Neuroinformatics Lecture (L3)

Prof. Dr. Gordon Pipa

Institute of Cognitive Science University of Osnabrück

Thomas Bayes (1701–1761)

Thomas Bayes was born in Tunbridge Wells and was a clergyman as well as an amateur scientist and a mathematician. He studied logic and theology at Edinburgh University and was elected Fellow of the Royal Society in 1742. During the 18th century, issues regarding probability arose in connection with gambling and with the new concept of insurance. One particularly important problem concerned so-called inverse probability. A solution was proposed by Thomas Bayes in his paper 'Essay towards solving a problem in the doctrine of chances', which was published in 1764, some three years after his death, in the Philosophical Transactions of the Royal Society. In fact, Bayes only formulated his theory for the case of a uniform prior, and it was Pierre-Simon Laplace who independently rediscovered the theory in general form and who demonstrated its broad applicability.

Bayes' Theorem

$$\boxed{p(Y|X)} = \frac{\boxed{p(X|Y)}\boxed{p(Y)}}{p(X)}$$

$$p(X) = \sum_Y p(X|Y)p(Y)$$

posterior $\propto$ likelihood × prior

using $\quad p(X) = \sum_Y p(X,Y) \qquad p(X,Y) = p(Y|X)p(X)$

$$p(X) = \sum_Y p(X,Y) = \sum_Y p(X|Y)p(Y)$$

Example HIV Testing:

Suppose:

- A blood test gives a false negative result with p=0.1
  (false negative: negative test if patient is HIV infected)
  (→ test power p=0.9 to detect HIV if patient has HIV)

$$p(negative\ test|has\ HIV) = 0.1 \quad and \ \ p(positive\ test|has\ HIV) = 0.9$$

- A blood test gives a false positive result with p=0.1
  (false positive: positive test if patient is not HIV infected)

  $p(positive\ test|has\ no\ HIV)$=0.1

- HIV infection is rare: p(has HIV)=0.006

What is the probability that a patient has HIV in the case that he receives a positive test result ? p(has HIV | positive test)

Suppose:

- A blood test gives has a test power
- False positive
- HIV infection is rare:

$p(positive\ test|has\ HIV) = 0.9$

$p(positive\ test|has\ no\ HIV)$=0.1

$p(has\ HIV)$=0.006

What is the probability that a patient has HIV in the case that he receives a positive test result ? p(has HIV | positive test)

$$p(has\ HIV\ |positive\ test) = \frac{p(positive\ test|has\ HIV)p(has\ HIV)}{p(positive\ test)}$$

We used Bayes

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

Suppose:

- A blood test gives has a test power
- False positive
- HIV infection is rare:

$p(\text{positive test}|\text{has HIV}) = 0.9$

$p(\text{positive test}|\text{has no HIV})$=0.1

$p(\text{has HIV})$=0.006

What is the probability that a patient has HIV in the case that he receives a positive test result ? p(has HIV | positive test)

$$p(\text{has HIV}\,|\text{positive test}) = \frac{p(\text{positive test}|\text{has HIV})p(\text{has HIV})}{p(\text{positive test})}$$

We use sum rule (‚we marginalize‘): Add all p for a positive test

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} \qquad p(X) = \sum_Y p(X,Y) = \sum_Y p(X|Y)p(Y)$$

Suppose:

- A blood test gives has a test power
- False positive
- HIV infection is rare:

$p(positive\ test|has\ HIV) = 0.9$

$p(positive\ test|has\ no\ HIV)$=0.1

$p(has\ HIV)$=0.006

What is the probability that a patient has HIV in the case that he receives a positive test result ? p(has HIV | positive test)

$$p(has\ HIV\ |positive\ test) = \frac{p(positive\ test|has\ HIV)p(has\ HIV)}{p(positive\ test)}$$

We use sum rule (‚we marginalize‘): Add all p for a positive test

$$p(positive\ test) = p(positive, HIV) + p(positive, no\ HIV) \qquad \text{sum rule}$$

$$= \underbrace{p(positive|HIV)p(HIV)}_{\text{product rule}} + \underbrace{p(positive|no\ HIV)p(no\ HIV)}_{\text{product rule}}$$

Suppose:

- A blood test gives has a test power
- False positive
- HIV infection is rare:

$p(positive\ test|has\ HIV) = 0.9$

$p(positive\ test|has\ no\ HIV)$=0.1

$p(has\ HIV)$=0.006

What is the probability that a patient has HIV in the case that he receives a positive test result ? p(has HIV | positive test)

$$p(has\ HIV\ |positive\ test) = \frac{p(positive\ test|has\ HIV)p(has\ HIV)}{p(positive\ test)}$$

$$p(has\ HIV\ |positive\ test) = \frac{p(positive|HIV)p(HIV)}{p(positive|HIV)p(HIV) + p(positive|no\ HIV)p(no\ HIV)}$$

$$p(has\ HIV\ |positive\ test) = \frac{(1-0.1)\cdot 0.006}{(0.9)\cdot 0.006 + 0.1\cdot(1-0.006)} = \frac{0.0054}{0.0054+0.0994}$$

Suppose:

- A blood test gives a false negative result with p=0.1 → test power 0.9
- A blood test gives a false positive result with p=0.1
- HIV infection is rare: p=0.006

What is the probability that a patient has HIV in the case that he receives a positive test result ?

$$p(has\ HIV\ |positive\ test) = \frac{(1-0.1)\cdot 0.006}{(1-0.1)\cdot 0.006 + 0.1\cdot(1-0.006)} = 0.0515$$

The probability that you have HIV when you get a positive test result is p=0.05. (Keep in mind the numbers a not real !)

Consequence: The HIV test alone is not sufficient but an important marker that indicates further medical check ups !

Neuroinformatics Lecture (L3)

Sometimes it is necessary to measure a variable indirectly.

Say, we want to know *x* but we can only access *y* via a measurement.

Knowing the functional relation, i.e. $y=g(x)$, between x and y we can use the probability $p(y)$ to determine the probability of $p(x)$.

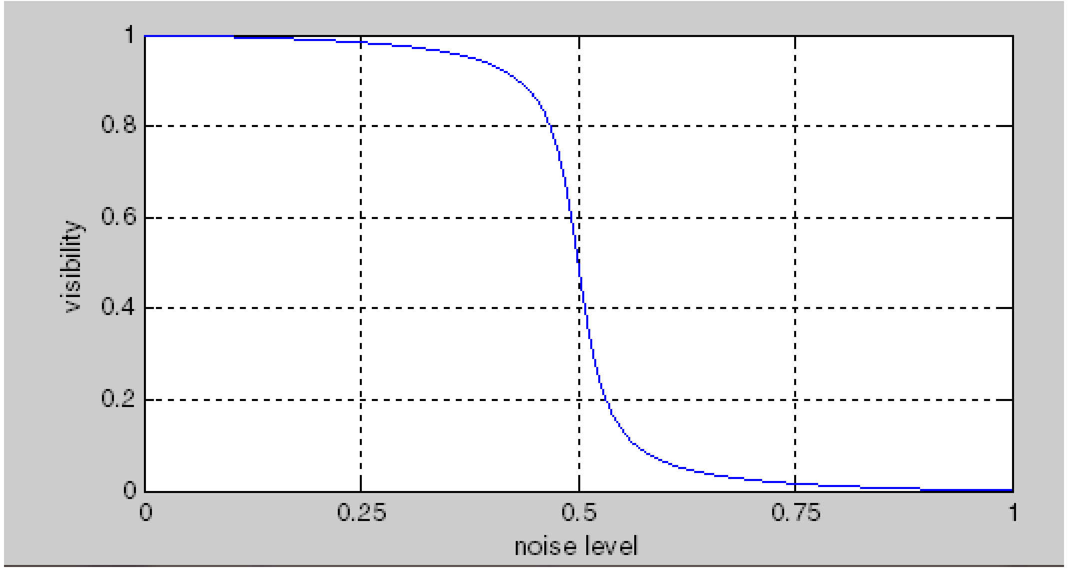
This is also called a “change of variable”.

In practice change of variable is often used to generate a random variable of arbitrary shape $f_{g(X)} = f_Y$ using a known (for instance uniform) random number generator.

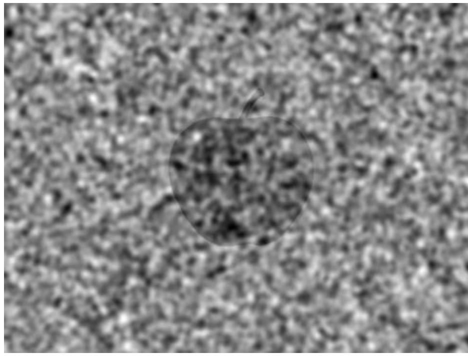
Example:

Say, we get a noisy Image with a certain noise level α.

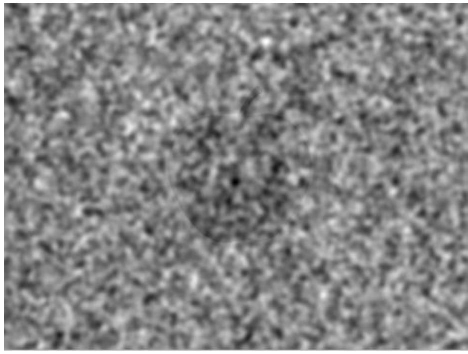
When we measure the visibility that we can identify the image as a function of the noise level we get a certain sigmoidal alike function.



$\alpha = 0$



$\alpha = 0.5$

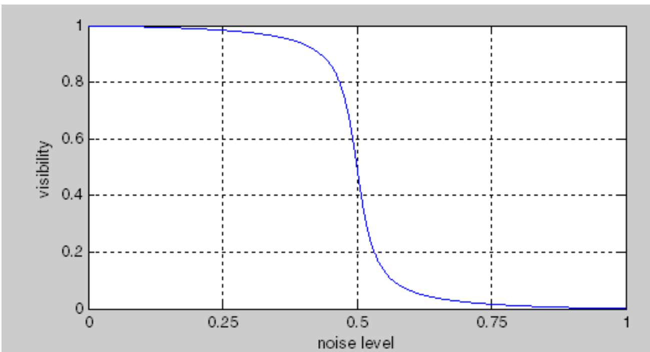


$\alpha = 1$

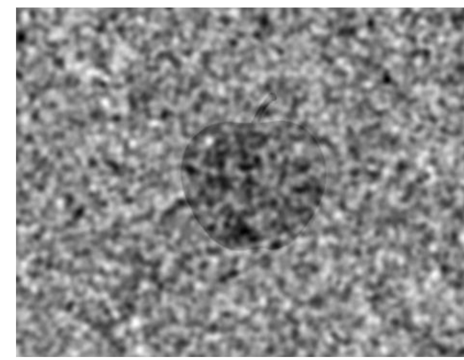
See example source code figures_adative_noise.m
to reproduce the figures

Example:

Say, we get a noisy Image with a certain noise level α.

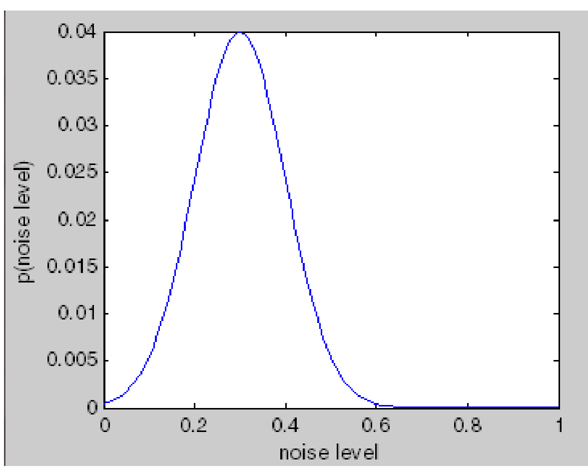


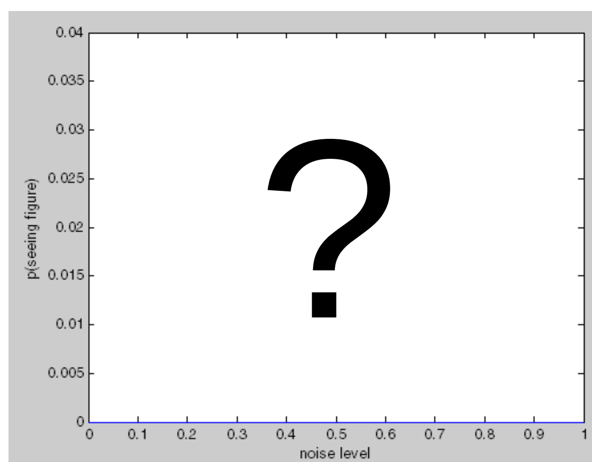
$\alpha = 0$



$\alpha = 0.5$

So what is the probability distribution of seeing an image when you know the probability distribution of the noise ?





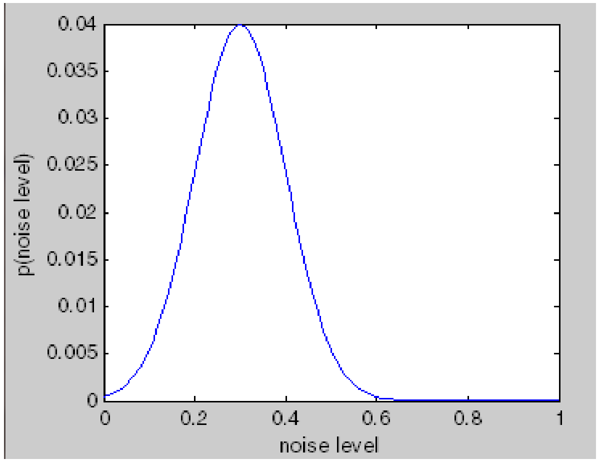
$\alpha = 1$

Sometimes it is necessary to measure a variable indirectly.

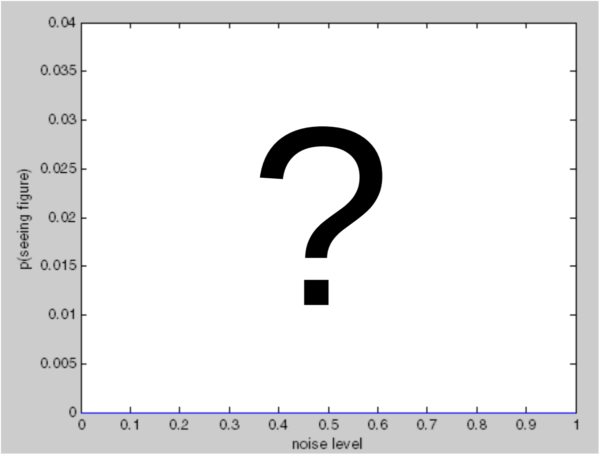
We know the probability for a certain noise level α : $p(x=\alpha)$
We also know how visibility for a given noise level : $y=g(x)$

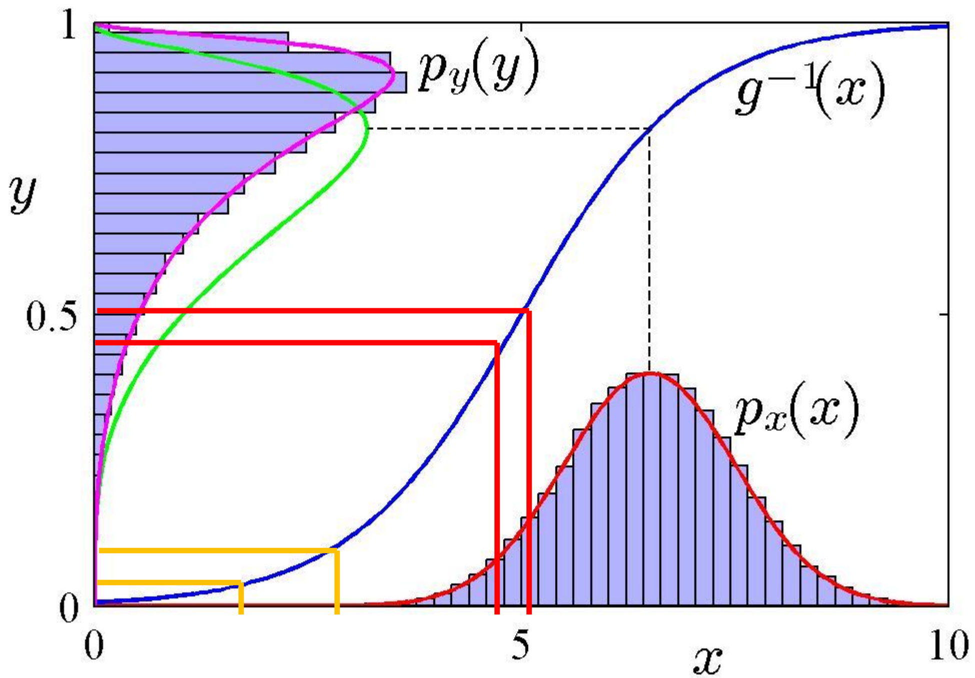
We want to know the probability of seeing the image given $p(x=\alpha)$, given $y=g(x)$.

We change from the random variable 'noise' to the random variable 'seen'



→





Change of variable

g: transformation

$$x = g(y)$$

y:noise
x:visibility

$$p_x(x \in (a,b)) = \int_a^b p(x)dx =$$

$$p_y(y \in (\alpha(a), \beta(b))) = \int_{\alpha(a)}^{\beta(b)} p(y)dy$$

This figure was taken from Solution 1.4 in the web edition of the solutions manual for PRML, available at http://research.microsoft.com/~cmbishop/PRML.

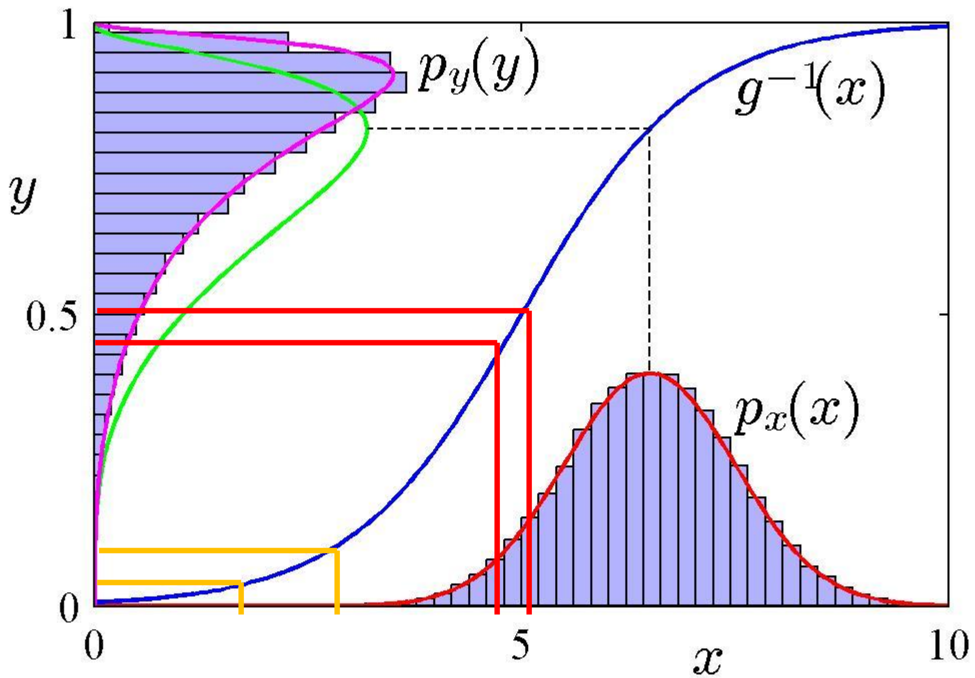
$$p_y(y) = p_x(x(y))\left|\frac{dx}{dy}\right|$$

$$= p_x(g(y))\left|g'(y)\right|$$

Change of variable

g: transformation

$$x = g(y)$$



wrong

$p_x(g(y))$

$p_x(g(y))|g'(y)|$

This figure was taken from Solution 1.4 in the web edition of the solutions manual for PRML, available at http://research.microsoft.com/~cmbishop/PRML.

Proof: Transformed Densities

We start with a density *p(x)* to compute the probability p_x

$$p_x(x \in (a, b) = \int_a^b p(x)dx$$

Now we transform with: $x = g(y)$ with g strictly monotonic: $y = g^{-1}(x)$

In general (substitution rule for integrals) :

$$with : \int_a^b f(x)dx = \int_\alpha^\beta f(x(u))\frac{dx}{du}du \text{ with } x(\alpha) = a \ , \ \ x(\beta) = b$$

here:

$$p_x(x \in (a,b)) = p_y(y \in (\alpha,\beta)) = \int_a^b p_x(x)dx = \int_\alpha^\beta p_x(x(y))\frac{dx}{dy}dy$$

Proof: Transformed Densities

We started with:

$$p_x(x \in (a,b)) = p_y(y \in (\alpha,\beta)) = \boxed{\int_a^b p_x(x)dx = \int_\alpha^\beta p_y(y)dy}$$

Than we rewrote p(x) using the substitution rule for integrals

$$\int_a^b p_x(x)dx = \int_\alpha^\beta \boxed{p_x(x(y))\frac{dx}{dy}} dy = \int_\alpha^\beta \boxed{p_y(y)} dy$$

Now we identify p(y) from the equation above

$$p_y(y) = p_x(x(y))\left|\frac{dx}{dy}\right|$$

That is the transformation rule for densities → q.e.d. („*quod erat demonstrandum*“)

Transformed Densities: Example 1 (proposed by the TAs)

You want to organize a party. You know the probability distribution for the number of people coming ($x = h_c$: head count): $p_x(x)$

Moreover you heard from a friend that he experienced that the number of drinks per person depends on the head count. From him you got the following functional relation between the head count and the number drinks (y) consumed.

$$g(x): y = e^{ax}$$

Given this you want to know the probability distribution of the number of drinks per person.

$$p_y(y)$$

Transformed Densities: Example 1 (proposed by the TAs)

We have: $p_x(x)$ $g(x): y = e^{ax}$ We want to know: $p_y(y)$

Step 1: We need to check whether g is a strictly monotonic function on the interval of interest.

We compute the derivative:

$$\frac{d}{dx} g(x) = \frac{d}{dx} e^{ax} = a e^{ax}$$

On the interval x>0 this function is always larger 0 → strictly monotonic on x>0

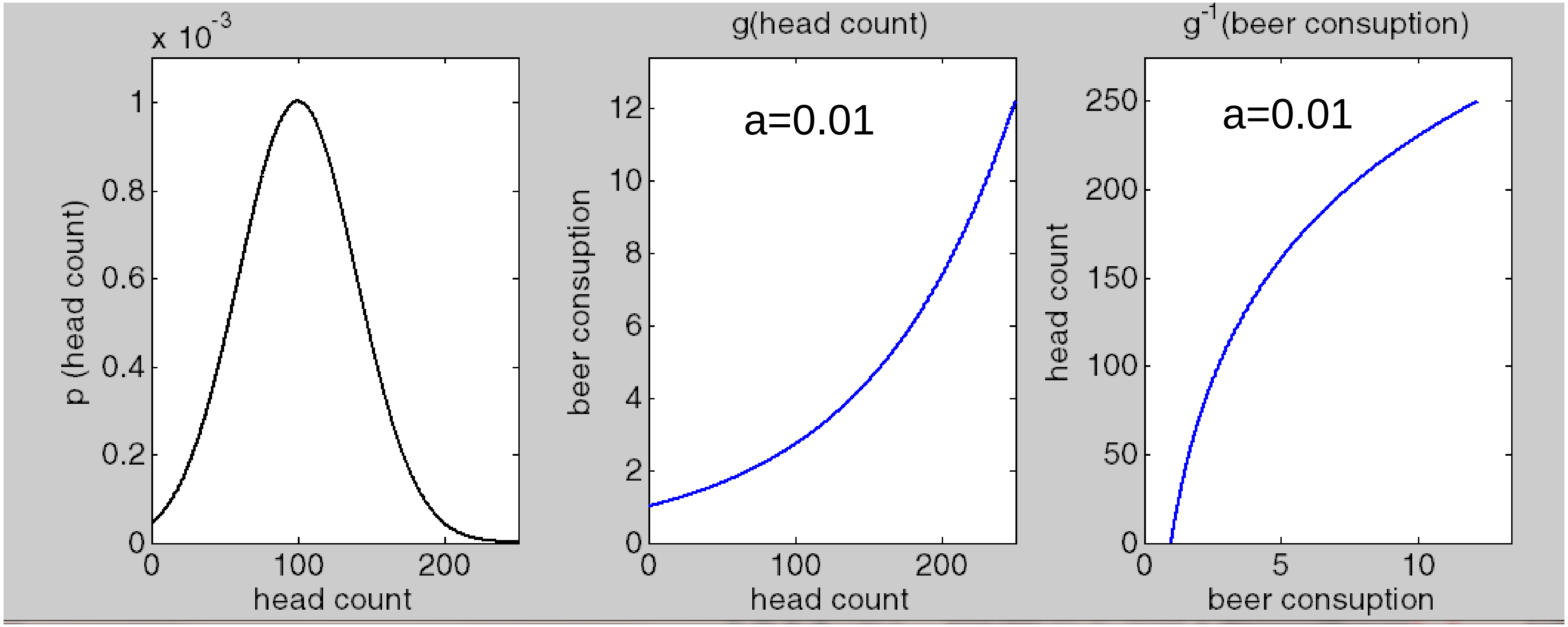
Transformed Densities: Example 1 (proposed by the TAs)

We have: $p_x(x)$ $g(x): y = e^{ax}$ We want to know: $p_y(y)$

Step 2: We know g(x). To compute the probability *p(y)* given *p(x)* we need to substitute *x* by *y* . → We need to compute the inverse function g^{-1} of *g*

$$g(x): y = e^{ax} \Leftrightarrow g(y)^{-1}: x = \ln(y)\frac{1}{a}$$

Check the interval of x,y !



Transformed Densities: Example 1 (proposed by the TAs)

We have: $p_x(x)$ $g(x): y = e^{ax}$ We want to know: $p_y(y)$

Step 2: We know y(x). To compute the probability *p(y)* given *p(x)* we need to substitute *x* by *y* . → We need to compute the inverse function g^{-1} of *g*

$$g(x): y = e^{ax} \Leftrightarrow g(y)^{-1}: x = \ln(y)\frac{1}{a}$$

Step 3: We substitute x by y, and multiply with the derivative of g^{-1}

$$p(y) = p_x(g^{-1}(y))\left|\frac{dg^{-1}(y)}{dy}\right| = p_x(x)\left|\frac{dg^{-1}(y)}{dy}\right|$$

$$p(y) = p_x\left(\frac{\ln(y)}{a}\right)\left|\frac{dg^{-1}(y)}{dy}\right|$$

$$p(y) = p_x\left(\frac{\ln(y)}{a}\right)\left|\frac{1}{ay}\right|$$

Transformed Densities: Example 1 (proposed by the TAs)

We have: $p_x(x)$ $\quad g(x): y = e^{ax}$ $\quad$ We want to know: $p_y(y)$

Lets say $p_x(x)$ is approximately Normal (Gaussian) distributed

$$p_x(x) = \frac{1}{norm} e^{\left(-\frac{(x-\mu_x)^2}{\sigma_x{}^2}\right)}$$

Step 3: Now we can compute p(y)

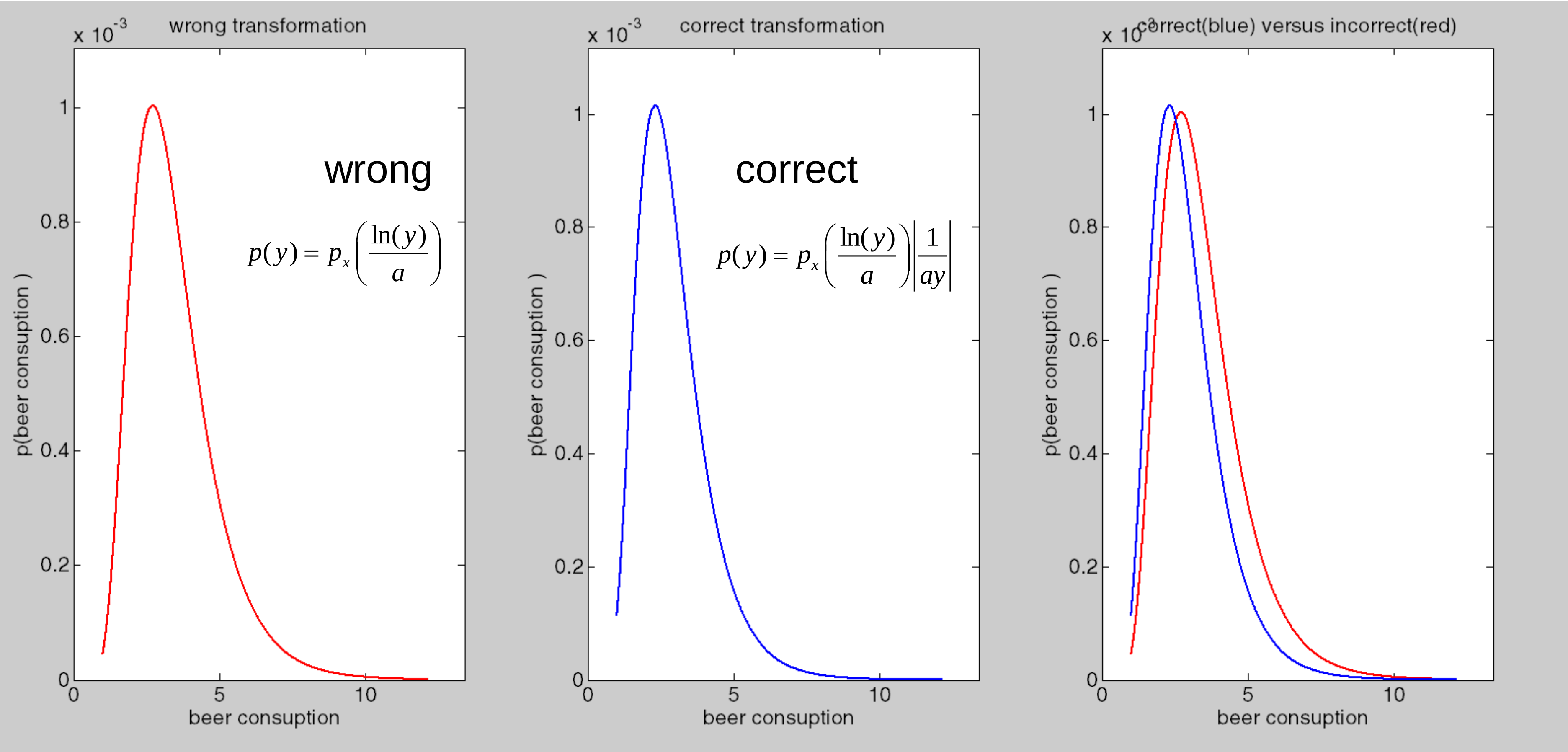
$$p(y) = p_x\left(\frac{\ln(y)}{a}\right)\left|\frac{1}{ay}\right|$$

$$p(y) = \left|\frac{1}{ay}\right| \frac{1}{norm} e^{\left(-\frac{(\ln(y)/a-\mu_x)^2}{\sigma_x{}^2}\right)}$$

Transformed Densities: Example 1 (proposed by the TAs)

We have: $p_x(x)$ $g(x): y = e^{ax}$ We want to know: $p_y(y)$

$$p(y) = \left|\frac{1}{ay}\right| \frac{1}{norm} e^{\left(\frac{(\ln(y)/a - \mu_x)^2}{\sigma_x{}^2}\right)}$$



Expectations

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

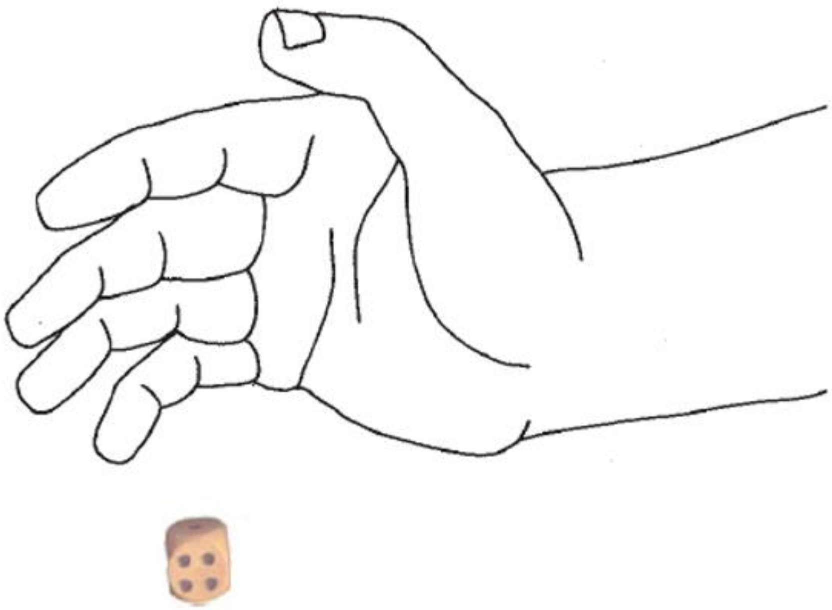
$$\mathbb{E}[f] = \int p(x) f(x) \, \mathrm{d}x$$

- The expected value of a random variable is the weighted average of all possible values that this random variable can take on. The weights correspond to the probabilities in case of a discrete random variable, or densities in case of a continuous random variable.
- In most cases the expected value is the sample mean as sample size grows to infinity. Than, it can be interpreted as the long-run average of many independent repetitions of an experiment (e.g. a dice roll).
- The value may not be expected in the ordinary sense—the "expected value" itself may be unlikely or even impossible (such as having 2.5 children), just like the sample mean.

Example: Roll Dice

Possible Outcomes: *1,2,3,4,5,6*

Fair dice: *p(x=1)=p(x=2) … = 1/6*



$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

E(x) = 1/6 * 1 + 1/6 * 2 + … + 1/6 * 6

E(x) = 1/6 * (1+2+3+4+5+6)

= 1/6 * (21)

= 3.5

Expectations

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

$$\mathbb{E}[f] = \int p(x) f(x) \, \mathrm{d}x$$

(Expected value or expectation, mathematical expectation, EV, mean, or the first moment)

$$E_x[f(x)] = \langle f(x) \rangle_x$$

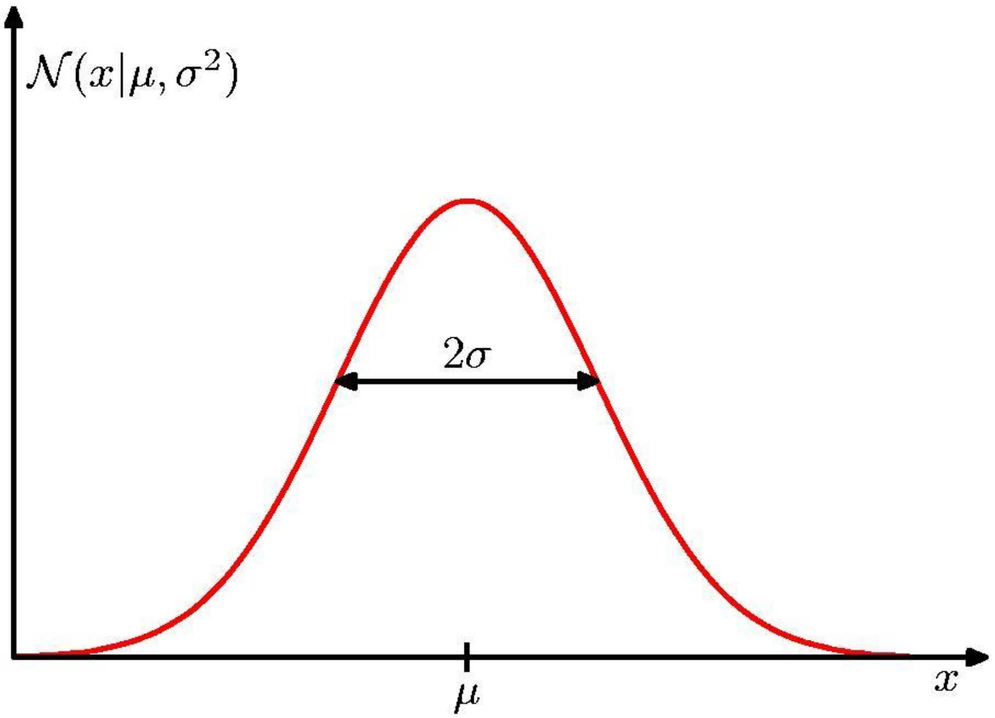
Alternative notation

$$\mathbb{E}_x[f|y] = \sum_x p(x|y) f(x)$$

Conditional Expectation (discrete)

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

$$\mathcal{N}(x|\mu,\sigma^2) > 0 \qquad \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1$$



Gaussian Mean and Variance

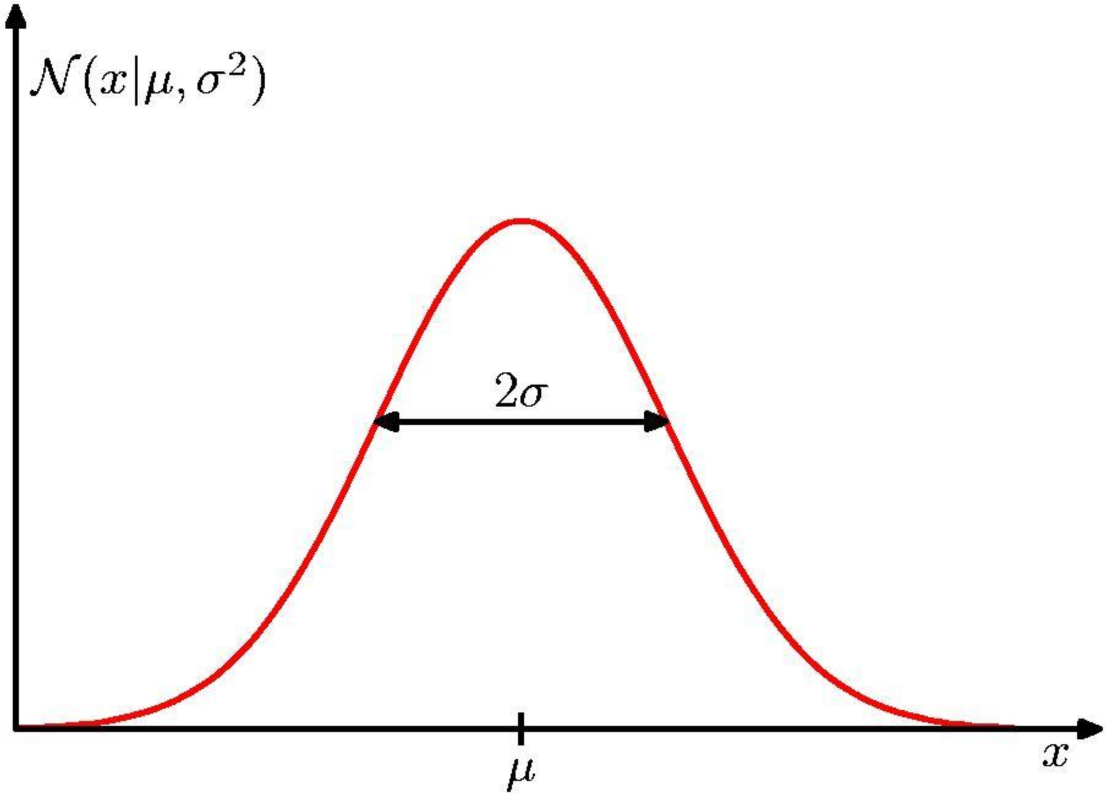
$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x \, \mathrm{d}x = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x^2 \, \mathrm{d}x = \mu^2 + \sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

$$\mathrm{var}[x] = E[(x - E[x])^2] = E[x^2] - E[x]^2 = \sigma^2$$



Q

1. Say we measure a random variable p(x) and would like to know p(y) when y=g(x). How do you calculate this.

   (a) (4 pts) There might be more than one correct statements.

   - [ ] We use $p(y) = p(g(x))$
   - [ ] We use $p(y) = p(g(x)) \cdot g(y)$
   - [ ] We use $p(y) = p(g(x)) \cdot |g'(x)|$
   - [ ] We use $p(y) = p(g(x)) \cdot \left|\frac{dg(x)}{dx}\right|$

2. We discussed a change of variables in the context of a set of provability distribution. Why is it that we have to use $p(y) = p(g(x)) \cdot |g'(x)|$ instead of just $p(y) = p(g(x))$.

Timer (5min): Start Stop

Say we measure a random variable p(x) and would like to know p(y) when y=g(x). How do you calculate this.

(a) (4 pts) There might be more than one correct statements.

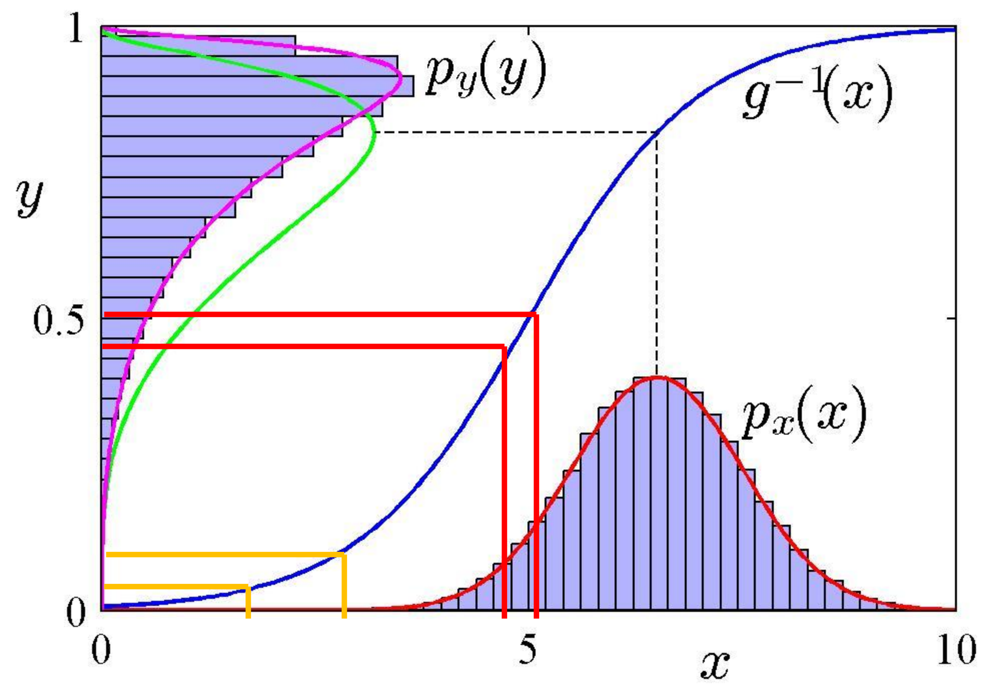
- ☒ We use $p(y) = p(g(x))$
- ☒ We use $p(y) = p(g(x)) \cdot g(y)$
- ☒ We use $p(y) = p(g(x)) \cdot |g'(x)|$
- ☒ We use $p(y) = p(g(x)) \cdot \left|\frac{dg(x)}{dx}\right|$

$$p_y(y) = p_x(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = p_x(g(y))\,|g'(y)|$$

Change of variable

g: transformation

$x = g(y)$



wrong

$p_x(g(y))$

$p_x(g(y))|g'(y)|$

Neuroinformatics Lecture

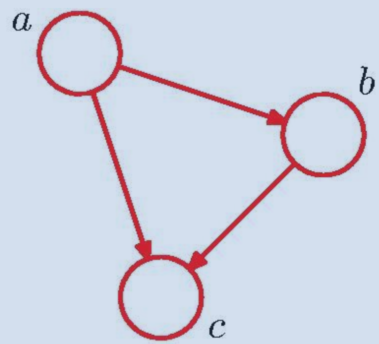
Topics:

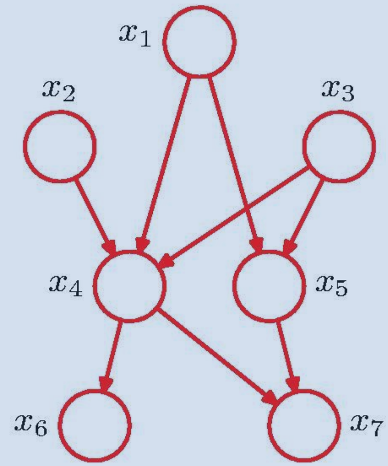
GRAPHICAL MODELS

Significance of Graphical Models

Graphical models: combinations of probability and graphs

- Probability theory provides the glue with which the parts are combined, ensuring that the system as a whole is consistent, and providing ways to interface models to data.
- The graph theoretic side of graphical models provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms.
- They provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering – **uncertainty** and **complexity**.





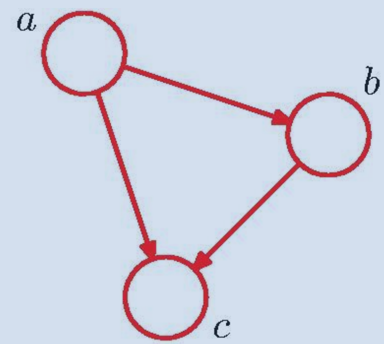
Significance of Graphical Models

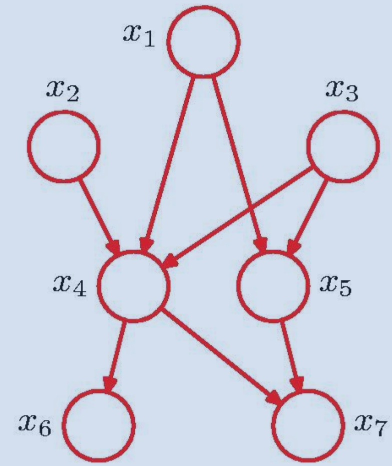
Graphical models: combinations of probability and graphs

- Fundamental to the idea of a graphical model is the **notion of modularity** – a complex system is built by **combining simpler parts**.

- Many of the classical multivariate probabilistic systems studied in fields such as statistics, systems engineering, information theory, pattern recognition and statistical mechanics are special cases of the general graphical model formalism -- **examples include mixture models, factor analysis, hidden Markov models, and Kalman filters**.

- The graphical model framework provides a way to view all of these systems as instances of a **common underlying formalism**.





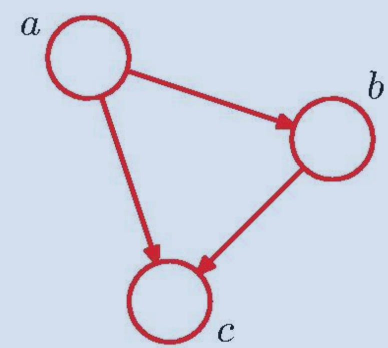
Significance of Graphical Models

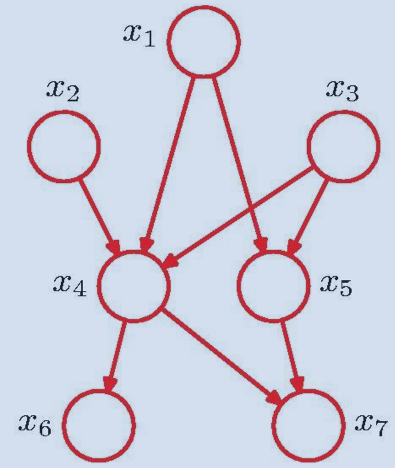
***Probabilistic graphical models* offer several useful properties:**

1. They provide a simple way to **visualize the structure of a probabilistic model** and they can be used to design and motivate new models.

2. **Insights into the properties of the model**, including **conditional independence properties**, can be obtained by inspection of the graph.

3. Complex computations, required to perform inference and learning in sophisticated models, can be expressed in terms of graphical manipulations, in which underlying mathematical expressions are carried along implicitly.

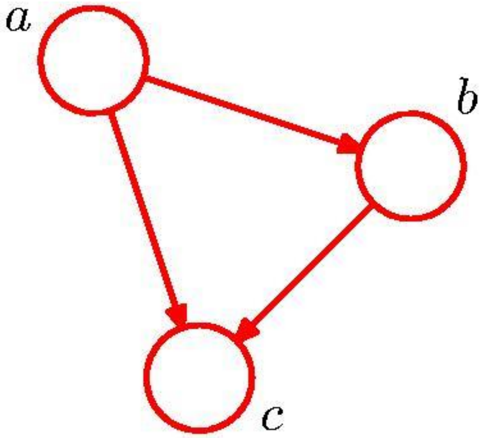




A graph comprises *nodes* (also called *vertices*) connected by *links* (also known as *edges* or *arcs*).

In a probabilistic graphical model, each node represents a random variable (or a group of random variables), and the links express probabilistic relationships between these variables.

The graph then captures the way in which the joint distribution over all of the random variables can be decomposed into a product of factors, each depending only on a subset of the variables.



Bayesian networks, also known as directed graphical models:

Links of the graphs have a particular directionality indicated by arrows.

Markov random fields, also known as undirected graphical models:

Links do not carry arrows and have no directional significance.

Directed graphs are useful for expressing causal relationships between random variables, whereas undirected graphs are better suited for expressing soft constraints between random variables.

In order to motivate the use of directed graphs to describe probability distributions, consider first an arbitrary joint distribution *p*(*a,b,c*) over three variables *a*, *b*, and *c*.

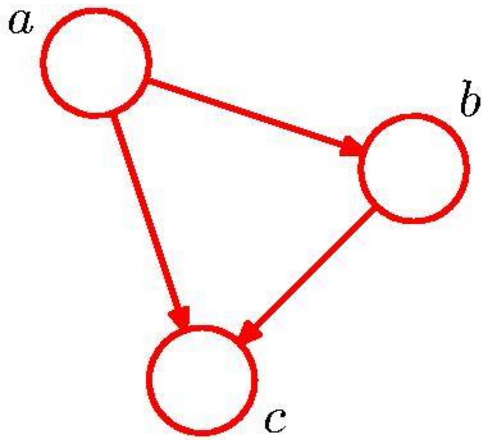
By application of the **product rule** of probability, we can write the joint distribution in the form:

$$p(a,b,c) = p(c|a,b)p(a,b) = p(c|a,b)p(b|a)p(a)$$

Note that this decomposition holds for any choice of the joint distribution.

$$p(a, b, c) = p(c|a, b)p(a, b) = p(c|a, b)p(b|a)p(a)$$

We now represent the right-hand side of this equation.
In terms of a simple graphical model as follows:



1. We introduce a node for each of the random variables *a*, *b*, and *c* and associate each node with the corresponding conditional distribution on the right-hand side of the eqn.

2. Then for each conditional distribution, we add directed links (arrows) to the graph from the nodes corresponding to the variables on which the distribution is conditioned.

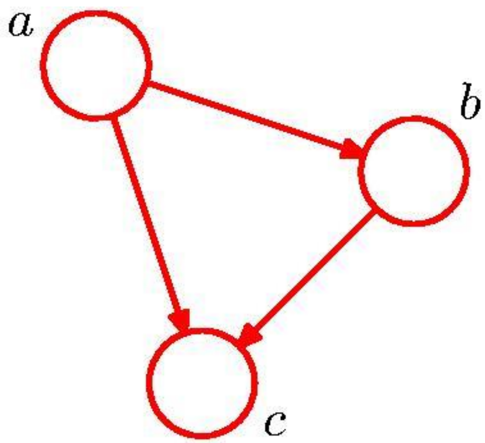
   Thus for the factor *p*(*c*|*a*,*b*), there will be links from nodes *a* and *b* to node *c*, whereas for the factor *p*(*a*) there will be no incoming links.

$$p(a,b,c) = p(c|a,b)p(a,b) = p(c|a,b)p(b|a)p(a)$$

If there is a link going from a node *a* to a node *b*, then we say that node *a* is the *parent* of node *b*, and we say that node *b* is the *child* of node *a*.

Note that we shall not make any formal distinction between a node and the variable to which it Corresponds, but will simply use the same symbol to refer to both.

Parent of c



Child of a

$$p(a,b,c) = p(c|a,b)p(a,b) = p(c|a,b)p(b|a)p(a)$$

An interesting point to note about the equation is that the left-hand side is symmetrical with respect to the three variables *a*, *b*, and *c*, whereas the right-hand side is not.

Indeed, in making the decomposition, we have implicitly chosen a particular ordering, namely *a, b, c*, and had we chosen a different Ordering, we would have obtained a different decomposition and hence a different graphical representation.

$$\begin{aligned} p(a,b,c) = {} & p(c\,|\,a,b)\,p(b\,|\,a)\,p(a) \\ & p(a\,|\,b,c)\,p(b\,|\,c)\,p(c) \\ & p(b\,|\,a,c)\,p(c\,|\,a)\,p(a) \end{aligned}$$

Conditional Models for K Variables

For K=3 variables:

$$p(a, b, c) = p(c|a, b)p(a, b) = p(c|a, b)p(b|a)p(a)$$

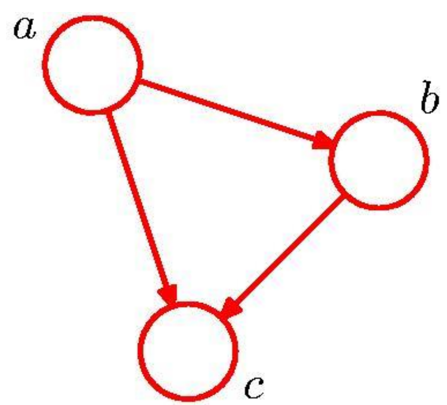
For K variables in general:

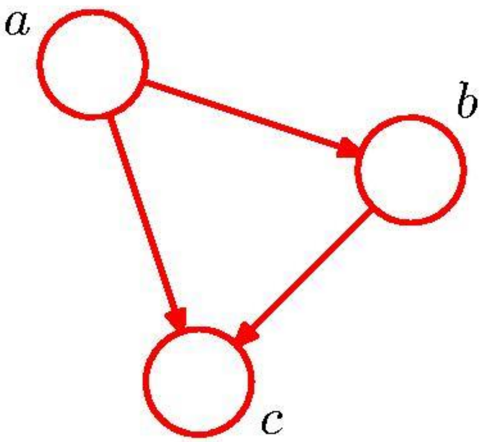
$$p(x_1, x_2, \ldots, x_K) = p(x_K|x_1, x_2, \ldots, x_{K-1}) \cdot p(x_{K-1}|x_1, x_2, \ldots, x_{K-2}) \cdot \ldots \cdot p(x_2|x_1)p(x_1)$$

For a given choice of *K*, we can again represent this as a directed graph having *K* nodes, one for each conditional distribution on the right-hand side, with each node having incoming links from all lower-numbered nodes. We say that this graph is fully connected because there is a link between every pair of nodes.

Conditional Models for K Variables

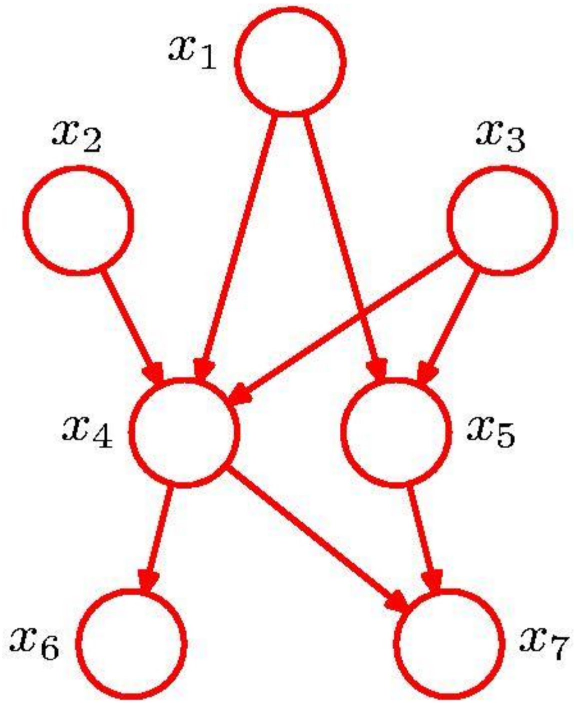
- The directed graphs that we are considering are subject to an important restriction namely that there must be no ***directed cycles.***
- In other words, there are no closed paths within the graph such that we can move from node to node along links following the direction of the arrows and end up back at the starting node.
- Such graphs are also called *directed acyclic graphs*, or *DAGs*. This is equivalent to the statement that there exists an ordering of the nodes such that there are no links that go from any node to any lower-numbered node.



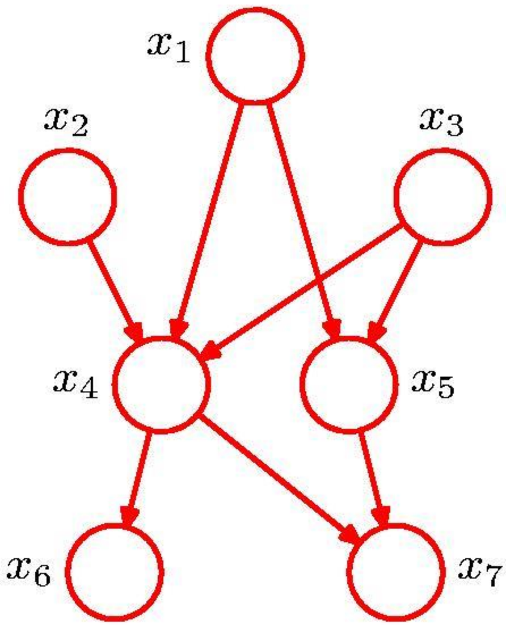


Joint distribution ?

$$p(a, b, c) = p(c|a, b)$$



$$p(x_1, \ldots, x_7) \quad = \quad p(x_1)p(x_2)p(x_3)$$



$$
\begin{aligned}
p(x_1, \ldots, x_7) \quad = \quad & p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3) \\
& p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)
\end{aligned}
$$

General Factorization

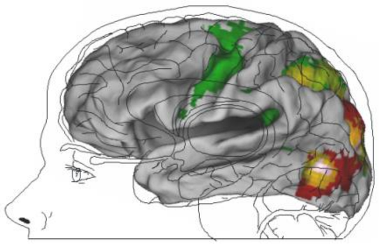
The joint distribution defined by a graph is given by the product, over all of the nodes of the graph, of a conditional distribution for each node conditioned on the variables corresponding to the parents of that node in the graph. Thus, for a graph with *K* nodes, the joint distribution is given by:

$$
p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)
$$

pa_k denotes the set of parents of x_k

We can interpret such models as expressing the processes by which the observed data arose.

Complex image



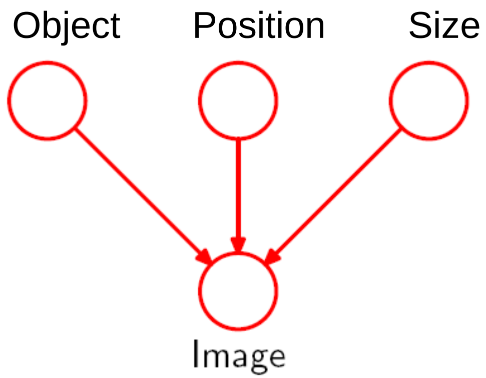
c.f. www.uniklinik-freiburg.de

For example, consider an object-recognition task in which each observed data point corresponds to an image (comprising a vector of pixel intensities) of one of the objects.

Image composed of building blocks

In this case, the latent variables might have an interpretation as the position and size of the objects.

Generative model



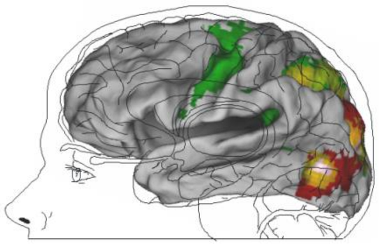
p(Image|Object,Position,Size) = p(Objects) · p(Positions) · p(Sizes)

Latent variables are variables that are not directly observed, **observed variables** are observed

The graphical model captures the *causal* process (Pearl, 1988) by which the observed data was generated. For this reason, such models are often called *generative* models.

Given a particular observed image, our goal is to find the posterior distribution over objects, in which we integrate over all possible positions and sizes.

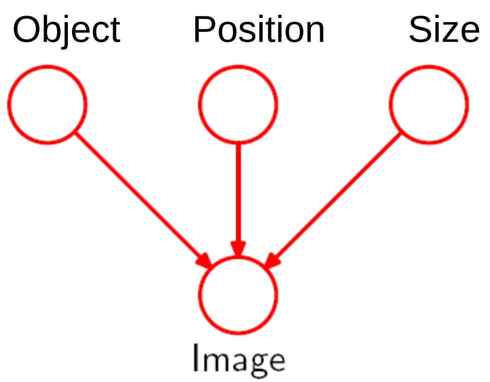
Complex image



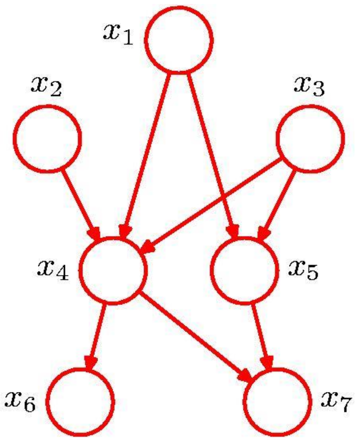
c.f. www.uniklinik-freiburg.de

Image composed of building blocks

Generative model



p(Image|Object,Position,Size) = p(Objects) · p(Positions) · p(Sizes)



$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|\mathrm{pa}_k)$$

3. Tick off correct statements.

(a) (4 pts) There might be more than one correct statements.

- ☐ A graph comprises nodes and links
- ☐ The entire graph describes how Joint distribution can be decomposed
- ☐ There is always only one unique graph per Joint distribution
- ☐ There can be more than one graph per joint distribution depending of the order that was used for decomposition into conditional probabilities.

4. Decompose $p(a, b, c)$ into a graphical model. Make a sketch of the model.

5. Write down the decomposed joint distribution given the graph shown on the slide.

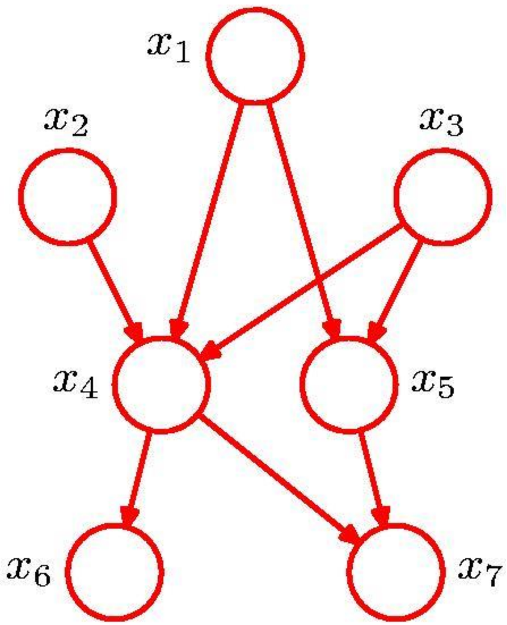
Timer (5min): Start Stop

S

3. Tick off correct statements.

(a) (4 pts) There might be more than one correct statements.

- [X] A graph comprises nodes and links
- [X] The entire graph describes how Joint distribution can be decomposed
- [ ] There is always only one unique graph per Joint distribution
- [X] There can be more than one graph per joint distribution depending of the order that was used for decomposition into conditional probabilities.



$$
\begin{aligned}
p(x_1, \ldots, x_7) \quad = \quad & p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3) \\
& p(x_5|x_1, x_3)p(x_6|x_4)p(x_7|x_4, x_5)
\end{aligned}
$$

$$
\begin{aligned}
p(a,b,c) = \; & p(c\,|\,a,b)\,p(b\,|\,a)\,p(a) \\
& p(a|b,c)p(b|c)p(c) \\
& p(b\,|\,a,c)\,p(c\,|\,a)\,p(a)
\end{aligned}
$$