COMP3206: A few scattered facts from probability theory: Mostly 2-dim Gaussians

Srinandan Dasmahapatra

November 10, 2017

Basic definitions from probability theory

- Probability of event A is $P(A)$.
- The set of all events is Ω
- $0 \leqslant P(A) \leqslant 1$ and $P(\Omega) = 1$

Basic definitions from probability theory

- Probability of event A is $P(A)$.
- The set of all events is Ω
- $0 \leqslant P(A) \leqslant 1$ and $P(\Omega) = 1$
- **Joint distribution** $P(A, B)$ is the probability that both events A and B occur.

Basic definitions from probability theory

- Probability of event A is $P(A)$.
- The set of all events is Ω
- $0 \leqslant P(A) \leqslant 1$ and $P(\Omega) = 1$
- **Joint distribution** $P(A, B)$ is the probability that both events A and B occur.
- If events A and B are **independent**, then $P(A, B) = P(A)P(B)$

Basic definitions from probability theory

- Probability of event A is $P(A)$.
- The set of all events is Ω
- $0 \leqslant P(A) \leqslant 1$ and $P(\Omega) = 1$
- **Joint distribution** $P(A, B)$ is the probability that both events A and B occur.
- If events A and B are **independent**, then $P(A, B) = P(A)P(B)$
- **Conditional probability**, $P(A|B)$ is the probability that event A occurs given that event B has occurred.

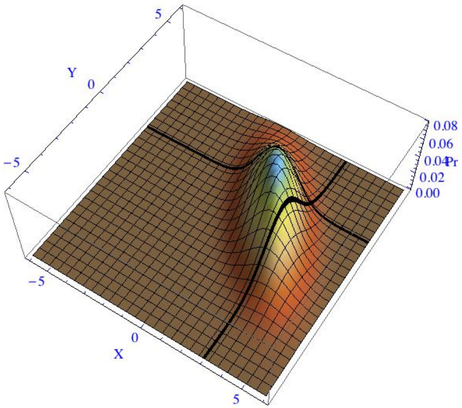
Basic definitions from probability theory

- Probability of event A is $P(A)$.
- The set of all events is Ω
- $0 \leqslant P(A) \leqslant 1$ and $P(\Omega) = 1$
- **Joint distribution** $P(A, B)$ is the probability that both events A and B occur.
- If events A and B are **independent**, then $P(A, B) = P(A)P(B)$
- **Conditional probability**, $P(A|B)$ is the probability that event A occurs given that event B has occurred.
- $P(A, B) = P(B|A)P(A) = P(A|B)P(B)$

Basic definitions from probability theory

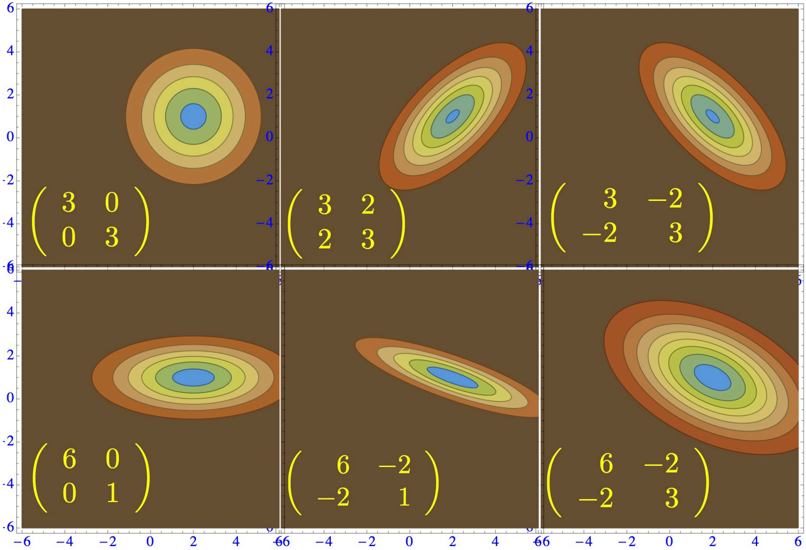
- Probability of event A is $P(A)$.
- The set of all events is Ω
- $0 \leqslant P(A) \leqslant 1$ and $P(\Omega) = 1$
- **Joint distribution** $P(A, B)$ is the probability that both events A and B occur.
- If events A and B are **independent**, then $P(A, B) = P(A)P(B)$
- **Conditional probability**, $P(A|B)$ is the probability that event A occurs given that event B has occurred.
- $P(A, B) = P(B|A)P(A) = P(A|B)P(B)$
- Chain rule: $P(x_4, x_3, x_2, x_1) = P(x_4|x_3, x_2, x_1)P(x_3|x_2, x_1)P(x_2|x_1)P(x_1)$

Two dimensional Gaussian distributions

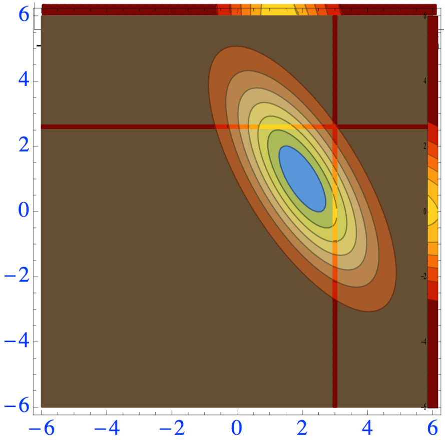


- The distribution on the left has mean $\mu = (2, 1)^T$ and covariance matrix $\Sigma = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$.
- The dark lines are for the conditional distributions $P(Y|X = 3.0)$ and $P(X|Y = 2.6)$. Notice that they are both Gaussian distributions.

Changing the covariance matrix of Gaussian - contour plots



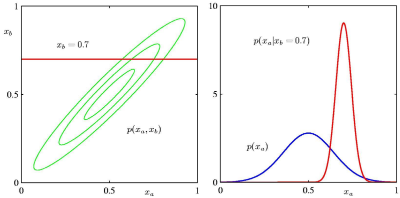
Conditionals on contour plot



- These are contour plots of the (same) Gaussian pdf with mean $\mu = (2, 1)^T$ and covariance matrix $\Sigma = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}$.
- Notice how the negative correlation between deviations from the means for X and Y.
- The light bands are for the conditional distributions $P(Y|X = 3.0)$ and $P(X|Y = 2.6)$. They are shown on the right and top, and they display a narrower distribution than the 2-d version.

Conditionals and Marginals of 2-D Gaussians

Partitioned Conditionals and Marginals



- Marginal $P(x_a)$ (blue) and conditional distribution $P(x_a|x_b = 0.7)$ (red) (from Bishop's book PRML)
- Given the mean and covariance matrix of the joint distribution
- Find mean $\mu_{a|b}$ and covariance matrix $\Sigma_{a|b}$ of the conditional distribution. Also μ_a and Σ_a.

Marginalisation: $p(x_a) = \int p(x_a, x_b) dx_b$

Choose $x_b = c$ with probability $p(x_b = c)$. For all possible values c, **"add"** $p(x_a|x_b = c)$ values **"weighted"** by $p(x_b = c)$:

$$p(x_a) = \int_{-\infty}^{\infty} p(x_a, x_b = c) dc = \int_{-\infty}^{\infty} p(x_a|x_b = c) p(x_b = c) dc$$

Linear dependence and random noise — covariance matrix

- Let $x = \eta_x$ and $y = ax + \eta_y$ where $\eta_x \sim \mathcal{N}(0, \sigma_x)$ and $\eta_y \sim \mathcal{N}(0, \sigma_y)$ are random numbers drawn from one-dimensional Gaussian distributions of 0 mean and standard deviations σ_x and σ_y respectively.

Linear dependence and random noise — covariance matrix

- Let $x = \eta_x$ and $y = ax + \eta_y$ where $\eta_x \sim \mathcal{N}(0, \sigma_x)$ and $\eta_y \sim \mathcal{N}(0, \sigma_y)$ are random numbers drawn from one-dimensional Gaussian distributions of 0 mean and standard deviations σ_x and σ_y respectively.
- Compute the expectation values $\langle x^2 \rangle$, $\langle xy \rangle$ and $\langle y^2 \rangle$ using the identity $\int dt\, t^2 \mathcal{N}(t; \mu, \sigma) = \sigma^2$. These are the components of the covariance matrix.

Linear dependence and random noise — covariance matrix

- Let $x = \eta_x$ and $y = ax + \eta_y$ where $\eta_x \sim \mathcal{N}(0, \sigma_x)$ and $\eta_y \sim \mathcal{N}(0, \sigma_y)$ are random numbers drawn from one-dimensional Gaussian distributions of 0 mean and standard deviations σ_x and σ_y respectively.
- Compute the expectation values $\langle x^2 \rangle$, $\langle xy \rangle$ and $\langle y^2 \rangle$ using the identity $\int dt\, t^2 \mathcal{N}(t; \mu, \sigma) = \sigma^2$. These are the components of the covariance matrix.

$$
\begin{aligned}
\langle x^2 \rangle &= \langle \eta_x^2 \rangle &&= \sigma_x^2 \\
\langle xy \rangle &= \langle a\eta_x^2 + \eta_x \eta_y \rangle &&= a\sigma_x^2 \\
\langle y^2 \rangle &= \langle a^2 \eta_x^2 + 2a\eta_x \eta_y + \eta_y^2 \rangle &&= a^2 \sigma_x^2 + \sigma_y^2
\end{aligned}
$$

$$
\boldsymbol{\Sigma} = \begin{pmatrix} {\sigma_x}^2 & a{\sigma_x}^2 \\ a{\sigma_x}^2 & a^2{\sigma_x}^2 + {\sigma_y}^2 \end{pmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_y^2} \begin{pmatrix} a^2 + \frac{\sigma_y^2}{\sigma_x^2} & -a \\ -a & 1 \end{pmatrix}.
$$

Inverse covariance matrix for $x = \eta_x$, $y = ax + \eta_y$

- Joint distribution

$$p(x,y) = \frac{1}{Z_x} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma_x^2}\right) \frac{1}{Z_y} \exp\left(-\frac{1}{2}\frac{(y-ax)^2}{\sigma_y^2}\right)$$

- Consider the exponent in the joint distribution

$$-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{(y-ax)^2}{\sigma_y^2}\right) = -\frac{1}{2}\left(\left(\frac{1}{\sigma_x^2} + \frac{a^2}{\sigma_y^2}\right)x^2 - 2\frac{a}{\sigma_y^2}xy + \frac{1}{\sigma_y^2}y^2\right)$$

Inverse covariance matrix for $x = \eta_x$, $y = ax + \eta_y$

- Joint distribution

$$p(x,y) = \frac{1}{Z_x}\exp\left(-\frac{1}{2}\frac{x^2}{\sigma_x^2}\right)\frac{1}{Z_y}\exp\left(-\frac{1}{2}\frac{(y-ax)^2}{\sigma_y^2}\right)$$

- Consider the exponent in the joint distribution

$$-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{(y-ax)^2}{\sigma_y^2}\right) = -\frac{1}{2}\left(\left(\frac{1}{\sigma_x^2}+\frac{a^2}{\sigma_y^2}\right)x^2 - 2\frac{a}{\sigma_y^2}xy + \frac{1}{\sigma_y^2}y^2\right)$$

- The exponent may be written as a quadratic form

$$-\frac{1}{2}\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} \frac{1}{\sigma_x^2}+\frac{a^2}{\sigma_y^2} & -\frac{a}{\sigma_y^2} \\ -\frac{a}{\sigma_y^2} & \frac{a^2}{\sigma_y^2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

Inverse covariance matrix for $x = \eta_x$, $y = ax + \eta_y$

- Joint distribution

$$p(x,y) = \frac{1}{Z_x}\exp\left(-\frac{1}{2}\frac{x^2}{\sigma_x^2}\right)\frac{1}{Z_y}\exp\left(-\frac{1}{2}\frac{(y-ax)^2}{\sigma_y^2}\right)$$

- Consider the exponent in the joint distribution

$$-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{(y-ax)^2}{\sigma_y^2}\right) = -\frac{1}{2}\left((\frac{1}{\sigma_x^2}+\frac{a^2}{\sigma_y^2})x^2 - 2\frac{a}{\sigma_y^2}xy + \frac{1}{\sigma_y^2}y^2\right)$$

- The exponent may be written as a quadratic form

$$-\frac{1}{2}(x \;\; y)\begin{pmatrix} \frac{1}{\sigma_x^2}+\frac{a^2}{\sigma_y^2} & -\frac{a}{\sigma_y^2} \\ -\frac{a}{\sigma_y^2} & \frac{a^2}{\sigma_y^2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} =: -\frac{1}{2}(x \;\; y)\left(\Sigma^{-1}\right)\begin{pmatrix} x \\ y \end{pmatrix}$$

with inverse covariance matrix Σ^{-1}.

Explicit form for 2-dimensional Gaussian integrals

To explicitly write out

$$\mathcal{N}(x, y; \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 6 & -2 \\ -2 & 1 \end{pmatrix}),$$

the term in the exponent is $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Lambda})(\mathbf{x} - \boldsymbol{\mu})$ (precision matrix $\boldsymbol{\Lambda}$)

$$\left(-\frac{1}{2 \cdot 2} \begin{pmatrix} x-2, & y-1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} x-2 \\ y-1 \end{pmatrix}\right),$$

where the inverse of the covariance matrix has been inserted and its determinant $= 2$ is in the denominator.

Explicit form for 2-dimensional Gaussian integrals

To explicitly write out

$$\mathcal{N}(x, y; \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 6 & -2 \\ -2 & 1 \end{pmatrix}),$$

the term in the exponent is $(\mathbf{x}-\mu)^T\mathbf{\Lambda})(\mathbf{x}-\mu)$ (precision matrix $\mathbf{\Lambda}$)

$$\left(-\frac{1}{2\cdot 2}\begin{pmatrix} x-2, & y-1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 6 \end{pmatrix}\begin{pmatrix} x-2 \\ y-1 \end{pmatrix}\right),$$

where the inverse of the covariance matrix has been inserted and its determinant $= 2$ is in the denominator. This evaluates to

$$\left(-\frac{x^2}{4} - xy + 2x - \frac{3y^2}{2} + 5y - \frac{9}{2}\right).$$

Explicit form for 2-dimensional Gaussian integrals

To explicitly write out

$$\mathcal{N}(x, y; \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 6 & -2 \\ -2 & 1 \end{pmatrix}),$$

the term in the exponent is $(\mathbf{x} - \mu)^T \boldsymbol{\Lambda})(\mathbf{x} - \mu)$ (precision matrix $\boldsymbol{\Lambda}$)

$$\left(-\frac{1}{2 \cdot 2} \begin{pmatrix} x-2, & y-1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} x-2 \\ y-1 \end{pmatrix}\right),$$

where the inverse of the covariance matrix has been inserted and its determinant $= 2$ is in the denominator. This evaluates to

$$\left(-\frac{x^2}{4} - xy + 2x - \frac{3y^2}{2} + 5y - \frac{9}{2}\right).$$

The normalisation factor is $1/(2\sqrt{2}\pi)$.

What are the distributions of subsets of variables that are jointly distributed as Gaussian?*

Let the components of $\mathbf{x} = (X_1, \ldots, X_D)$ be grouped into two sets:

$$\mathbf{x} = (\underbrace{X_1, \ldots, X_k}_{a_1, \ldots, a_k}, \underbrace{X_{k+1}, \ldots, X_D}_{b_1, \ldots, b_{D-k}})$$

which we write as $\mathbf{x} = (\mathbf{a}, \mathbf{b})$, with $\mathbf{a} = (a_1, \ldots, a_k)$ and $\mathbf{b} = (b_1, \ldots, b_{D-k})$. The mean of $\mathbf{x}$ can also be split thus: $\mu = (\mu_a, \mu_b)$. Further, we can write the covariance matrix in the block matrix form

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix},$$

where Σ_{aa} is the covariance matrix of the $\mathbf{a}$ variables, Σ_{bb} is the covariance matrix of the $\mathbf{b}$ variables and $\Sigma_{ab} = \Sigma_{ba}^T$ is the matrix of covariances between the $\mathbf{a}$ and $\mathbf{b}$ variables.

Gaussian conditional and marginal distributions*

- Marginal probability denisity function (pdf) of $\mathbf{a}$, $\mathbf{b}$ is Gaussian, $\mathbf{a} \sim \mathcal{N}(\mu_a, \Sigma_A)$,$\mathbf{b} \sim \mathcal{N}(\mu_b, \Sigma_B)$.
- Conditional pdf $p(\mathbf{b}|\mathbf{a})$ also Gaussian with mean and covariance:

$$\begin{aligned} \mu_{\mathbf{b}|\mathbf{a}} &= \mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(\mathbf{a} - \mu_a) \\ \Sigma_{\mathbf{b}|\mathbf{a}} &= \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \end{aligned}$$

Gaussian conditional and marginal distributions*

- Marginal probability denisity function (pdf) of $\mathbf{a}, \mathbf{b}$ is Gaussian, $\mathbf{a} \sim \mathcal{N}(\mu_a, \Sigma_A)$,$\mathbf{b} \sim \mathcal{N}(\mu_b, \Sigma_B)$.
- Conditional pdf $p(\mathbf{b}|\mathbf{a})$ also Gaussian with mean and covariance:

$$
\begin{aligned}
\mu_{\mathbf{b}|\mathbf{a}} &= \mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(\mathbf{a} - \mu_a) \\
\Sigma_{\mathbf{b}|\mathbf{a}} &= \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}
\end{aligned}
$$

- **Example**: random variables $\mathbf{X} = (x, y)$: $x = \eta_x$, $y = ax + \eta_y$. joint distribution proportional to $\exp(-\frac{1}{2}\mathbf{X}^T\Sigma^{-1}\mathbf{X})$ with

$$
\Sigma^{-1} = \frac{1}{\sigma_y^2}\begin{pmatrix} a^2 + \frac{\sigma_y^2}{\sigma_x^2} & -a \\ -a & 1 \end{pmatrix}.
$$

Condition on y: reduced 1-d distribution of x has quadratic form:

$$
\frac{1}{\sigma_y^2}\left(a^2 + \frac{\sigma_y^2}{\sigma_x^2}\right) \longrightarrow \frac{1}{\sigma_x^2}
$$