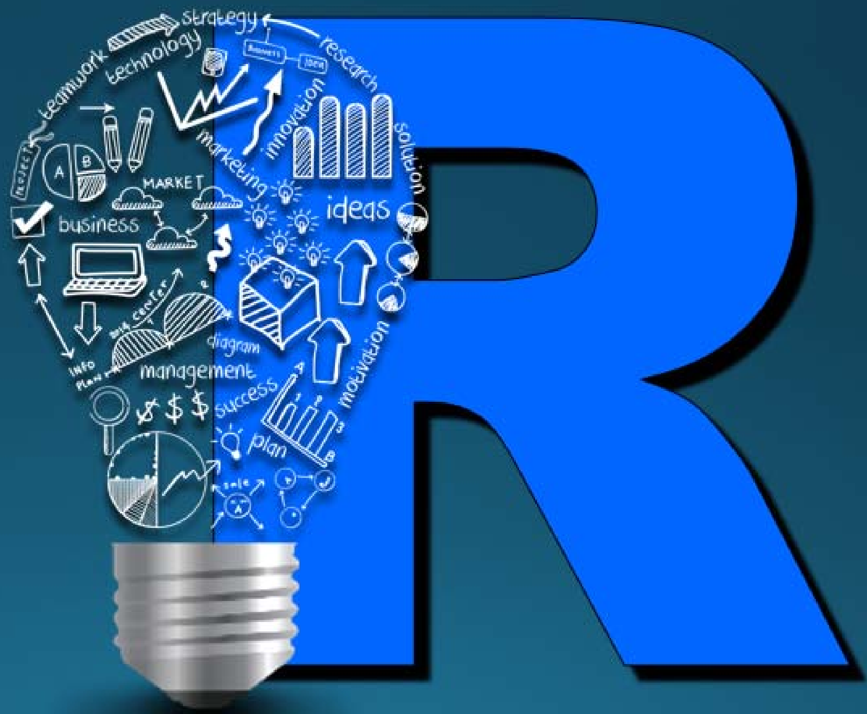
B03

參數估計



吳漢銘
國立政治大學 統計學系

http://www.hmwu.idv.tw

本章大綱

- **參數估計** (parameter estimation)
  (利用樣本統計量及其抽樣分配來對母體參數進行推估, 以瞭解母體的特性)
  - **點估計** (動差法、最大概似法、最小平方法)
    - 評斷準則: 不偏性、有效性、一致性、最小變異不偏性、充份性。
  - **區間估計**
  - **貝式估計法**

可能性函數、概似函數 (The Likelihood Function)

1. Suppose the sample are iid from a distribution with density function $f(X|\theta)$, where θ is a parameter.

2. The **likelihood function** is the conditional probability of observing the sample, given θ

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta) .$$

   (a) The parameter could be a vector of parameters, $\theta = (\theta_1, \cdots, \theta_p)$.

   (b) The likelihood function regards the data as a function of the parameter θ.

   (c) The **log likelihood** function

$$l(\theta) = \log(L(\theta)) = \sum_{i=1}^{n} \log f(x_i|\theta) .$$

最大概似估計法
Maximum Likelihood Estimation

1. The method of maximum likelihood was introduced by **R.A. Fisher** (1890-1962, English statistician).

   (a) By maximizing the likelihood function $L(\theta)$ with respect to θ, we are looking for the most likely value of θ given the sample data.

   (b) Θ: parameter space of possible values of θ.

   (c) If the $\max L(\theta)$ exists and it occurs at a unique point $\hat{\theta} \in \Theta$, then $\hat{\theta}$ is called maximum likelihood estimator of θ.

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \quad 且 \quad \frac{\partial^2 L(\theta)}{\partial \theta^2} < 0$$

點估計步驟：
1. 抽取代表性樣本
2. 選擇一個較佳的樣本統計量當估計式
3. 計算估計式的估計值
4. 以該估計值推論母體參數並作決策

範例: 估計最有可能中獎的機率

假設有一台抽獎機，每次抽的中獎機率都不會改變，也就是說每次抽中與否，都與前一次是否抽中無關，表示每次抽都是獨立事件。

假設此抽獎機連抽 5 次，只有第 1 次和第 4 次中獎，其他 3 次沒有中獎。若每次中獎機率為 p，請推測最有可能的 p 值為多少？

抽獎機的機率模型

將隨機 變數 X_i 定義為： $X_i = \begin{cases} 1\text{ (中獎)} \\ 0\text{ (沒中獎)} \end{cases}$

每次中獎的機率為 p

沒中獎的機率為 $(1-p)$

想要推估的參數就是 p 的值

則抽 5 次的中獎機率可分別寫為：

$$
\begin{aligned}
&P(X=X_1)\cdot P(X=X_2)\cdot P(X=X_3)\cdot P(X=X_4)\cdot P(X=X_5)\\
&= p\cdot(1-p)\cdot(1-p)\cdot p\cdot(1-p)\\
&= p^2\cdot(1-p)^3 \qquad (6.3.2)
\end{aligned}
$$

- 式子(6.3.2)稱為**概似函數 (L ikelihood function)**。
- 只要找出能讓概似函數出現極大值的p就是最能符合此抽獎機率模型的答案。
- 要找出極大值，就是找出概似函數微分後等於 0的 p, 且此p可以讓概似函數出現極大值。
- 概似函數習慣上會用 L (Likelihood)做為函數名稱，但許多機器學習的書中習慣用L表示損失函數 (Loss function), 應避免還淆。

對數概似函數

$$P(X=X_1)\cdot P(X=X_2)\cdot P(X=X_3)\cdot P(X=X_4)\cdot P(X=X_5)$$
$$= p\cdot(1-p)\cdot(1-p)\cdot p\cdot(1-p)$$
$$= p^2\cdot(1-p)^3$$

$$\log(p^2(1-p)^3) = 2\log p + 3\log(1-p)$$

$$\frac{2}{p}+\frac{3\cdot(-1)}{1-p}=0$$
$$\Leftrightarrow\ 2(1-p)-3p=0$$
$$\Leftrightarrow\ 5p=2$$
$$\Leftrightarrow\ p=\frac{2}{5}=0.4$$

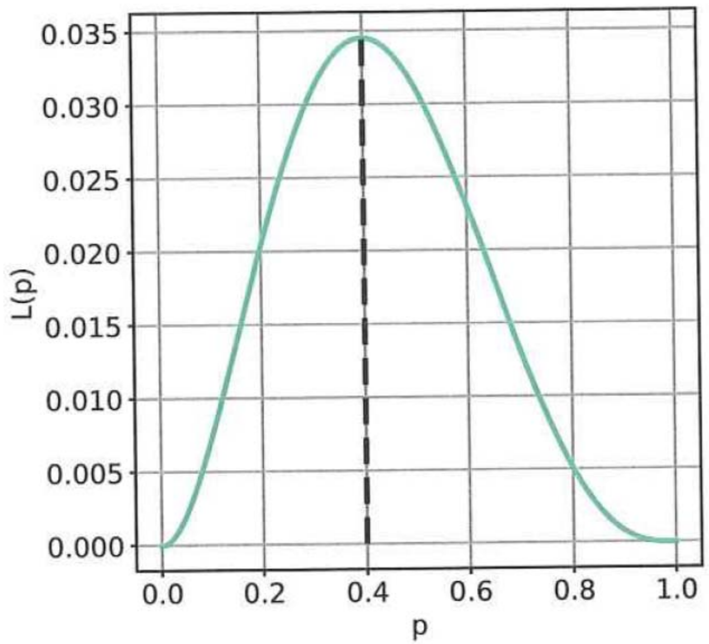


圖 6-13　橫軸為 p，縱軸為概似函數的值

最大概似估計量
(maximum likelihood estimator, MLE)

為何概似函數的極值是求最大值，而不是最小值?

- 最大概似估計法是找出「概似函數微分等於 0」的參數值。照理講，找出的參數也有可能讓概似函數出現極小值或無極值。
- 概似函數是由各已知事件的機率(介於0~1)相乘而來，數值只會大於等於0，而等於 0就是極小值，也就是此機率模型最不可能發生的情況。
- 我們希望的是此機率模型最可能發生的情況，因此能產生極大值的參數才是我們要的。

MLE using mle {stats4}

Suppose Y_1, Y_2 are iid with density $f(y) = \theta e^{-\theta y}$, $y > 0$. Find the MLE of θ.

By independence, $L(\theta) = \underline{(\theta e^{-\theta y_1})(\theta e^{-\theta y_2})} = \underline{\theta^2 e^{-\theta(y_1+y_2)}}$.

(a) Thus $\ell(\theta) = \underline{2\log\theta - \theta(y_1+y_2)}$ and the log-likelihood equation to be solved is

$$\frac{d}{d\theta}\ell(\theta) = \underline{\frac{2}{\theta} - (y_1+y_2)} = 0, \quad \theta > 0$$

(b) The unique solution is $\underline{\hat{\theta} = 2/(y_1+y_2)}$, which maximizes $L(\theta)$.

(c) Therefore the MLE is the <u>reciprocal</u> of the sample mean in this example.

(a) The **mle** function takes as its first argument the function that evaluates $-\ell(\theta) = -\log(L(\theta))$.

(b) The negative log-likelihood is minimized by a call to **optim**, an optimization routine.

```
> y <- c(0.04304550, 0.50263474)
> theta_hat <- length(y) / sum(y)
> theta_hat
[1] 3.66515
>
> mlogL <- function(theta = 1) {
+   n <- length(y)
+   f <- -(n * log (theta) - theta * sum(y))
+   f
+ }
>
> library(stats4)
> fit <- mle(mlogL)
```

```
> summary(fit)
Maximum likelihood estimation

Call:
mle(minuslogl = mlogL)

Coefficients:
      Estimate Std. Error
theta  3.66515   2.591652

-2 log L: -1.195477
```

求 MLE of (μ, σ^2) from a normal population

題目: 若 $X_1, \ldots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2)$. 求($\mu$, σ^2) 之MLE。

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

解:

The probability density function for a sample of *n* independent identically distributed (iid) normal random variables (the likelihood) is

$$f(x_1, \ldots, x_n \mid \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right),$$

$$\mathcal{L}(\mu, \sigma) = f(x_1, \ldots, x_n \mid \mu, \sigma)$$

$$\log(\mathcal{L}(\mu, \sigma)) = (-n/2)\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$0 = \frac{\partial}{\partial\mu}\log(\mathcal{L}(\mu, \sigma)) = 0 - \frac{-2n(\bar{x}-\mu)}{2\sigma^2}. \Rightarrow \boxed{\hat{\mu} = \bar{x} = \sum_{i=1}^{n}\frac{x_i}{n}.} \qquad E\left[\widehat{\mu}\right] = \mu$$

https://en.wikipedia.org/wiki/Maximum_likelihood_estimation

求MLE of (μ, σ^2) from a normal population

$$0 = \frac{\partial}{\partial \sigma} \log\left(\left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right) \right)$$

$$= \frac{\partial}{\partial \sigma} \left(\frac{n}{2} \log\left(\frac{1}{2\pi\sigma^2} \right) - \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right)$$

$$= -\frac{n}{\sigma} + \frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{\sigma^3}$$

$$\boxed{E\left[\widehat{\sigma}^2\right] = \frac{n-1}{n}\sigma^2.}$$

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2. \qquad \mu = \widehat{\mu} \quad \Rightarrow \quad \boxed{\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

The maximum likelihood estimator (MLE) for $\theta = (\mu, \sigma^2)$ is

$$\boxed{\hat{\mu} = \bar{x} = \sum_{i=1}^n \frac{x_i}{n}.} \qquad \boxed{\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

MLE using optim {stats}


```
> loglikefun <- function(x, par){
+   mu <- par[1]
+   sigma <- par[2]
+   n <- length(x)
+   loglikelihood <- - (n / 2)*(log(2 * pi * sigma^2)) +
+     (-1/(2 * sigma^2)) * sum((x - mu)^2)
+   # return the negative to maximize rather than minimize
+   - loglikelihood
+ }
>
> set.seed(1123)
> x <- rnorm(100)
> x <- x/sd(x) * 8 # sd of 8
> x <- x - mean(x) + 10 # mean of 10
> cat("mean(x) =", mean(x), ", sd(x) =", sd(x))
mean(x) = 10 , sd(x) = 8
> optim(par = c(0.5, 0.5), fn = loglikefun, x = x)
$par
[1] 10.001693  7.975965

$value
[1] 349.3359

$counts
function gradient
      95       NA

$convergence
[1] 0

$message
NULL
```

$$\log(\mathcal{L}(\mu,\sigma)) = (-n/2)\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

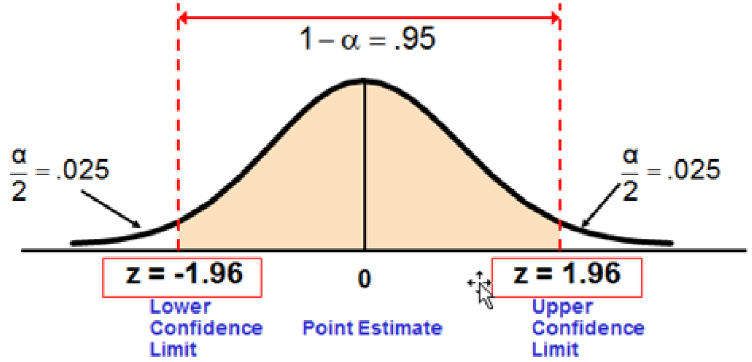
區間估計
(Interval Estimation)

- 區間估計是先對未知的母體參數求點估計值，然後在一信賴水準 (Confidence Level) 下，導出一個上下區間，此區間稱為信賴區間 (Confidence Interval)，信賴水準是指該區間包含母體參數的可靠度。
- 95% 信賴區間表示，做100 次信賴區間，區間約包含母體參數95 次

Interval Estimate of Population Mean

若大樣本(n> 30)、 母體σ已知,
由中央極限定理知 $\bar{X} \sim N(\mu, \sigma^2/n)$ ⇨ $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

$$P(-z \le Z \le z) = 1 - \alpha = 0.95.$$

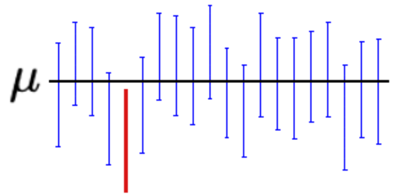


$$\Phi(z) = P(Z \le z) = 1 - \tfrac{\alpha}{2} = 0.975,$$

$$z = \Phi^{-1}(\Phi(z)) = \Phi^{-1}(0.975) = 1.96,$$

$$0.95 = 1 - \alpha = P(-z \le Z \le z) = P\left(-1.96 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le 1.96\right)$$

$$= P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right).$$



A 95% confidence interval indicates that 19 out of 20 samples (95%) from the same population will produce confidence intervals that contain the population parameter.

範例: 老年人看電視的時間

根據行政院主計處調查，台灣地區15歲以上的人口中，以老年人(65歲以上)看電視的時間最長。現在新立傳播公司計畫推出老年人的電視節目，因此想要了解老年人看電視的時間，以決定電視節目的數量。新立公司於是採隨機抽樣法抽取台北市100位老人調查看電視的時數，結果得知，每星期看電視的平均時間為 21.2小時。假設根據過去數次調查的資料，已知每星期看電視時間的標準差為8小時，問在95%信賴水準下，每星期看電視平均時間的信賴區間為何？

信賴水準為95%，$\bar{X}=21.2$小時，$\sigma=8$小時，$n=100$

$\bar{X}$ 的抽樣分配為常態分配 $N \sim (\mu, \sigma_{\bar{X}}^2)$ ➡ $P(|\bar{X}-\mu| \le 1.96\sigma_{\bar{X}}) = 0.95$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{100}} = 0.8$$

在 $1-\alpha$ 信賴水準下，母體平均數的信賴區間為

$$\bar{X} \pm Z_{\alpha/2}\sigma_{\bar{X}}$$

$$\bar{X} \pm Z_{\alpha/2}\sigma_{\bar{X}} = 21.2 \pm 1.96 \times 0.8$$

➡

$$19.632 \le \mu \le 22.768$$

```
> alpha <- 0.05
> xbar <- 21.2
> sigma <- 8
> n <- 100
> v <- qnorm(1-alpha/2)*(sigma/sqrt(n))
> c(xbar - v, xbar + v)
[1] 19.63203 22.76797
```

可推論：「老年人每星期平均看電視的時間在19.632~22.768 小時之間，而此一區間的可信度(信賴水準)為95%。」

Bayesian Statistics
貝式統計

1. In the **frequentist approach** to statistics, the parameters of a distribution are considered to be fixed but unknown constants.

2. The **Bayesian approach** views the unknown parameters of a distribution as random variables.

   (a) In Bayesian analysis, probabilities can be computed for parameters as well as the sample statistics.

   (b) Bayes' Theorem allows one to revise the prior belief about an unknown parameter based on observed data.

Bayes' Theorem

1. If A and B are events and $P(B) > 0$, then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

2. The distributional form of Bayes' Theorem for continuous random variables is

$$f_{X|Y=y}(x) = \frac{f_{Y|X=x}(y)f_X(x)}{f_Y(y)} = \frac{f_{Y|X=x}(y)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X=x}(y)f_X(x)\ dx}$$

貝氏定理 (Bayes' Theorem)

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(B|A)\times P(A)}{P(B)}$$

$$後驗機率 = \frac{可能性 \times 先驗機率}{標准化常量}$$

- $P(A|B)$: 已知在事件 B 發生的情況下事件 A 發生的機率。
  (稱作 A 的事後機率或後驗機率)(posterior probability)。
- $P(A)$, $P(B)$: A, B 的事前機率或
  先驗機率 (prior probability)。
  $(P(A)\neq 0,\ P(B)\neq 0)$
- $P(B|A)$: 已知 A 發生後，B 的條件機率。
  (稱作概似函數 likelihood function)。

例子: 假設有兩個甕，第一個甕裡面有 3 顆紅球，第二個甕裡面有 2 顆紅球和 1 顆白球。我們隨機選擇一個甕，然後從中抽出 2 顆球。假設結果是 2 顆紅球，留在甕裡的那顆球是紅球的機率是多少？(https://ccjou.wordpress.com/)

樣本空間 $\Omega = \{r_1, r_2, r_3, r_4, r_5, w_1\}$。

令 $U_1 = \{r_1, r_2, r_3\}$ 和 $U_2 = \{r_4, r_5, w_1\}$。

A: 從一個甕中抽出 2 顆紅球之事件。

$$\begin{aligned} P(U_1|A) &= \frac{P(A|U_1)P(U_1)}{P(A|U_1)P(U_1)+P(A|U_2)P(U_2)} \\ &= \frac{1\cdot\frac{1}{2}}{1\cdot\frac{1}{2}+\frac{1}{3}\cdot\frac{1}{2}} = \frac{3}{4}. \end{aligned}$$

Bayes' Theorem

1. If A and B are events and $P(B) > 0$, then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

2. The distributional form of Bayes' Theorem for continuous random variables is

$$f_{X|Y=y}(x) = \frac{f_{Y|X=x}(y)f_X(x)}{f_Y(y)} = \frac{f_{Y|X=x}(y)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X=x}(y)f_X(x)\ dx}$$

1. In the **frequentist approach** to statistics, the parameters of a distribution are considered to be fixed but unknown constants.

2. The **Bayesian approach** views the unknown parameters of a distribution as random variables.

   (a) In Bayesian analysis, probabilities can be computed for parameters as well as the sample statistics.

   (b) Bayes' Theorem allows one to revise the prior belief about an unknown parameter based on observed data.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

3. Suppose that X has the density $f(x|\theta)$.

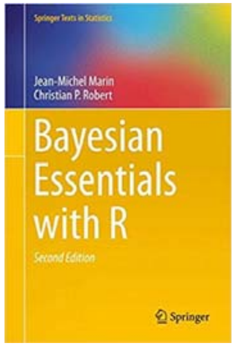
   (a) $f_\theta(\theta)$: the pdf of the prior distribution of θ.

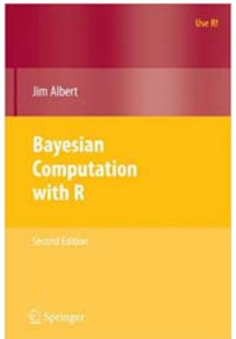
   (b) The conditional density of θ given the sample observations $x_1, \cdots, x_n$ is called the posterior density

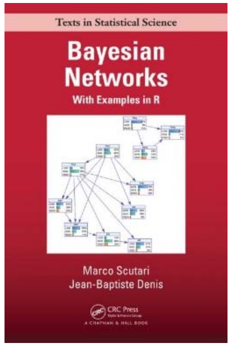
   $$f_{\theta|x}(\theta) = \frac{f(x_1, \cdots, x_n|\theta) f_\theta(\theta)}{\int f(x_1, \cdots, x_n|\theta) f_\theta(\theta)\ d\theta}.$$

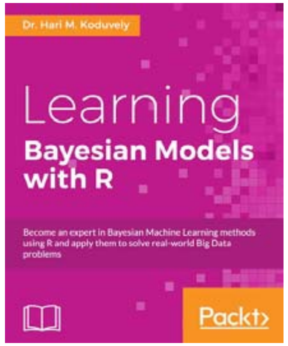
   (c) The posterior distribution summarizes our modified belief about the unknown parameters, taking into account the observed data.

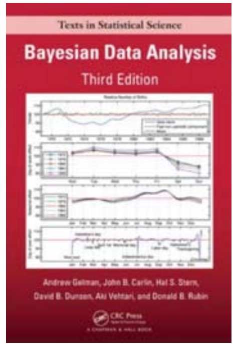
   (d) One is interested in computing posterior quantities such as posterior means, posterior modes, posterior standard deviations.











Bayes Estimator for the Mean of a Normal Distribution

$X_1, X_2, \ldots, X_n$ be a random sample $\quad X_1, \ldots, X_n \sim$ i.i.d. $N(\mu, \sigma^2)$.

μ is unknown and σ^2 is known.

prior distribution for μ is normal with mean μ_0 and variance σ_0^2

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(\mu-\mu_0)^2/(2\sigma_0^2)} = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(\mu^2-2\mu_0+\mu_0^2)/(2\sigma_0^2)}$$

The joint probability distribution of the sample

$$f(x_1, x_2, \ldots, x_n | \mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)}$$

the joint probability distribution of the sample and μ is

$$f(x_1, x_2, \ldots, x_n, \mu) = \frac{1}{(2\pi\sigma^2)^{n/2}\sqrt{2\pi}\sigma_0} e^{-(1/2)\left[\left(1/\sigma_0^2+n/\sigma^2\right)\mu^2-\left(2\mu_0/\sigma_0^2+2\sum x_i/\sigma^2\right)\mu+\sum x_i^2/\sigma^2+\mu_0^2/\sigma_0^2\right]}$$

$$= e^{-(1/2)\left[\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}\right)\mu^2-2\left(\frac{\mu_0}{\sigma_0^2}+\frac{\bar{x}}{\sigma^2/n}\right)\mu\right]} h_1(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

Bayes Estimator for the Mean of a Normal Distribution

$$f(x_1, x_2, \ldots, x_n, \mu) = e^{-(1/2)\left[\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}\right)\mu^2 - 2\left(\frac{\mu_0}{\sigma_0^2}+\frac{\bar{x}}{\sigma^2/n}\right)\mu\right]} h_1(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

$$\Rightarrow \quad f\left(x_1, x_2, \ldots, x_n, \mu\right) = e^{-(1/2)\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}\right)\left[\mu - \left(\frac{(\sigma^2/n)\,\mu_0}{\sigma_0^2+\sigma^2/n}+\frac{\bar{x}\sigma_0^2}{\sigma_0^2+\sigma^2/n}\right)\right]^2} h_2(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

$h_i(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2\ )$ is a function of the observed values and the parameters σ^2, μ_0, and σ_0^2.

because $f(x_1, \ldots, x_n)$ does not depend on μ,

$$\Rightarrow \quad f\left(\mu | x_1, \ldots, x_n\right) = e^{-(1/2)\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}\right)\left[\mu - \left(\frac{\left(\sigma^2/n\right)\mu_0 + \sigma_0^2\bar{x}}{\sigma_0^2+\sigma^2/n}\right)\right]^2} h_3\left(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2\right)$$

a normal probability density function

posterior mean $\dfrac{\left(\sigma^2/n\right)\mu_0 + \sigma_0^2\,\bar{x}}{\sigma_0^2+\sigma^2/n}$

posterior variance $\left(\dfrac{1}{\sigma_0^2}+\dfrac{1}{\sigma^2/n}\right)^{-1} = \dfrac{\sigma_0^2\left(\sigma^2/n\right)}{\sigma_0^2+\sigma^2/n}$

Bayes Estimator for the Mean of a Normal Distribution

posterior mean $\dfrac{(\sigma^2/n)\mu_0 + \sigma_0^2 \bar{x}}{\sigma_0^2 + \sigma^2/n}$

suppose that we have a sample of size $n = 10$ from

from a normal distribution with unknown mean μ and variance $\sigma^2 = 4$.

Assume that the prior distribution for μ is normal with mean $\mu_0 = 0$ and variance $\sigma_0^2 = 1$.

If the sample mean is 0.75, the Bayes estimate of μ is

$$\frac{(4/10)0 + 1(0.75)}{1 + (4/10)} = \frac{0.75}{1.4} = 0.536$$