Bayes Rule and its Applications

Bayes Rule: $P(B_k|A) = P(A|B_k)P(B_k)/\sum_{i=1}^{n} P(A|B_i)P(B_i)$

Example 1: In a certain factory, machines *A, B, and C* are all producing springs of the same length. Of their production, machines *A, B, and C* produce 2%, 1%, and 3% defective springs, respectively. Of the total production of springs in the factory, machine A produces 35%, machine B produces 25%, and machine C produces 40%. Then we have

$P(D|A) = 0.02,\ P(A) = 0.35;$

$P(D|B) = 0.01,\ P(B) = 0.25;$

$P(D|C) = 0.03,\ P(C) = 0.40.$

If one spring is selected at random from the total springs produced in a day, the probability that it is defective equals

$P(D) = \sum_{X \in \{A,B,C\}} P(D|X)P(X) = 215/10000$

If the selected spring is defective, the conditional probability that it was produced by machine *A, B, or C* can be calculated by

$P(A|D) = P(D|A)P(A)/P(D) = 70/215$

$P(B|D) = P(D|B)P(B)/P(D) = 25/215$

$P(C|D) = P(D|C)P(C)/P(D) = 120/215$

Foundation for Normal Distributions

- $\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt, \quad for\ \alpha > 0$
- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
- $\Gamma(1) = 1$
- $\Gamma(n+1) = n!\ \ \forall\ integer\ n \geq 0$ where $0! \equiv 1$.
- $\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $X \ \sim\ N(\mu, \sigma^2)$ means that $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}, \ \ -\infty < x < \infty$
- $Sampling\ \ X\ \sim\ N(\mu, \sigma^2)$ by Matlab
- $Y\ =\ random('Normal',\ \mu,\ \sigma,\ SampleSize,\ 1)$

♠ ♡ ♢ ♣ ♠ ♡ ♢ ♣ ♠ ♡ ♢ ♣

Expectation and Covariance Matrix

Let $X_1,\ X_2,\ \cdots,\ X_n$ be random variables such that the expectation, variance, and covariance are defined as follows.

$$\mu_j \;=\; E[X_j], \quad \sigma_j^2 \;=\; Var(X_j) = E[(X_j - \mu_j)^2] \tag{1}$$

$$\rho_{ij}\sigma_i\sigma_j \;=\; Cov(X_i, X_j) \;=\; E[(X_i - \mu_i)(X_j - \mu_j)] \tag{2}$$

Suppose $\mathbf{X} = [X_1,\ X_2,\ \cdots,\ X_n]^t$ be a random vector, then the expectation and covariance matrix of $\mathbf{X}$ is defined as

$$E[\mathbf{X}] = [\mu_1,\ \mu_2,\ \cdots,\ \mu_n]^t = \mu \tag{3}$$

$$Cov(\mathbf{X}) = [E[(X_i - \mu_i)(X_j - \mu_j)]] \tag{4}$$

Bayes Decision Theory

(1) $p(\omega_i)$: a priori probability

(2) $p(\mathbf{x}|\omega_i)$: class conditional density function

(3) $p(\omega_i|\mathbf{x})$: a posteriori probability

(4) $\alpha(\mathbf{x})$: an action (a decision)

(5) $\lambda(\alpha(\mathbf{x})|\omega_j)$: the loss function

$p(error) \; = \; \sum_{\mathbf{x}} p(error|\mathbf{x})p(\mathbf{x}) \; = \; \sum_{\mathbf{x}} p(\alpha(\mathbf{x}) \in \omega_i, \; \mathbf{x} \in \omega_j, \; i \neq j)p(\mathbf{x})$

$R(\alpha(\mathbf{x})|\mathbf{x}) \; = \; \sum_{j=1}^{C} \lambda(\alpha(\mathbf{x})|\omega_j)p(\omega_j|\mathbf{x})$: conditional risk for pattern $\mathbf{x}$

$\sum_{\mathbf{x}} R(\alpha(\mathbf{x})|\mathbf{x})p(\mathbf{x})$: average error of probability (error rate)

- **Bayes Decision Rule**

For each $\mathbf{x}$, find $\alpha(\mathbf{x})$ which minimizes $R(\alpha(\mathbf{x})|\mathbf{x})$

For the 0-1 loss function, i.e. $\lambda(\alpha(\mathbf{x})|\omega_j) = \begin{cases} 0 & if \;\; \alpha(\mathbf{x}) = \omega_j, \\ 1 & otherwise \end{cases}$

Then the Bayes decision rule can be reduced to

$$min \; [R(\alpha(\mathbf{x})|\mathbf{x})] \; = \; min_j[1 - p(\omega_j|\mathbf{x})] \; = \; max_i \; p(\omega_i|\mathbf{x})$$

or

Assign $\mathbf{x}$ to class ω_i if $p(\omega_i|\mathbf{x}) \; > \; p(\omega_j|\mathbf{x})$ for $j \neq i$

Example 2: $X|\omega_i \sim N(\mu_i, \sigma_i^2)$

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} exp[-(x-\mu_i)^2/2\sigma_i^2]$$

The Bayes decision rule is to *assign* x *to class* ω_i *if*

$$p(\omega_i|x) > p(\omega_j|x) \ for \ j \neq i$$

iff

$$p(x|\omega_i)p(\omega_i) > p(x|\omega_j)p(\omega_j)$$

iff

$$\frac{1}{2}ln[\sigma_j^2/\sigma_i^2] + \frac{1}{2\sigma_j^2}(x-\mu_j)^2 - \frac{1}{2\sigma_i^2}(x-\mu_i)^2 > ln[p(\omega_j)/p(\omega_i)]$$

In particular,

(a) if $\sigma_i = \sigma$ for each i, then the Bayes rule is to assign x to class ω_i if

$$(\mu_i - \mu_j)x - \frac{1}{2}(\mu_i^2 - \mu_j^2) > \sigma^2 ln[p(\omega_j)/p(\omega_i)]$$

(b) if $\sigma_i = \sigma$ and $p(\omega_i) = 1/C$ for each i, then the Bayes rule is to assign x to class ω_i if

$$(\mu_i - \mu_j)[x - \frac{1}{2}(\mu_i + \mu_j)] > 0$$

Example 3: $X|\omega_i \;\sim\; N(\mu_i, C_i)$

$$p(\mathbf{x}|\omega_i) \;=\; \frac{1}{(2\pi)^{d/2}|C_i|^{1/2}} exp[-(\mathbf{x}-\mu_i)^t C_i^{-1}(\mathbf{x}-\mu_i)/2]$$

The Bayes decision rule is to assign $\mathbf{x}$ to class ω_i if

$$p(\omega_i|x) \;>\; p(\omega_j|x) \; for \; j \neq i$$

iff

$$\frac{1}{2}\left[ln(|C_j|/|C_i|) + \left(\mathbf{x}^t C_j^{-1}\mathbf{x} - 2\mathbf{x}^t C_j^{-1}\mu_j + \mu_j^t C_j^{-1}\mu_j\right) - \left(\mathbf{x}^t C_i^{-1}\mathbf{x} - 2\mathbf{x}^t C_i^{-1}\mu_i + \mu_i^t C_i^{-1}\mu_i\right)\right]$$

$$> \; ln[p(\omega_j)/p(\omega_i)]$$

In particular,

(a) if $C_i = C$ for each i, then the Bayes decision rule is to assign $\mathbf{x}$ to class ω_i if

$$(\mu_i - \mu_j)^t C^{-1}\mathbf{x} - \frac{1}{2}[\mu_i^t C^{-1}\mu_i - \mu_j^t C^{-1}\mu_j] \;>\; ln[p(\omega_j)/p(\omega_i)]$$

(b) if $C_i = \sigma^2 I$ for each i, then the Bayes decision rule is to assign $\mathbf{x}$ to class ω_i if

$$(\mu_i - \mu_j)^t\mathbf{x} - \frac{1}{2}(\|\mu_i\|^2 - \|\mu_j\|^2) \;>\; \sigma^2 ln[p(\omega_j)/p(\omega_i)]$$

Example 4: $X|\omega_i \sim N(\mu_i, C), \quad i = 1, 2$

Let $\Lambda(\mathbf{x}) = ln[p(\mathbf{x}|\omega_1)/p(\mathbf{x}|\omega_2)] = [\mathbf{x} - \frac{1}{2}(\mu_1 + \mu_2)]^t C^{-1}(\mu_1 - \mu_2)$,
and let $r = ln[p(\omega_2)/p(\omega_1)]$.

The Bayes decision rule is to assign $\mathbf{x}$ to class ω_1 if $\Lambda(\mathbf{x}) > r$.

Note that

$$\mathbf{x} \in \omega_1 \;\Rightarrow\; \Lambda(X) \sim N(\frac{\Delta}{2}, \Delta)$$

$$\mathbf{x} \in \omega_2 \;\Rightarrow\; \Lambda(X) \sim N(\frac{-\Delta}{2}, \Delta)$$

where $\Delta = (\mu_1 - \mu_2)^t C^{-1}(\mu_1 - \mu_2)$ is called the square Mahalanobis distance.

Hint: Prove that $\int \Lambda(\mathbf{x}) p(\mathbf{x}|\omega_1) d\mathbf{x} = \frac{\Delta}{2}$ and $\int [\Lambda(\mathbf{x}) - \frac{\Delta}{2}]^2 p(\mathbf{x}|\omega_1) d\mathbf{x} = \Delta$

Then the Bayes error rate can be computed by

$$\begin{aligned} E^* &= p[\mathbf{x} \in \omega_2, \Lambda(\mathbf{x}) > r] + p[\mathbf{x} \in \omega_1, \Lambda(\mathbf{x}) < r] \\ &= p(\omega_2) \int_r^\infty \frac{1}{\sqrt{2\pi\Delta}} exp[-(y + \frac{\Delta}{2})^2/2\Delta] dy + p(\omega_1) \int_{-\infty}^r \frac{1}{\sqrt{2\pi\Delta}} exp[-(y - \frac{\Delta}{2})^2/2\Delta] dy \end{aligned}$$

Proof of $\Lambda(X)|\omega_1 \sim N(\frac{\Delta}{2}, \Delta)$

X is a multivariate normal distribution so is its linear mapping $\Lambda(X)$.

$$\begin{aligned}
E[\Lambda(X)|\omega_1] &= \int \Lambda(\mathbf{x})p(\mathbf{x}|\omega_1)d\mathbf{x} \\
&= \int \{[\mathbf{x} - \tfrac{1}{2}(\mu_1+\mu_2)]^t C^{-1}(\mu_1-\mu_2)\}p(\mathbf{x}|\omega_1)d\mathbf{x} \\
&= \int \mathbf{x}^t C^{-1}(\mu_1-\mu_2)p(\mathbf{x}|\omega_1)d\mathbf{x} \ - \ \tfrac{1}{2}\int (\mu_1+\mu_2)^t C^{-1}(\mu_1-\mu_2)p(\mathbf{x}|\omega_1)d\mathbf{x} \\
&= \int [\mathbf{x}^t p(\mathbf{x}|\omega_1)d\mathbf{x}]C^{-1}(\mu_1-\mu_2) \ - \ \tfrac{1}{2}[(\mu_1+\mu_2)^t C^{-1}(\mu_1-\mu_2)]\int p(\mathbf{x}|\omega_1)d\mathbf{x} \\
&= \mu_1^t C^{-1}(\mu_1-\mu_2) \ - \ \tfrac{1}{2}(\mu_1+\mu_2)^t C^{-1}(\mu_1-\mu_2) \\
&= \tfrac{1}{2}(\mu_1-\mu_2)^t C^{-1}(\mu_1-\mu_2) \ = \ \tfrac{\Delta}{2}
\end{aligned}$$

$$\begin{aligned}
Var[\Lambda(X)|\omega_1] &= \int [\Lambda(\mathbf{x}) - \tfrac{\Delta}{2}]^2 p(\mathbf{x}|\omega_1)d\mathbf{x} \\
&= \int [(\mathbf{x}-\mu_1)^t C^{-1}(\mu_1-\mu_2)]^2 p(\mathbf{x}|\omega_1)d\mathbf{x} \\
&= (\mu_1-\mu_2)^t C^{-1}\{\int [(\mathbf{x}-\mu_1)(\mathbf{x}-\mu_1)]^t p(\mathbf{x}|\omega_1)d\mathbf{x}\}C^{-1}(\mu_1-\mu_2) \\
&= (\mu_1-\mu_2)^t C^{-1} C C^{-1}(\mu_1-\mu_2) \\
&= (\mu_1-\mu_2)^t C^{-1}(\mu_1-\mu_2) = \Delta
\end{aligned}$$

In Example 4, let

$$\mu_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad C = I_2, \text{ and let } p(\omega_1) = p(\omega_2) = 1/2$$

Then

$$\Lambda(\mathbf{x}) = (\mathbf{x} - \mathbf{0})^t I^{-1} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2(x_1 + x_2)$$

$$r = ln\,[p(\omega_2)/p(\omega_1)] = 0$$

$$\Delta = (\mu_1 - \mu_2)^t I^{-1} (\mu_1 - \mu_2) = 8$$

The Bayes decision rule is to assign $\mathbf{x}$ to ω_1 if $2x_1 + 2x_2 > 0$

The Bayes error rate is $p(error) = \Phi(-\frac{\sqrt{\Delta}}{2}) = \Phi(-\frac{\sqrt{8}}{2}) = \Phi(-\sqrt{2})$

Example 5: *Bayes decision theory in a discrete case*

Consider a two-class problem and assume that each component x_j of the pattern $\mathbf{x}$ is either 0 or 1 with the conditional probabilities

$$p_j = p(x_j = 1|\omega_1) \textit{ and } q_j = p(x_j = 1|\omega_2) \tag{5}$$

Suppose that the components of $\mathbf{x}$ are conditionally independent. Then

$$p(\mathbf{x}|\omega_1) = \prod_{j=1}^{d} p_j^{x_j}(1-p_j)^{1-x_j}, \quad p(\mathbf{x}|\omega_2) = \prod_{j=1}^{d} q_j^{x_j}(1-q_j)^{1-x_j} \tag{6}$$

Let the log-likelihood ratio $\Lambda(\mathbf{x}) = ln[p(\mathbf{x}|\omega_1)/p(\mathbf{x}|\omega_2)]$, and $r = ln[p(\omega_2)/p(\omega_1)]$, then

$$\Lambda(\mathbf{x}) = \sum_{j=1}^{d} x_j ln[p_j(1-q_j)/q_j(1-p_j)] + \sum_{j=1}^{d} ln[(1-p_j)/(1-q_j)] \tag{7}$$

The Bayes decision rule is to assign $\mathbf{x}$ to ω_1 if $\Lambda(\mathbf{x}) > r$.

The decision boundary is $\sum_{j=1}^{d} w_j x_j + w_0 = 0$, where

$w_j = ln[p_j(1-q_j)/q_j(1-p_j)], \ 1 \le j \le d,$

$w_0 = \sum_{j=1}^{d} ln[(1-p_j)/(1-q_j)] + ln[p(\omega_1)/p(\omega_2)]$

In particular, if $p(\omega_1) = p(\omega_2) = 1/2$, $p_j = p$, $q_j = q = 1-p$ for $1 \le j \le d$, d is odd and $p > q$, then the Bayes error rate is

$$E^* = \sum_{k=0}^{(d-1)/2} \frac{d!}{(d-k)!k!} p^k (1-p)^{d-k} \tag{8}$$

◇ $X|\omega_1 \sim N(3, 2^2)$ and $X|\omega_2 \sim N(6, 1^2)$

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} exp[-(x-\mu_i)^2/2\sigma_i^2], \quad i = 1, 2$$

The Maximum Likelihood (ML) decision is to *assign x to* ω_1 *if* $p(x|\omega_1) > p(x|\omega_2)$

The Bayes decision is to *assign x to* ω_1 *if* $p(x|\omega_1)p(\omega_1) > p(x|\omega_2)p(\omega_2)$

Note that ML is the special case by assuming $p(\omega_1) = p(\omega_2) = \frac{1}{2}$ which need not be true in practical applications. We shall show the effect of $p(\omega_1) = \frac{2}{3}$, $p(\omega_2) = \frac{1}{3}$.

ML Decision: $y \in \omega_1$ if $y < 7 - \sqrt{4 + (8ln2)/3}$ or $y > 7 + \sqrt{4 + (8ln2)/3}$

Bayes Decision: $y \in \omega_1$ if $y < 5$ or $y > 9$

The error probability can be computed by

$$\begin{aligned} Err &= p(\omega_1) \int_a^b \frac{1}{\sqrt{2\pi}2} \exp[-(x-3)^2/(2 \cdot 2^2)]dx \\ &+ p(\omega_2)\{\int_{-\infty}^a \frac{1}{\sqrt{2\pi}} \exp[-(x-6)^2/2]dx \\ &\quad + \int_b^\infty \frac{1}{\sqrt{2\pi}} \exp[-(x-6)^2/2]dx\} \end{aligned}$$

where a=4.5817, b=9.4183 for ML decision, and a=5, b=9 for Bayes decision, then

♣ $Err_{ML} = 0.1650$ and $Err_{Bayes} = 0.1532$.

◇ $X|\omega_1 \sim Rayleigh(1^2)$ and $X|\omega_2 \sim Rayleigh(3^2)$

$$p(x|\omega_i) = \frac{x}{\sigma_i^2}\exp(-\frac{x^2}{2\sigma_i^2}), \quad i = 1, 2$$

The Maximum Likelihood (ML) decision is to *assign x to* ω_1 *if* $p(x|\omega_1) > p(x|\omega_2)$

The Bayes decision is to *assign x to* ω_1 *if* $p(x|\omega_1)p(\omega_1) > p(x|\omega_2)p(\omega_2)$

Note that ML is the special case by assuming $p(\omega_1) = p(\omega_2) = \frac{1}{2}$ which need not be true in practical applications. We shall show the effect of $p(\omega_1) = \frac{3}{4}$, $p(\omega_2) = \frac{1}{4}$ with $\sigma_1 < \sigma_2$.

ML Decision: $y \in \omega_1$ if $0 \le y < \sqrt{\frac{2\sigma_1^2\sigma_2^2}{\sigma_2^2-\sigma_1^2} \times ln(\frac{\sigma_2^2}{\sigma_1^2})}$

Bayes Decision: $y \in \omega_1$ if $0 \le y < \sqrt{\frac{2\sigma_1^2\sigma_2^2}{\sigma_2^2-\sigma_1^2} \times [ln(\frac{\sigma_2^2}{\sigma_1^2}) + ln(\frac{p(\omega_1)}{p(\omega_2)})]}$

The error probability can be computed by

$$\begin{aligned} Err &= p(\omega_1)\int_t^\infty \frac{x}{\sigma_1^2}\exp(-\frac{x^2}{2\sigma_1^2})dx \\ &\quad + p(\omega_2)\int_0^t \frac{x}{\sigma_2^2}\exp(-\frac{x^2}{2\sigma_2^2})dx \\ &= p(\omega_1)\cdot\exp(-\frac{t^2}{2\sigma_1^2}) + p(\omega_2)\cdot[1-\exp(-\frac{t^2}{2\sigma_2^2})] \end{aligned}$$

where t=2.2235 for ML decision, and t=2.7232 for Bayes decision, then

♣ $Err_{ML} = 0.1234$ and $Err_{Bayes} = 0.1028$.

◇ $X|\omega_1 \sim \chi^2(r_1)$ and $X|\omega_2 \sim \chi^2(r_2)$

$$p(x|\omega_i) = \frac{1}{\Gamma(r_i/2)2^{r_i/2}} x^{(r_i/2)-1} e^{-x/2}, \; x > 0, \quad i = 1, 2$$

The Maximum Likelihood (ML) decision is to *assign* x *to* ω_1 *if* $p(x|\omega_1) > p(x|\omega_2)$

The Bayes decision is to *assign* x *to* ω_1 *if* $p(x|\omega_1)p(\omega_1) > p(x|\omega_2)p(\omega_2)$

Note that ML is the special case by assuming $p(\omega_1) = p(\omega_2) = \frac{1}{2}$ which need not be true in practical applications. We shall show the effect of $p(\omega_1) = \frac{1}{7}$, $p(\omega_2) = \frac{6}{7}$ with $r_1 < r_2$.

ML Decision:

$$y \in \omega_1 \;\; if \;\; \frac{\Gamma(r_2/2)}{\Gamma(r_1/2)} 2^{(r_2-r_1)/2} > x^{(r_2-r_1)/2}$$

Bayes Decision:

$$y \in \omega_1 \;\; if \;\; \frac{\Gamma(r_2/2)}{\Gamma(r_1/2)} 2^{(r_2-r_1)/2} \frac{p(\omega_1)}{p(\omega_2)} > x^{(r_2-r_1)/2}$$

The error probability can be computed by

$$\begin{aligned} Err &= p(\omega_1) \int_t^\infty \frac{1}{\Gamma(r_1/2)2^{r_1/2}} x^{(r_1/2)-1} e^{-x/2} dx \\ &+ p(\omega_2) \int_0^t \frac{1}{\Gamma(r_2/2)2^{r_2/2}} x^{(r_2/2)-1} e^{-x/2} dx \end{aligned}$$

When $r_1 = 4, r_2 = 8, \; p(\omega_1) = 1/7, \; p(\omega_2) = 6/7$, we have
t=4.899 for ML decision, and t=2.0 for Bayes decision, and

♣ $Err_{ML} = \mathbf{0.2411}$ and $Err_{Bayes} = \mathbf{0.1214}$.

$\diamond$ $X|\omega_1 \sim N(\mu_1, \sigma_1^2)$ and $X|\omega_2 \sim N(\mu_2, \sigma_2^2)$

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} exp[-(x-\mu_i)^2/2\sigma_i^2], \quad i = 1, 2$$

The Maximum Likelihood (ML) decision is to *assign* x *to* ω_1 *if* $p(x|\omega_1) > p(x|\omega_2)$

The Bayes decision is to *assign* x *to* ω_1 *if* $p(x|\omega_1)p(\omega_1) > p(x|\omega_2)p(\omega_2)$

Note that ML is the special case by assuming $p(\omega_1) = p(\omega_2) = \frac{1}{2}$ which need not be true in practical applications.

The error probability can be computed by

$$\begin{aligned} Err &= p(\omega_1) \textstyle\int_T^\infty \frac{1}{\sqrt{2\pi}\sigma_1} \exp[-(x-\mu_1)^2/(2\sigma_1^2)]dx \\ &+ p(\omega_2) \textstyle\int_0^T \frac{1}{\sqrt{2\pi}\sigma_2} \exp[-(x-\mu_2)^2/(2\sigma_2^2)]dx \end{aligned}$$

$\square$ As soon as T is chosen, Parameters $p(\omega_i), \mu_i, \sigma_i^2,\ i = 1, 2$ could be estimated, so is *Err*.

- Otsu (1979) and Fisher (1936) chose T to maximize the following criterion, respectively

$$\begin{aligned} \sigma_B^2 &= p(\omega_1)p(\omega_2)(\mu_1 - \mu_2)^2 \\ Fisher &= \tfrac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2} \end{aligned}$$

A Simple Thresholding Algorithm

By Otsu, IEEE Trans. on SMC, 62-66, 1979

(1) $p_i \ \leftarrow \ \frac{n_i}{n}$, where $n = \sum_{i=0}^{G-1} n_i$

(2) $u_T \ = \ \sum_{k=0}^{G-1} k p_k$

(3) Do for $k = 0, \quad G-1$

$\omega(k) \ \leftarrow \ \sum_{i=0}^{k} p_i$

$u(k) \ \leftarrow \ \sum_{i=0}^{k} i p_i$

$\sigma_B^2(k) \ \leftarrow \ \frac{[u_T \omega(k) - u(k)]^2}{\omega(k)[1-\omega(k)]}$

(4) Select k^* such that $\sigma_B^2(k)$ is maximized

Note that

- $\omega_0 = \omega(k)$, $u_0 = \sum_{i=0}^{k} iP(i|C_0) = u(k)/\omega(k)$
- $\omega_1 = 1 - \omega(k)$, $u_1 = \sum_{i=k+1}^{G-1} iP(i|C_1) = [u_T - u(k)]/[1-\omega(k)]$
- $\sigma_\omega^2 = \omega_0 \sigma_0^2 + \omega_1 \sigma_1^2$
- $\sigma_B^2 = \omega_0 (u_0 - u_T)^2 + \omega_1 (u_1 - u_T)^2$
- $\sigma_T^2 = \sum_{i=0}^{G-1} (i - u_T)^2 p_i$

♣ $\sigma_\omega^2 + \sigma_B^2 = \sigma_T^2$

♣ $\epsilon = \frac{\sigma_B^2}{\sigma_T^2}, \quad \kappa = \frac{\sigma_T^2}{\sigma_\omega^2}, \quad \lambda = \frac{\sigma_B^2}{\sigma_\omega^2}$

Gamma Function and the Volumes of High Dimensional Spheres

1. Define $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ for $x > 0$ and let $\gamma = \int_0^\infty e^{-x^2} dx$. Then

(a) $\Gamma(x) = \int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$

(b) Show that $\Gamma(\frac{1}{2}) = 2\gamma = \sqrt{\pi}$

(c) $\Gamma(x+1) = x\Gamma(x),\ for\ x > 0,\ \Gamma(n) = (n-1)!$ if $n \in N$.

(d) The volume of a $d - dimensional$ unit sphere is $\pi^{d/2}/\Gamma(\frac{d}{2}+1)$.

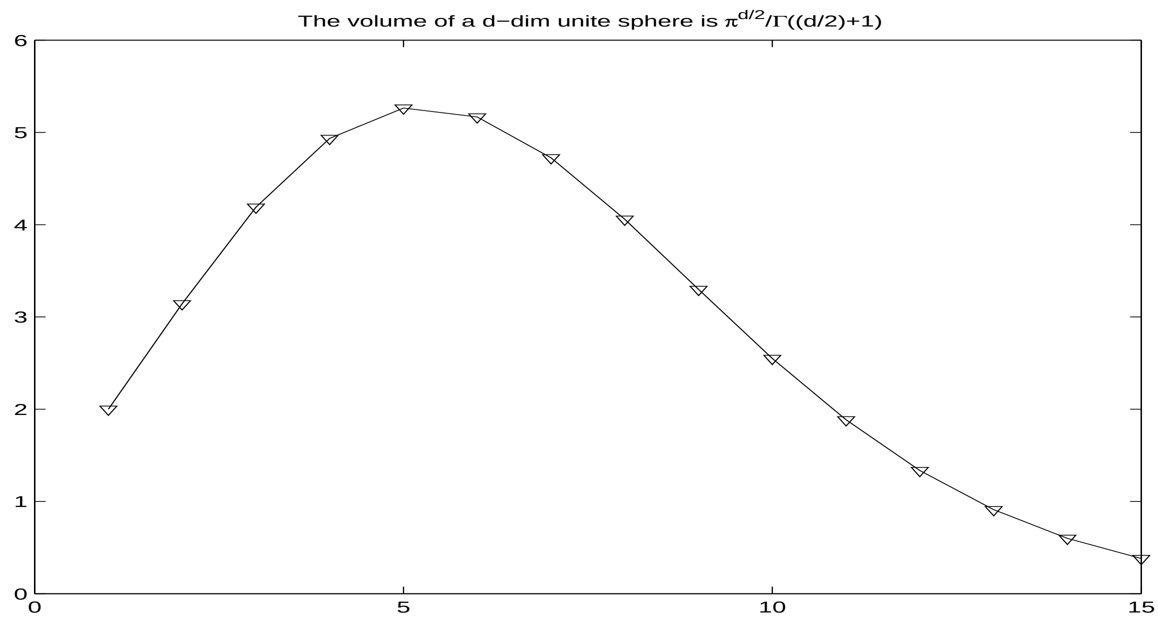

Figure 1: The Volume of a High Dimensional Sphere.

```
V(1)=2;      V(2)=pi;     V(3)=4*pi/3;
V(4)=pi*pi/2;             V(5)=8*pi*pi/15;       V(6)=pi*pi*pi/6;
V(7)=16*pi*pi*pi/105;     V(8)=(pi)^4/24;        V(9)=32*(pi)^4/945;
V(10)=(pi)^5/120;         V(11)=64*(pi)^5/10395; V(12)=(pi)^6/720;
V(13)=128*(pi)^6/135135; V(14)=(pi)^7/5040;     V(15)=256*(pi)^7/2027025;
D=1:15;
plot(D,V,'b-v')
title('The volume of a d-dim unit sphere is \pi^{d/2}/\Gamma((d/2)+1)')
```

The Derivation of Volumn for an n-Dimensional Sphere

□ For $n = 2$,

$$\begin{aligned} x_1 &= r\cos\theta, \quad 0 \le \theta \le 2\pi \\ x_2 &= r\sin\theta, \quad 0 \le r \le R \end{aligned} \quad , \quad J_2 = \frac{\partial(x_1, x_2)}{\partial(r, \theta)} = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_2}{\partial r} \\ \frac{\partial x_1}{\partial \theta} & \frac{\partial x_2}{\partial \theta} \end{vmatrix} = r$$

The volumn is computed by

$$V_2 = \int_0^R \int_0^{2\pi} J_2 dr d\theta = \int_0^R \int_0^{2\pi} r dr d\theta = \pi R^2$$

□ For $n = 3$,

$$\begin{aligned} x_1 &= r\cos\theta_1\cos\theta_2, \quad 0 \le \theta_2 \le 2\pi \\ x_2 &= r\cos\theta_1\sin\theta_2, \quad -\tfrac{\pi}{2} \le \theta_1 \le \tfrac{\pi}{2} \\ x_3 &= r\sin\theta_1, \quad 0 \le r \le R \end{aligned} \quad , \quad J_3 = \frac{\partial(x_1, x_2, x_3)}{\partial(r, \theta_1, \theta_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_2}{\partial r} & \frac{\partial x_3}{\partial r} \\ \frac{\partial x_1}{\partial \theta_1} & \frac{\partial x_2}{\partial \theta_1} & \frac{\partial x_3}{\partial \theta_1} \\ \frac{\partial x_1}{\partial \theta_2} & \frac{\partial x_2}{\partial \theta_2} & \frac{\partial x_3}{\partial \theta_2} \end{vmatrix} = r^2\cos\theta_1$$

The volumn is computed by

$$V_3 = \int_0^R \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} J_3 dr d\theta_1 d\theta_2 = \int_0^R \int_{-\pi/2}^{\pi/2} \int_0^{2\pi} r^2\cos\theta_1 dr d\theta_1 d\theta_2 = \frac{4\pi R^3}{3}$$

□ For $n \geq 4$,

$$
\begin{aligned}
x_1 &= r\cos\theta_1\cos\theta_2\cdots\cdots\cos\theta_{n-2}\cos\theta_{n-1}, && 0 \le \theta_{n-1} \le 2\pi \\
x_2 &= r\cos\theta_1\cos\theta_2\cdots\cdots\cos\theta_{n-2}\sin\theta_{n-1}, && -\tfrac{\pi}{2} \le \theta_{n-2} \le \tfrac{\pi}{2} \\
x_3 &= r\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-3}\sin\theta_{n-2}, && -\tfrac{\pi}{2} \le \theta_{n-3} \le \tfrac{\pi}{2} \\
&\vdots && \vdots \\
x_j &= r\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-j}\sin\theta_{n-j+1}, && -\tfrac{\pi}{2} \le \theta_{n-j} \le \tfrac{\pi}{2} \\
&\vdots && \vdots \\
x_{n-1} &= r\cos\theta_1\sin\theta_2, && -\tfrac{\pi}{2} \le \theta_1 \le \tfrac{\pi}{2} \\
x_n &= r\sin\theta_1, && 0 \le r \le R
\end{aligned}
$$

Note that $\sum_{i=1}^{n} x_i^2 = r^2$ and denote $c_i = \cos\theta_i, \quad s_i = \sin\theta_i$ for $1 \le i \le n-1$. Then the Jacobian $J_n = \frac{\partial(x_1, x_2, \cdots, x_n)}{\partial(r, \theta_1, \cdots, \theta_{n-1})}$ is computed as

$$
J_n = r^{n-1}
\begin{vmatrix}
c_1c_2\cdots c_{n-2}c_{n-1} & c_1c_2\cdots c_{n-2}s_{n-1} & c_1c_2\cdots c_{n-3}s_{n-2} & \cdots & s_1 \\
-s_1c_2\cdots c_{n-2}c_{n-1} & -s_1c_2\cdots c_{n-2}s_{n-1} & -s_1c_2\cdots c_{n-3}s_{n-2} & \cdots & c_1 \\
-c_1s_2\cdots c_{n-2}c_{n-1} & -c_1s_2\cdots c_{n-2}s_{n-1} & -c_1s_2\cdots c_{n-3}s_{n-2} & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
-c_1c_2\cdots s_{n-2}c_{n-1} & -c_1c_2\cdots s_{n-2}s_{n-1} & c_1c_2\cdots c_{n-3}c_{n-2} & \cdots & 0 \\
-c_1c_2\cdots c_{n-2}s_{n-1} & c_1c_2\cdots c_{n-2}c_{n-1} & 0 & \cdots & 0
\end{vmatrix}
$$

Then

$$J_n = r^{n-1}c_1^{n-1}c_2^{n-2}\cdots c_{n-1}^1 s_1 s_2 \cdots s_{n-1} \begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ -t_1 & -t_1 & -t_1 & \cdots & \frac{1}{t_1} \\ -t_2 & -t_2 & -t_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{n-2} & -t_{n-2} & \frac{1}{t_{n-2}} & \cdots & 0 \\ -t_{n-1} & \frac{1}{t_{n-1}} & 0 & \cdots & 0 \end{vmatrix}, \quad where \;\; t_i = \frac{sin\theta_i}{\cos\theta_i}$$

Subtracting each column from the preceding one and do further simplifications, we obtain

$$\begin{aligned} J_n &= r^{n-1}c_1^{n-1}c_2^{n-2}\cdots c_{n-1}^1 s_1 s_2 \cdots s_{n-1}\Pi_{j=1}^{n-1}(t_j + \frac{1}{t_j}) \\ &= r^{n-1}c_1^{n-1}c_2^{n-2}\cdots c_{n-1}^1 s_1 s_2 \cdots s_{n-1}\Pi_{j=1}^{n-1}(\frac{1}{s_j \cdot c_j}) \\ &= r^{n-1}c_1^{n-2}c_2^{n-3}\cdots c_{n-3}^2 c_{n-2}^1 \\ &= r^{n-1}\cos^{n-2}\theta_1 \cos^{n-3}\theta_2 \cdots \cos^2\theta_{n-3}\cos^1\theta_{n-2} \end{aligned}$$

Therefore, the volumn of an n-dimensional sphere can be calculated by

$$\begin{aligned} V_n &= \int_0^R \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} (J_n) d\theta_1 d\theta_2 \cdots d\theta_{n-2} d\theta_{n-1} dr \\ &= \int_0^R \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} [r^{n-1} \cos^{n-2}\theta_1 \cos^{n-3}\theta_2 \cdots \cos\theta_{n-2}] d\theta_1 d\theta_2 \cdots d\theta_{n-2} d\theta_{n-1} dr \\ &= \tfrac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} \cdot R^n \end{aligned}$$

Note that the above computations exploit the properties of the following Gamma and Beta functions, and trignometry.

$$\begin{aligned} \Gamma(\alpha) &= \int_0^\infty e^{-t} t^{\alpha-1} dt \quad for\ \alpha > 0 \\ Beta(\alpha, \beta) &= \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx, \quad where\ \alpha, \beta > 0 \\ Beta(\alpha, \beta) &= \tfrac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)} \end{aligned}$$

- *Relationship Between Gamma and Beta Functions*

$$\begin{aligned} \Gamma(x)\Gamma(y) &= \int_0^\infty e^{-u} u^{x-1} du \int_0^\infty e^{-v} v^{y-1} dv \\ &= \int_0^\infty \int_0^\infty e^{-u-v} u^{x-1} v^{y-1} du dv \\ &= \int_{z=0}^\infty \int_{t=0}^1 e^{-z} (zt)^{x-1} [z(1-t)]^{y-1} z dt dz \quad by\ putting\ u = zt,\ v = z(1-t) \\ &= \int_{z=0}^\infty e^{-z} z^{x+y-1} dz \int_{t=0}^1 t^{x-1}(1-t)^{y-1} dt \\ &= \Gamma(x+y) Beta(x, y) \end{aligned}$$

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \ for\ \alpha > 1$$

$$\Gamma(1) = 1, \qquad \Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

$$\int_0^{2\pi} d\theta_{n-1} = 2\pi, \qquad \int_{-\pi/2}^{\pi/2} \cos\theta_{n-2} d\theta_{n-2} = 2,$$

$$\int_{-\pi/2}^{\pi/2} \cos^2\theta_{n-3} d\theta_{n-3} = \frac{\pi}{2}, \quad \int_{-\pi/2}^{\pi/2} \cos^m\theta d\theta = \int_0^{\pi/2} 2\cos^m\theta d\theta, \quad 3 \le \alpha \le n-2$$

Now

$$\begin{aligned}
\int_0^{\pi/2} 2\cos^m\theta d\theta &= \int_1^0 2x^m \frac{1}{-\sqrt{1-x^2}}dx \quad \textit{by letting } x=\cos\theta \\
= \int_0^1 2x^m(1-x^2)^{-1/2}dx &= \int_0^1 2y^{m/2}(1-y)^{-1/2}\frac{1}{2\sqrt{y}}dy, \quad \textit{where} \quad x=\sqrt{y} \\
= \int_0^1 y^{\frac{m}{2}-\frac{1}{2}}(1-y)^{-1/2}dy &= \int_0^1 2y^{\frac{m+1}{2}-1}(1-y)^{\frac{1}{2}-1}dy \\
= Beta(\tfrac{m+1}{2},\tfrac{1}{2}) &= \tfrac{\Gamma(\frac{m+1}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{m}{2}+1)}
\end{aligned}$$

Therefore,

$$\begin{aligned}
V_n = &\int_0^R\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}\cdots\int_{-\pi/2}^{\pi/2}\int_{-\pi/2}^{\pi/2}[r^{n-1}\cos^{n-2}\theta_1\cos^{n-3}\theta_2\cdots\cos\theta_{n-2}]d\theta_1 d\theta_2\cdots d\theta_{n-2}d\theta_{n-1}dr \\
= &\tfrac{R^n}{n}\cdot(2\pi)\cdot(2)\cdot(\tfrac{\pi}{2})\cdot\Pi_{m=3}^{n-2}Beta(\frac{m+1}{2},\ \frac{1}{2}) \ = \ \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}\cdot R^n
\end{aligned}$$