第三章

估計與假說檢定

周賓凰

July 8, 2011

母體和樣本的關係

- 母體是指「某一特定事物或變數的所有可能結果所成的集合」
- 樣本是母體的子集合
- 樣本擁有母體的資訊, 因此可透過有限的樣本來推估母體的特性

關於「母體」與「樣本」的一點評論

- 就擲銅板而言 (令H、T代表銅板的二面), 其「母體」為何呢?

關於「母體」與「樣本」的一點評論

- 就擲銅板而言 (令H、T代表銅板的二面), 其「母體」為何呢?
- 是$\{H, T\}$嗎? 但這是「出象空間」, 不是母體

較好的定義如下:

令A_i代表某一序列 (sequence) 擲銅板的結果所成的集合, 亦即:

$$A_i = \{\omega_{i1}, \omega_{i2}, \cdots\},$$

$Pr(\omega_{ij} = H) = p$, 所以母體定義為:

$$\Omega \equiv \bigcup_{i=1}^{\infty} A_i,$$

而所謂的樣本, 則定義為某一序列的子集合,

$$S_i(T) = \{\omega_{i1}, \omega_{i2}, \cdots, \omega_{iT}\}.$$

隨機抽樣與隨機樣本

最常見的是所謂的「隨機樣本(random sample)」。

例如: $x_1, x_2, \cdots, x_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

每個x_i, $i = 1, \cdots, n$, 都是一個隨機變數。

'$\overset{iid}{\sim}$'代表這n個隨機變數為獨立同態, 構成一個隨機樣本。

- 目標: 嘗試藉此樣本去推估母體的特性, 並進而作預測之用。
- 古典統計的作法是假設樣本由某一個分配所產生, 通常也將分配的型態假設為已知, 但分配的參數則為未知, 因此需要一些估計方法。
- 各種估計方法各有各的優缺點, 因此需要一些準則來判定一估計子或估計式是否夠好。

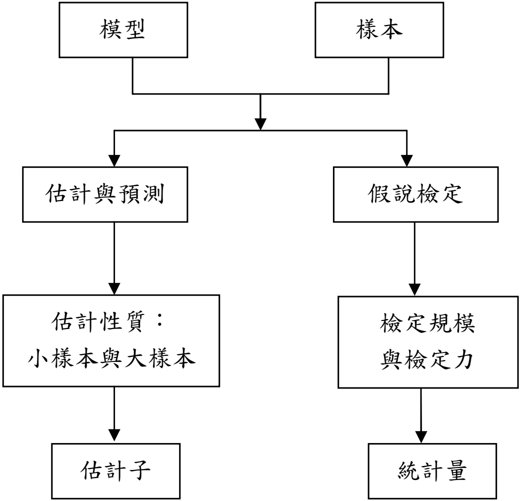


Figure: 3.1: 估計、預測與假說檢定的架構

定義3.1: 若$y_1, \cdots, y_T \sim f(\theta)$, 則估計子$\hat{\theta} = h(y_1, \cdots, y_T, X)$為樣本$(y_1, \cdots, y_T)$與其他已知變數$X$(可以多於一個變數) 的函數, 其中不含任何的未知參數或變數。$h(\cdot)$也是確定的函數。

- 估計方面, 我們是以樣本來代入估計子 (estimator) 中, 而得到估計值 (estimate)。
- 因為樣本是隨機的, 所以估計子也是隨機的。估計子 (或統計量) 的實際分配就稱為其樣本分配或抽樣分配 (sampling distribution), 表示其值會因樣本的不同而有所不同。

- 在假說檢定方面, 則是以樣本代入統計量 (statistic) 中, 以決定所關心的假說 (hypothesis) 是否為真。
- 跟估計子一樣, 統計量也是樣本的函數, 同樣其中不含任何的未知參數或變數。

估計子抽樣性質: 小樣本性質

- 所謂的小樣本性質 (small-sample properties) 指的是樣本規模 (sample size) 有限的情況下的統計性質。
- 主要的小樣本性質包括「不偏性 (unbiasedness」與「效率性 (efficiency)」。

令$\hat{\theta}$為某一參數θ的估計子

- 定義3.2 不偏性: 若$\hat{\theta}$的期望值等於真正的參數值, 則稱$\hat{\theta}$為θ的不偏估計子 (unbiased estimator), 亦即$E(\hat{\theta}) = \theta$。

令$\hat{\theta}$為某一參數θ的估計子

- 定義3.2 不偏性: 若$\hat{\theta}$的期望值等於真正的參數值, 則稱$\hat{\theta}$為θ的不偏估計子 (unbiased estimator), 亦即$E(\hat{\theta}) = \theta$。
- 定義偏誤 (bias) 為估計子期望值與參數值的差異,

$$Bias \equiv E(\hat{\theta}) - \theta,$$

則不偏性表示此偏誤為零。「不偏性」表示平均而言, 估計子的值會等於真正的參數值。

- 定義3.3 效率性: 一不偏估計子$\hat{\theta}$如果較另一個不偏估計子$\tilde{\theta}$變異數小, 亦即

$$var(\hat{\theta}) < var(\tilde{\theta}),$$

  則我們說$\hat{\theta}$比$\tilde{\theta}$有效率。

- 如果一估計子是所有不偏估計子裡變異數最小的, 則我們稱此估計子為效率估計子。

- 實務上, 不偏性估計子並不總是存在。
- 在這樣的情況下, 則估計子的選擇可能需要在「偏誤程度」與「變異程度」上作取捨。
- 均方誤 (mean-squared error, MSE):

$$
\begin{aligned}
MSE(\hat{\theta}) &\equiv E(\hat{\theta}-\theta)^2 \\
&= E[\hat{\theta}-E(\hat{\theta})]^2+[E(\hat{\theta})-\theta]^2 \\
&= var(\hat{\theta})+Bias(\hat{\theta})^2.
\end{aligned}
$$

MSE是估計子變異數與偏誤的平方之和, 因此變異數與 (或) 偏誤越大都會導致 MSE 越大。估計子的 MSE 越小也就表示整體而言, 估計的「績效」較佳。

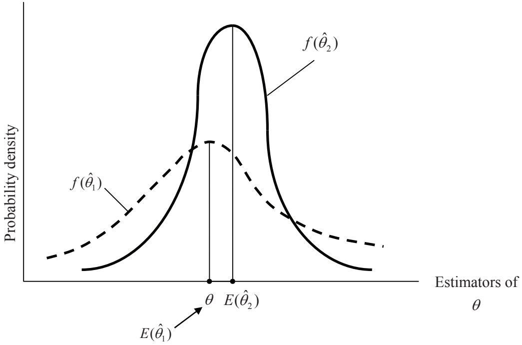


Figure: 3.2: 準確性與不偏性的取捨

3.3 估計子抽樣性質: 大樣本性質

- 大樣本性質 (large-sample properties) 探討估計子或統計量在樣本規模趨近無窮大時的統計性。
- 大數法則 (law of large numbers, LLN)
- 中央極限定理 (central limit theorem, CLT)

- 大數法則多指一估計子或統計量「收斂至某個常數」，而中央極限定理則多指「收斂至常態或卡方分配」。
- 因為估計子或統計量是由樣本計算而得，所以是樣本的函數，同樣會是隨機變數，而且統計性質會受到樣本規模大小的影響。

例如,$x_1, x_2, \cdots, x_T$代表一個隨機樣本，平均值為μ而變異數為σ^2，則其樣本平均值:

$$\bar{x}_T \equiv \frac{1}{T}\sum_{t=1}^{T} x_t$$

的分配的平均值與變異數分別為:

$$\bar{x}_T \sim (\mu, \frac{1}{T}\sigma^2).$$

當樣本規模趨近無窮大，則其變異數會趨近於零，所有的機率都集中在一點。這表示當樣本規模是無窮大時，不同的樣本所計算出來的樣本平均值都會等於真正的參數值μ。

- 定義3.4 機率上收斂 (convergence in probability): 如果對任一$\epsilon > 0$而言, 下式成立:

$$\lim_{T\to\infty} Pr(|x_T - a| > \epsilon) = 0,$$

則我們說隨機變數x_T在機率上收斂到一常數a。

- 定義3.4 機率上收斂 (convergence in probability): 如果對任一$\epsilon > 0$而言, 下式成立:

$$\lim_{T\to\infty} Pr(|x_T - a| > \epsilon) = 0,$$

則我們說隨機變數x_T在機率上收斂到一常數a。

- 一般會將機率上的收斂表達為:

$$plim\ x_T = a.$$

或是

$$x_T \xrightarrow{p} a.$$

- 定義3.5 $\hat{\theta}_T$是θ的一致性估計子, 假若

$$plim\ \hat{\theta}_T = \theta.$$

- 定義3.5 $\hat{\theta}_T$是θ的一致性估計子, 假若

$$plim\ \hat{\theta}_T = \theta.$$

- 定理3.1 弱式大數法則 (Kinchine's weak law of large numbers, WLLN): 若$\{x_t\}$一序列的 iid 隨機變數, 而平均值μ是有限的, 則

$$\bar{x}_T \xrightarrow{p} \mu.$$

這個 LLN 只要求μ 為有限, 變異數可以是無窮大。

定義3.6 幾乎確定地收斂 (convergence almost surely): 如果對任一$\epsilon > 0$而言, 下式成立:

$$Pr(\lim_{T\to\infty} |x_T - x| \leq \epsilon) = 1,$$

則我們說隨機變數x_T幾乎確定地收斂到隨機變數x。

一般會表達為:

$$x_T \overset{a.s.}{\to} x.$$

定理3.2 強式大數法則 (Kolmogorov's strong law of large numbers, SLLN)：若$\{x_t\}$一序列的獨立隨機變數，平均值為$\mu_t < \infty$，而變異數為σ_t^2，滿足$\sum_{t=1}^{\infty} \frac{\sigma_t^2}{t^2} < \infty$，則

$$\bar{x}_T - \bar{\mu}_T \xrightarrow{a.s.} 0,$$

其中$\bar{\mu}_T = \frac{1}{T} \sum \mu_t$

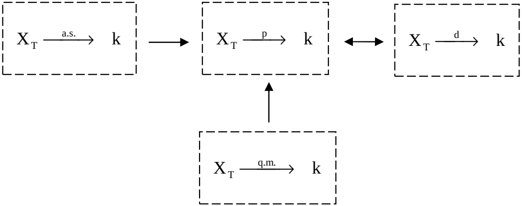


Figure: 3.3: 收斂到一常數

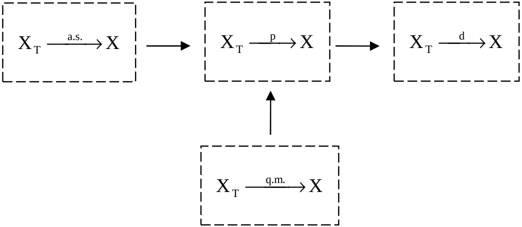


Figure: 3.4: 收斂到一隨機變數

- 定理3.3 Slutsky 定理: 對任一連續函數$g(\cdot)$而言, 下列成立

$$plim\ g(x_T) = g(plim\ x_T).$$

- 定理3.3 Slutsky 定理: 對任一連續函數$g(\cdot)$而言, 下列成立

$$plim\ g(x_T) = g(plim\ x_T).$$

- 定理3.4 如果一估計子$\hat{\theta}_T$滿足下面二個條件,

$$\lim_{T\to\infty} E(\hat{\theta}_T) = \theta, \tag{1}$$

$$\lim_{T\to\infty} var(\hat{\theta}_T) = 0, \tag{2}$$

則$plim\ \hat{\theta}_T = \theta$, 亦即$\hat{\theta}_T$是$\theta$的一致性估計子。

定義3.7 分配上收斂 (convergence in distribution):
如果當樣本趨近無窮大時，$\hat{x}_T$的 cdf $F_T(x)$趨近於x的 cdf $F(x)$，亦即對所有$x \in \Re$而言，下式成立

$$\lim_{T\to\infty} |F_T(x) - F(x)| = 0,$$

則稱$\hat{x}_T$在分配上收斂至x。
這個情況下，我們稱$F(x)$為$\hat{x}_T$的極限分配 (limiting distribution)。

- 定理3.5 Lindberg-Levy 中央極限定理:
  如果$\{x_t\}$是一序列的 iid 隨機變數, 而平均值與變異數分別為μ與σ^2, 二者都是有限的, 則$\sqrt{T}(\bar{x}-\mu)$在分配上收斂至常態,

$$\sqrt{T}(\bar{x}-\mu) \xrightarrow{d} \mathcal{N}(0,\sigma^2).$$

- 定理3.5 Lindberg-Levy 中央極限定理:
  如果$\{x_t\}$是一序列的 iid 隨機變數, 而平均值與變異數分別為μ與σ^2, 二者都是有限的, 則$\sqrt{T}(\bar{x}-\mu)$在分配上收斂至常態,

$$\sqrt{T}(\bar{x}-\mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

- 在此我們並未對x_t做任何特定分配假設, 而只要求其為 iid, 且前二個動差存在而已。

定理3.6 Lindberg-Feller 中央極限定理:

如果$\{x_t\}$是一序列獨立的隨機變數, 而平均值與變異數分別為μ與σ_t^2, 二者都是有限的, 且

$$\bar{\sigma}^2 = \lim_{T\to\infty} \frac{1}{T} \sum \sigma_t^2 < \infty,$$

則$\sqrt{T}(\bar{x}-\mu)$在分配上收斂至常態,

$$\sqrt{T}(\bar{x}-\mu) \xrightarrow{d} \mathcal{N}(0, \bar{\sigma}^2).$$

這個 CLT 允許變數間的變異數是不同的。

定理3.7 (多變量) 中央極限定理:

如果$\{x_t\}$是一序列 iid 的$(n \times 1)$的隨機變數向量, 而平均值與變異數分別為μ與Σ, 二者都是有限的,
則$\sqrt{T}(\bar{x} - \mu)$在分配上收斂至常態,

$$\sqrt{T}(\bar{x} - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma).$$

3.4 常用估計方法

- 動差法 (method of moments)
- 最小平方法 (least squares)
- 最大概似法 (maximum likelihood estimation)

3.4.1 動差法

- 在一些條件下，樣本動差會在機率上趨近其母體動差。而由於動差又是參數的函數，因此就可以利用倒推其反函數來求得參數估計值。

3.4.1 動差法

- 在一些條件下, 樣本動差會在機率上趨近其母體動差。而由於動差又是參數的函數, 因此就可以利用倒推其反函數來求得參數估計值。
- 令$\{x_t\}$為一序列的 iid 隨機變數, 假設其第r階動差存在, 亦即$m_r = E(x^r) < \infty$。當第r階動差存在, 根據大數法則可知其樣本平均在機率上會趨近母體動差:

$$\hat{m}_r = \frac{1}{T}\sum_{t=1}^{T} x_t^r \xrightarrow{p} m_r.$$

3.4.1 動差法

如果m_r與m_{2r}都存在的話,x_t^r的平均值與變異數如下:

$$
\begin{aligned}
E(x_t^r) &= m_r \\
var(x_t^r) &= m_{2r} - m_r^2,
\end{aligned}
$$

根據中央極限定理, 可知其極限分配為:

$$
\sqrt{T}(\hat{m}_r - m_r) \xrightarrow{d} \mathcal{N}(0, m_{2r} - m_r^2).
$$

- 動差法估計子並不一定是唯一的 (unique), 因為利用不同的動差有可能得到不同的估計子。
- 動差法估計子基本上是一致性估計子, 但一般而言卻不是有效率的估計子。尤其是當分配為已知的情況下, 動差法估計子只使用到一部分的資訊 (因為只有用到少數幾個動差條件), 所以效率性的確較差。

- 動差法的缺點, 也恰是其優點。
- 動差法利用到大數法則與中央極限定理, 而大數法則與中央極限定理的成立並不總是需要 iid 或是獨立。
- 當分配未知時 (尤其是變數並不服從任何特定分配時), 動差法有時所加 (impose) 的假設反而是較少的。

- Hansen (1982) 的一般化動差法 (generalized method of moments, GMM):
  Hansen將所謂的動差條件一般化到所有的期望值關係, 例如
  $$E(g(x_t)) = 0,$$
  其中$g(\cdot)$可以是線性或非線性的連續函數。
- GMM 法的另一個特色是動差條件的數目可以比要估計的參數數目多, 同時也允許資料有相當的序列相關與異質變異。

3.4.2 最小平方法

若$x_1, \cdots, x_T \overset{\text{iid}}{\sim} (\mu, \sigma^2)$, 則其實此關係可表達為:

$$x_t = \mu + e_t,$$

其中$e_t = x_t - \mu$為一誤差項 (error term), 代表實際上觀察到的隨機變數與其期望值之差。顯然$e_t \overset{\text{iid}}{\sim} (0, \sigma^2)$。
最小平方法的概念就是尋找一個參數估計值, 以使得誤差項 (事實上是殘差項) 的平方和最小, 亦即:

$$\min_{\{\mu\}} \sum_{t=1}^{T} (x_t - \mu)^2.$$

- 最小平方法的好處是計算簡單, 而且很多情況下 (尤其是對線性模型而言), 存在解析解。
- 缺點則是估計結果容易受到極端值 (或稱離群值 (outliers)) 的影響。

- 最小平方法的好處是計算簡單，而且很多情況下 (尤其是對線性模型而言)，存在解析解。
- 缺點則是估計結果容易受到極端值 (或稱離群值 (outliers)) 的影響。
- 一個解決的辦法就是不要將誤差值項平方，而是取絕對值，如下：

$$\min \sum_{t=1}^{T} |x_t - \mu|.$$

這就稱為「最小絕對變異 (miminum-absolute-deviation (MAD) 估計子」。

- 因為 MAD 估計子是最小化絕對變異, 因此極端值較不會因為「平方」而過份影響估計結果, 所以是比較「穩健」的估計法。
- 不過因為是對誤差取絕對值, 因此在求解上也比較困難 (因為求一階條件時需考慮正負號), 連帶地, 估計子的統計性質也比較複雜。

3.4.3 最大概似法 (MLE)

定義3.8 概似函數: 若$f(y_1, ..., y_T|\theta)$為樣本$(y_1, ..., y_T)$的聯合 pdf, 而未知參數為θ, 我們定義概似函數為任何與其聯合 pdf 同比例的函數:

$$L(\theta|y_1, ..., y_T) \propto f(y_1, ..., y_T|\theta).$$

概似函數在函數形式上不一定要完全等於其 pdf, 而只要是有同比例的形式就可。

- 雖然概似函數看起來跟樣本的聯合 pdf 長得一模一樣，但是前者是參數θ的函數，而後者是樣本$(y_1, ..., y_T)$的函數。

- 雖然概似函數看起來跟樣本的聯合 pdf 長得一模一樣, 但是前者是參數θ的函數, 而後者是樣本$(y_1, ..., y_T)$的函數。
- 所謂的最大概似法估計子 (maximum liklihood estimator, MLE) 就是尋找一組參數的估計值$\tilde{\theta}$, 以極大化其概似函數:

$$\max_{\{\theta\}} L(\theta|y_1, ..., y_T).$$

- 所以 MLE 在概念上是視樣本$(y_1, ..., y_T)$為已知 (或是給定的 (fixed)), 而 "最佳" 的參數值就是使得觀察到的樣本機率最大者。

- 極大化概似函數等同於極大化其對數概似函數 (log-likelihood function) $\ell(\theta|X) = lnL(\theta|X)$, 因為

$$\begin{aligned}\frac{\partial \ell}{\partial \theta} &= \frac{\partial \ell}{\partial L}\frac{\partial L}{\partial \theta}\\ &= \frac{1}{L}\frac{\partial L}{\partial \theta}.\end{aligned}$$

- 概似函數常是個別 pdf 的連乘的函數, 因此微分較困難, 而取對數之後,「乘號」變為「加號」, 所以處理起來較簡單。
- log-likelihood 的一階條件$\frac{\partial lnL(.)}{\partial \theta} = 0$稱為「概似方程式 (likelihood equation)」。

在一些常規條件下,MLE 的統計性質如下:

1. 一致性: $plim\ \tilde{\theta}_{ML} = \theta$。
2. 漸近常態
$$\tilde{\theta}_{ML} \overset{a}{\sim} \mathcal{N}(\theta, I(\theta)^{-1}),$$
其中$I(\theta) = -E[\frac{\partial^2 l}{\partial\theta\partial\theta'}]$為訊息矩陣 (information matrix)。
3. 漸近有效率: 當樣本趨近無窮大時, 其變異數會趨近 Cramer-Rao 下限 (Cramer-Rao lower bound; CRLB), 是有效率估計子可能達到的最小變異數。
4. 不變性 (invariance): $g(\theta)$的 MLE 是 $g(\tilde{\theta}_{ML})$。

- MLE 一般只有大樣本統計性質, 小樣本性質是未知的, 也就是一般而言, 我們並不知道 MLE 是否為不偏、或是效率估計子。
- 漸近常態」也是很重要的性質, 因為 MLE 的小樣本分配一般為未知, 因此「漸近常態」的性質允許我們在樣本夠大時, 作區間估計與假說檢定。

漸近效率性

- 當樣本趨近無窮大時, MLE的共變異矩陣會是訊息矩陣的反矩陣, 而此矩陣就是所謂的 Cramer-Rao 下限, 是所有效率估計子可能達到的變異數下限。
- 換言之, 只要一不偏估計子的變異數 (或是共變異矩陣) 等於 CRLB, 那麼它就是效率估計子。
- 就 MLE 而言, 我們未必有不偏性, 但有一致性, 而其變異數又會趨近 CRLB, 所以我們說 MLE 具備「漸近效率性」。

訊息矩陣

- 為什麼$I(\theta) = -E[\frac{\partial^2 \ell}{\partial\theta\partial\theta'}]$稱為訊息矩陣呢？先忽略期望值$E(\cdot)$，只看其中的元素。
- log-likelihood 函數對參數的一階微分代表斜率，對有極大值的函數 (也就是凹函數) 而言，斜率隨著參數值遞增，先為正，而後遞減；當斜率為零時，即為 MLE。
- 至於第二階微分則代表斜率的變化 (也稱為 Hessian)，對凹函數而言，其值為負 (因為斜率是遞減的)。

- Hessian 的絕對值越大表示斜率的變化越大。這表示如果在 MLE(也就是頂點) 的‘$|Hessian|$’越大，那一點的曲率也就越大，同時也就表示在 MLE 那一點的「機率」要比其他估計值 (點) 大的多，表示 MLE 估計值是變精確的。
- MLE 極限分配的變異數 (共變異矩陣) 正是訊息矩陣取倒數 (反矩陣)。所以，當曲率越大時，表示 MLE 這一點是越有資訊內涵的 (informative)。
- 如果 log-likelihood 函數是比較平坦的，則 MLE 那一點所對應的機率與其他點所對應的機率差別不大，表示這個 MLE 並不是太有資訊內涵。因為這樣的原因，所以稱$I(\cdot)$為訊息矩陣。

和前兩種方法比較

- 相較於動差法,MLE 的優點是將整個 (或一部分) 分配的資訊包含進來, 而動差法則僅利用到一小部分的資訊。
- MLE 與最小平方法最大的不同是對 MLE 而言, 樣本的分配 (通常是常態分配) 必須是已知的。如果分配假設有誤的話, 那麼上述的統計性質也就不一定成立了。

3.5 假說檢定

在給定樣本$Y \sim f(y \mid \theta)$下, 檢定的程序可分為幾個步驟。

1. 設立虛無假說H_0與對立假說H_1。概念上, 令Ω為θ的參數空間, 而Ω_0與Ω_1皆為Ω的子集合, 且$\Omega_0 \cap \Omega_1 = \phi$, 虛無假說與對立假說可表達為

$$H_0 : \theta \in \Omega_0, \quad \text{vs} \quad H_1 : \theta \in \Omega_1.$$

2. 計算統計量 $h = h(y_1, \cdots, y_T)$.
3. 決定拒絕域 C, 若$h \in C$, 則拒絕H_0, 否則不拒絕 H_0。

- 一些教科書說我們感到興趣的假說應設為對立假說。例如, 我們想檢驗每包餅乾的平均重量是否大於300克, 則H_0與 H_1分別為

$$H_0 : \mu = 300, \quad H_1 : \mu > 300.$$

- 假如樣本的平均重量$\bar{x} = 256$克, 這個時候, 研究者會設立假說為:

$$H_0 : \mu = 300, \quad H_1 : \mu < 300.$$

- 當我們觀察到 $\bar{x} < 300$, 可知資料絕對不會支持$\mu > 300$的假說, 如果我們至此就下μ不大於 300 的結論, 那麼看來就完全沒有做統計檢定的必要。
- 但是, 現在我們雖觀察到$\bar{x} < 300$, 卻可轉而檢驗μ是否顯著地小於 300, 因此可設$H_1 : \mu < 300$。如果證據支持H_1, 則我們不僅可說μ不大於 300, 還可說μ是顯著小於 300, 因此有更強的證據不支持我們原先的想法!

型I誤差 (Type I error) 與型 II 誤差 (Type II error)

- 型 I 誤差: 虛無假說H_0為真, 但拒絕H_0的機率:

$$Pr(\text{Type I error}) = Pr(\text{reject } H_0 \mid H_0 \text{ is true}).$$

- 型 II 誤差: 對立假說H_1為真, 但不拒絕H_0的機率,

$$Pr(\text{Type II error}) = Pr(\text{do not reject } H_0 \mid H_1 \text{ is true}).$$

- 因為H_1都是以複合假說(composite hypothesis) 的型式, 所以要確切知道型 II 誤差的大小較困難。
- 相反地, 因為在虛無假說為真下, 分配多為已知, 所以較能控制型 I 誤差的大小。所謂的「顯著水準(significance level)」, 就是型 I 誤差的機率。
- 正確拒絕虛無假說的機率, 也就是當H_1為真, 而我們正確地拒絕了H_0, 則稱為「檢定力(power)」,

$$\text{power} = 1 - Pr(\text{Type II error}).$$

- 若$x_1, x_2, \cdots x_T \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, 而我們想檢定的假說為:

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0.$$

因為樣本來自常態分配, 其樣本平均也會服從常態分配:

$$\overline{x} \sim \mathcal{N}(\mu, \frac{1}{T}\sigma^2)$$

- 簡單運算, 得:

$$\begin{aligned} t &= \frac{\frac{\bar{x}-\mu}{\sqrt{\frac{\sigma^2}{T}}}}{\sqrt{\frac{(T-1)\hat{\sigma}^2/\sigma^2}{T-1}}} \\ &= \frac{\overline{x} - \mu}{\sqrt{\frac{\hat{\sigma}^2}{T}}} \end{aligned} \tag{3}$$

- 在H_0為真下, 我們知道$\mu = \mu_0$, 可以計算出其值:

$$t_0 = \frac{\overline{x} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{T}}}.$$

- 令顯著水準為α, 且令$t_{T-1,\frac{\alpha}{2}}$為自由度為$T-1$的 t 分配的$1-\frac{\alpha}{2}$的百分位(percentile),

$$Pr(t \leq t_{T-1,\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}.$$

$t_{T-1,\frac{\alpha}{2}}$是對應此一百分位的臨界值。所以拒絕域C為:

$$C = \{t \leq -t_{T-1,\frac{\alpha}{2}}\} \cup \{t \geq t_{T-1,\frac{\alpha}{2}}\}.$$

只要$t_0 \in C$, 就拒絕虛無假說, 否則就不拒絕。

- 另一個方式是構建選定顯著水準下的「信賴區間」, 並就此區間是否包含虛無假說所指定的參數值。
- 實證上更常使用的是「P 值 (p value)」, 代表的是「可拒絕虛無假說的最低顯著水準。假如一個統計量h的樣本估計值為h_0, 一統計量的 P 值是指「在虛無假說為真下, 統計量h可能落於h_0之外而支持對立假說的機率值」。
- 計算 P 值的好處是不用去查表。我們只要知道 P 值就可以知道在怎樣的顯著水準下會拒絕或支持虛無假說。

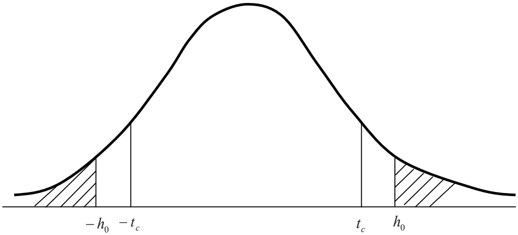


Figure: 3.5: P 值

補充說明

- 一些教科書會用α與$\alpha/2$來作為單尾與雙尾檢定時,P 值的比較基礎。這個觀念是不正確的。「正統」的作法裡,沒有所謂的0.5%或2.5%的顯著水準。正確的作法應是就選定的對立假說去計算相關的 P 值，並與1%或5%比較。
- 雖然假說檢定時,「顯著水準」是由我們自己決定的，但實際上這個名目上的顯著水準未必是真正我們所承擔的風險。這是因為實際上統計量的分配未必與我們假設的符合。因而可能拒絕虛無假說過多或不足。

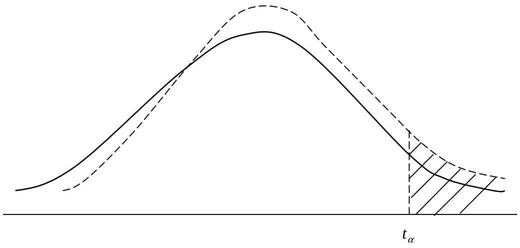


Figure: 3.6 實際分配與假設分配下的拒絕機率

- 標準差 (standard deviation)」一般是指某個樣本變數的標準差的估計值, 是觀察不到的。「標準誤 (standard error)」則是指一估計子的標準差的估計值, 是觀察的到的。
- 就常態分配變數而言, 將均數平移, 再除以其標準差後, 會有標準常態分配。一估計子經平移除以其標準誤後, 通稱為t統計量 (t statistic), 但不一定會有t分配, 除非原本估計子就服從常態分配。
- 在作假說檢定時, 有一個常用的經驗法則是在5%的顯著水準下, 如果t統計量的絕對值大於2, 我們就拒絕虛無假說。

- 一篇實證論文在報告實證結果前，有必要對資料的來源與特性做一些說明。例如，資料取自哪一資料庫？樣本期間為何？樣本是橫斷面、時間序列或是合併資料型態呢？如果是時間序列資料，必須再說明資料的頻率 (frequency)，亦即是日、月、季或者年資料。另外，近年來財務的研究已發展到研究使用日內 (intraday) 資料，也就是短於一日的資料，例如是五分鐘、十分鐘或是逐筆交易資料 (tick data)，也稱為高頻率 (high frequency) 或超高頻率 (ultra-high frequency) 資料。
- 實證研究的論文在呈現正式的實證結果之前，一般都會報告樣本的一些敘述統計量，稱為「summary statistics」，以便讓讀者對所使用的資料的特性有一些初步的概念。

- 「取樣原則」(sample selection criteria)：較嚴謹的作法應是說明原始樣本、選樣標準，以及最終留下的樣本。報告原始樣本與最終樣本有助於讀者瞭解研究所使用的樣本的代表性。此外，尤其要注意的是，在「選樣原則」上應避免造成樣本的選擇偏誤 (selectivity bias)。
- 避免 (或降低) 這類因素所衍生的問題有兩個作法，第一是盡量完整地將樣本納入，第二是以市值加權投資組合 (value-weighted portfolio) 來做分析。

單變量敘述統計量

- 包括變數的集中趨勢統計量如樣本平均 (sample mean)、中位數 (median) 與衆數 (mode) 等。另外, 一般常會報告的還包括樣本標準差或變異數、偏態係數、峰態係數、最小值與最大值等。在統計檢定上, 最常見的是檢定研究變數的分布是否服從常態的「常態性檢定」(normality test) 以及檢定研究變數的平均值是否異於 0。如果是時間序列資料, 則在必要的情況下, 也會報告不同階次 (orders) 的自我相關係數 (autocorrelation coefficient)。

- 離散趨勢主要包括樣本變異數或樣本標準差，以及全距 (range) 與四分位距 (interquartile range)，而樣本標準差 $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ 只是將樣本變異數開根號。為何離散趨勢會是重要的統計指標呢？讓我們看一個例子。

例子

- 以近來討論頗多的地球暖化 (global warming) 問題為例。美國前任副總統高爾在「不願面對的真相」(An Inconvenient Truth) 紀錄片中指出, 二氧化碳排放的增加是造成地球暖化的主因。有些人認為地球平均溫度上升1度有那麼嚴重嗎? 全球溫度的平均上升只是第一個動差的趨勢, 此外還伴隨著第二個動差的上升。雖然均溫上升1度, 但在南北極的溫度卻可能上升超過10度!根據報導,2007年7月北極竟然出現攝氏22度的高溫, 進一步造成冰山、冰棚的急速溶化, 全球氣候也產生劇變。

問題

- 以資產報酬而言, 第二動差不存在是否表示資產風險無窮大呢? 而在資產報酬存在「風險報酬抵換」(risk-return trade-off) 關係的概念下, 風險無窮大不就隱含投資人需要無窮大的報酬才願意承擔風險嗎? 進一步來說, 因為資產報酬不可能是無窮大, 沒有投資人會願意去交易金融資產, 因此股市也將瓦解 (collapse)! 這樣的推論與論述有什麼問題嗎?

- 在二階動差不存在的情況下, 有時研究者會將重點放在「全距」(range) 或是樣本四分位距 (sample interquartile range)。「全距」(range) 是樣本中, 最大值與最小值的差距。令 $x = (x_1, \cdots, x_T)$, 則:

$$\text{range}(\ x\ ) = \max(x_1, \cdots, x_T) - \min(x_1, \cdots, x_T)$$

- 另一個常見的指標則是樣本四分位距, 定義為樣本百分之七十五分位數與百分之二十五分位數的差距。可見樣本四分位距捕捉的是樣本百分之五十的變動範圍。

- 樣本偏態與峰態的公式分別如下:

$$\text{skewness} = \frac{\hat{m}_3}{\hat{\sigma}^3} \tag{4}$$

$$\text{kurtosis} = \frac{\hat{m}_4}{\hat{\sigma}^4} \tag{5}$$

其中, $\hat{m}_k$ 是樣本的第 k 階中央動差, 估計如下:

$$\hat{m}_k = \frac{1}{T-1}\sum_{i=1}^{T}(x_t - \bar{x})^k.$$

不過, 許多統計軟體在報告峰態時會再減去3, 以與常態分配比較。

不同階次動差不存在的含意

- 例如, 變數 x 存在前二階動差 μ_x 與 σ_x^2 , 但不存在第四階動差, 而變數 y 則只存在一階動差 μ_y , 但不存在二階動差(亦即 $\sigma_y^2 = \infty$)。那麼 x 與 y 兩變數在統計性質上有何不同呢? 根據動差法, 動差 m_r 的估計子如下:

$$\hat{m}_r \overset{a}{\sim} (m_r, \frac{1}{T}(m_{2r} - m_r^2))$$

這表示:

$$\hat{\mu} \overset{a}{\sim} (\mu, \frac{\sigma^2}{T})$$

$$\hat{\sigma}^2 \overset{a}{\sim} (\sigma^2, \frac{1}{T}(m_4 - \sigma^4))$$

以此例而言, $\hat{\mu}_x$ 的變異數存在, 但 $\hat{\mu}_y$ 的變異數不存在, 這表示 $\hat{\mu}_y$ 的波動會較 $\hat{\mu}_x$ 為大。

- 如果 $\{x_t\}$ 與 $\{y_t\}$ 都是iid, 則大數法則都適用, 亦即:

$$\bar{x} \xrightarrow{p} \mu_x$$

$$\bar{y} \xrightarrow{p} \mu_y$$

  這表示隨著樣本趨近於無窮大, $\bar{x}$ 與 $\bar{y}$ 都會收斂到其母體平均值, 但是中央極限定理則不一定適用。

- 根據 Lindberg-Levy 定理可知, $\bar{x}$ 會有常態的極限分配:

$$\sqrt{T}(\bar{x} - \mu_x) \xrightarrow{d} \mathcal{N}(0, \sigma_x^2)$$

  但 $\bar{y}$ 的極限分配不會趨近常態; 連帶地, 這顯示相關的 t 、χ^2 或 F 等檢定都不適用於y。

- 概念上, 我們知道 y_t 與 $\bar{y}$ 都會比 x_t 與 $\bar{x}$ 的波動來得大; 同理, 如果我們計算 $\hat{\sigma}_x^2$ 與 $\hat{\sigma}_y^2$ ，則可知:

$$\hat{\sigma}_x^2 \xrightarrow{p} \sigma_x^2$$

  但 $\hat{\sigma}_y^2$ 並不會趨近於 σ_y^2 ，因為 $\sigma_y^2 = \infty$ ，這也表示即使樣本趨近於無窮大, $\hat{\sigma}_y^2$ 的變異仍不會消失。

- 同理, 我們不知道 $\hat{\sigma}_x^2$ 的分配為何, 因為Lindberg-Feller CLT並不適用。當然, 在有更明確的條件下, 仍有可能找出 $\hat{\sigma}_x^2$ 的極限分配。至於 y 呢? 由於 σ_y^2 不存在, 更不用說 $\hat{\sigma}_y^2$ 的分配了。

常態性檢定

- 實證上最常見的檢定包括 Kolmogorov-Smirnov 的無母數檢定與 Jarque-Bera (JB) 的檢定。JB 檢定的公式如下：

$$\text{JB} = T\left[\frac{s^2}{6} + \frac{(\kappa-3)^2}{24}\right] \overset{a}{\sim} \chi^2_2 \tag{6}$$

其中，s 與 κ 分別為偏態與峰態係數的估計值。在常態的虛無假設下，偏態係數為0，而峰態係數為3，因此可知 JB 統計量會趨近於0。也就是當樣本愈偏離常態，則其 JB 統計量的值也會愈大。

時間序列資料的敘述統計量

- 兩個常見的名詞: 自我共變異數 (autocovariance) 與自我相關 (autocorrelation) 係數。

$$\gamma_s \equiv E(y_t y_{t+s}), \qquad s = 0, 1, 2, \cdots$$

$$\rho_s \equiv \frac{cov(y_t, y_{t+s})}{\sqrt{var(y_t)var(y_{t+s})}} = \frac{\gamma_s}{\gamma_0}$$

其中, γ_s 與 ρ_s 分別代表變數 y_t 的第 s 階的自我變異數與自我相關係數。其對應的樣本估計為:

$$\hat{\gamma}_s = \frac{1}{T-s}\sum_{t=1}^{T-s} \hat{y}_t \hat{y}_{t+s},$$

$$\hat{\rho}_s = \frac{\hat{\gamma}_s}{\hat{\gamma}_0}$$

其中, $\hat{y}_t = y_t - \bar{y}$ 為平移(demeaned) 後的 y_t 。

Ljung-Box (1979) 檢定

- Ljung-Box (1979) 檢定常被用來檢定一個序列是否有序列相關:

$$Q = T(T+2)\sum_{j=1}^{p}\frac{\hat{\rho}_j^2}{T-j}$$

此統計量有漸近卡方分配, 自由度為 p 。

均數的母數與非母數檢定

- 如果樣本服從 iid 常態的話, 那麼均數為0的假說就可以使用一般的 t 檢定; 但如果不服從常態分配呢? 概念上來說, 在樣本不服從常態分配的假設下, 可知 t 統計量將不再有 t 分配, 因此所做的假說檢定也是錯的。不過, 在樣本夠大時, 即使樣本不服從常態, 但 t 統計量還是會趨近於標準常態分配, 而由於 t 分配在自由度趨近於無窮大時也會趨近標準常態, 所以 t 檢定還是適用的。
- 無母數方法: 一般最常用的大概就是符號檢定 (sign test) 與 Wilcoxon 的順序符號檢定 (Wilcoxon's signed rank test)。基本上, 這兩種方法只應用到樣本的正負號與排序, 而無須假設特定分配, 因此, 在樣本偏離常態時, 也較可靠。

- 假設有兩樣本 $\{x_1, \cdots, x_m\}$ 與 $\{y_1, \cdots, y_n\}$，兩者的分配如下:

$$x_i \overset{iid}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$$
$$y_j \overset{iid}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$$

且 x_i 與 x_j 為統計獨立。我們關心的虛無假說為: $H_0 : \mu_1 = \mu_2$ 或是 $H_0 : \mu_1 - \mu_2 = 0$。

- 在初統的教科書中，一般會將檢定分為變異數已知與變異數未知兩類做討論。事實上，現實世界中變異數從來就不是已知的，因此真正只有第二類需要討論，但在 σ_1^2 與 σ_2^2 未知下，又得分兩類進行：
  1. $\sigma_1^2 = \sigma_2^2$
  2. $\sigma_1^2 \neq \sigma_2^2$

  以 $\sigma_1^2 = \sigma_2^2$ 而言，則只需估計一個共同變異數就好。此一合併變異數估計子(pooled variance estimator) $\hat{\sigma}^2_{pool}$ 估計為：

$$\hat{\sigma}^2_{pool} = \frac{1}{m+n-2}\Big((m-1)\hat{\sigma}_1^2 + (n-1)\hat{\sigma}_2^2\Big)$$

  如果兩者變異數不同，則需用另一個較複雜的檢定，因為一般統計教科書都已詳述，在此不多做說明。

「配對檢定」(paired test)

- 我們面對的問題多半可以表達為:

$$x_t \overset{iid}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$$

$$y_t \overset{iid}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$$

其中, $t = 1, \cdots, T$ ，而 x_t 與 y_t 因為是同一時點的變數, 所以是有相關的, 其共變異數$cov(x_t, y_t) = \sigma_{12} \neq 0$ 。

- 因為 x_t 與 y_t 不獨立, 所以前面討論的檢定都不適用。因為我們關心的是 $H_0 : \mu_1 - \mu_2 = 0$ ，所以只要定義一新變數 $z_t = x_t - y_t$ ，則可知 $z_t \overset{iid}{\sim} \mathcal{N}(\mu_z, \sigma_z^2)$, 其中

$$\mu_z = \mu_1 - \mu_2$$

$$\sigma_z^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$$

- 前述平均數檢定的問題除了樣本彼此不獨立外, 另一個問題就是樣本未必服從常態。以配對資料而言, 如果 x_t 、y_y 不服從常態, 則只要以無母數檢定方法對 z_t 的均數做檢定即可。
- 最後一個問題是樣本不是服從 iid, 而是存在序列相關或異質變異。在這樣的情況下, 通常又可分兩種方式作分析: 一是對非 iid 的型態做調整, 二是以較一般化的「一般化動差法」(generalized method of moment (GMM)) 來做檢定; 關於這部分將留待稍後的章節再詳加說明。

相關係數

- 假如樣本包含兩個變數: $\{x_t, y_t\}$, $t = 1, \cdots, T$, 則兩者的共變異數估計子 (covariance estimate) 與樣本相關係數分別如下:

$$\widehat{cov(x,y)} \equiv \hat{\sigma}_{xy} = \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y}) \qquad (7)$$

$$\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} \qquad (8)$$

相關係數

- 報告變數間的相關係數有兩個主要的目的: 一、可讓研究者對變數間是否有相關性有一個初步的瞭解。可預見的是, 如果相關係數不顯著, 則後續的分析, 如迴歸估計等, 也大概不會顯著。二、如果我們嘗試以多個變數來解釋另一個變數, 則有可能需要報告各個解釋變數間的相關性, 因為如果各個解釋變數間有太高的相關性, 則可能導致變數在解釋上的「糾纏不清」(「共線性」(multicollinearity) 問題)。

- 資產 (如股票) 報酬率的計算可說是財務研究的最基本, 即便是一般的經濟變數, 也常以報酬率的型態呈現, 本節僅以資產報酬率來說明。令 p_t 為一資產在時點t的價格 (也可以是經濟變數如 GNP 或物價指數等), 則報酬率可分為離散 (discrete) 與連續複利 (continuously compounded) 兩類, 分別定義如下:

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}}$$

$$r_t = \ln\left(\frac{p_t}{p_{t-1}}\right)$$

其中, R_t 為簡單的離散報酬率, r_t 則為連續複利下的報酬率, $\ln$ 代表「自然對數」。以下簡稱 r_t 為連續報酬。離散報酬與連續報酬在實務上的意義不同, 在此我們僅從統計的角度來探討。

- 根據定義可知, 在不同報酬下, 資產的價格為:

$$p_t = p_{t-1}(1 + R_{t-1})$$
$$p_t = p_{t-1}e^{r_t}$$

從報酬率倒推去計算資產的價格, 有些研究會將報酬率誤植為 $p_t = p_{t-1}(1 + r_t)$, 這是不對的。另外, R_t 與 r_t 有如下關係:

$$r_t = \ln(1 + R_t) \tag{9}$$
$$R_t = e^{r_t} - 1 \tag{10}$$

r_t 與 R_t 的關係

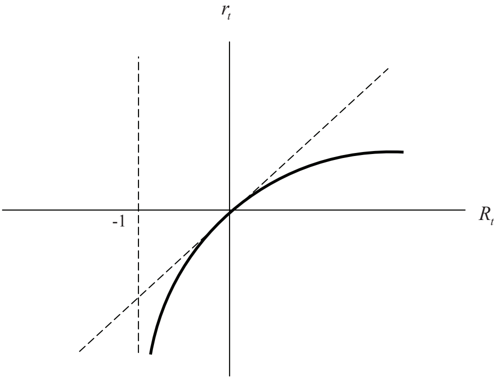


Figure: r_t 與 R_t 的關係

- 由圖7可知, 離散報酬最小為-1, 上界則無限制; 相對地, r_t 則無此限制:

$$R_t \in (-1, \infty)$$

$$r_t \in (-\infty, \infty)$$

- 要進一步分析 r_t 與 R_t 的關係, 可以利用「泰勒展開式」。根據泰勒展開式, 如果將 $f(x)$ 對某一個值 x_0 展開, 可得:

$$f(x) = f(x_0) + \sum_{i=1}^{\infty} \frac{f(i)}{i!}\bigg|_{x=x_0} (x - x_0)^i$$

其中, $f^{(i)}$ 代表 f 對 x 的第 i 階微分。

- 所以, 我們將 $r_t = \ln(1+R_t)$ 對0(也就是 $x_0 = 0$) 展開, 取前兩項近似, 得:

$$\begin{aligned} r_t &\approx \ln(1+0) + \frac{(1+R_t)^{-1}}{1!}\bigg|_{R_t=0}(R_t - 0) \\ &\quad + \frac{-(1+R_t)^{-2}}{2!}\bigg|_{R_t=0}(R_t - 0)^2 \\ &= R_t - \frac{1}{2}R_t^2 \end{aligned}$$

由此可見, r_t 總是比 R_t 小, 而且 R_t^2 愈大, 則兩者差距愈大。

另外兩個現象

- $r_t \le R_t$。連續報酬總是比離散報酬小, 除了在原點 $(r_t = R_t = 0)$。此外, 隨著 R_t 愈大 (愈小), 兩者的差距也愈大 (愈小)。
- 在報酬率接近 0 的情況下, r_t 與 R_t 非常接近, 亦即 $r_t \approx R_t$。

離散報酬 R_t 與連續報酬還有幾點不同:

- 在離散報酬下, 多期報酬是「連乘」的非線性關係, 而連續報酬則有線性加總之性質:

$$R(t,t+T) \equiv \frac{p_{t+T}-p_t}{p_t} = \prod_{i=1}^{T}(1+R_{t+i}) - 1$$

$$r(t,t+T) \equiv \ln\left(\frac{p_{t+T}}{p_t}\right) = \sum_{i=1}^{T} r_{t+i}$$

- 令 $\bar{R}$ 與 $\bar{r}$ 分別為離散與連續報酬之平均報酬, 亦即

$$\bar{R} = \sqrt[T]{1+R(t,t+T)} - 1$$

$$\bar{r} = \frac{1}{T} r(t,t+T)$$

幾何平均與算術平均報酬率

- 一般而言我們很少使用連續報酬。因為離散的多期間報酬不是單期報酬的加總，而是一非線性連乘關係:

$$R(t,t+T)=\prod_{i=1}^{T}(1+R_{t+i})-1$$

因此，要計算一期間，如 $R(t,t+T)$，的每期平均報酬，也不應用算術平均，而是幾何平均:

$$\bar{R}=\sqrt[T]{\prod_{i=1}^{T}(1+R_{t+i})}-1$$

- 不論是實務上，或是多數的實證研究都仍是使用算術平均報酬如下:

$$\bar{R}^a=\frac{1}{T}\sum_{i=1}^{T}R_{t+i}$$

- $\ln(1+\bar{R})$ 的泰勒展開式如下:

$$\ln(1+\bar{R}) \approx \bar{R} - \frac{1}{2}\bar{R}^2$$

所以可知:

$$\begin{aligned}\bar{R} - \frac{1}{2}\bar{R}^2 &\approx \frac{1}{T}\Big(\sum (R_{t+i} - \frac{1}{2}R_{t+i}^2)\Big) \\ &= \frac{1}{T}\sum R_{t+i} - \frac{1}{2T}\sum R_{t+i}^2\end{aligned}$$

因為 $\bar{R}^2$ 較小, 若忽略, 則得:

$$\bar{R} \approx \bar{R}^a - \frac{1}{2}\Big(\frac{1}{T}\sum R_{t+i}^2\Big)$$

將上式取期望值, 得:

$$\begin{aligned}E(\bar{R}) &\approx E(\bar{R}^a) - \frac{1}{2}\Big(\frac{1}{T}\sum R_{t+i}^2\Big) \\ &= \mu - \frac{1}{2}\Big(\sigma^2 + \mu^2\Big)\end{aligned}$$

- 同樣地, 因為 μ^2 較小, 所以可知:

$$E(\bar{R}) - E(\bar{R}^a) \approx -\frac{1}{2}\sigma^2$$

也就是當報酬率波動愈大(亦即 σ^2 愈大), 則 $\bar{R}$ 與 $\bar{R}^a$ 的差距也愈大。

投資組合報酬率

- 若一投資組合 p 是由 N 個資產所構成的, 則其報酬為:

$$R_{pt} = \sum_{i=1}^{N} w_i R_{it}$$

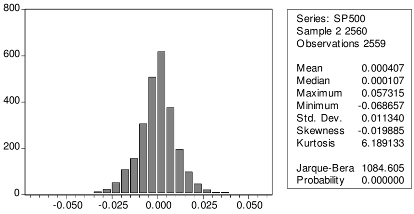
其中, w_i 代表權重(weights), 也就是投資於資產 i 占總金額的比率, 但這樣的表達只有在離散報酬下才成立, 在連續報酬下則是不成立的。

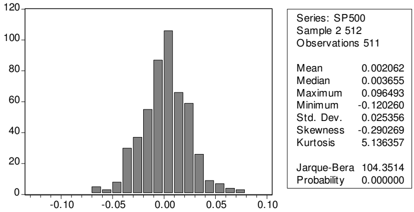
資產報酬的實證特性

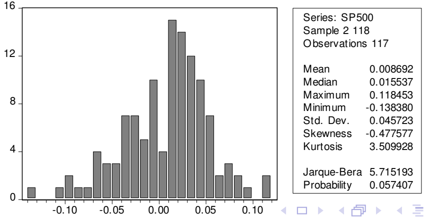
- 報酬自我相關很小: 一般而言, 資產報酬的自我相關是不顯著的, 但有幾個例外: 在高頻率的日內報酬間有負自我相關, 但這主要是因為市場微結構 (market microstructure) 因素如買賣價間的跳動 (bid-ask bounce) 與交易不頻繁 (infrequent trading) 所造成的, 在小公司股票的報酬上尤其顯著。市場指數, 尤其是平均加權 (equal-weighted) 的指數, 其報酬間存在正自我相關。
- 寬尾 (或厚尾): 資產報酬不符合常態, 且其峰態較常態為高。早期一些學者如 Mendelbrot 與 Fama 都認為, 資產報酬符合 Pareto 分配, 但變異數是無窮大。近年來在發現動差的階次大概都在 2 與 5 之間; 一般而言資產報酬的變異數存在, 但超過峰態之後就不存在了。

- 加總常態性 (aggregational Gaussianity): 隨著計算報酬的期限愈長, 報酬的分配愈趨近常態分配。例如, 日報酬普遍存在厚尾與負偏, 但月報酬的分配則很接近常態。不過, 不同頻率分配的型態 (shape) 並不同, 顯示不完全符合所謂的 Pareto law(也就是變數相加後, 仍然維持原有分配特性)。此外, 由於長期間報酬 (如年報酬) 不是短期間報酬 (如日報酬) 之相加, 而是類似連乘的複利關係, 這似乎使得長期間報酬反而出現正偏的情況。
- 波動聚集 (volatility clustering): 不論是報酬率的平方或絕對值, 其自我相關性都很高。進一步的討論, 請參閱本書第九章。

- 槓桿效果 (leverage effect): 波動與報酬呈負向關係, 這是 Fisher Black 所發明的用語。為什麼稱報酬與波動的負向關係為槓桿效果呢? 概念上, 因為負的報酬會造成股價降低。由於股價代表公司股東權益 (equity) 的價值, 較低的權益價值代表公司的負債較高, 也就是較高的公司槓桿, 因此代表公司的風險變得較大。
- 交易量與波動存在顯著的正相關。報酬與價格的波動與交易量存在正相關。傳統認為這與資訊有關, 但近年的研究則認為市場參與者的過度自信與異質信念等行為偏誤可能是主要原因。


800
600
400
200
0
-0.050
-0.025
0.000
0.025
0.050
Series: SP500
Sample 2 2560
Observations 2559
Mean 0.000407
Median 0.000107
Maximum 0.057315
Minimum -0.068657
Std. Dev. 0.011340
Skewness -0.019885
Kurtosis 6.189133
Jarque-Bera 1084.605
Probability 0.000000


120
100
80
60
40
20
0
-0.10
-0.05
0.00
0.05
0.10
Series: SP500
Sample 2 512
Observations 511
Mean 0.002062
Median 0.003655
Maximum 0.096493
Minimum -0.120260
Std. Dev. 0.025356
Skewness -0.290269
Kurtosis 5.136357
Jarque-Bera 104.3514
Probability 0.000000


16
12
8
4
0
-0.10
-0.05
0.00
0.05
0.10
Series: SP500
Sample 2 118
Observations 117
Mean 0.008692
Median 0.015537
Maximum 0.118453
Minimum -0.138380
Std. Dev. 0.045723
Skewness -0.477577
Kurtosis 3.509928
Jarque-Bera 5.715193
Probability 0.057407