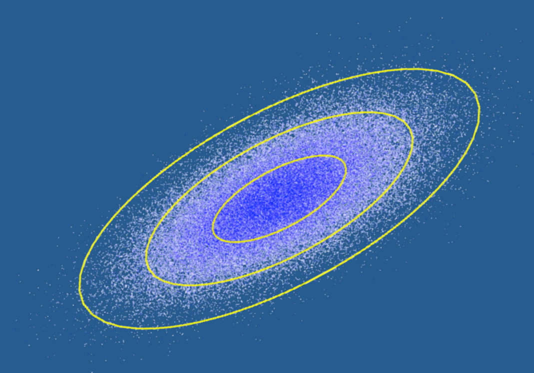
1



L19. GAUSSIAN DISTRIBUTIONS

EECS 492
March 17, 2011

Administrative

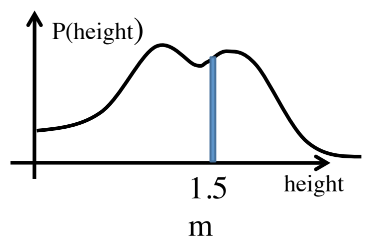
2

- PS4 due tonight
- PS5 & PS6 groups will be computed tonight
  - Update your group preferences!
- Midterm 2:
  - A week from today!

Continuous Probability

3

- What is the probability that someone is exactly 1.5 m?



- Probability is *area under the curve*!
- Corollaries:
  - Total area under curve = 1.
  - Magnitude of probability density can be greater than 1.

Probability Basics

Discrete Probability	Continuous Probability
P(x) = Probability of event occurring	P(x) = Probability *density* at x
$Prob(x) = P(x)$	$Prob(x) = 0$
$0 \leq P(x) \leq 1$	$0 \leq P(x) \leq \infty$
$\sum_{-\infty}^{\infty} P(x) = 1$	$\int_{-\infty}^{\infty} P(x)dx = 1$

Probability Basics: Expectation

- Weighted average according to probability

$$E[x] = \int_{-\infty}^{\infty} xP(x)dx$$

- Basic properties of expectation

$$E[\alpha] = \alpha$$
$$E[\alpha x] = \alpha E[x]$$
$$E[\alpha + x] = \alpha + E[x]$$
$$E[x + y] = E[x] + E[y]$$

Variance

6

- How much does a variable vary around its average value?

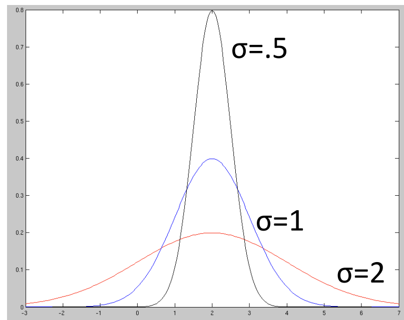
$$E[(x - E[x])^2]$$

- Suppose you have a stream of data coming in and you want to compute the "running" mean and variance?
  - Do you have to store all the samples in memory?

Gaussian Distribution

- Specified by mean and variance
- Structure: exponential quadratic loss

$$P(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}}$$



- Why do we like Gaussian distributions?
  - It's its own conjugate prior—
  - Gaussians in ➔ Gaussians out

Where is Jill?

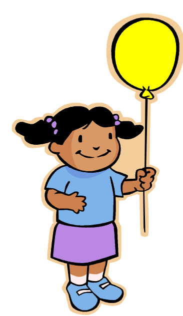
- Where is Jill standing?
  - Our initial belief:

$$x \sim N(\mu_x = 2, \sigma_x^2 = 2)$$

  - Bob sees Jill, but his vision isn't so great.

$$P(z|x) \sim N(x, 1)$$

    - And we'll suppose that he says "z=1".



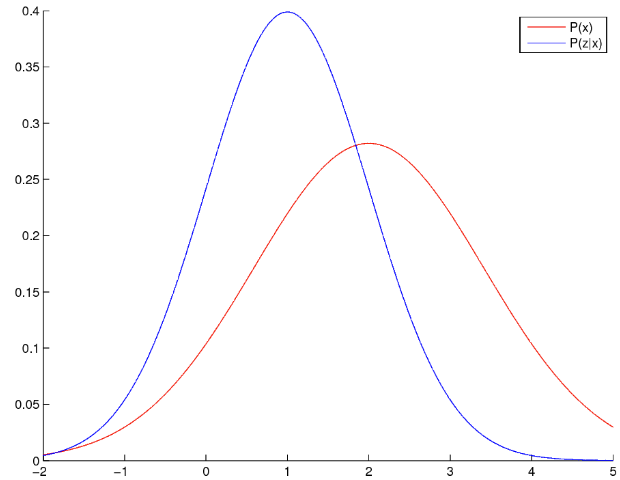
  - What is our posterior distribution?

$$P(x|z)$$

Prediction

9

- What will the posterior look like?



Where is Jill?

10

$$P(x|z) = \frac{P(z|x)P(x)}{P(z)}$$ Normalizing constant

$$P(x|z) \propto \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}} \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{\frac{-(z-x)^2}{2\sigma_z^2}}$$

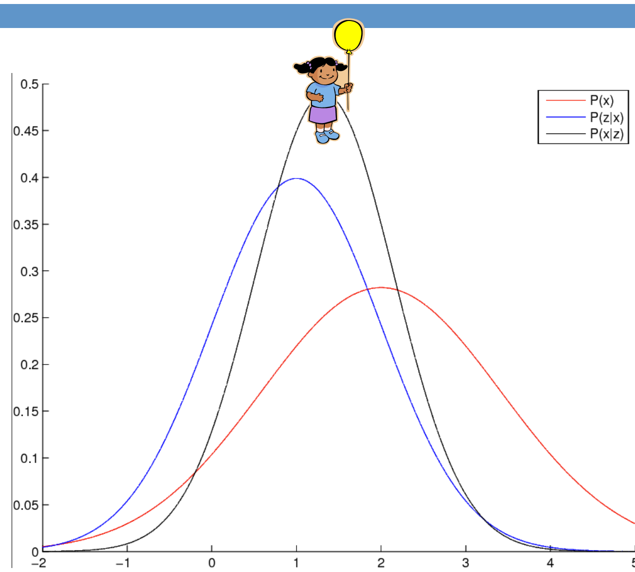
- Let's manipulate the exponential part:

$$\begin{aligned} & \frac{-x^2+4x-4}{4} + \frac{-1+2x-x^2}{2} \\ = & \frac{-x^2+4x-4-2+4x-2x^2}{4} \\ = & \frac{-3x^2+8x-6}{4} \\ = & \frac{-x^2+8/3x-2}{4/3} \\ = & \frac{-(x-4/3)^2+16/9-2}{4/3} \end{aligned}$$

Mean=4/3,
Variance=2/3

Where is Jill? (Result)

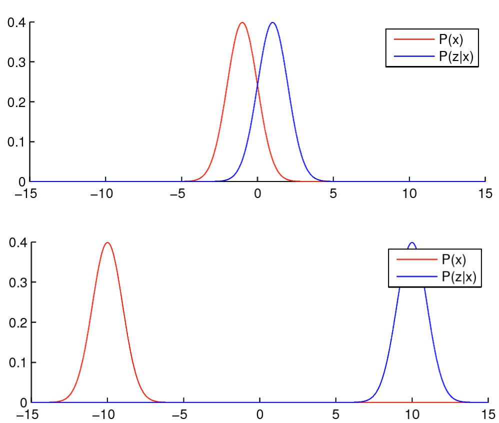
11



Building Intuition

12

- Conditioning on variables can never increase uncertainty.



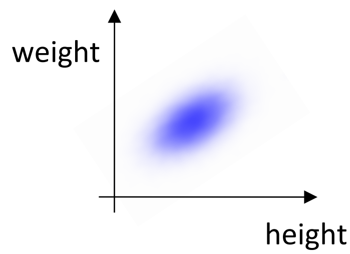
Multiple random variables

13

- We can stack several random variables together, forming a column vector:

$$x = \begin{bmatrix} height \\ weight \end{bmatrix}$$

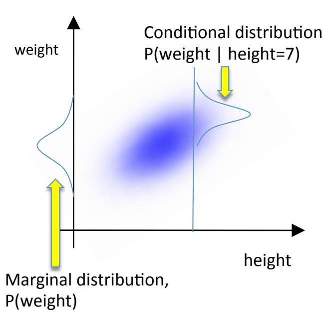
- It has a N-dimensional probability density:



Are weight and height independent?

Correlations

- Density function can exhibit correlations in the functions
  - (They're dependent!)
- Marginal distributions and conditional distributions can be computed from the joint distribution



Multiple random variables

15

- Most operations extend naturally:

$$E[x] = \int_{-\infty}^{\infty} xP(x)dx$$

- Conditional, Joint, Marginal rules all work.
- Variance changes a bit:

$$E[(x - E[x])^2] \Rightarrow E[(x - E[x])(x - E[x])^T]$$

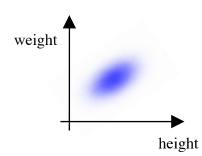
Co-variance

16

- When computing variance of a vector, we get a *covariance:*

$$\Sigma = E\left(\begin{bmatrix} (h-\bar{h})^2 & (h-\bar{h})(w-\bar{w}) \\ (w-\bar{w})(h-\bar{h}) & (w-\bar{w})^2 \end{bmatrix}\right)$$

- Diagonal terms are just the variances of the marginal distributions.
- What do the off-diagonal terms mean?



Visualizing Gaussians

- Recall our PDF:

$$P(x) = \alpha e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

- Find contours of constant probability

$$K^2 = (x-\mu)^T \Sigma^{-1}(x-\mu) \qquad K = 1, 2, 3, \ldots$$

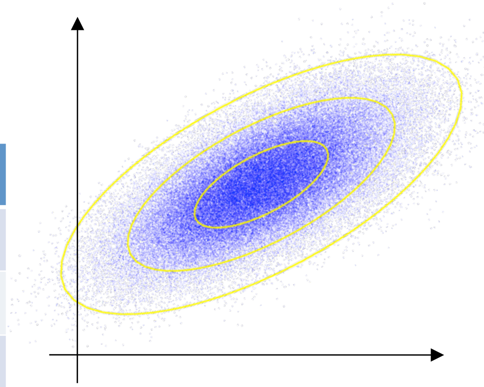
  - K is also known as the "Mahalanobis distance"

- Expand these terms, we end up with quadratic curve
  - An ellipse!

Visualizing Gaussians

- Number of particles within each ellipse can be computed based on properties of Gaussian distributions

Sigma	1D	2D
1	0.68	0.39
2	0.96	0.87
3	0.997	0.99

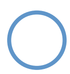


Covariance Matrices: Intuition

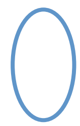
19

- Sketch equi-potential curves for the matrices below (to scale):

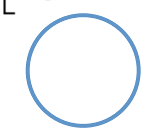
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



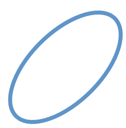
$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$



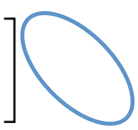
$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$



$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$



$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$



Covariance Matrices: Intuition

20

- What about:

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

- What about:

$$\begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} \qquad \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$$

Gaussian Distributions

21

- Mean and covariance are meaningful for *any* distribution
  - But they do not *define* the distribution– a incomplete description
  - … Unless it's a Gaussian distribution.
- The Gaussian distribution is *exactly* parameterized by mean and covariance.
  - Compact (low memory)
  - Conjugate prior
- **Central Limit Theorem**: Distribution of the sum (or average) of N independent and identically distributed (IID) random variables approaches a normal distribution.
  - In other words, even if you start off with something non-Gaussian, you're likely to end up with one!

Multi-Variate Gaussian Distributions

- Here's the multi-variable distribution
  - Note the structure!

$$P(x) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

- Characterized by mean & covariance

$$\mu_x = E[x]$$

$$\Sigma_x = E[(x - E[x])(x - E[x])^T]$$

Functions of random variables

23

- Suppose we know something about random variable x:

$$x \sim N(\mu_x, \Sigma_x)$$

- And suppose I know a function y:

$$y = f(x)$$

- What is the distribution of y?
  - Let's derive μ_y, Σ_y

Linear functions of random variables

24

- Let's start with the linear case:

$$y = f(x)$$
$$y = Ax + b$$

- What is E(y)?

$$\begin{aligned} \mu_y &= E(y) \\ &= E(Ax+b) \end{aligned}$$

- Simplify:

$$\mu_y = AE(x) + b$$

Linear functions of random variables (2)

25

- Reminders:

$$\Sigma_y = E[(y - E[y])(y - E[y])^T]$$
$$\mu_y = AE(x) + b$$

$$\begin{aligned}
\Sigma_y &= E[(Ax + b - A\mu_x - b)(Ax + b - A\mu_x - b)^T] \\
&= E[(Ax - A\mu_x)(Ax - A\mu_x)^T] \\
&= E[A(x - \mu_x)(x - \mu_x)^T A^T] \\
&= AE[(x - \mu_x)(x - \mu_x)^T]A^T \\
&= A\Sigma_x A^T
\end{aligned}$$

Non-linear functions of random variables

- Again, suppose:

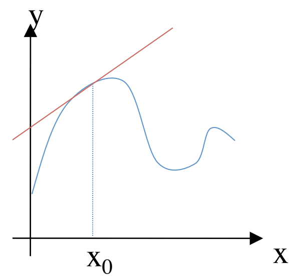
$$x \sim N(\mu_x, \Sigma_x)$$

$$y = \cancel{Ax} + b \qquad y = f(x)$$

- Approach: approximate f(x) with Taylor expansion
  - What point should we approximate f(x) around?

Linearizing functions: Taylor expansions

- First-order Taylor expansion
  - Let's review 1D case



$$y \approx \left.\frac{df}{dx}\right|_{x_0} (x - x_0) + f(x_0)$$

Linearizing functions (Generalization)

- Generalized case:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2, \ldots) \\ f_2(x_1, x_2, \ldots) \\ \ldots \end{bmatrix}$$

$$y \approx \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \ldots \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix} \begin{bmatrix} x_1 - x_{1_0} \\ x_2 - x_{2_0} \\ \ldots \end{bmatrix} + \begin{bmatrix} f_1(x_{1_0}, x_{2_0}) \\ f_2(x_{1_0}, x_{2_0}) \\ \ldots \end{bmatrix}$$

"Jacobian"

$$\vec{y} \approx J|_{\vec{x_0}} (\vec{x} - \vec{x_0}) + f(\vec{x_0})$$

Projecting means and covariances (ta da!)

$$y \approx J|_{x_0}(x - x_0) + f(x_0)$$

$$y \approx \underbrace{J|_{x_0}}_{\mathbf{A}} x \underbrace{- J|_{x_0} x_0 + f(x_0)}_{\mathbf{b}}$$

$$\boxed{\begin{aligned} y &= Ax + b \\ \Sigma_y &= A \Sigma_x A^T \end{aligned}}$$

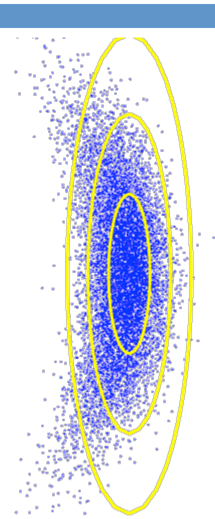
Non-linear case is reduced to linear case via first-order Taylor approximation.

What do we lose by dropping higher order terms?

Linearization Error

30

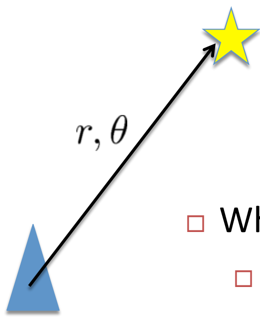
- Mean and covariance are computed around the expected value
- Non-linear behavior away from expected value is not well approximated.
- More on this later…



Covariance Projection: Example

31

- A robot observes a landmark
  - Sensor measures range and theta
  - Uncertainty in range and theta



$$\sigma_r^2 = 1$$
$$\sigma_\theta^2 = 0.01$$

- What is the uncertainty in x and y?
  - Step one: write x,y as f(r, theta)

Covariance Projection: Example

32

- Write our desired quantities as function of other random variables

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f_x(r,\theta) \\ f_y(r,\theta) \end{bmatrix} = \begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \end{bmatrix}$$

- What is $\mu_{x,y}$?
  - Suppose we observe r = 10, theta = π/2

$$\mu_{x,y} = \begin{bmatrix} 10\cos(\pi/2) \\ 10\sin(\pi/2) \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$

Covariance Projection: Example

33

- Now on to covariance… our equations from before:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f_x(r,\theta) \\ f_y(r,\theta) \end{bmatrix} = \begin{bmatrix} r\cos(\theta) \\ r\sin(\theta) \end{bmatrix}$$

- We linearize the function:

$$f(r,\theta) = J \begin{bmatrix} r - r_0 \\ \theta - \theta_0 \end{bmatrix} + f(r_0,\theta_0)$$

- What is J?

$$J = \left. \begin{bmatrix} \frac{\partial f_x}{\partial r} & \frac{\partial f_x}{\partial \theta} \\ \frac{\partial f_y}{\partial r} & \frac{\partial f_y}{\partial \theta} \end{bmatrix} \right|_{r=r_0,\theta=\theta_0}$$

Covariance Projection: Example

34

$$J = \begin{bmatrix} \cos(\theta_0) & -r_0\sin(\theta_0) \\ \sin(\theta_0) & r_0\cos(\theta_0) \end{bmatrix}$$

↓

r = 10, theta = 0

↓

$$J = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}$$

Covariance Projection: Example

35

$$\Sigma_{x,y} = J\Sigma_{r,\theta}J^T$$

$$\Sigma_{x,y} = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix} \Sigma_{r,\theta} \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}^T$$

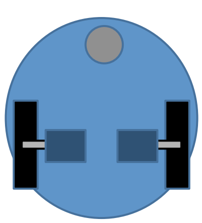
- But what is $\Sigma_{r,\theta}$?

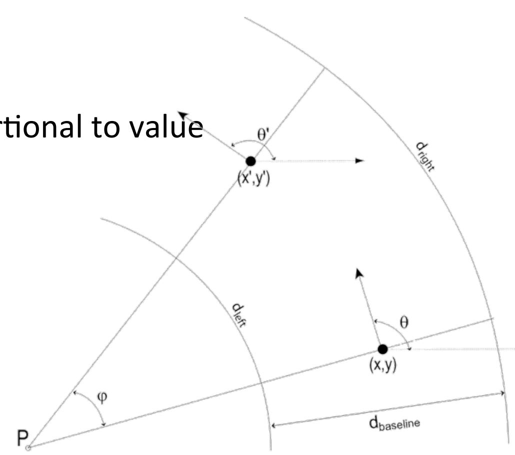
$$\sigma_r^2 = 1$$
$$\sigma_\theta^2 = 0.01$$

Covariance Projection: Your turn!

36

- Consider a differentially-driven robot
  - We observe d_R, d_L
  - d_b is a constant
  - Std. deviation proportional to value
    - e.g., $\sigma_{d_R}^2 = \alpha^2 d_R$



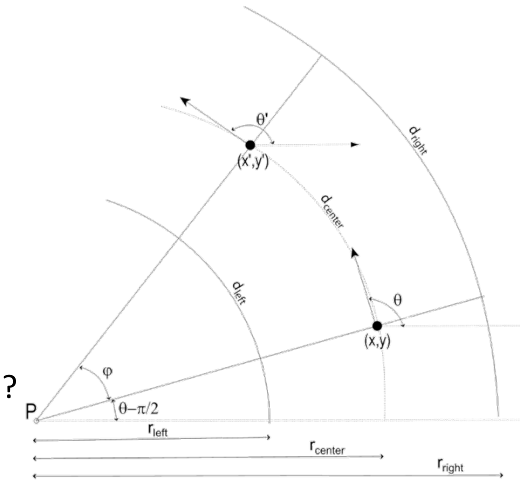


Your turn!

- How to convert left/ right distances to a change in position?

$$
\begin{aligned}
\Delta x &= \frac{d_R + d_L}{2} \\
\Delta \theta &= \frac{d_R - d_L}{d_B}
\end{aligned}
$$

- What is $\Sigma_{\Delta x, \Delta\theta}$?
  - What is the Jacobian?



(Solution)

38

$$\Sigma_{d_R, d_L} = \begin{bmatrix} \alpha^2 d_R & 0 \\ 0 & \alpha^2 d_L \end{bmatrix}$$

$$J = \begin{bmatrix} 1/2 & 1/2 \\ 1/d_B & -1/d_B \end{bmatrix}$$

$$\Sigma_{\Delta x, \Delta\theta} = J \Sigma_{d_R, d_L} J^T$$

$$\Sigma_{\Delta x, \Delta\theta} = \begin{bmatrix} 1/2 & 1/2 \\ 1/d_B & -1/d_B \end{bmatrix} \begin{bmatrix} \alpha^2 d_R & 0 \\ 0 & \alpha^2 d_L \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 1/d_B & -1/d_B \end{bmatrix}^T$$

Next Time

39

- Extended Kalman Filter
  - Efficient, recursive inference for continuous-valued problems