beta$覆蓋率$\gamma$信賴水準的單邊有上界容忍區間

$S(\underset{\sim}{X})=(-\infty,\overline{X}+ks)$，假設

$C_{X_{n+1},\mu,\sigma^2}(S(\underset{\sim}{X}))=P_{X_{n+1}|\mu,\sigma^2}(X_{n+1}<\overline{X}+ks)=\Phi(\dfrac{\overline{X}+ks-\mu}{\sigma})$，給定

$\beta$(0<$\beta$<1)，則$C_{X_{n+1},\mu,\sigma^2}(S(\underset{\sim}{X}))\geq\beta$就相當於$\dfrac{\overline{X}+ks-\mu}{\sigma}\geq z_{1-\beta}$，其中$z_{1-\beta}$是

N(0,1)的$\beta$分位數，即$\beta=\dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{z_{1-\beta}}e^{-\frac{t^2}{2}}dt$。$k$要滿足於

$P(C_{X_{n+1},\mu,\sigma^2}(S(\underset{\sim}{X}))\geq\beta)=\gamma$，也就是

$$\begin{aligned}P_{\underset{\sim}{X}|\mu,\sigma^2}(\overline{X}+ks\geq\mu+\sigma z_{1-\beta})&=P_{\underset{\sim}{X}|\mu,\sigma^2}\left(\frac{(\overline{X}-\mu)\sqrt{n}\Big/\sigma-z_{1-\beta}\sqrt{n}}{s\Big/\sigma}\geq-k\sqrt{n}\right)\\&=P_{\underset{\sim}{X}|\mu,\sigma^2}(t'\geq-k\sqrt{n})=\gamma\end{aligned}$$

這裡的$t'=\dfrac{\overline{X}-\mu-\sigma z_{1-\beta}}{S}\sqrt{n}\sim t'(n-1,-z_{1-\beta}\sqrt{n})$，我們可以取

$k=-\dfrac{t_{\gamma}{}'(n-1,-z_{1-\beta}\sqrt{n})}{\sqrt{n}}=\dfrac{t_{1-\gamma}{}'(n-1,z_{1-\beta}\sqrt{n})}{\sqrt{n}}$。

所以單邊有上界的$\beta$覆蓋率$\gamma$信賴水準的單邊有上界容忍區間為

$(-\infty,\overline{X}+\dfrac{t_{1-\gamma}{}'(n-1,z_{1-\beta}\sqrt{n})}{\sqrt{n}}s)$，可以用同樣的方法得到$\beta$覆蓋率$\gamma$信賴水準的

單邊有下界容忍區間為$(\overline{X}-ks,\infty)$，其中

$$k=\frac{t_{1-\gamma}{}'(n-1,z_{1-\beta}\sqrt{n})}{\sqrt{n}}=-\frac{t_{\gamma}{}'(n-1,-z_{1-\beta}\sqrt{n})}{\sqrt{n}}$$ 。

### 2-3. 貝氏的容忍區間

除了以上我們從 frequentist 的觀點來定義容忍區間，我們也可以用貝氏的觀點來看容忍區間。在傳統信賴區間的估計上，貝氏方法得到的區間是要求在樣本資料固定下，參數落在某一區間的機率要達到某一信賴水準。在 Aitchison(1964) 的論文中也是利用相同的觀點。在 frequentist 的方面，我們要求 $\Pr_{\underset{\sim}{X}|\theta}[C_{X_{n+1},\theta}(S(\underset{\sim}{X}))\ge\beta]\ge\gamma$ 其中 $C_{X_{n+1},\theta}(S(\underset{\sim}{X}))=P_\theta(X_{n+1}\in S(\underset{\sim}{X}))$，而貝氏的觀點則是要求 $\Pr_{\theta|\underset{\sim}{X}}[C_{X_{n+1},\theta}(S(\underset{\sim}{X}))\ge\beta]\ge\gamma$。若 $S(\underset{\sim}{X})=(-\infty,b(\underset{\sim}{X}))$ 藉由這裡的 $C_{X_{n+1},\theta}(S(\underset{\sim}{X}))\ge\beta$，可得到 $F_\theta(b(\underset{\sim}{X}))\ge\beta$，所以我們可以知道 $b(\underset{\sim}{X})\ge F_\theta^{-1}(\beta)$ 因此，給定一組樣本資料 $\underset{\sim}{X}$，我們可以計算 $\theta$ 的後驗分配，藉由 $\theta$ 的後驗分配再求 $F_\theta^{-1}(\beta)$ (單一樣本 $x_i$ 的 $\beta$ 分位數)的後驗分配的 $\gamma$ 分位數，$b(\underset{\sim}{X})$ 就是此分位數單邊有上界的貝氏 $\beta$ 覆蓋率 $\gamma$ 信賴水準的貝氏容忍區間 $(-\infty,b(\underset{\sim}{X}))$。

# <u>第三章　信賴區間</u>

### 3-1. 尋找信賴區間的方法

在找信賴區間時，我們可以先找一個樞紐量(pivotal)，其定義為:若一個量 $Q(\theta,x_1,x_2,\cdots x_n)$ 是未知參數 $\theta$ 和 $(x_1,x_2,\cdots x_n)$ 的聯合函數，而且其分配和 $\theta$ 無關，則我們稱 $Q(\theta,x_1,x_2,\cdots x_n)$ 為一個樞紐量，利用這個樞紐量的分配我們就可以得到

一個信賴區間。

舉例一:一組樣本 $\underset{\sim}{X}=(x_1,x_2,\cdots x_n)$ ，其分配為 $\exp(\theta)$ ，這裡的 $\theta$ 是 rate，$E(X)=1/\theta$，為了得到 $\theta$ 的 $\alpha$ 信賴水準的單邊上界信賴區間，我們先找到一個樞紐量 $Q(\theta,x_1,x_2,\cdots x_n)=2\theta\sum_{i=1}^{n}x_i$ 其分配為 $\chi^2(2n)$ ，其分配和參數無關，然後可以得到 $P(\theta\le\frac{\chi^2_{1-\alpha}(2n)}{2\sum_{i=1}^{n}x_i})=P(2\theta\sum_{i=1}^{n}x_i\le\chi^2_{1-\alpha}(2n))=\alpha$ ，則 $\theta$ 的 $\alpha$ 信賴水準單邊上界信賴區間就是 $(0,\frac{\chi^2_{1-\alpha}(2n)}{2\sum_{i=1}^{n}x_i})$ ，同樣的我們也可以藉由 $P(\frac{\chi^2_{(1+\alpha)/2}(2n)}{2\sum_{i=1}^{n}x_i}\le\theta\le\frac{\chi^2_{(1-\alpha)/2}(2n)}{2\sum_{i=1}^{n}x_i})=\alpha$ 得到估計 $\theta$ 的 $\alpha$ 信賴水準的雙邊信賴區間 $(\frac{\chi^2_{(1+\alpha)/2}(2n)}{2\sum_{i=1}^{n}x_i},\frac{\chi^2_{(1-\alpha)/2}(2n)}{2\sum_{i=1}^{n}x_i})$ 。

舉例二: 一組樣本 $\underset{\sim}{X}=(x_1,x_2,\cdots x_n)$ ，其分配為 $N(\theta,1)$ ，為了得到 $\theta$ 的 $\alpha$ 信賴水準的單邊上界信賴區間，我們先找到一個樞紐量 $Q(\theta,x_1,x_2,\cdots x_n)=\frac{\overline{X}-\theta}{1/\sqrt{n}}$ ，為標準常態分配和參數無關，然後可以得到 $P(\theta\le\overline{X}-\frac{z_\alpha}{\sqrt{n}})=P(z_\alpha\le\frac{\overline{X}-\theta}{1/\sqrt{n}})=\alpha$ ，則 $\theta$ 的 $\alpha$ 信賴水準單邊上界信賴區間就是 $(-\infty,\overline{X}-\frac{z_\alpha}{\sqrt{n}})$，我們也可以利用相同的方法得到估計 $\theta$ 的 $\alpha$ 信賴水準的雙邊信賴區間 $(\overline{X}+\frac{z_{(1+\alpha)/2}}{\sqrt{n}},\overline{X}+\frac{z_{(1-\alpha)/2}}{\sqrt{n}})$ 。

## 3-2. 信賴區間的近似區間

從上面我們得知信賴區間要藉由樞紐量來求得，但要是樞紐量沒有這麼容易得到。因此，我們想從貝氏的方法中找到一個近似的方法。Lindley(1958)提出了

$h(S,\alpha)$ 滿足 $\dfrac{\int^{h(S,\alpha)} p(S,t)w(t)dt}{\int p(S,t)w(t)dt}=\alpha$，B. L. Welch 和 H. W. Peers(1963)在文中提到如果讓 $w(t)$ 為 $\sqrt{\text{Fisher information}}$，則 $\Pr\{\theta<h(S,\alpha)\,|\,\theta\}$ 和 $\alpha$ 的誤差會最小。其中 Fisher information $I=E((\frac{\partial \log L}{\partial\theta})^2)=E(-\frac{\partial^2 \log L}{\partial\theta^2})$，這裡的 $L$ 指的是概似函數(likelihood)，所以 $\log L=\sum_{i=1}^{n}\log f(x_i,\theta)$，使用這個方法我們來計算一下之前的例子:假設一組樣本 $\underset{\sim}{X}=(x_1,x_2,\cdots x_n)$ 來自分配 $\exp(\theta)$，這裡的 $\theta$ 指的是 rate，使用 Lindley(1958)的方法得到估計 $\theta$ 的信賴區間，首先我們求得

$w(t)=\sqrt{\text{Fisher information}}=\sqrt{\dfrac{n}{\theta^2}}$，把 w(t)帶入 $\dfrac{\int^{h(S,\alpha)} p(S,t)w(t)dt}{\int p(S,t)w(t)dt}=\alpha$ 中，可以得到 $\alpha=\int_0^{h(S,\alpha)}\dfrac{(\sum_{i=1}^{n}x_i)^n}{\Gamma(n)}\theta^{n-1}e^{-\theta\sum_{i=1}^{n}x_i}d\theta$，也就是說可以把 $h(S,\alpha)$ 想成是參數 $\theta$ 服從 $\Gamma(n,\sum_{i=1}^{n}x_i)$ 分配的 $\alpha$ 分位數，我們可以得到一個估計 $\theta$ 近似的 $\alpha$ 信賴水準的單邊有上界信賴區間 $(0,\dfrac{\chi^2_{1-\alpha}(2n)}{2\sum_{i=1}^{n}x_i})$，和之前樞紐量求得的信賴區間 $(0,\dfrac{\chi^2_{1-\alpha}(2n)}{2\sum_{i=1}^{n}x_i})$ 是一樣的。

## 第四章　容忍區間

假設觀測值為X=($x_1\cdots x_n$)，若定義一個隨機的區間 $S(\underset{\sim}{X})=(a(\underset{\sim}{X}),b(\underset{\sim}{X}))$，再定義這個區間的覆蓋率為 $C_{X_{n+1},\theta}(S(\underset{\sim}{X}))=\Pr_\theta(X_{n+1}\in S(\underset{\sim}{X}))$，假如

$\Pr_{\underset{\sim}{X}|\theta}[C_{X_{n+1},\theta}(S(\underset{\sim}{X}))\ge\beta]=\gamma$ 對所有的 $\theta$，則我們說 $S(\underset{\sim}{X})$ 是一個 $\beta$ 覆蓋率 $\gamma$ 信賴水準的容忍區間(這裡的 $X_{n+1}$ 是未來的觀測值)，也就是說我們有 $\gamma$ 比率的信心，

這個區間會包含新資料的比率會大於等於$\beta$。倘若我們考慮的是平均而言新資料會落在此區間之機率，亦即當$S(\underset{\sim}{X})=(a(\underset{\sim}{X}),b(\underset{\sim}{X}))$決定後，假如$E_{X_{n+1}|\theta}(C_{X_{n+1},\theta}(S(\underset{\sim}{X})))=\beta$對所有的$\theta$，(這個期望值是對於在$\theta$分配$F_\theta$下重複抽取的樣本$\underset{\sim}{X}$即$X_{n+1}$)，則我們稱此區間為$\beta$ expectation 容忍區間。(從另外一種觀點來看，此區間即為預測區間)

4-1. 利用分位數尋找容忍區間

我們首先對每一個參數求出一個$\beta$分位數（此分位數即為參數之函數），然後再求取這個分位數的$\gamma$信賴水準的信賴區間來當作單邊有上界的$\beta$覆蓋率$\gamma$信賴水準的容忍區間，如以下步驟:

步驟一:找$q_\beta$滿足$F_\theta(q_\beta)=\beta$，$q_\beta$為$\theta$的函數。

步驟二: 找$q_\beta$的$\gamma$信賴水準的信賴區間，也就是$P_{\underset{\sim}{X}|\theta}(q_\beta \le T(\underset{\sim}{X}))\ge\gamma$。在此步驟$T(\underset{\sim}{X})$可用樞紐量的方法求得。

則我們知道$P_{\underset{\sim}{X}|\theta}(P_{X_{n+1}|\theta}(X_{n+1}<T(\underset{\sim}{X}))\ge\beta)=P_{\underset{\sim}{X}|\theta}(F_\theta(T(\underset{\sim}{X}))\ge\beta)$，又$P_{\underset{\sim}{X}|\theta}(F_\theta(T(\underset{\sim}{X}))\ge\beta)=P_{\underset{\sim}{X}|\theta}(T(\underset{\sim}{X})\ge q_\beta)$，所以$P_{\underset{\sim}{X}|\theta}(P_{X_{n+1}|\theta}(X_{n+1}<T(\underset{\sim}{X}))\ge\beta)\ge\gamma$。也就是$(-\infty,T(\underset{\sim}{X}))$是$\beta$覆蓋率$\gamma$信賴水準的容忍區間。

舉例一:假設 location family 的機率密度函數為$f(x-\theta)$，所以它的分位數應為$\theta+c$，c 為$\beta$的函數，則要求取這個分位數的信賴區間，其樞紐量應為$X-(\theta+c)$，舉例: $N(\theta,1)$，其$\beta$分位數為$\theta+z_{1-\beta}$，藉由分位數$\theta+z_{1-\beta}$的樞紐量$X-(\theta+z_{1-\beta})$可得到$\gamma$信賴水準的信賴區間，此樞紐量的分配為$N(-z_{1-\beta},1)$，又

$P(X-\theta-z_{1-\beta} \geq z_\gamma - z_{1-\beta}) = \gamma$ ，所以它的$\gamma$信賴水準的信賴區間為

$(-\infty, X - z_\gamma + z_{1-\beta})$ 即為所求容忍區間。

舉例二:scale family 的機率密度函數通常為$\frac{1}{\theta} f(\frac{1}{\theta}x)$ 或者是$\theta f(\theta x)$，所以它的分位數應為$\theta c$ 或$\frac{c}{\theta}$，c 為$\beta$的函數，則要求取這個分位數的信賴區間，其樞紐量應為 $X/(\theta c)$ 或$\theta X/c$ ，舉例: $\exp(\theta)$ ，其$\beta$分位數為$-\ell \mathrm{n}(1-\beta)/\theta$ ，求取這個分位數的信賴區間，其樞紐量應為$-\theta X/\ell n(1-\beta)$ ，此樞紐量的分配為 $\exp(-\ell n(1-\beta))$ ，又$P(\frac{\theta X}{-\ell n(1-\beta)} \geq \frac{-\ell n(\gamma)}{-\ell n(1-\beta)}) = \gamma$ ，所以它的$\gamma$信賴水準的信賴區間為$(-\infty, \frac{\ell n(1-\beta)}{\ell n(\gamma)} X)$ 即為所求容忍區間。

舉例三:對於 location-scale family，其機率密度函數為$\frac{1}{\sigma} f(\frac{x-\theta}{\sigma})$ ，其$\theta$和$\sigma$均未知，所以它的分位數應為$\theta + c\sigma$ ，c 為已知是$\beta$的函數，則要求取這個分位數的信賴區間，其樞紐量應為$\frac{X-(\theta+c\sigma)}{s}$ ，舉例: $N(\mu,\sigma^2)$ ，其$\mu$和$\sigma^2$均未知，這個例子可在下節中看到，所以就不多做解釋。

4-2. 利用分位數尋找常態分配的容忍區間

在文獻回顧 2-2.中，作者I. Janiga 和 I. Garaj（2005）一開始便假設單邊有上界的$\beta$覆蓋率$\gamma$信賴水準的容忍區間的形式為$(-\infty, \overline{X} + ks)$ ，也就是

$\mathrm{P}_{\underset{\sim}{X}|\theta}(P_{X_{n+1}|\theta}(X < \overline{X} + ks) \geq \beta) = \gamma$ ，其中 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 、$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$ ，

這裡的$k$是要被決定的值，只要決定出$k$值就能得到一個單邊有上界的$\beta$覆蓋率$\gamma$信賴水準的容忍區間，但是在前一節，我們提出一個比較有系統的方法來得到

單邊有上界的$\beta$覆蓋率$\gamma$信賴水準的容忍區間，以下我們就利用這個方法來找常態分配的容忍區間。假設一組樣本$\underset{\sim}{X}=(x_1,x_2,\cdots x_n)$，來自$N(\mu,\sigma^2)$，我們想得到一個估計$\mu$單邊有上界的$\beta$覆蓋率$\gamma$信賴水準的容忍區間，這裡的$\mu$和$\sigma^2$是未知參數，首先我們找出$\beta$分位數為$\mu+z_{1-\beta}\sigma$，再取這個分位數$\mu+z_{1-\beta}\sigma$ 的$\gamma$信賴水準的信賴區間。為了求取$\mu+z_{1-\beta}\sigma$的$\gamma$信賴水準的信賴區間，我們先找一個樞紐量（此樞紐量為$\mu+z_{1-\beta}\sigma$與資料之函數，且樞紐量之分配與參數無關），在此我們選取 $Q(\underset{\sim}{X},\mu+z_{1-\beta}\sigma)=\sqrt{n}\dfrac{\overline{X}-\left(\mu+z_{1-\beta}\sigma\right)}{s}$， 其中$\overline{X}=\sum_{i=1}^{n}X_i$，

$s^2=\dfrac{\sum_{i=1}^{n}(x_i-\overline{X})^2}{n-1}$，其分配為

$$\sqrt{n}\frac{\overline{X}-\left(\mu+z_{1-\beta}\sigma\right)}{s}=\frac{\frac{\overline{X}-\mu}{\sigma}\sqrt{n}-z_{1-\beta}\sqrt{n}}{\frac{s}{\sigma}}\sim t'(n-1,-z_{1-\beta}\sqrt{n})$$ 。

所以藉由$P(\dfrac{\frac{\overline{X}-\mu}{\sigma}\sqrt{n}-z_{1-\beta}\sqrt{n}}{\frac{s}{\sigma}}>t'_{\gamma}(n-1,-z_{1-\beta}\sqrt{n}))=\gamma$可以推得

$P(\overline{x}-\dfrac{t'_{\gamma}(n-1,-z_{1-\beta}\sqrt{n})}{\sqrt{n}}s>\mu+z_{1-\beta}\sigma)=\gamma$，最後得到這個分位數$\mu+z_{1-\beta}\sigma$的$\gamma$信賴水準的信賴區間為$(-\infty,\overline{X}-\dfrac{t_{\gamma}'(n-1,-z_{1-\beta}\sqrt{n})}{\sqrt{n}}s)$，也就是

$(-\infty,\overline{X}+\dfrac{t_{1-\gamma}'(n-1,z_{1-\beta}\sqrt{n})}{\sqrt{n}}s)$。亦即$\beta$覆蓋率$\gamma$信賴水準的容忍區間為

$(-\infty,\overline{X}+\dfrac{t_{1-\gamma}'(n-1,z_{1-\beta}\sqrt{n})}{\sqrt{n}}s)$。把這個結果拿來和和文獻回顧中的2-2.常態分配的容忍區間比較一下，發現雖然我們一開始沒有和這篇論文做一樣的假設，但是

結果卻是相同的。

在以下的部分，我們將會運用這種方法來計算容忍區間，由於我們下面的文章會用到常態分配的容忍區間，所以在這裡我們只舉出常態分配來當例子，但是實際上還有很多例子可以用此方法求得容忍區間，有multivariate normal distribution、exponential、Weibull 和其它連續或離散的分配。

4-3. 利用分位數尋找迴歸的容忍區間

若利用上面文獻回顧中 2-2. Normal 分配的容忍區間來求迴歸的容忍區間，假設 $\underset{\sim}{X}=(x_1,x_2,\cdots x_n)$，$\underset{\sim}{Y}=(y_1,y_2,\cdots y_n)$ 服從迴歸模型 $y_i=\beta_0+\beta_1 x_i+\varepsilon_i$，這裡的 $\varepsilon_i$ 服從 $N(0,\sigma^2)$ 且獨立，若有一新資料 $x_0$，我們想得到估計 $y_0=\beta_0+\beta_1 x_0+\varepsilon_0$ 的單邊有上界的 $\beta$ 覆蓋率 $\gamma$ 信賴水準的容忍區間，在此模型中 $\beta_0$、$\beta_1$ 和 $\sigma^2$ 都是未知，首先我們找出 $y_0$ 的 $\beta$ 分位數為 $\beta_0+\beta_1 x_0+\sigma z_{1-\beta}$，再求取這個分位數 $\beta_0+\beta_1 x_0+\sigma z_{1-\beta}$ 的 $\gamma$ 信賴水準的信賴區間。 由迴歸分析的結論，假設 $(b_0,b_1,s^2)$ 為 $(\beta_0,\beta_1,\sigma^2)$ 的最小平方法估計量，我們有(1) $b_0+b_1 x_0$ 分配為 $N(\beta_0+\beta_1 x_0,(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})\sigma^2)$，其中 $S_{xx}=\sum_{i=1}^{n}(x_i-\overline{X})^2$ (2) $\frac{(n-2)s^2}{\sigma^2}$ 分配為 $\chi^2(n-2)$ (3) $(b_0,b_1)$ 和 $s^2$ 獨立。因此我們知道 $\frac{\frac{(b_0+b_1 x_0)-(\beta_0+\beta_1 x_0+\sigma z_{1-\beta})}{\sqrt{(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})}}}{s}$ 的

分配為 $t'(n-2,-\frac{z_{1-\beta}}{\sqrt{(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})}})$ 是一個樞紐量。而且

$$P((b_0+b_1x_0)-\sqrt{(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})}t'_{\gamma}(n-2,-\frac{z_{1-\beta}}{\sqrt{(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})}})s>(\beta_0+\beta_1x_0)-z_{1-\beta}\sigma)=\gamma$$

（在這裡的 $t'_{\gamma}(n-2,-\frac{z_{1-\beta}}{\sqrt{(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})}})$ 表示 $t'$ 的 $(1-\gamma)$ 分位數），最後我們得到

這個分位數 $\beta_0+\beta_1x_0+\sigma z_{1-\beta}$ 的 $\gamma$ 信賴水準的信賴區間為

$(-\infty,(b_0+b_1x_0)-\sqrt{(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})}t'_{\gamma}(n-2,-\frac{z_{1-\beta}}{\sqrt{(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})}})s)$，也就是

$(-\infty,(b_0+b_1x_0)+\sqrt{(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})}t_{1-\gamma}'(n-2,\frac{z_{1-\beta}}{\sqrt{(\frac{1}{n}+\frac{(x_0-\overline{X})^2}{S_{xx}})}})s)$。這種計算方法可輕易推至多維。舉例: $X'=(X_1,X_2,\cdots X_n)_{n\times p}$，$Y'=(y_1,y_2,\cdots y_n)$ 的迴歸模型 $Y=X\beta_{p\times 1}+\varepsilon$，這裡的 $\varepsilon$ 服從 $N_n(O_n,\sigma^2 I_{p\times p})$，我們想得到估計 $y_0=x_0'\beta+\varepsilon_0$ 的單邊有上界的 $\alpha$ 覆蓋率 $\gamma$ 信賴水準的容忍區間，在此模型中 $\beta$ 和 $\sigma^2$ 都是未知，首先我們找出 $y_0$ 的 $\alpha$ 分位數為 $x_0'\beta+z_{1-\alpha}\sigma$，再求取這個分位數的 $\gamma$ 信賴水準的信賴區間。 由迴歸分析的結論，我們用 $\hat{\beta}=(X'X)^{-1}X'Y$ 估計 $\beta$，同樣的我們知道 (1) $\hat{\beta}$ 分配為 $N_p(\beta,\sigma^2(X'X)^{-1})$ (2) $\frac{s^2(n-(p+1))}{\sigma^2}$ 分配為 $\chi^2_{n-p}$，其中 $s^2=\frac{1}{n-p}(Y-X\hat{\beta})'(Y-X\hat{\beta})$ (3) $\hat{\beta}$ 和 $s^2$ 獨立。因此我們知道

$$\frac{\frac{x_0'\hat{\beta}-x_0'\beta}{\sigma\sqrt{x_0'(X'X)^{-1}x_0}}-z_{1-\alpha}\frac{1}{\sqrt{x_0'(X'X)^{-1}x_0}}}{\frac{s}{\sigma}}$$ 分配為 $t'(n-(p+1),-z_{1-\alpha}\frac{1}{\sqrt{x_0'(X'X)^{-1}x_0}})$

是一個樞紐量。而且

$P(x_0'\hat{\beta}-t'_{\gamma}(n-(p+1),-z_{1-\alpha}\frac{1}{\sqrt{x_0'(X'X)^{-1}x_0}})\hat{\sigma}\sqrt{x_0'(X'X)^{-1}x_0})>x_0'\beta+z_{1-\alpha}\sigma)=\gamma$，最後我們得到這個分位數 $x_0'\beta+z_{1-\alpha}\sigma$ 的 $\gamma$ 信賴水準的信賴區間為

$(-\infty,\ x_0'\hat{\beta}-t'_{\gamma}(n-(p+1),-z_{1-\alpha}\frac{1}{\sqrt{x_0'(X'X)^{-1}x_0}})s\sqrt{x_0'(X'X)^{-1}x_0}))$，也就是

$(-\infty,\ x_0'\hat{\beta}+t'_{1-\gamma}(n-(p+1),z_{1-\alpha}\frac{1}{\sqrt{x_0'(X'X)^{-1}x_0}})s\sqrt{x_0'(X'X)^{-1}x_0}))$。

4-4. 利用分位數尋找貝氏的容忍區間

在第三章我們提到的信賴區間是由一組給定的樣本計算出來的範圍，由於樣本是隨機的，所以我們可以知道這個區間也是隨機的，而要估計的參數卻是一個定值，所以當我們得到一個 90%信賴區間，意思是說知道大約會有 90%的樣本點蓋住這個參數，但是當樣本固定時參數落在裡面的機率不是 0 就是 1，但是相反的在貝氏的想法中，參數本身並不是一個定值，它是有分配的，通常我們稱這個分配叫做先驗分配，所以就像樣本點一樣參數本身是隨機的，所以這個假設使得我們可以說這個參數會有某個機率落在估計區間裡面，而不像信賴區間一樣只能說機率不是 0 就是 1，由於他們是如此的不一樣，所以在貝氏的假設下我們稱估計區間為 credible(可信的、可靠的) interval 而非信賴區間。那麼在貝氏的假設下，我們要怎麼得到參數的容忍區間呢?其實方法還是差不多的，參考文獻回顧中 2-3.貝氏的容忍區間，比起之前參數固定時的方法，只是多了一個步驟，我們必須先

得到參數的後驗分配。給定一組樣本資料$\underset{\sim}{X}$，我們可以計算$\theta$的後驗分配，再藉由$\theta$的後驗分配求$F_{\theta}^{-1}(\beta)$(單一樣本$x_i$的$\beta$分位數)的後驗分配，則這個分配的$\gamma$分位數$b(\underset{\sim}{X})$就是單邊有上界的貝氏$\beta$覆蓋率$\gamma$信賴水準的容忍區間$(-\infty, b(\underset{\sim}{X}))$。舉例:一組樣本$\underset{\sim}{X}=(x_1, x_2, \cdots x_n)$其分配為$\exp(\theta)$，這裡的$\theta$的先驗分配為$\exp(\alpha)$，想得到估計新觀察資料貝氏單邊有上界的$\beta$覆蓋率$\gamma$信賴水準的容忍區間，所以利用上面的方法，我們先找出$\theta$的後驗分配為$\Gamma(n+1, \alpha+\sum_{i=1}^{n} x_i)$，藉由$\theta$的後驗分配求$F_{\theta}^{-1}(\beta)=-\ell\mathrm{n}(1-\beta)/\theta$的γ信賴水準的 credible interval 為

$$\frac{-2\ell n(1-\beta)(\alpha+\sum_{i=1}^{n} x_i)}{\chi_{\gamma}^{2}(2n+2)}$$

，因此，新觀察資料的貝氏單邊有上界的$\beta$覆蓋率$\gamma$信賴水準的容忍區間就是$(0, \frac{-2\ell n(1-\beta)(\alpha+\sum_{i=1}^{n} x_i)}{\chi_{\gamma}^{2}(2n+2)})$。

4-5. 利用單邊容忍區間尋找雙邊容忍區間

上面的文章內容和所舉的例子都只有單邊$\beta$覆蓋率$\gamma$信賴水準的容忍區間，但是若我們希望找的是雙邊$\beta$覆蓋率$\gamma$信賴水準的容忍區間，那該怎麼辦呢?其實作法就像找雙邊$\gamma$信賴水準的信賴區間一樣。要找一個雙邊$\gamma$信賴水準的信賴區間只要讓$\gamma_1+\gamma_2=1-\gamma$，然後選取區間$(a(\underset{\sim}{X}), b(\underset{\sim}{X}))$滿足$\mathrm{P}(\theta \le a(\underset{\sim}{X}))=\gamma_1$並且$\mathrm{P}(\theta \ge b(\underset{\sim}{X}))=\gamma_2$，這樣子我們就可以得到$\mathrm{P}(a(\underset{\sim}{X}) \le \theta \le b(\underset{\sim}{X})) \ge \gamma$，所以區間$(a(\underset{\sim}{X}), b(\underset{\sim}{X}))$就是一個雙邊$\gamma$信賴水準的信賴區間。把這個想法和4-1.(利用分位數的信賴區間找常態分配的容忍區間)的方法綜合一下，我們就會得到一個雙邊的容忍區間。計算的步驟是:首先求出單一觀察值的$\beta_1$分位數和$\beta_2=\beta_1+\beta$分位

數，再分別求$\beta_1$分位數的單邊有下界的$\gamma_1$信賴水準的信賴區間和$\beta_2$分位數的單邊有上界的$\gamma_2 = 1+\gamma-\gamma_1$信賴水準的信賴區間，把這兩個區間綜合起來就會成為一個$\beta$覆蓋率$\gamma$信賴水準的容忍區間。我們用數學式子來說明:假設

$\underset{\sim}{X} = \left(X_1, X_2, ..., X_n\right)$且$X_1, X_2, ..., X_n, X_{n+1}$為iid，這裡的$X_{n+1}$為新的觀測值。若

$P_{X_{n+1}|\theta,\underset{\sim}{X}}\left(T_1(\underset{\sim}{X}) < X_{n+1}\right) > 1-\beta_1$且$P_{X_{n+1}|\theta,\underset{\sim}{X}}\left(X_{n+1} < T_2(\underset{\sim}{X})\right) > \beta_2$，則我們可以得到

$P_{X_{n+1}|\theta,\underset{\sim}{X}}\left(X_{n+1} \in \left(T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})\right)\right) > 1-\beta_1-(1-\beta_2) = \beta_2-\beta_1$，若

$P_{\underset{\sim}{X}|\theta}\left(P_{X_{n+1}|\theta,\underset{\sim}{X}}\left(T_1(\underset{\sim}{X}) < X_{n+1}\right) > 1-\beta_1\right) \geq \gamma_1$且$P_{\underset{\sim}{X}|\theta}\left(P_{X_{n+1}|\theta,\underset{\sim}{X}}\left(X_{n+1} < T_2(\underset{\sim}{X})\right) > \beta_2\right) \geq \gamma_2$，那麼$P_{\underset{\sim}{X}|\theta}\left(P_{X_{n+1}|\theta,\underset{\sim}{X}}\left(X_{n+1} \in \left(T_1(\underset{\sim}{X}), T_2(\underset{\sim}{X})\right)\right) > \beta_2-\beta_1\right) > \gamma_1+\gamma_2-1$。因為希望

$\beta_2-\beta_1=\beta$，則$\beta_2=\beta_1+\beta$，所以我們可以取$\beta_1=(1-\beta)/2$，

$\beta_2=\beta_1+\beta=(1+\beta)/2$，所以取$T_1(\underset{\sim}{X})$為樣本的$(1-\beta)/2$分位數，$T_2(\underset{\sim}{X})$為樣本的$(1+\beta)/2$分位數。同樣的希望$\gamma_1+\gamma_2-1=\gamma$，則取$\gamma_1=(1+\gamma)/2$，

$\gamma_2=1+\gamma-\gamma_1=(1+\gamma)/2$，所以取$(1-\beta)/2$分位數的單邊有下界的$(1+\gamma)/2$信賴水準的信賴區間$(T_1(\underset{\sim}{X}),\infty)$和$(1+\beta)/2$分位數的單邊有上界的$(1+\gamma)/2$信賴水準的信賴區間$(-\infty, T_2(\underset{\sim}{X}))$，把這兩個區間綜合起來就會成為一個$\beta$覆蓋率$\gamma$信賴水準的容忍區間。

舉個例子來說: $\underset{\sim}{X}=(x_1, x_2, \cdots x_n)$來自$\mathrm{N}(\mu,\sigma^2)$，想得到一個估計新的觀測值雙邊$\beta$覆蓋率$\gamma$信賴水準的容忍區間，這裡的$\mu$和$\sigma^2$皆是未知，首先我們找出這個分配的$(1-\beta)/2$分位數為 $\mu+z_{(1+\beta)/2}\sigma$和$(1+\beta)/2$分位數為$\mu+z_{(1-\beta)/2}\sigma$，可以發現區間$(\mu+z_{(1+\beta)/2}\sigma, \mu+z_{(1-\beta)/2}\sigma)$會包含這個分配$\beta$比率，則由樞紐量求出$(1-\beta)/2$分位數的單邊有下界的$(1+\gamma)/2$信賴水準的信賴區間為

$(T_1(\underset{\sim}{X}),\infty)=(\overline{X}-\frac{t_{(1-\gamma)/2}{}'(n-1,z_{(1+\beta)/2}\sqrt{n})}{\sqrt{n}}s,\infty)$ 和 $(1+\beta)/2$ 分位數的單邊有上界的

$(1+\gamma)/2$ 信賴水準的信賴區間為

$(-\infty,T_2(\underset{\sim}{X}))=(-\infty,\overline{X}+\frac{t_{(1-\gamma)/2}{}'(n-1,z_{(1-\beta)/2}\sqrt{n})}{\sqrt{n}}s)$ ，也就是

$P_{\underset{\sim}{X}|\theta}(T_1(\underset{\sim}{X})\le\mu+z_{(1+\beta)/2}\sigma)=(1+\gamma)/2$ 且 $P_{\underset{\sim}{X}|\theta}(\mu+z_{(1-\beta)/2}\sigma_1\le T_2(\underset{\sim}{X}))=(1+\gamma)/2$ ，再把

這兩個區間綜合起來就會成為一個 $\beta$ 覆蓋率 $\gamma$ 信賴水準的容忍區間

$(T_1(\underset{\sim}{X}),T_2(\underset{\sim}{X}))$ 為 $(\overline{X}-\frac{t_{(1-\gamma)/2}{}'(n-1,z_{(1+\beta)/2}\sqrt{n})}{\sqrt{n}}s,\overline{X}+\frac{t_{(1-\gamma)/2}{}'(n-1,z_{(1-\beta)/2}\sqrt{n})}{\sqrt{n}}s)$ ，這是

因為藉由

$P_{\underset{\sim}{X}|\theta}(T_1(\underset{\sim}{X})\le\mu+z_{(1+\beta)/2}\sigma\le\mu+z_{(1-\beta)/2}\sigma\le T_2(\underset{\sim}{X}))\ge(1+\gamma)/2-(1-(1+\gamma)/2)=\gamma$ ，可以

推出 $P_{\underset{\sim}{X}|\theta}(P_{X_{n+1}|\theta}(X_{n+1}\in(T_1(\underset{\sim}{X}),T_2(\underset{\sim}{X})))>\beta)\ge\gamma$ 。

4-5. 關於非工業統計方面容忍區間的運用

在文章的最前面我們提過容忍區間比較常被使用在工業統計，也就是說容忍區間不僅僅只能用在工業統計，除了工業統計之外，容忍區間在別的方面也很有貢獻，以下就是介紹兩篇不是工業統計方面而使用容忍區間的論文。當然，除了我們介紹的這兩篇論文，還有很多有關於容忍區間使用在非工業統計的文章，只是礙於文章篇幅，所以在本文內我們只介紹這兩篇論文中和本文主要討論的容忍區間有關係的內容。

4-5.1 Prediction Approaches to Sequentially Searching for an Optimal Dose

第一篇論文是，Prediction Approaches to Sequentially Searching for an Optimal Dose，這是一篇 1989 年的論文，作者是 W. J. Shih，本篇論文主要是在討論如何用容忍區間找出某些用藥的適當的劑量。在醫學上所使用的藥劑雖然可以治癒我們的疾病，但是如果大量的使用同一種藥劑或許反而會有反效果，可能是藥劑本身含有毒性，若是太大量的使用反而會破壞人體器官正常的功能，舉例來說:嗎啡有止痛的效果，適當的使用可以紓緩病人身體的疼痛，但是如果過度的增加病人的使用劑量，反而會讓病人上癮，所以這篇文章的目標就是找出在可以容忍的範圍內，讓藥劑療效達到最高的藥劑使用量。雖然也可以找出單邊有下界的區間，這樣子所得到的就會是在有藥效的前提下藥劑的最小用量，但是文章中主要討論的區間是單邊有上界的容忍區間，找出可以容忍的藥劑使用量的最大值。文章中提到 Eichhorn 和 Zacks (1973)兩個人定義出藥劑最大使用量，在藥劑毒性沒有超出某個已知數值($\zeta$)下的機率至少為某個比率($\gamma$)的前提下，他們所使用的統計模型有兩個，在這裡我們只介紹其中一個和本文相關的模型:讓 $y(x)$ 表示在藥劑使用量 $x$ 下所產生的毒性反應，$x$ 和 $y$ 都是測量得到，必要時還會對他們做數學轉換，例如:對於觀測值取 log，再讓 $x_0$ 表示身體可以自然消化的毒性的藥劑量，所以在藥劑使用量 $x$ 在給定範圍內，$y(x)$ 服從一個常態分配，其期望值為 $E(y(x)\,|\,x)=\beta(x-x_0)$，變異數為 $Var(y(x)\,|\,x)=\sigma^2(x-x_0)^2$，這裡的 $\beta$ 和 $\sigma$ 為未知參數，在參數固定下，若我們希望 $P(y\le\zeta\,|\,x)\ge\gamma$，令 $X^{(n)}=(x_1,x_2,\cdots x_n)$，$Y^{(n)}=(y_1,y_2,\cdots y_n)$ 代表已收集到的資料，則在藥劑使用量為最大值的前提下，因

為 $P(y \le \zeta \mid x) = P(\frac{y-\beta(x-x_0)}{\sigma(x-x_0)} \le z_{1-\gamma} \mid x) \ge \gamma$，所以藥劑最佳使用量為

$\xi_\gamma = x_0 + \zeta/(\beta + \sigma z_{1-\gamma})$，$P(y \le \zeta \mid x)$ 表示毒性在可容忍範圍 $(-\infty, \zeta)$ 的機率，假設我們從 n 個病人身上調查到了 $X^{(n)}$ 和 $Y^{(n)}$，根據這些收集到的資料，我們希望找到 $x_{n+1}$(為 $X^{(n)}$ 和 $Y^{(n)}$ 的函數)使得 $P_{Y^{(n)}}(P_{y_{n+1}}(y_{n+1} \le \zeta \mid x_{n+1},(X^{(n)},Y^{(n)})) \ge \gamma) \ge \alpha$，此為給定歷史資料 $(X^{(n)},Y^{(n)})$ 和在藥劑使用量 $(x_{n+1})$ 為最大值的前提下，毒性 $(y_{n+1})$ 在可容忍範圍 $(-\infty, \zeta)$ 的機率，文章中用了兩種模型尋找 $x_{n+1}$，其中一種模型為: $P_{Y^{(n)}}(P_{y_{n+1}}(y_{n+1}(x) \le \zeta \mid x_{n+1},(X^{(n)},Y^{(n)})) \ge \gamma) \ge \alpha$。，這個模型和我們前面所討論的容忍區間一樣。W. J. Shih（1989）沒有寫下推導的過程而只呈現結論，我們現在利用前面 4-1.提過的利用分位數的信賴區間找常態分配的容忍區間來推導這個模型。令 $K = y/(x-x_0)$，所以 $K \sim N(\beta, \sigma^2)$，則

$P_{Y^{(n)}}(P_{y_{n+1}}(y_{n+1}(x) \le \zeta \mid x_{n+1},(X^{(n)},Y^{(n)})) \ge \gamma) \ge \alpha$ 等同於

$P_{Y^{(n)}}(P_{y_{n+1}}(\frac{y_{n+1}(x)}{(x-x_0)} \le \frac{\zeta}{(x-x_0)} \mid x_{n+1},(X^{(n)},Y^{(n)})) \ge \gamma) \ge \alpha$，在此模型中 $\beta$ 和 $\sigma^2$ 未知，首先我們找出 $K$ 的 $\gamma$ 分位數為 $\beta + z_{1-\gamma}\sigma$，再取這個分位數 $\beta + z_{1-\gamma}\sigma$ 的 $\alpha$ 信賴水準的信賴區間，在此 $\beta$ 和 $\sigma^2$ 都是未知參數，利用 $K = y/(x-x_0)$ 服從標準常態分配， 再利用論文前面的結果我們可以得到 $K = y/(x-x_0)$ 的單邊有上界的 $\gamma$ 覆蓋率 $\alpha$ 信賴水準的容忍區間為 $(-\infty, \overline{K} - \frac{t_\alpha{}'(n-1,-z_{1-\gamma}\sqrt{n})}{\sqrt{n}}s)$，也就是

$(-\infty, \overline{K} + \frac{t_{1-\alpha}{}'(n-1,z_{1-\gamma}\sqrt{n})}{\sqrt{n}}s)$，藉由這個區間我們可以知道 $\zeta/(x-x_0)$ 也就是

$\overline{K} + \frac{t_{1-\alpha}{}'(n-1,z_{1-\gamma}\sqrt{n})}{\sqrt{n}}s$，所以可以再進一步推得

$$x_{n+1} = x_0 + \frac{\zeta}{\overline{K} + \frac{t_{1-\alpha}'(n-1, z_{1-\gamma}\sqrt{n})}{\sqrt{n}} s}$$ 。

4-5.2 SPEAKER VERIFICATION WITHOUT BACKGROUND SPEAKER MODELS

第二篇論文是， SPEAKER VERIFICATION WITHOUT BACKGROUND SPEAKER MODELS(2003)，作者是 Chun-Nan Hsu、Hau-Chung Yu 和 Bo-Hou Yang，這篇論文的出處是 Institute of Information Science, Academia, Sinica, Nankang 115, Taipei City, Taiwan，這篇文章主要是找一個聲音辨識系統 OSCILLO，這個系統可以依照接收到的音調辨識說話的人，也就是依照所接收到的聲波，這個系統會自動的辨識說話者的身分，舉例來說:若是現在市面上便於攜帶式的手機或是 PDA 都可以具備一個辨識音調的系統，那麼除了擁有者以外的人就不可以使用它。藉由音調辨識說話者身分的系統就可以產生這個功能，這篇文章提供的這個聲音辨識系統把被判斷說話者的一段說話音調分成數個樣本，再建構出這些樣本的單邊有下界$\beta$覆蓋率$\gamma$信賴水準的容忍區間 $(b(\underset{\sim}{X}),\infty)$，然後把$b(\underset{\sim}{X})$當做確認門檻，若是再收集到的一段說話音調落在這個容忍區間內就確認這個說話者的身分無誤。

## 第五章　結論

在意義上，信賴區間和容忍區間是不相同的。信賴區間是我們期望一個給定參數

(例如:期望值)落在裡面的區間,而容忍區間則是我們期望一個未來資料有特定的比率落在裡面的區間，當樣本數越來越大，信賴區間趨近於 0，而容忍區間則會趨近於一個定區間。但是在推導 $\beta$ 覆蓋率 $\gamma$ 信賴水準的容忍區間的數學式中，容忍區間可以藉由計算某些特定分位數的信賴區間得到,我們發現若不和文獻回顧 2-2.(常態分配的容忍區間)中對容忍區間做一樣的事先假設，而是先求出此分配的 $\beta$ 分位數再求取分位數的 $\gamma$ 信賴水準的信賴區間，會得到相同的結果，所以本篇論文對於容忍區間給予讀者另一種推導的想法。

對於未來研究方向大致上可分為三個方向:

(1) 如何比較不同的容忍區間？在不同的準則下何謂最好的容忍區間？

(2) 在不同的準則下如何找到覆蓋區間最好的估計量？何謂最好？

(3) 對於不同的樞紐量求得的區間該如何比較？

以上的問題我們尚未討論，未來可以加以討論，有些討論方向可見交通大學統計研究所黃景業(2003)之博士論文-眾數型態區間。

## 參考文章和書籍


1. R. Willink. (2004). Coverage intervals and statistical coverage intervals. *Metrologia* **41** (3):L5-L6 .

2. John W. Pratt, Jean D. Gibbons. (1981). Concepts of nonparametric theory. *New York : Springer-Verlag.*

3. George Casella, Roger L. Berger. (June 2001). Statistical inference 2nd edition.

*Wadsworth Pub Co.*

4. J. Aitchison, I.R. Dunsmore. (September 18, 1975). Statistical prediction analysis. *Cambridge University Press.*

5. B. L.Welch, H. W. Peers. (1963). On Formulate for Confidence Points Based on Integrals of Weighted Likelihood. *University of Leeds.*

6. Michael Hamada, Valen Johnson, Leslie M. Moore, Joanne Wendelberger. (2004). Bayesian prediction intervals and Their Relationship to Tolerance Intervals. *Technometrics, Volume 46, Number 4, 1 November 2004, pp. 452-459(8).*

7. Aitchison, J. (1964). Bayesian Tolerance Regions. *Journal of Royal Statistical Society*, Ser. B, 26, 161-175.

8. M. S. Bartlett. (1953). Approximate confidence interval. *JSTOR 12-19.*

9. Kendall M. G. (1979). The advanced theory of statistics. *London: Griffin, 4th ed.* 19.10~19.12.

10. D. V. Lindley. (1957). Fiducial distributions and Bayes' theorem. *JSTOR 102-107.*

11. D. T. Shirke, R. R. Kumbhar, D. Kundu. (2005). tolerance intervals for exponentiated scale family of distributions. *Journal of Applied Statistics. Routledge, part of the Taylor & Francis Group.* Volume 32.

12. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining. (2003). Introduction to Linear Regression Analysis 3rd edition. *The American*

*Statistician, Volume 57, pp. 67-67(1).*

13. I. Janiga, I. Garaj. (2005). On one-sided tolerance intervals of Normal distribution With Unknown Parameters. *Journal of the Applied Mathematics, Statistics and Informatics (JAMSI), No. 1.*

14. W. J. Shih. (1989). Prediction Approaches to Sequentially Searching for an Optimal Dose. *JSTOR. Biometrics.*

15. Chun-Nan Hsu, Hau-Chung Yu, Bo-Hou Yang. (2003). SPEAKER VERIFICATION WITHOUT BACKGROUND SPEAKER MODELS. *Institute of Information Science, Academia, Sinica, Nankang 115, Taipei City, Taiwan*.

16. BE Ellison. (1964). On Two-Sided Tolerance Intervals for a Normal Distribution. *JSTOR. The Annals of Mathematical Statistics.*