Estimator(估計量)

Data D={$x_1, x_2, ..., x_n$}

estimator 就是利用統計的方法，從抽樣樣本中近似母體樣本的各種參數θ (e.g. mean, variance)

對抽樣樣本來說

$$\mu_{MLE} = \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

在這裡μ是母體樣本的 mean

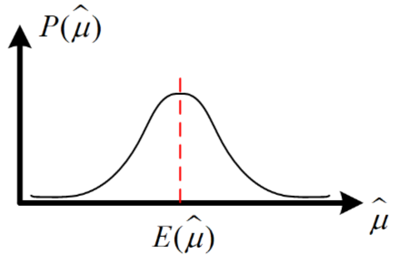
$$\sigma_{MLE}{}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

為了接下來推導用，我們要推導$E(\hat{\mu})$及$Var(\hat{\mu})$

$E(\hat{\mu})$：是取很多抽樣樣本後，每個樣本取其 mean，將所有的 mean 再取 mean(mean of sample mean)

$Var(\hat{\mu})$：取很多抽樣樣本後，每個樣本取其 mean，再將所有的 mean 取 variance(variance of sample mean)

note1: 根據中央極限定理，當我們取很多抽樣樣本後，每個樣本的 mean 會呈現 Gaussian distribution



note2: 這裡的$\hat{\mu}$是個隨機變數，並不是只"某一次"抽樣試驗得到的 mean，就像是我們令值一顆骰子的隨機變數 X，X 並不是"擲某一次"骰子得到的點數

$$E(\hat{\mu}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n}\cdot n\mu = \mu$$

$$Var(\hat{\mu}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = Var(\frac{X_1 + X_2 + ... + X_n}{n})$$

推到這裡，我們要先推導，若隨機變數X=$kY \Rightarrow \mathrm{V}ar(X) = k^2Var(Y)$

deviation：

$$E(X) = \sum X_i P(X_i) = \sum kY_i P(Y_i) = k\sum Y_i P(Y_i) = kE(Y)$$
$$Var(\mathrm{X}) = \mathrm{E}(\mathrm{X}^2) - E^2(X) = \sum X_i{}^2 P(X_i) - (\sum X_i P(X_i))^2$$
$$= \sum k^2 Y_i^2 P(Y_i) - k^2 E^2(Y) = k^2(E(Y^2) - E^2(Y)) = k^2 Var(Y)$$

參考網站：
http://onlinecourses.science.psu.edu/stat414/node/166
https://onlinecourses.science.psu.edu/stat414/node/167

則

$$Var(\hat{\mu}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = Var(\frac{X_1 + X_2 + ... + X_n}{n}) = \frac{1}{n^2}Var(X_1 + X_2 + ... + X_n) = \frac{1}{n^2}(Var(X_1) + Var(X_2) + ... + Var(X_n))$$
$$= \frac{1}{n^2} nVar(X) = \frac{Var(X)}{n} = \frac{\sigma^2}{n}$$

這裡的σ^2是母體樣本的 variance

note:2; 取樣次數 n 很大時，mean 的 variance 分布就會越陡峭，我們越能精準估計母體樣本的 mean

bias

是我們用 estimator 得到的統計參數，其期望值和實際母體樣本的差距，也就是$E(\hat{\theta}) - \theta$

若是 unbias 的統計方法，其 bias=0

以前我們計算 mean 和 variance 都很簡單，因為我們已經假設目前發生的事件就能代表整個母體，但是實際上我們通常沒辦法能得到一個可以完全和母體相同分布的樣本，所以我們只能用估計值來近似，如果用我們以前計算 variance 的方式計算抽樣樣本的 variance，是會有 bias 的，我們需要對 variance 做修正，接下來就是要說明如何修正可得到無 bias 的 variance

首先，我們要先證明先前對於 variance 的定義在抽樣樣本下是會有 bias 的

對於 mean 而言是不會有誤差的，前面我們也證過，$E(\hat{\mu}) = \mu$，故有可能會有可能會有誤差的是 variance

推導

$$E(\sigma_{MLE}{}^2) = E(\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2) = \frac{1}{n}E(\sum_{i=1}^{n}(x_i{}^2 - 2\hat{\mu}x_i + \hat{\mu}^2))$$
$$= \frac{1}{n}E(\sum_{i=1}^{n}(x_i{}^2) - 2\hat{\mu}(\sum_{i=1}^{n}x_i) + \sum_{i=1}^{n}\hat{\mu}^2)$$
$$= \frac{1}{n}E(\sum_{i=1}^{n}(x_i{}^2) - 2\hat{\mu}n\hat{\mu} + n\hat{\mu}^2) = \frac{1}{n}E(\sum_{i=1}^{n}(x_i{}^2) - n\hat{\mu}^2) = \sum_{i=1}^{n}\mathrm{E}(x_i{}^2) - nE(\hat{\mu}^2)$$

而 $E(\sum_{i=1}^{n}x_i)$ 為一次抽樣樣本和的 mean，應為 $n\hat{\mu}$

而 $\sum_{i=1}^{n}E(\hat{\mu}^2)$ ，lession3 中有推導過 $Var(x) = E((x - E(x))^2) = E(x^2) - E^2(x)$ ，故 $E(x^2) = Var(x) + E^2(x)$

$$E(\hat{\mu}^2) = Var(\hat{\mu}) + E^2(\hat{\mu}) = \frac{\sigma^2}{n} + \mu^2$$

同理， $E(x_i{}^2) = Var(x_i) + E^2(x_i) = \sigma^2 + \mu^2$

$$E(\sigma_{MLE}{}^2) = \sum_{i=1}^{n}\mathrm{E}(x_i{}^2) - nE(\hat{\mu}^2)$$
$$= \frac{1}{n}(n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \hat{\mu}^2))$$
$$= \frac{1}{n}(n\sigma^2 - \sigma^2) = \sigma^2 - \frac{\sigma^2}{n}$$

$$E(\hat{\theta}) - \theta = \sigma^2 + \frac{\sigma^2}{n} - \sigma^2 \neq 0$$

note:估計值和實際值是不一樣的，因為我們實際進行試驗時，我們是無法預先得知母體的 mean 和 variance，我們唯一可以得到的是目前抽樣樣本的 mean 和 variance，來作暫時(估計)的 mean 和 variance。但像是 $E(x_i)$ ，雖然 x_i 是發生第 i 次事件的 outcome，不過這裡的 x_i 是一個隨機變量。像是如果我們設 X 為擲骰子的隨機變量， $E(x_i)$ 並不是指我隨便值一次骰子擲到 i 次是擲到幾點，是指”我們擲到第 i 次骰子時，平均會擲出幾點”

故如果是抽樣樣本，我們的 variance 公式需要修正，修正為

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$$

這樣就不會有 bias，證明如下

$$E(\sigma_*^{2})=E(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\hat{\mu})^2)=\frac{1}{n-1}E(\sum_{i=1}^{n}(x_i^{2}-2\hat{\mu}x_i+\hat{\mu}^{2}))$$

$$=\frac{1}{n-1}E(\sum_{i=1}^{n}(x_i^{2})-2\hat{\mu}(\sum_{i=1}^{n}x_i)+\sum_{i=1}^{n}(\hat{\mu}^{2}))$$

$$=\frac{1}{n-1}E(\sum_{i=1}^{n}(x_i^{2})-2\hat{\mu}n\hat{\mu}+n\hat{\mu}^{2})$$

$$=\frac{1}{n-1}E(\sum_{i=1}^{n}(x_i^{2})-n\hat{\mu}^{2})$$

$$=\frac{1}{n-1}(n(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\hat{\mu}^{2}))$$

$$=\frac{1}{n-1}(n\sigma^2-\sigma^2)=\sigma^2$$

$$E(\sigma_*^{2})-\sigma^2=0$$

note:

$$E(\hat{\mu}\sum_{i=1}^{n}x_i)\neq\hat{\mu}E(\sum_{i=1}^{n}x_i)$$

因為 $\hat{\mu}$ 是一個隨機變數，是 n 個隨機變數 x_i 加總除以 n，故不能直接搬到 E 外面

若是用自由度來想這個問題(不懂自由度就不用看了)，若是我們已經得知 sample mean 時，我們只需要得到 n-1 個 data 就能得到全部 data 的資訊了，故最後一個資料的資訊是完全不重要，也就是完全沒有資訊可言的，故指需要除以 n-1 即可

參考網站：

http://web.math.sinica.edu.tw/math_media/d291/29102.pdf

https://www.zhihu.com/question/20099757