

The Gaussian Distribution

Chris Williams

School of Informatics, University of Edinburgh

October 2007

- Probability density functions
- Univariate Gaussian
- Multivariate Gaussian
- Mahalanobis distance
- Properties of Gaussian distributions
- Graphical Gaussian models
- Read: Bishop sec 2.3 (to p 93)

1 / 19

2 / 19

Continuous distributions

- Probability density function (pdf) for a continuous random variable X

$$P(a \leq X \leq b) = \int_a^b p(x) dx$$

therefore

$$P(x \leq X \leq x + \delta x) \simeq p(x) \delta x$$

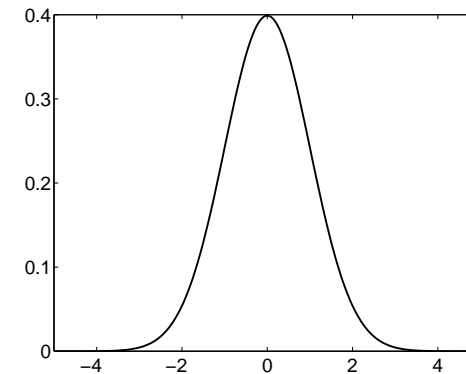
- **Example:** Gaussian distribution

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

shorthand notation $X \sim N(\mu, \sigma^2)$

- Standard normal (or Gaussian) distribution $Z \sim N(0, 1)$
- Normalization

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



- Cumulative distribution function

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z p(z') dz'$$

3 / 19

4 / 19

Bivariate Gaussian I

- Expectation

$$E[g(X)] = \int g(x)p(x)dx$$

- mean, $E[X]$
- Variance $E[(X - \mu)^2]$
- For a Gaussian, mean = μ , variance = σ^2
- Shorthand: $x \sim N(\mu, \sigma^2)$

- Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$
- If X_1 and X_2 are independent

$$p(x_1, x_2) = \frac{1}{2\pi(\sigma_1^2\sigma_2^2)^{1/2}} \exp -\frac{1}{2} \left\{ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right\}$$

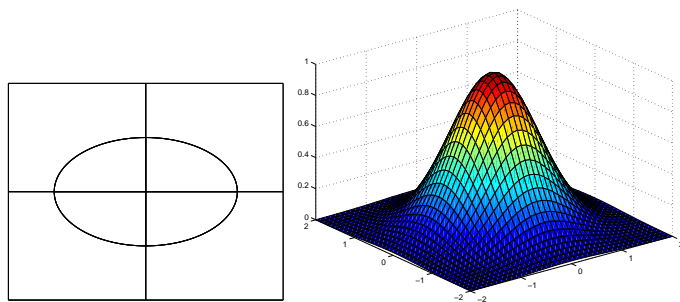
- Let $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$

$$p(\mathbf{x}) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}} \exp -\frac{1}{2} \left\{ (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

5/19

6/19

Bivariate Gaussian II

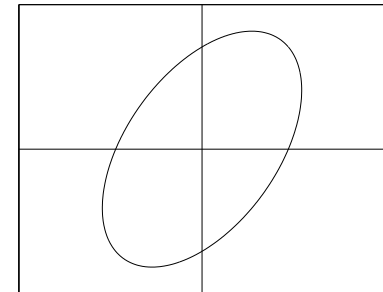


- Covariance
- $\boldsymbol{\Sigma}$ is the covariance matrix

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T]$$

$$\Sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$$

- Example: plot of weight vs height for a population



7/19

8/19

- $P(\mathbf{x} \in \mathcal{R}) = \int_{\mathcal{R}} p(\mathbf{x}) d\mathbf{x}$
- Multivariate Gaussian

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

- Σ is the covariance matrix

$$\Sigma = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T]$$

$$\Sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$$

- Σ is symmetric
- Shorthand $\mathbf{x} \sim N(\boldsymbol{\mu}, \Sigma)$
- For $p(\mathbf{x})$ to be a density, Σ must be positive definite
- Σ has $d(d+1)/2$ parameters, the mean has a further d

9 / 19

$$d_{\Sigma}^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \Sigma^{-1} (\mathbf{x}_i - \mathbf{x}_j)$$

- $d_{\Sigma}^2(\mathbf{x}_i, \mathbf{x}_j)$ is called the Mahalanobis distance between \mathbf{x}_i and \mathbf{x}_j
- If Σ is diagonal, the contours of d_{Σ}^2 are axis-aligned ellipsoids
- If Σ is not diagonal, the contours of d_{Σ}^2 are *rotated* ellipsoids

$$\Sigma = U \Lambda U^T$$

where Λ is diagonal and U is a rotation matrix

- Σ is positive definite \Rightarrow entries in Λ are positive

10 / 19

- Fully general $\Sigma \implies$ variables are correlated
- Spherical or isotropic. $\Sigma = \sigma^2 I$. Variables are independent
- Diagonal $[\Sigma]_{ij} = \delta_{ij} \sigma_i^2$ Variables are independent
- Rank-constrained: $\Sigma = WW^T + \Psi$, with W being a $d \times q$ matrix with $q < d - 1$ and Ψ diagonal. This is the factor analysis model. If $\Psi = \sigma^2 I$, then we have the probabilistic principal components analysis (PPCA) model

11 / 19

- Linear transformations of Gaussian RVs are Gaussian

$$\mathbf{X} \sim N(\boldsymbol{\mu}_x, \Sigma)$$

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

$$\mathbf{Y} \sim N(\mathbf{A}\boldsymbol{\mu}_x + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}^T)$$

- Sums of Gaussian RVs are Gaussian

$$Y = X_1 + X_2$$

$$E[Y] = E[X_1] + E[X_2]$$

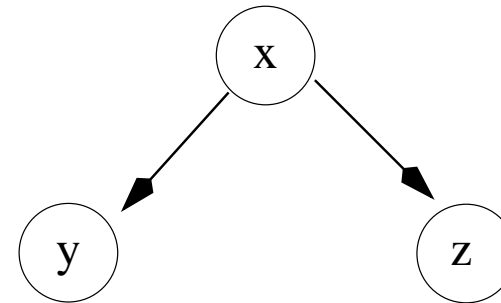
$$\text{var}[Y] = \text{var}[X_1] + \text{var}[X_2] + 2\text{covar}[X_1, X_2]$$

$$\text{if } X_1 \text{ and } X_2 \text{ are independent } \text{var}[Y] = \text{var}[X_1] + \text{var}[X_2]$$

12 / 19

- Gaussian has relatively simple analytical properties
- Central limit theorem. Sum (or mean) of M independent random variables is distributed normally as $M \rightarrow \infty$ (subject to a few general conditions)
- Diagonalization of covariance matrix \implies rotated variables are independent
- All marginal and conditional densities of a Gaussian are Gaussian
- The Gaussian is the distribution that maximizes the entropy $H = - \int p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}$ for fixed mean and covariance

Example:



- Let X denote pulse rate
- Let Y denote measurement taken by machine 1, and Z denote measurement taken by machine 2

- Model
 - $X \sim N(\mu_x, v_x)$
 - $Y = \mu_y + w_y(X - \mu_x) + N_y$
 - $Z = \mu_z + w_z(X - \mu_x) + N_z$
 - noise $N_y \sim N(0, v_y^N)$, $N_z \sim N(0, v_z^N)$, independent
- (X, Y, Z) is jointly Gaussian; can do inference for X given $Y = y$ and $Z = z$

As before

$$P(x, y, z) = P(x)P(y|x)P(z|x)$$

Show that

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} v_x & w_y v_x & w_z v_x \\ w_y v_x & w_y^2 v_x + v_y^N & w_y w_z v_x \\ w_z v_x & w_y w_z v_x & w_z^2 v_x + v_z^N \end{pmatrix}$$

- Partition variables into two groups, \mathbf{X}_1 and \mathbf{X}_2

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

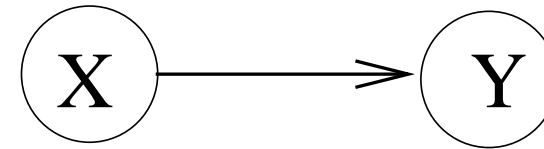
$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\boldsymbol{\mu}_{1|2}^c = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

$$\Sigma_{1|2}^c = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

- For proof see §2.3.1 of Bishop (2006) (not examinable)
- Formation of joint Gaussian is analogous to formation of joint probability table for discrete RVs. Propagation schemes are also possible for Gaussian RVs

17/19



$$Y = 2X + 8 + N_y$$

- Assume $X \sim N(0, 1/\alpha)$, so $w_y = 2$, $\mu_y = 8$, and $N_y \sim N(0, 1)$
- Show that

$$\mu_{x|y} = \frac{2}{4 + \alpha}(y - 8)$$

$$\text{var}(x|y) = \frac{1}{4 + \alpha}$$

18/19

Hybrid (discrete + continuous) networks

- Could discretize continuous variables, but this is ugly, and gives large CPTs
- Better to use parametric families, e.g. Gaussian
- Works easily when continuous nodes are children of discrete nodes; we then obtain a *conditional Gaussian* model