SYDE 372 - Winter 2011
Introduction to Pattern Recognition

Pattern Representation: Random Vectors

Alexander Wong

Department of Systems Design Engineering
University of Waterloo

Outline

Representing features

- In statistical pattern recognition, patterns are represented using random vectors (i.e., vector of random variables)
- Behaviour of patterns follows joint probabilistic behaviour of associated measurements making up random vector
- Therefore, important to review basics of:
  - univariate (one variable) probability theory
  - multivariate (multiple variables in random vector) probability theory

Random variables

- **Random variable** is a single scalar (e.g., x) that is random.
- Random variables either discrete (e.g., random integer) or continuous (e.g., random real number)
- Random variable characterized by its cumulative distribution function (CDF) (F_x):

$$F_x(\tau) = Pr(x < \tau) \tag{1}$$

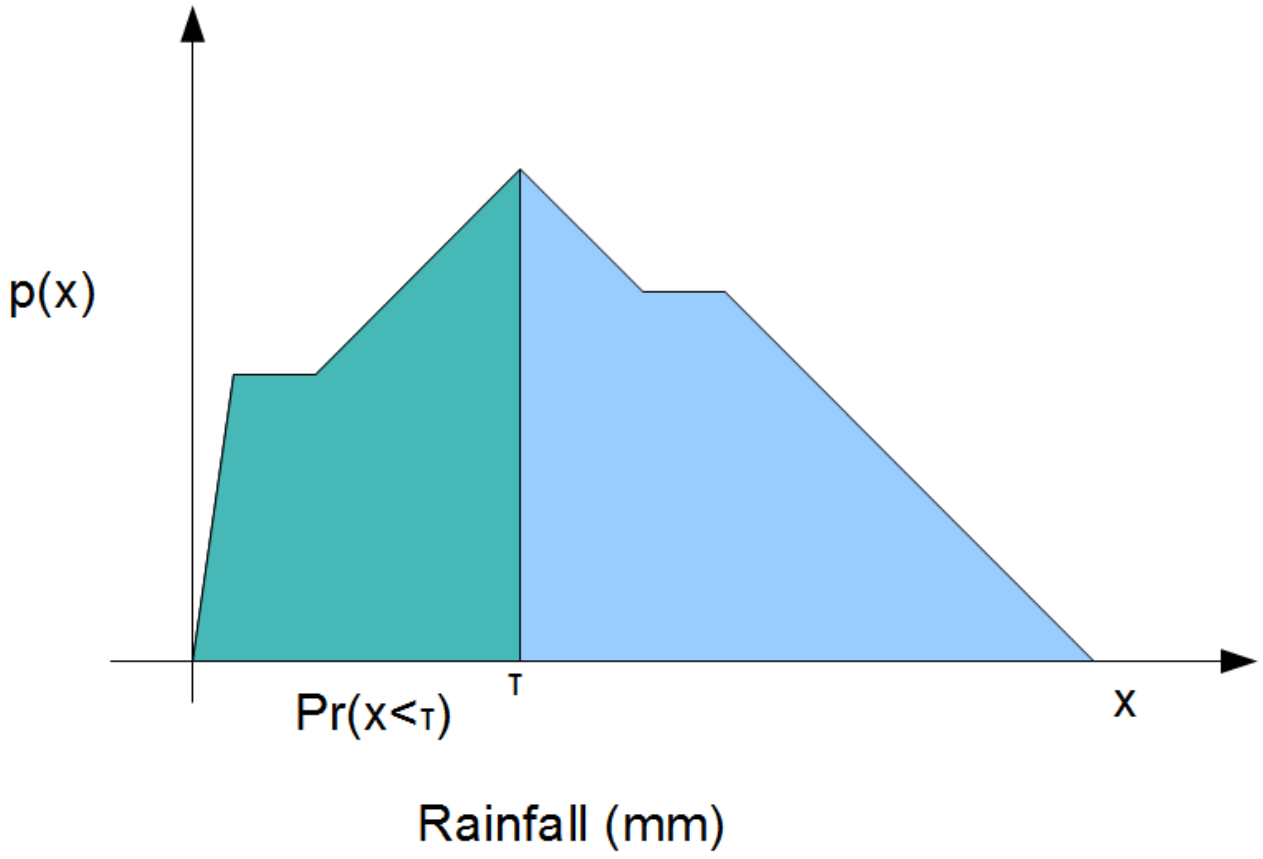
Random variables

- More convenient to characterize using probability distribution (PDF) $p(x)$:

$$F_x(\tau) = Pr(x < \tau) = \int_{-\infty}^{\tau} p(x)dx \tag{2}$$

$$\int_{-\infty}^{\infty} p(x)dx = 1 \tag{3}$$

Example: Probability of Rainfall Amount



Expectation

- Expectation:

$$E[f(x)] = \int_{-\infty}^{\infty} p(x)f(x)dx \tag{4}$$

- Choice of f leads to definitions of mean (μ_x) and variance (σ_x^2):

$$\mu_x = E[x] \tag{5}$$

$$\sigma_x^2 = E[x^2] - E[x]^2 = E[(x - E[x])^2] \tag{6}$$

- Such statistical measures are very important for compactly defining statistical class models!

Common distribution models

- Gaussian (by far the most common):

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp -\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2 \tag{7}$$

- Uniform:

$$p(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & b < x \end{cases} \tag{8}$$

- Exponential:

$$p(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases} \tag{9}$$

- Allows for compact representations of statistical models using a couple of parameters

Joint statistics

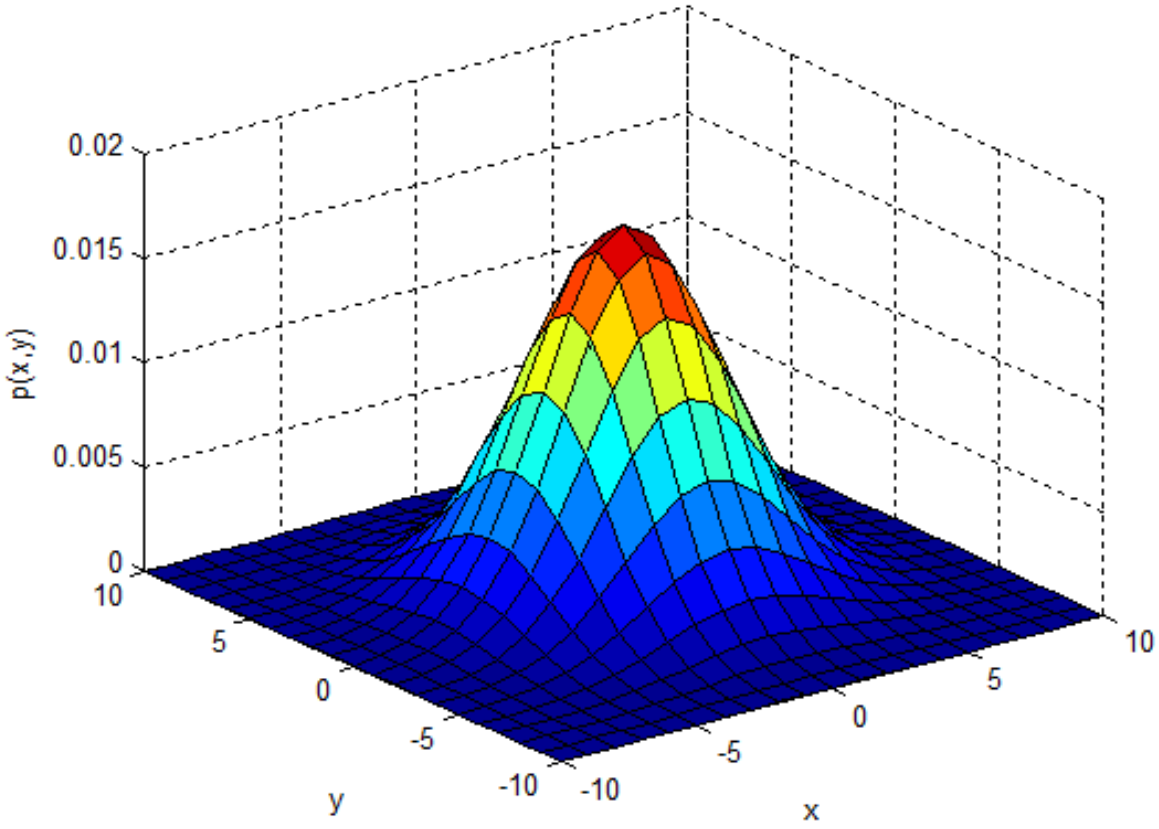
- Now that we understand the statistics of a single random variable x, how about characterizing the relationship of two random variables x and y?
- Can be characterized using joint probability distribution ($p(x, y)$):

$$Pr(x < \alpha, y < \beta) = \int_{-\infty}^{\alpha} \int_{-\infty}^{\beta} p(x, y) dx dy \tag{10}$$

- Marginal distribution ($p(x)$):

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dx dy \tag{11}$$

Example: Joint Probability



Joint statistics

- Expectation:

$$E[f(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)f(x,y)dxdy \tag{12}$$

Because of linearity of integral operator:

$$E[x+y] = E[x] + E[y] \tag{13}$$

Independence and Correlation

- Two random variables x and y are independent if:

$$p(x, y) = p(x)p(y) \tag{14}$$

- Means that knowing x tells us nothing about y.
- x and y are uncorrelated if:

$$E[xy] = E[x]E[y] \tag{15}$$

- If x and y are independent, they are uncorrelated.
- Note: uncorrelatedness does not imply independence!

Correlation

- Correlation:

$$E[(x - \mu_x)(y - \mu_y)] \tag{16}$$

- Correlation coefficient:

$$\rho_{x,y} = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y} \tag{17}$$

- Measures ability to predict y as linear function of x
  - $\rho_{x,y} = 0$ implies no predictability (x and y are uncorrelated)
  - $\rho_{x,y} = \pm 1$ implies perfect predictability (x and y are deterministically linearly related)

Conditional statistics

- $p(x|y)$ is the PDF for x conditioned on y
- $p(x|A)$ is the PDF for x given event A
- Conditional statistics very important for statistical pattern recognition, since we are assessing the unknown (e.g., identity of pattern) conditioned on what's known (e.g., measurement). For example:
  - $p(x|A)$: distribution of measurements given a class (likelihood distribution)
  - $p(A|x)$: distribution of a class given measurements (posterior distribution)
- Very important when building statistical classifiers.

Relationship between conditional statistics and joint statistics

- Related by Bayes' rule (most important equation in pattern recognition!)

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(y|x)p(x)}{p(y)} \tag{18}$$

- In the context of classification:

$$p(x|A) = \frac{p(A|x)p(x)}{p(A)} \tag{19}$$

where $p(A)$ is the prior/marginal of A and $p(B)$ is the prior/marginal of B

Random vectors

- A random vector $\underline{x}$ of dimension n is a column vector of random variables:

$$\underline{x} = [x_1 \quad x_2 \quad \dots \quad x_n]^T \tag{20}$$

- PDF of $\underline{x}$ is joint density function of random variables in vector:

$$Pr(x_1 < \tau_1, \dots, x_n < \tau_n) = \int_{-\infty}^{\tau_1} \dots \int_{-\infty}^{\tau_n} p(\underline{x}) d\underline{x} \tag{21}$$

Random vectors

- Marginal distribution of a sub-set:

$$P(x_1, \ldots, x_{l-1}, x_{l+1}, x_n) = \int_{-\infty}^{\infty} p(\underline{x}) dx_l \tag{22}$$

- Expectation

$$E[f(\underline{x})] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p(\underline{x}) f(\underline{x}) d\underline{x} \tag{23}$$

- Mean:

$$\underline{u}_x = E[\underline{x}] \tag{24}$$

Covariance

- Covariance:

$$\Sigma_x = E[(\underline{x} - \mu_x)(\underline{x} - \mu_x)^T] \tag{25}$$

- Unlike variance, which is a scalar, covariance is an $n \times n$ matrix:

$$(\Sigma_x)_{i,j} = E[(x_i - \mu_i)(x_j - \mu_j)^T] \tag{26}$$

- Just indicates correlation between x_i and x_j.
- Diagonal terms are the variances of the random variables (hence, always positive)
- $\Sigma_x = {\Sigma_x}^T$

Correlation, Independence, Bayes' Rule for Random Vectors

- Two random vectors $\underline{x}$ and $\underline{y}$ are independent if:

$$p(\underline{x}, \underline{y}) = p(\underline{x})p(\underline{y}) \tag{27}$$

- x and y are uncorrelated if:

$$E[\underline{x}\underline{y}^T] = E[\underline{x}]E[\underline{y}^T] \tag{28}$$

- Bayes' Rule:

$$p(\underline{x}|A) = \frac{p(A|\underline{x})p(\underline{x})}{p(A)} \tag{29}$$

where $p(A)$ is the prior/marginal of A and $p(B)$ is the prior/marginal of B

Sample statistics

- Given probability density p, you can work out any expectation or correlation you want.
- **Issue**: In reality, often don't know the probability density, or even the mean or covariance of a random vector!
- We need to infer them from actual observations $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_N$.
- Such inferred statistics are called **sample statistics**.

Sample statistics

- Sample mean:

$$\underline{m}_x = \frac{1}{N}\sum_{i=1}^{N}\underline{x}_i \tag{30}$$

- Recall that the covariance is $E[\underline{x}\underline{x}^T] - E[\underline{x}]E[\underline{x}]^T$.
- Sample covariance:

$$S_x = \frac{1}{N}\sum_{i=1}^{N}(\underline{x}_i - \underline{m})(\underline{x}_i - \underline{m})^T = \frac{1}{N}\sum_{i=1}^{N}(\underline{x}_i\underline{x}_i^T) - \underline{m}\underline{m}^T. \tag{31}$$

- Computing empirical pdf is challenging and will be discuss later...

Sample statistics example

- Suppose we are given the following data: $\underline{x}_1 = [2 \;\; 1]^T, \underline{x}_2 = [3 \;\; 2]^T, \underline{x}_3 = [2 \;\; 3]^T, \underline{x}_4 = [1 \;\; 2]^T$.
- Sample mean:

$$\begin{aligned} \underline{m}_x &= \tfrac{1}{4}\left\{[2 \;\; 1]^T + [3 \;\; 2]^T + [2 \;\; 3]^T + [1 \;\; 2]^T\right\} \\ \underline{m}_x &= [2 \;\; 2]^T. \end{aligned} \tag{32}$$

- Sample covariance:

$$\begin{aligned} S_x = \; & \tfrac{1}{4}\{[2 \;\; 1]^T[2 \;\; 1] + [3 \;\; 2]^T[3 \;\; 2] + [2 \;\; 3]^T[2 \;\; 3] \\ & + [1 \;\; 2]^T[1 \;\; 2]\} - [2 \;\; 2]^T[2 \;\; 2] \\ & \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \end{aligned} \tag{33}$$

- Based on S_x, the features are uncorrelated!

Multivariate Gaussian Distribution

- Since the Gaussian distribution is by far the most common statistical model used in pattern recognition, it is important to understand how it can be used in the multivariate case (e.g., more than one feature)
- Definition:

$$p(\underline{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}((\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu}))\right\} \tag{34}$$

Important special cases

- **Case 1: n = 1**
- For only one variable,
  - $\Sigma = \sigma^2$
  - $|\Sigma| = \Sigma = \sigma^2$
- Therefore, we have

$$p(\underline{x}) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{-\frac{1}{2}((x-\mu)^2/\sigma^2)\right\} \tag{35}$$

Important special cases

- **Case 2: Covariance is diagonal**
- This means that all components of $\underline{x}$ are uncorrelated!
- Much easier to deal with since the following then holds true:
  - $\Sigma = diag(1/\sigma^2)$, since $diag(\sigma^2)diag(1/\sigma^2) = I$
  - $|\Sigma| = \prod_{i=1}^{n} \sigma_{ii}^2$
- Therefore, we have

$$p(\underline{x}) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}\sigma_{ii}} \exp\left\{-\frac{1}{2}((x_i - \mu_i)^2/\sigma_{ii}^2)\right\} \tag{36}$$

- So we have just the product of univariate Gaussian distributions.

Important special cases

- **Case 3: n=2**
- Very useful for visualizing and developing intuition about n-dimensional case.
- One important way of visualizing bivariate Gaussian distributions is to sketch the **equiprobability contour** (all points along contour have equal probability), defined as:

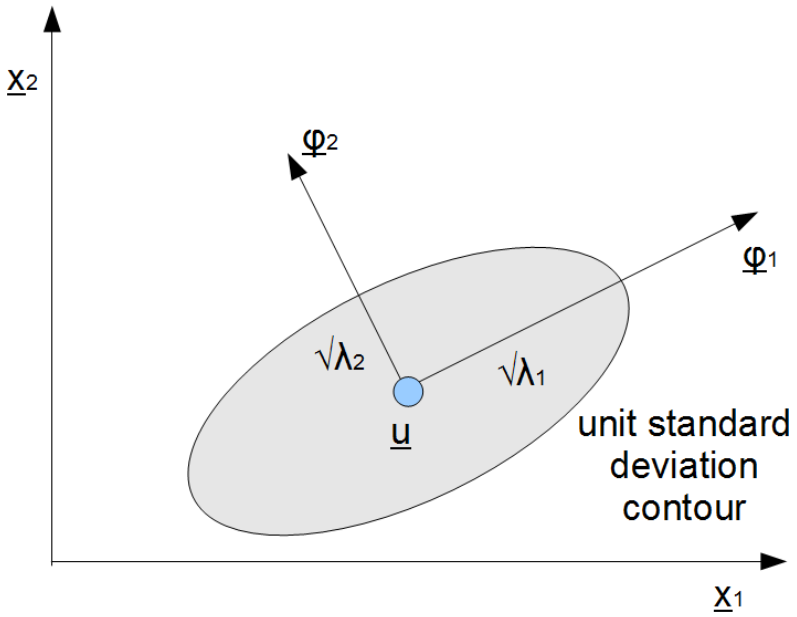
$$p(\underline{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}((\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu}))\right\} = \mathrm{C} \tag{37}$$

- Which is the same as saying

$$(\underline{x}-\underline{\mu})^T\Sigma^{-1}(\underline{x}-\underline{\mu}) = C. \tag{38}$$

Equiprobability contour

- Just an ellipse with the following properties:
  - Centered at $\underline{\mu}$,
  - Axes are eigenvectors of Σ,
  - Lengths of axes equal to the square roots of eigenvalues of Σ

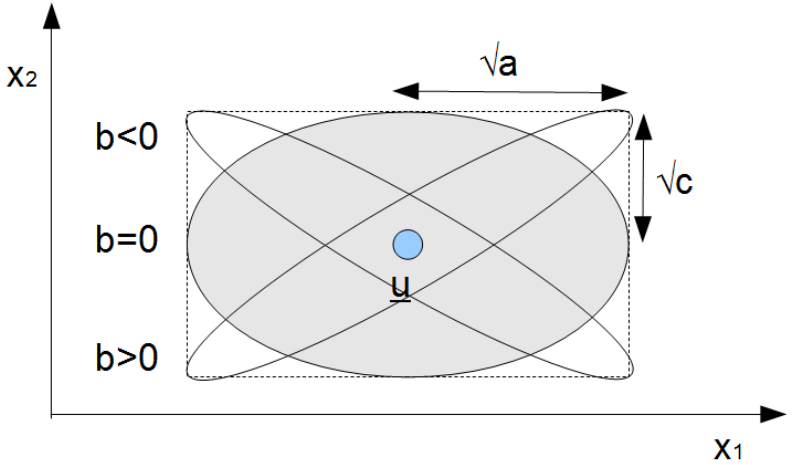


Equiprobability contour quick tips

- 2D covariance in the form of:

$$\Sigma = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad a, c \geq 0 \tag{39}$$

- So...



Equiprobability contour quick tips

- Given mean $\underline{m}$ and covariance matrix S,
  1. Compute eigenvalues λ_1 and λ_2

$$\det(S - \lambda I) = 0 \tag{40}$$

  2. Compute eigenvectors Φ_1 and Φ_2

$$Sv = \lambda \underline{v} \tag{41}$$

$$(S - \lambda I)\underline{v} = \underline{0} \tag{42}$$

  3. Sketch ellipse, centered at $\underline{\mu}$, with axes as Φ_1 and Φ_2 and length of axes as $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$.

Equiprobability contour: Example

- Suppose we are given the following data: $\underline{x}_1 = [2 \;\; 1]^T, \underline{x}_2 = [3 \;\; 2]^T, \underline{x}_3 = [2 \;\; 7]^T, \underline{x}_4 = [5 \;\; 2]^T$.
- Sketch the equiprobability contour.

Equiprobability contour: Example

- Step 1: Find sample mean and sample covariance:

$$\begin{aligned} \underline{m}_x &= \tfrac{1}{4}\left\{[2\ \ 1]^T + [3\ \ 2]^T + [2\ \ 7]^T + [5\ \ 2]^T\right\} \\ \underline{m}_x &= [3\ \ 3]^T. \end{aligned} \tag{43}$$

$$\begin{aligned} S_x &= \tfrac{1}{4}\{[2\ \ 1]^T[2\ \ 1] + [3\ \ 2]^T[3\ \ 2] + [2\ \ 7]^T[2\ \ 7] \\ &\quad + [5\ \ 2]^T[5\ \ 2]\} - [3\ \ 3]^T[3\ \ 3] \\ S_x &= \tfrac{1}{4}\begin{bmatrix} 10.5 & 8 \\ 8 & 14.5 \end{bmatrix} - \begin{bmatrix} 9 & 9 \\ 9 & 9 \end{bmatrix} \\ S_x &= \begin{bmatrix} 3/2 & -1 \\ -1 & 11/2 \end{bmatrix} \end{aligned} \tag{44}$$

Equiprobability contour: Example

- Observations:
  - the features are correlated (non-zero non-diagonal elements),
  - the features have different variances (the diagonal elements are not equal)
  - the equiprobability contour is rotated clockwise (the non-diagonal elements are negative)

Equiprobability contour: Example

- Compute eigenvalues λ_1 and λ_2

$$\det(S - \lambda I) = 0 \tag{45}$$

$$\det(\begin{bmatrix} 3/2 & -1 \\ -1 & 11/2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}) = 0 \tag{46}$$

$$(3/2 - \lambda)(11/2 - \lambda) - 1 = 0 \tag{47}$$

$$4\lambda^2 - 28\lambda + 29 = 0 \tag{48}$$

- Therefore, the eigenvalues are $\lambda_1 = 1956/341$ and $\lambda_2 = 431/341$

Equiprobability contour: Example

- Compute eigenvectors Φ_1 and Φ_2

$$(S - \lambda I)\underline{v} = \underline{0} \tag{49}$$

- For $\lambda_1 = 1956/341$:

$$\left(\begin{bmatrix} 3/2 & -1 \\ -1 & 11/2 \end{bmatrix} - \begin{bmatrix} 1956/341 & 0 \\ 0 & 1956/341 \end{bmatrix}\right)\underline{v} = \underline{0} \tag{50}$$

$$\begin{bmatrix} -1597/377 & -1 \\ -1 & -89/377 \end{bmatrix} \underline{v} = \underline{0} \tag{51}$$

$$\Phi_1 = \begin{bmatrix} 139/605 \\ -764/785 \end{bmatrix} \tag{52}$$

Equiprobability contour: Example

- Compute eigenvectors Φ_1 and Φ_2

$$(S - \lambda I)\underline{v} = \underline{0} \tag{53}$$

- For $\lambda_2 = 431/341$:

$$\left(\begin{bmatrix} 3/2 & -1 \\ -1 & 11/2 \end{bmatrix} - \begin{bmatrix} 431/341 & 0 \\ 0 & 431/341 \end{bmatrix}\right)\underline{v} = \underline{0} \tag{54}$$

$$\begin{bmatrix} 89/377 & -1 \\ -1 & 1597/377 \end{bmatrix}\underline{v} = \underline{0} \tag{55}$$

$$\Phi_2 = \begin{bmatrix} 764/785 \\ 139/605 \end{bmatrix} \tag{56}$$

Equiprobability contour: Example

- Sketch equiprobability contour