PRML

Chapter 2(~2.3.9)

2018.09.13

2.3.6 Bayesian inference for the Gaussian

~~**2.3.7 Student's t-distribution**~~

2.3.8 Periodic variables

2.3.9 Mixtures of Gaussians

2.3.6 Bayesian inference for the Gaussian

- MLE는 μ와Σ 를 추론하는데 사용하는 frame work이며, 데이터에 기반해 **<u>점 추정</u>**을 함
- 베이지안은 사전분포를 활용하여 두개의 모수의 **<u>분포를 추정</u>**함

We shall suppose that the variance **<u>σ_2 is known</u>**

When **X** = $\{x_1, \ldots, x_N\}$, The likelihood function that is the probability of the observed data given μ, viewed as a function of μ, is given by

$$p(\mathbf{X}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2\right\}$$

2.3.6 Bayesian inference for the Gaussian

베이지안 공식을 이용하여 분포를 추정해 봅시다.

- μ 의사전분포 $\boldsymbol{P}(\mu) = \mathrm{N}(\mu|\mu_0, \sigma_0)$ …. 아랫 첨자 0는 초기 믿음에 의해 설정한 값
- $\boldsymbol{P}(\mu|\boldsymbol{X}) \propto \boldsymbol{P}(\mu)\boldsymbol{P}(\boldsymbol{X}|\mu)$ … 사후 확률분포
- 그리고 사후 확률 분포는 **<u>공액적 특성</u>**에 따라 정규분포를 따른다.

$$\boldsymbol{P}(\mu|\boldsymbol{X}) = \boldsymbol{N}(\mu|\mu_N, \sigma_N)$$

Conjugate Distribution:

컴퓨터가 없던 시절, 사후 분포 추정은 매우 어려운 일이었음.

통계학자들은 **사전분포의 특성**을 이어 받는 사후 분포를 생각해 계산 부담을 줄여주는 방법을 생각하게 됨.

표본분포(우도함수)	공액사전분포	사후분포
$X \sim B(n, p)$	$p \sim Beta(\alpha, \beta)$	$p \mid X \sim Beta(\alpha + X, n - X + \beta)$
$X \sim poisson(\theta)$	$\theta \sim Gamma(\alpha, \beta)$	$\theta \mid X \sim Gamma(\alpha + X, (1 + 1/\beta)^{-1})$
$X \sim \mathrm{Exp}(\theta)$	$\frac{1}{\theta} = \lambda \sim Gamma(\alpha, \beta)$	$\lambda \mid X \sim Gamma(\alpha + 1, (X + 1/\beta)^{-1})$
$X \sim N(\mu, \sigma^2)$ (σ가 알려져 있을 때)	$\mu \sim N(a, b^2)$	$\mu \mid X \sim N(\frac{Xb^2 + a\sigma^2}{b^2 + \sigma^2}, \frac{b^2\sigma^2}{b^2 + \sigma^2})$
$X \sim N(\mu, \sigma^2)$ (μ가 알려져 있을 때)	$\frac{1}{\sigma^2} = \tau \sim IG(\alpha, \beta)$	$\tau \mid X \sim IG(\frac{1}{2} + \alpha, [\frac{(X - \mu)^2}{2} + \frac{1}{\beta}]^{-1})$

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

2.3.6 Bayesian inference for the Gaussian

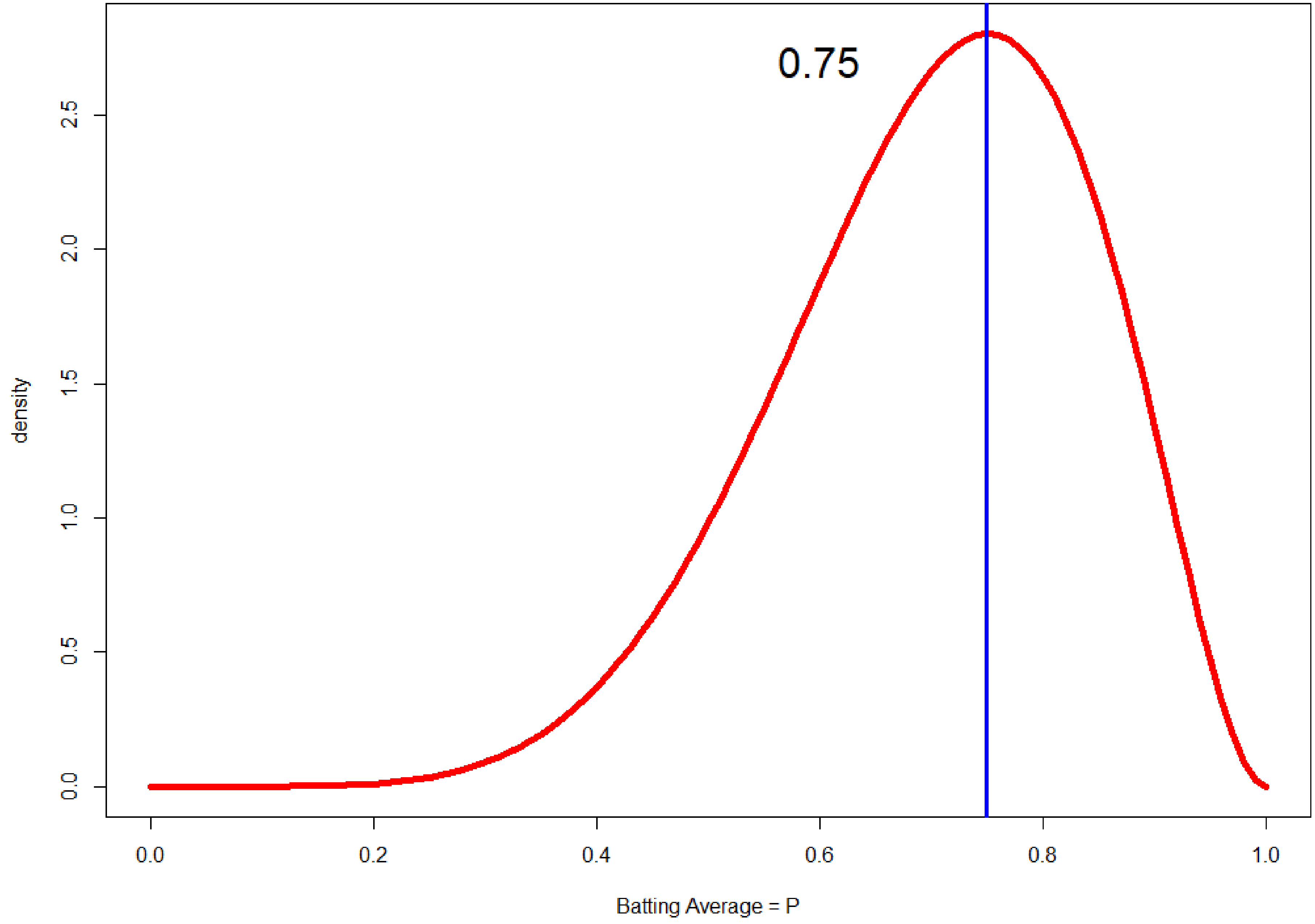
- μ 의사전분포 $\boldsymbol{P}(\mu) = \mathrm{N}(\mu|\mu_0, \sigma_0)$ …. 아랫 첨자 0는 초기 믿음에 의해 설정한 값
- $\boldsymbol{P}(\mu|\boldsymbol{X}) \propto \boldsymbol{P}(\mu)\boldsymbol{P}(\boldsymbol{X}|\mu)$ … 사후 확률분포
- 그리고 사후 확률 분포는 **공액적 특성**에 따라 정규분포를 따른다.

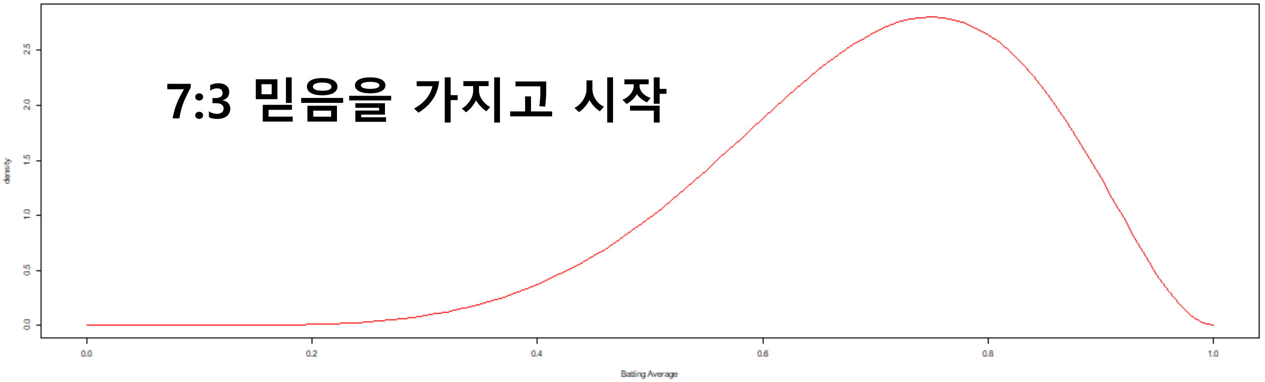
$$\boldsymbol{P}(\mu|\boldsymbol{X}) = \boldsymbol{N}(\mu|\mu_N, \sigma_N)$$

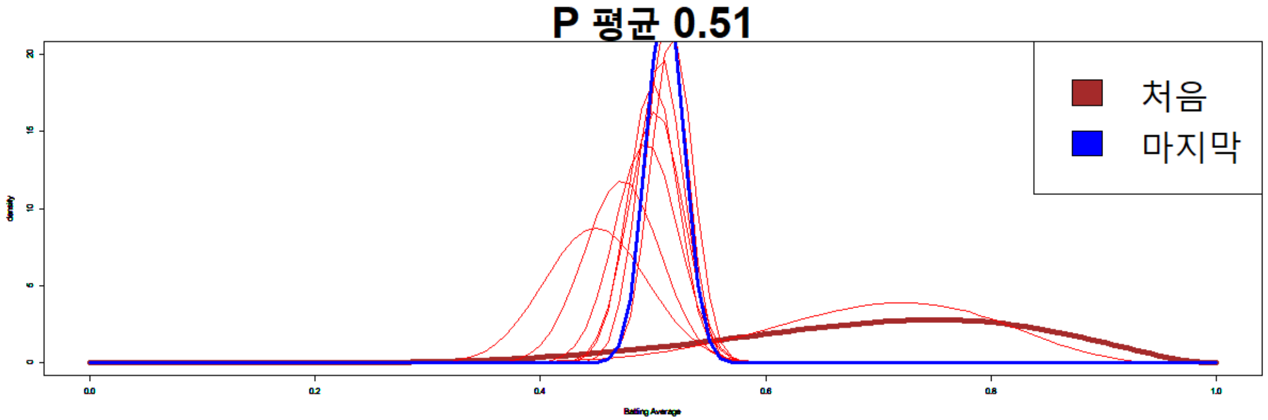
- **구체적인 예시**로 어떤 동전을 두고 아래 표와 같이 추론할 경우

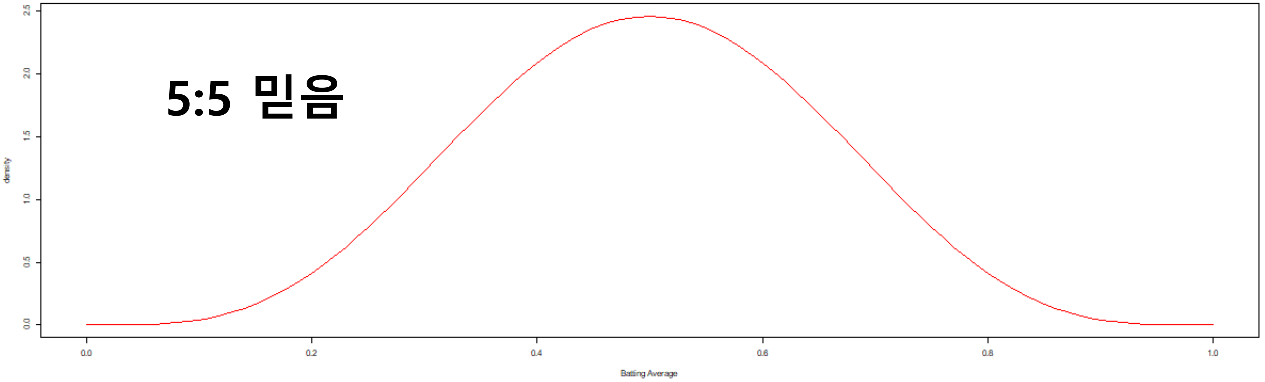
앞면이 나올 확률 p를 구하기 위해선, Beta(a,b)분포를 이용합니다.

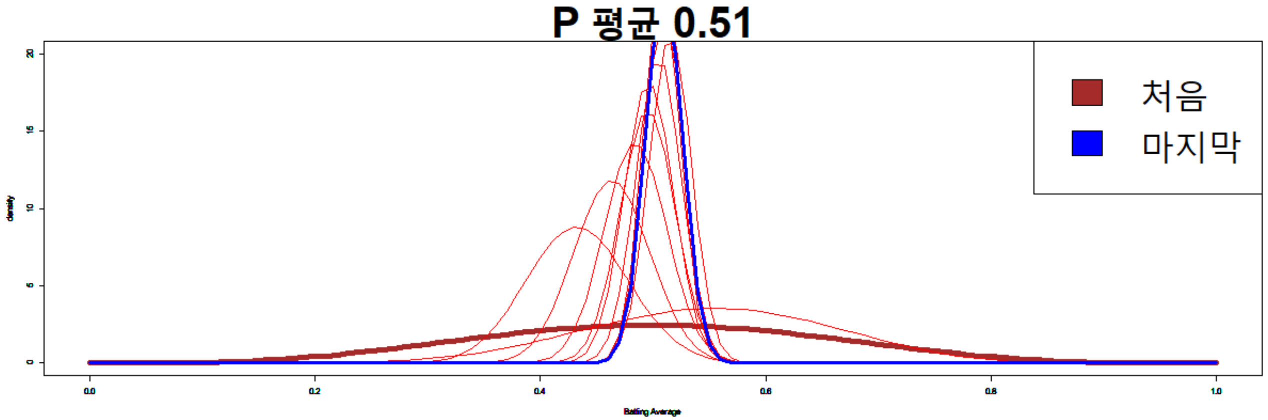
	P(앞면)	P(뒷면)
현진	7(a)	3(b)
대훈	2	8
이삭	5	5

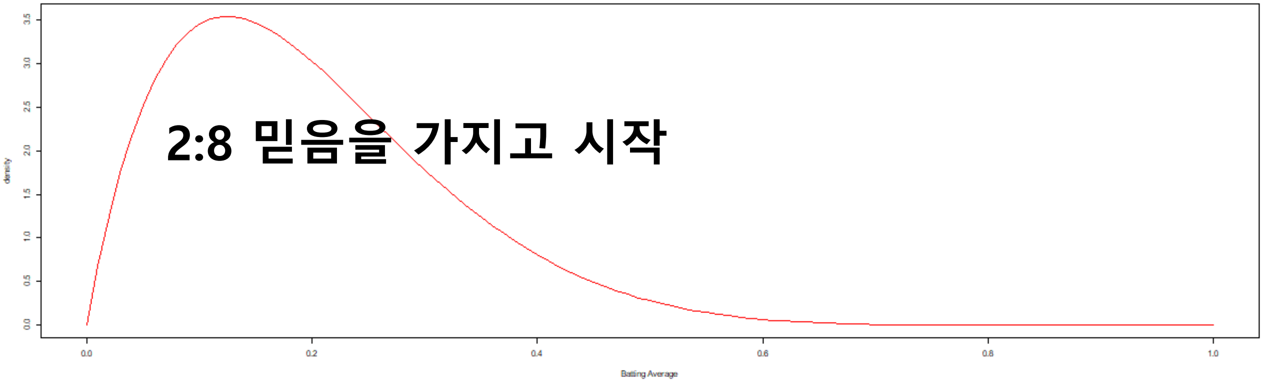

0.75
density
Batting Average = P
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5

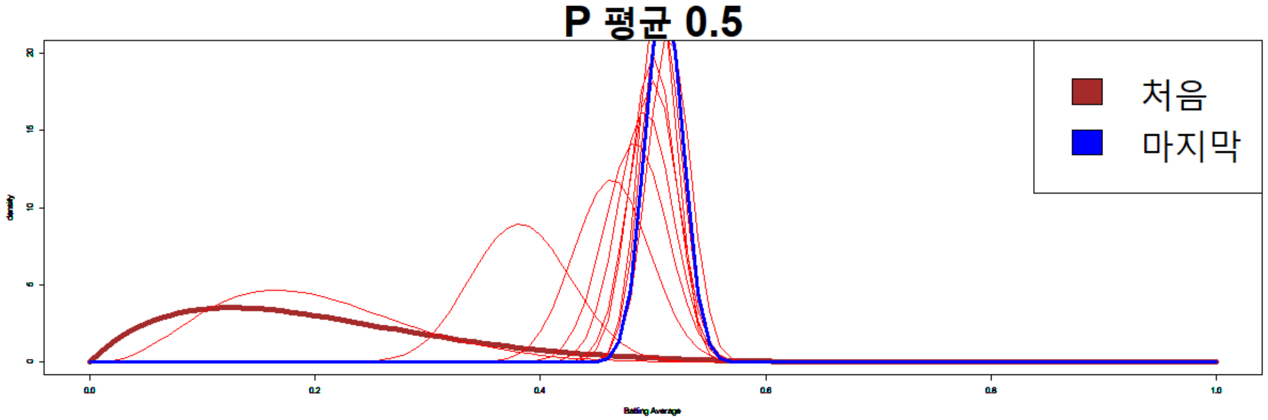

7:3 믿음을 가지고 시작
density
Batting Average


P 평균 0.51
처음
마지막
density
Batting Average


5:5 믿음
density
Batting Average


P 평균 0.51
처음
마지막
density
Batting Average


2:8 믿음을 가지고 시작
density
Batting Average


P 평균 0.5
처음
마지막
density
Batting Average

2.3.6 Bayesian inference for the Gaussian

$\boldsymbol{P}(\mu|\boldsymbol{X}) \propto \boldsymbol{P}(\mu)\boldsymbol{P}(\boldsymbol{X}|\mu)$ 을 전개하면 다음과 같은 식을 얻을 수 있다.

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\mu_{\mathrm{ML}}$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2}$$ ...가법적인 성질 (모 분산은 알고 있다.)

μ_N 의 분자 부분을 보면,

$$\mu_N = \frac{\boldsymbol{\frac{\mu_0}{\sigma_0^2} + \frac{N\mu_{ML}}{\sigma^2}}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}}$$

사전 분포의 분산을 아주 작게 설정했다면

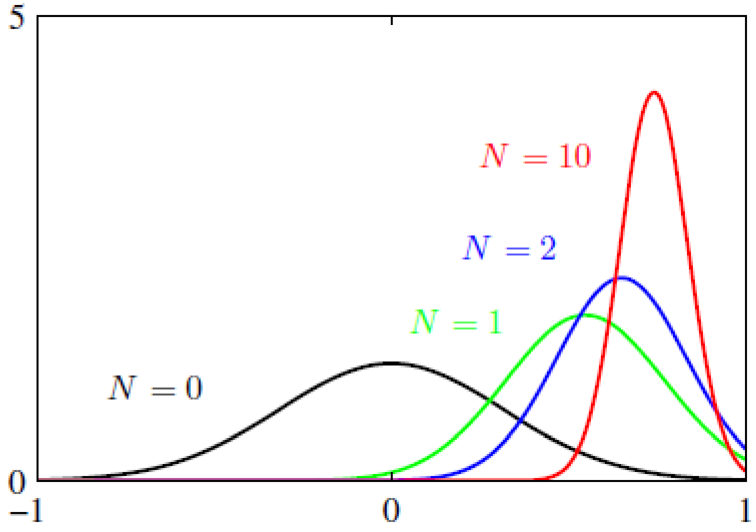
→ precision이 크다면, 사전 평균이 미치는 영향이 크고

→ 반대의 상황도 마찬가지

사전분포나 자료에 대한 확신이 강할수록 그것이 사후적 판단에 끼치는 영향이 더 커야 한다는 우리의 직관을 잘 반영한다.

2.3.6 Bayesian inference for the Gaussian

Figure 2.12 Illustration of Bayesian inference for the mean μ of a Gaussian distribution, in which the variance is assumed to be known. The curves show the prior distribution over μ (the curve labelled $N = 0$), which in this case is itself Gaussian, along with the posterior distribution given by (2.140) for increasing numbers N of data points. The data points are generated from a Gaussian of mean 0.8 and variance 0.1, and the prior is chosen to have mean 0. In both the prior and the likelihood function, the variance is set to the true value.



사전 확률 분포를 $N(0,0.1)$로 설정한 뒤, 평균의 분포를 N이 증가함에 따라 관찰한 그림.
N=10 일 때, 붉은 선을 보면 0.8 부근에서 밀도가 가장 높은 것을 확인 할 수 있다.

2.3.6 Bayesian inference for the Gaussian

Section 2.3.5

We have already seen how the maximum likelihood expression for the mean of a Gaussian can be re-cast as a sequential update formula in which the mean after observing N data points was expressed in terms of the mean after observing $N-1$ data points together with the contribution from data point $\mathbf{x}_N$. In fact, the Bayesian paradigm leads very naturally to a sequential view of the inference problem. To see this in the context of the inference of the mean of a Gaussian, we write the posterior distribution with the contribution from the final data point $\mathbf{x}_N$ separated out so that

$$p(\boldsymbol{\mu}|D) \propto \left[p(\boldsymbol{\mu}) \prod_{n=1}^{N-1} p(\mathbf{x}_n|\boldsymbol{\mu}) \right] p(\mathbf{x}_N|\boldsymbol{\mu}). \tag{2.144}$$

이전에 배웠던 순차 추정을 당연히 이용 할 수 있다는 내용.

2.3.6 Bayesian inference for the Gaussian

- MLE는 μ와Σ 를 추론하는데 사용하는 frame work이며, 데이터에 기반해 **점 추정**을 함
- 베이지안은 사전분포를 활용하여 두개의 모수의 **분포를 추정**함

We shall suppose that the variance **μ is known**

$\lambda \equiv 1/\sigma 2$

When $\mathbf{X} = \{x_1, \ldots, x_N\}$, The likelihood function that is the probability of the observed data given λ, viewed as a function of λ, is given by

$$p(\mathbf{X}|\lambda) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \lambda^{-1}) \propto \lambda^{N/2} \exp\left\{-\frac{\lambda}{2}\sum_{n=1}^{N}(x_n - \mu)^2\right\}$$

2.3.6 Bayesian inference for the Gaussian

베이지안 공식을 이용하여 분포를 추정해 봅시다.

- λ 사전분포 $\boldsymbol{P}(\lambda) = \boldsymbol{Gam}(\mu|a_0, b_0)$ .... 아랫 첨자 0는 초기 믿음에 의해 설정한 값
- $\boldsymbol{P}(\lambda|\boldsymbol{X}) \propto \boldsymbol{P}(\lambda)\boldsymbol{P}(\boldsymbol{X}|\lambda)$ ... 사후 확률분포
- 그리고 사후 확률 분포는 **<u>공액적 특성</u>**에 따라 감마분포를 따른다.

$$\boldsymbol{P}(\lambda|\boldsymbol{X}) = \boldsymbol{Gam}(\lambda|a_N, b_N)$$

$$a_N = a_0 + \frac{N}{2} \tag{2.150}$$

$$b_N = b_0 + \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2 = b_0 + \frac{N}{2}\sigma_{\mathrm{ML}}^2 \tag{2.151}$$

2.3.6 Bayesian inference for the Gaussian

$$a_N = a_0 + \frac{N}{2} \tag{2.150}$$

$$b_N = b_0 + \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2 = b_0 + \frac{N}{2}\sigma^2_{\mathrm{ML}} \tag{2.151}$$

- a_N 을 구하는 식으로부터 N 개의 관찰 데이터가 a_N 값에 미치는 영향을 확인해볼 수 있다.
  - a_N 의 식을 보면 $N/2$ 만큼 값이 보정되고 있다.
  - 만약 $N = 2a_0$ 라면 어떻게 될까? 이러면 $a_N = a_0 + \frac{2a_0}{2} = 2a_0$ 가 된다.
    - 즉, 초기 값에서 a_0 만큼 증가함을 알 수 있음.
  - b_N 도 마찬가지인데 $N = 2a_0$ 라면 사전 확률의 분산 값은 $2b_0/(2a_0) = b_0/a_0$ 이고 이를 대입하면 $b_N = 2b_0$ 가 된다.
    - 즉, 초기 값에서 b_0 만큼 증가함을 알 수 있음
  - 결국 effective number 는 $N = 2a_0$ 지점임
    - 여기서 effective number는 관찰 데이터의 영향력이 지정된 사전 확률의 영향력을 넘어서는 지점에서의 관찰 데이터의 수라고 생각하면 된다.
- 참고로 지금까지 분산을 구하기 위해 정확도(precision)를 이용하여 식을 전개하였는데, 반대로 공분산을 이용하여 식을 전개할수도 있다.
  - 이 때에는 공액 분포로 감마 분포가 아니라 역감마 분포 (*inverse gamma distribution*)를 사용한다.

2.3.6 Bayesian inference for the Gaussian

만약, μ,Σ 둘다 모를 때는 어떻게 해야 할까?   방법은 동일하다.

1. 사전분포를 정의 : $P(\mu,\lambda) = P(\mu|\lambda)P(\lambda)$ ...결합확률 법칙

$$p(\mu, \lambda) = \mathcal{N}(\mu|\mu_0, (\beta\lambda)^{-1})\mathrm{Gam}(\lambda|a, b)$$

$$\begin{aligned} p(\mu, \lambda) &\propto \left[\lambda^{1/2} \exp\left(-\frac{\lambda\mu^2}{2}\right)\right]^{\beta} \exp\{c\lambda\mu - d\lambda\} \\ &= \exp\left\{-\frac{\beta\lambda}{2}(\mu - c/\beta)^2\right\} \lambda^{\beta/2} \exp\left\{-\left(d - \frac{c^2}{2\beta}\right)\lambda\right\} \end{aligned} \quad (2.153)$$

where c, d, and β are constants. Since we can always write $p(\mu, \lambda) = p(\mu|\lambda)p(\lambda)$,

2. 사후 확률분포를 구한다.

2.3.6 Bayesian inference for the Gaussian

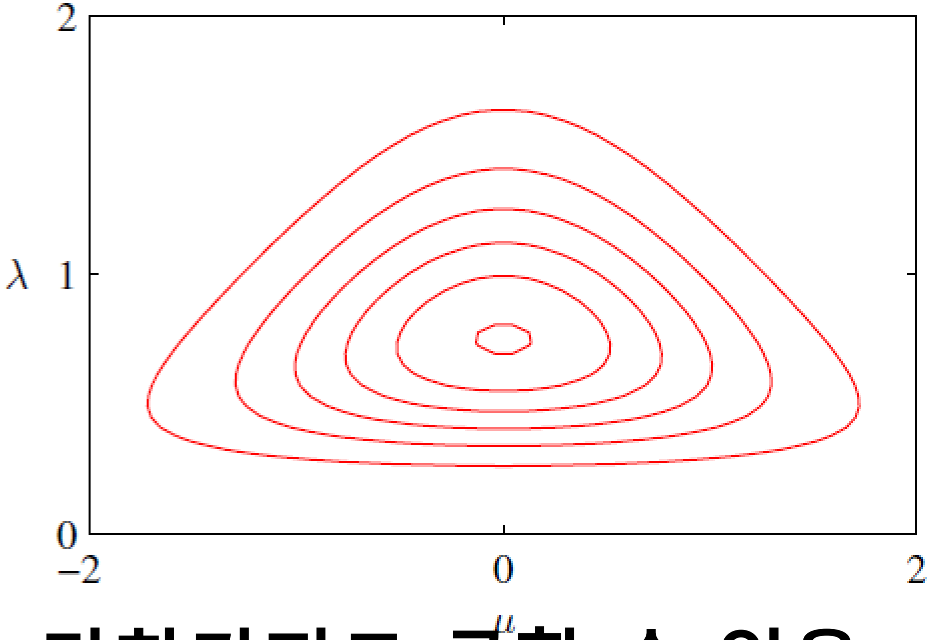
2. 사후 확률분포를 구한다.

$$\mu_0 = \frac{c}{\beta}$$

$$a = (1+\beta)/2$$

$$b = d - \frac{c^2}{2\beta}$$

Figure 2.14 Contour plot of the normal-gamma distribution (2.154) for parameter values $\mu_0 = 0$, $\beta = 2$, $a = 5$ and $b = 6$.



다변량 데이터일 경우에도 이전과 마찬가지로 구할 수 있음

2.3.8 Periodic variables

정규분포는 대체로 잘 맞기만, 항상 그렇지는 않다.

An example of a periodic variable would be the **wind direction** :
measure values of wind direction on a number of days and wish to summarize this using a parametric distribution

Another example is **calendar time**, where we may be interested in modelling quantities that are believed to be periodic over 24 hours or over an annual cycle

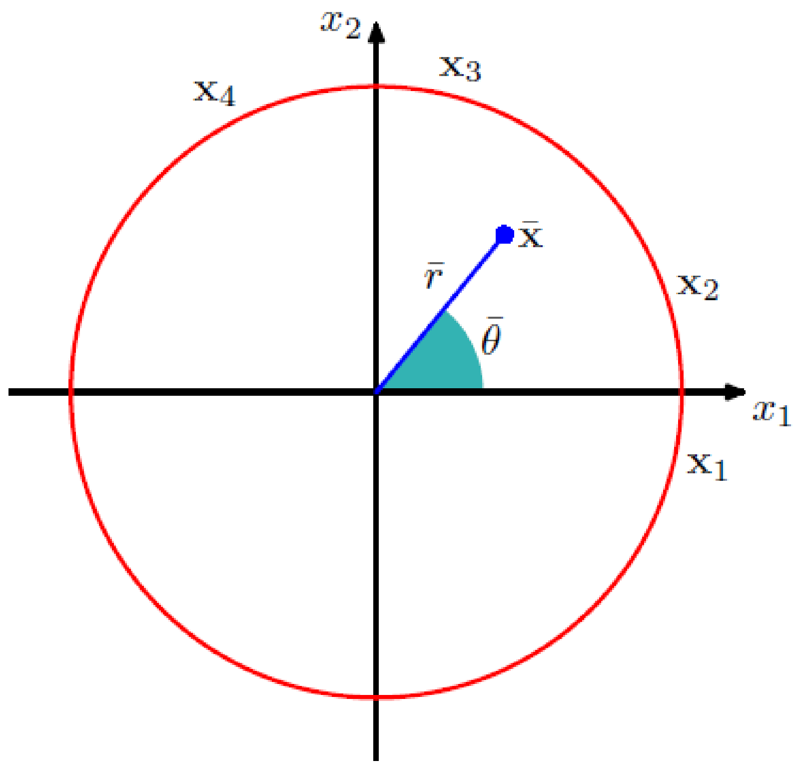
We might be tempted to treat periodic variables by choosing **some direction as the origin** and then applying a conventional distribution such as the Gaussian.

데이터가 $\theta_1 = 1^o$, $\theta_2 = 359^o$ 일 때,

- 만약, 원점을 0^o로 설정한 경우 평균은 180, 표준편차는 179가 됨
- 만약, 원점을 180^o로 설정한 경우 평균은 0, 표준편차는 1이 됨

기준에 관계없이 언제나 동일한 결과가 나오는 방법 연구 필요

2.3.8 Periodic variables



단위 원 위에 있는 x들의 평균은 다음과 같이 구할 수 있다.

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$$

이 때, 필요한 정보는 θ 뿐이니 극좌표로 변환!!

$x_n = (cos\theta_n, sin\theta_n)$ 각 좌표들의 평균은 아래와 같이 구할 수 있으며,

$\bar{x} = (\bar{r}cos\bar{\theta}, \bar{r}sin\bar{\theta})$ $\bar{\theta}$만을 얻기 위해 간단한 삼각함수 공식을 이용

$$\tan\bar{\theta} = \frac{\bar{r}\sin\bar{\theta}}{\bar{r}\cos\bar{\theta}} \qquad \bar{\theta} = \tan^{-1}\left\{\frac{\sum_n \sin\theta_n}{\sum_n \cos\theta_n}\right\}$$

이번 절에서는 방금 구한 $\bar{\theta}$가 베이지안으로도 구할 수 있는지 알아보려고 함.

2.3.8 Periodic variables

- ***von Mises*** 분포는 '정규 분포로 주기성 확률 변수에 대한 일반화'를 의미
- $P(\theta)$가 주기 2π를 가진 확률변수라고 할 때, 다음의 성질이 존재

  1) $P(\theta) \geq 0$ ... 항상 0보다 커야 하며

  2) $\int_0^{2\pi} P(\theta)d\theta = 1$ ... 확률의 합은 1

  3) $P(\theta) = P(\theta + 2\pi)$ ...2pi마다 부기를 가지고 있음

방금 예시로 돌아가 2차원에서 정규분포를 고려

$$p(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{(x_1-\mu_1)^2 + (x_2-\mu_2)^2}{2\sigma^2}\right\}$$

$$\left.\begin{aligned} x_1 &= r cos\theta \\ x_2 &= r sin\theta \end{aligned}\right\} \text{좌측 식에 대입}$$

$$\mu_1 = r_0 \cos\theta_0$$

$$\mu_2 = r_0 \sin\theta_0$$

2.3.8 Periodic variables

$$
\begin{aligned}
& -\frac{1}{2\sigma^2}\left\{(r\cos\theta - r_0\cos\theta_0)^2 + (r\sin\theta - r_0\sin\theta_0)^2\right\} \\
= \quad & -\frac{1}{2\sigma^2}\left\{1 + r_0^2 - 2r_0\cos\theta\cos\theta_0 - 2r_0\sin\theta\sin\theta_0\right\} \\
= \quad & \frac{r_0}{\sigma^2}\cos(\theta - \theta_0) + \mathrm{const}
\end{aligned}
$$

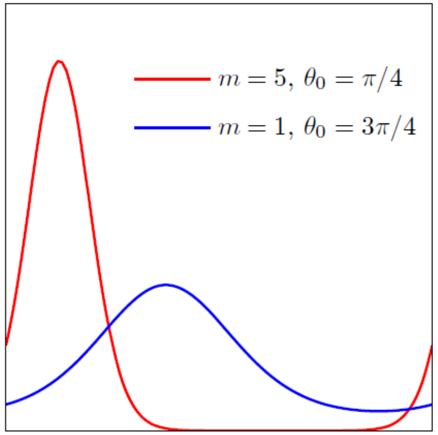
2.3.8 Periodic variables

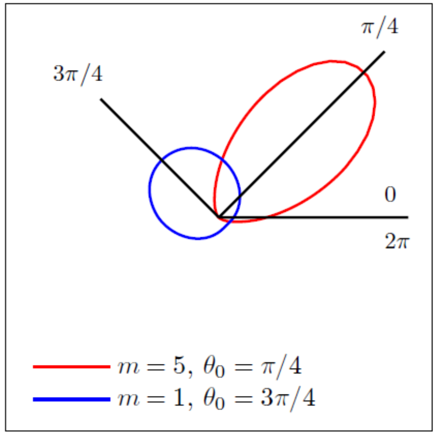
$$p(x_1, x_2) = \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{(x_1-\mu_1)^2 + (x_2-\mu_2)^2}{2\sigma^2} \right\}$$

$$p(\theta|\theta_0, m) = \frac{1}{2\pi I_0(m)} \exp\{m\cos(\theta-\theta_0)\}$$

이 분포를 ***von Mises*** 분포라고 함

기존의 정규 분포처럼 평균을 사용하고 분산 대신 m 을 사용하는 분포이다. $I_0(m) = \frac{1}{2\pi}\int_0^{2\pi} \exp\{m\cos\theta\}\, d\theta.$



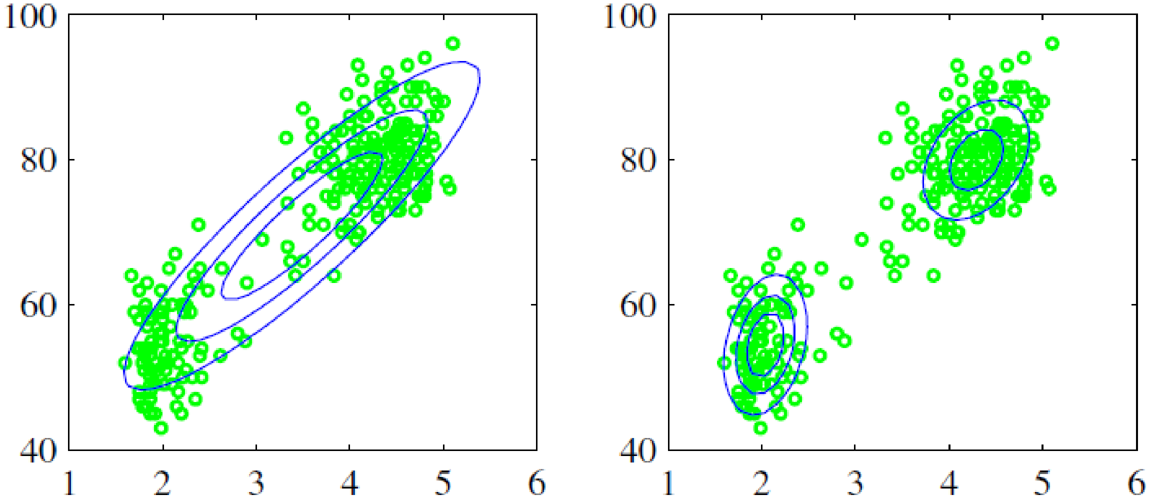


Left: 데카르트 좌표
Right: 극 좌표

2.3.9 Mixtures of Gaussians

정규분포가 항상 맞을 수 없기에 **<u>선형결합</u>**을 고려.

- 왼쪽 그림을 보면 무언가 잘못되 보임.
  → 데이터가 평균 근처에 있지 않음
- 오른쪽 그림에서 2개의 군집을 각각의 분포로 표현
  →평균 주변에 잘 뭉침



Y: 폭발 지속 시간, X: 다음 폭발까지의 시간

오른쪽 그림과 같이 데이터 상황에 맞는 정규분포들의 선형 결합을 찾고 싶음

2.3.9 Mixtures of Gaussians

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

이름: Mistures of Gaussians

특징:

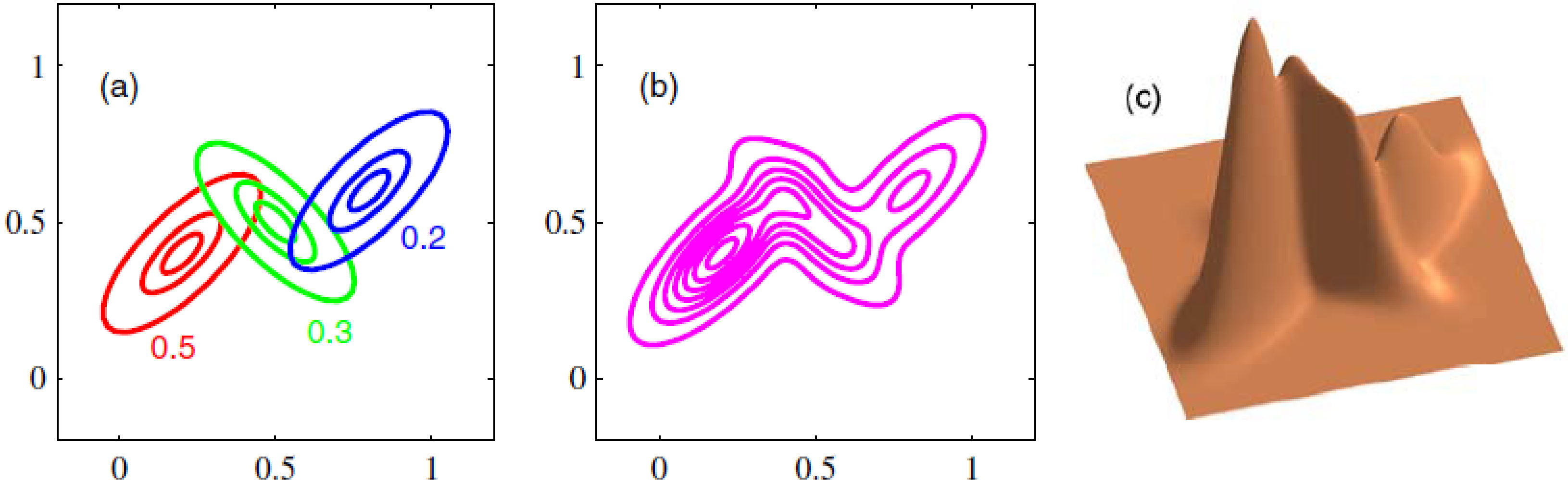
1) π_k는 *mixing coefficients* 라고 하며, 합은 1이 됨 $$\sum_{k=1}^{K} \pi_k = 1$$

2) $N(x|\mu_k\Sigma_k) \geq 0$ 이므로, $P(x) \geq 0$이다. 따라서 $0 \leq \pi_k \leq 1$

3) $P(x)$는 곱의 법칙을 만족하므로, $P(x) = \sum_k P(k)P(x|k)$ 로 전개

맨 처음 식과 정확히 같아야 하므로, $P(k) = \pi_k$, $P(x|k) = N(x|\mu_k\Sigma_k)$ 을 만족

2.3.9 Mixtures of Gaussians

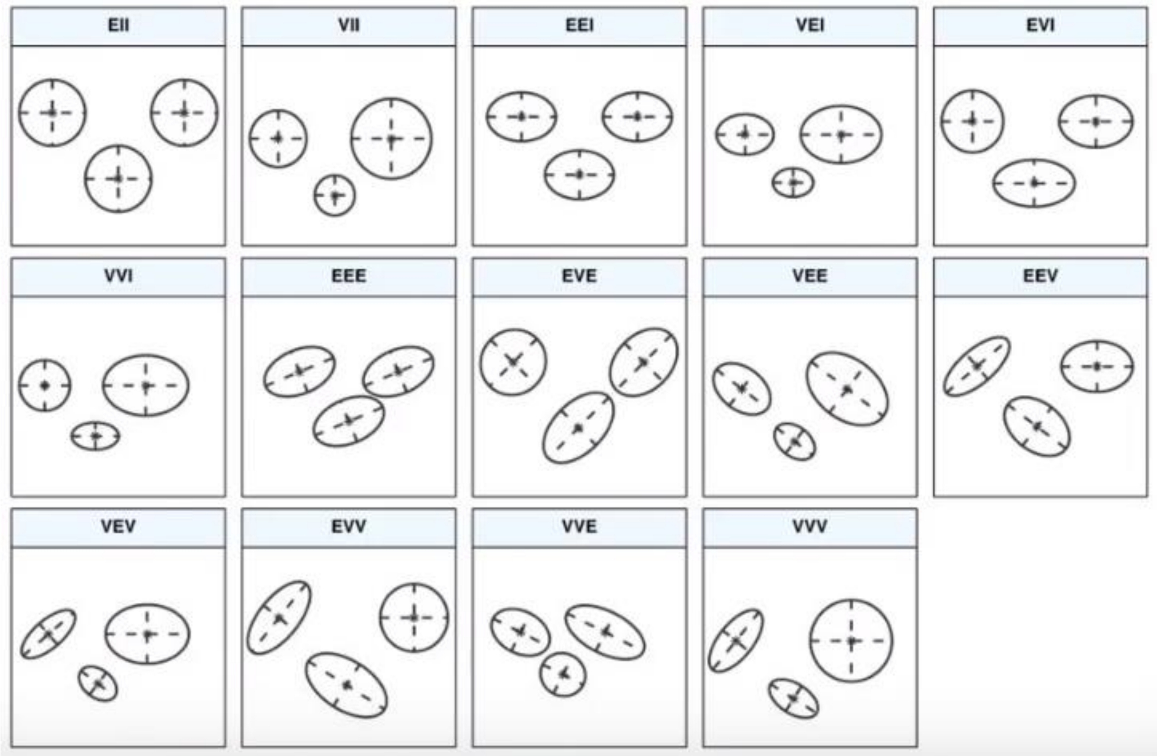


같은 정규분포를 이용 했음에도 그림 C를 보면, 왼쪽부터 높이 솟아 있는 것을 알 수 있고

그 차이는 앞선 계수에 의해 발생함.

자세한 내용은 9장에 나온다고 함.

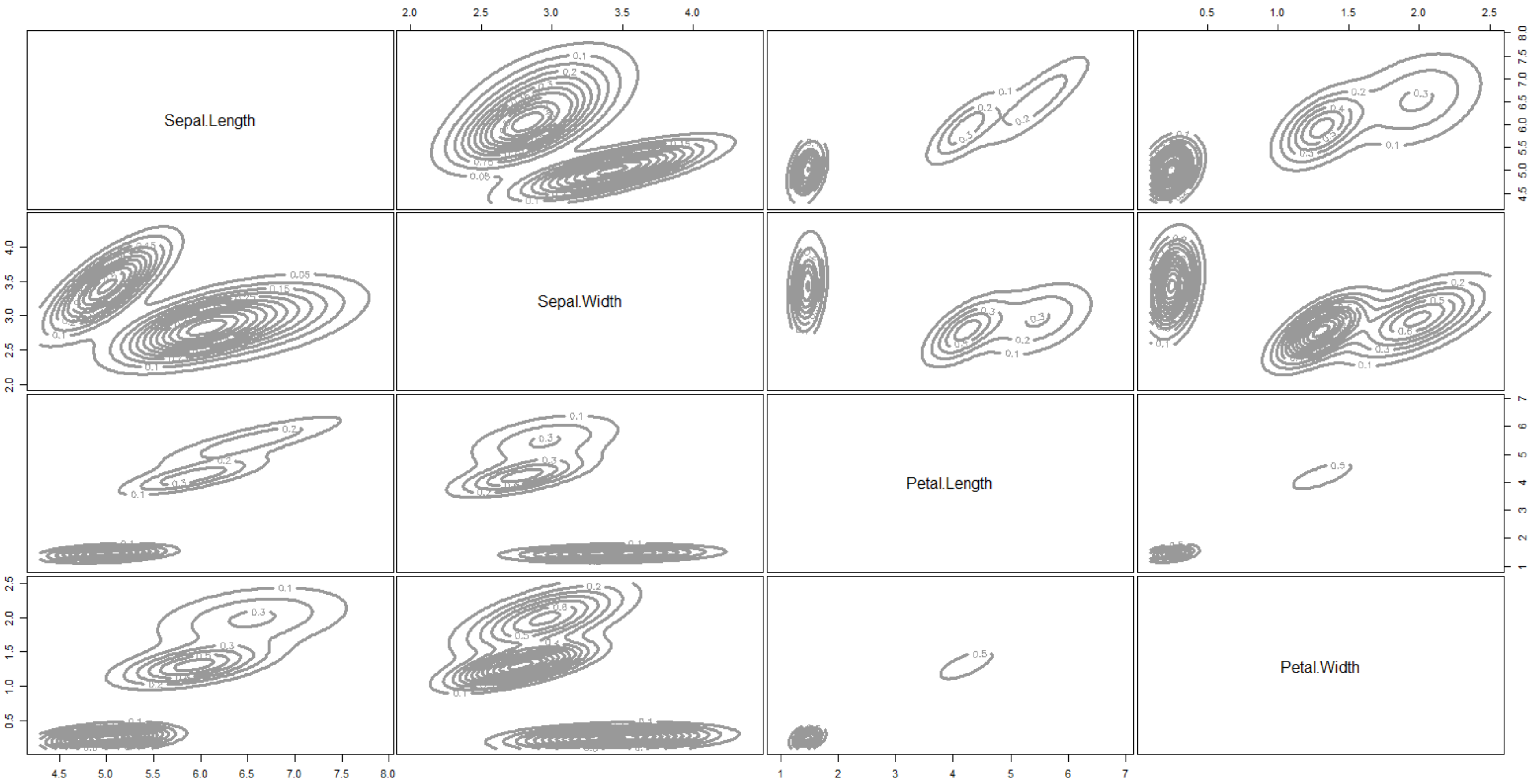
2.3.9 Mixtures of Gaussians



Model	$\boldsymbol{\Sigma}_k$	Distribution	Volume	Shape	Orientation
E	σ	Univariate	equal		
V	σ_k	Univariate	variable		
EII	$\lambda \boldsymbol{I}$	Spherical	equal	equal	
VII	$\lambda_k \boldsymbol{I}$	Spherical	variable	equal	
EEI	$\lambda \boldsymbol{A}$	Diagonal	equal	equal	coordinate axes
VEI	$\lambda_k \boldsymbol{A}$	Diagonal	variable	equal	coordinate axes
EVI	$\lambda \boldsymbol{A}_k$	Diagonal	equal	variable	coordinate axes
VVI	$\lambda_k \boldsymbol{A}_k$	Diagonal	variable	variable	coordinate axes
EEE	$\lambda \boldsymbol{D}\boldsymbol{A}\boldsymbol{D}^\top$	Ellipsoidal	equal	equal	equal
EEV	$\lambda \boldsymbol{D}_k\boldsymbol{A}\boldsymbol{D}_k^\top$	Ellipsoidal	equal	equal	variable

https://www.youtube.com/watch?v=nktiUUd6X_U

2.3.9 Mixtures of Gaussians



코드를 통해 자세히 알아 보겠음