ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Graphics Lab
ML Group

Prof. M. Gross / Prof. J. Buhmann

Remo Ziegler, Christian Voegeli, Daniel Cotting,
Christian Sigg, Jens Keuchel

Solve before: June 27, 2006

Visual Computing
Gaussian Distribution, Maximum Likelihood Estimation
Solution

General Remarks

It is not necessary to hand in your results. You will find an exemplary solution on the lecture's web page.

1) Sampling a d-dimensional Gaussian distribution

We consider the problem of drawing samples from a multivariate normal (Gaussian) distribution $\mathcal{N}(\mu, \Sigma)$ with arbitrary parameters $\mu \in \mathbb{R}^d$ (expectation) and $\Sigma \in \mathbb{R}^{d \times d}$ (covariance). For the special case of $\mu = (0, \ldots, 0)$ and $\Sigma = I$ (the identity matrix), Matlab provides the function `randn` to produce pseudo-random samples of the corresponding normal distribution $\mathcal{N}(0, I)$. These samples can be transformed into samples for a general Gaussian by using the *eigenvalue decomposition* of the symmetric covariance matrix Σ:

$$\Sigma = V D V^\top ,$$

where D denotes the diagonal matrix containing the eigenvalues of Σ on the diagonal, and the columns of V contain the corresponding eigenvectors.

If g is a sample vector from $\mathcal{N}(0, I)$, then a sample $\tilde{g}$ from $\mathcal{N}(\mu, \Sigma)$ is obtained by computing

$$\tilde{g} = V \sqrt{D} g + \mu ,$$

where $\sqrt{D}$ denotes the diagonal matrix containing the values $\sqrt{D_{ii}}$. This corresponds to first applying an axis transformation according to the covariance matrix Σ and then adding the expectation vector μ.

a) Implement a Matlab-function `x = gsample(mu, Sigma, n)` to produce `n` draws from a d-dimensional Gaussian distribution. The dimension d is implicitly specified by `mu` and `Sigma`.

Hint: The eigenvalue decomposition of a matrix is computed in Matlab by `[V,D] = eig(Sigma)`.

Solution:

```
function x = gsample(mu, Sigma, n)

[V,D] = eig(Sigma);
d = length(mu);
g = randn(n,d)';

% instead of a loop, use repmat
x = V*sqrt(D)*g + repmat(mu,[1,n]);
```

b) Test your implementation: Apply the matlab functions `mean` and `cov` to a sufficiently large sample generated by your program. They should approximately reproduce your input parameters.

Solution: For example, for $\mu = (1,1)^t$, $\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ the function gives

n	μ	Σ
10	$\begin{pmatrix} 1.14 \\ 0.86 \end{pmatrix}$	$\begin{pmatrix} 3.04 & 1.11 \\ 1.11 & 1.71 \end{pmatrix}$
100	$\begin{pmatrix} 0.87 \\ 1.03 \end{pmatrix}$	$\begin{pmatrix} 1.97 & 0.84 \\ 0.84 & 2.06 \end{pmatrix}$
1000	$\begin{pmatrix} 0.90 \\ 0.94 \end{pmatrix}$	$\begin{pmatrix} 2.03 & 1.11 \\ 1.11 & 2.15 \end{pmatrix}$
10000	$\begin{pmatrix} 1.00 \\ 1.01 \end{pmatrix}$	$\begin{pmatrix} 2.03 & 1.01 \\ 1.01 & 1.99 \end{pmatrix}$
100000	$\begin{pmatrix} 0.997 \\ 0.997 \end{pmatrix}$	$\begin{pmatrix} 2.004 & 1.007 \\ 1.007 & 2.002 \end{pmatrix}$

c) Produce 100 samples each in two and three dimensions, using the parameter values $\mu = (1,1)^t$, $\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ and $\mu = (1,1,1)^t$, $\Sigma = \begin{pmatrix} 5 & 2 & 0 \\ 2 & 3 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, respectively. Plot your results using the functions `plot` and `plot3`.

Solution:
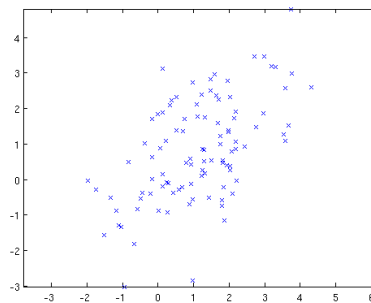

```
plot(x(1,:),x(2,:),'x');
```
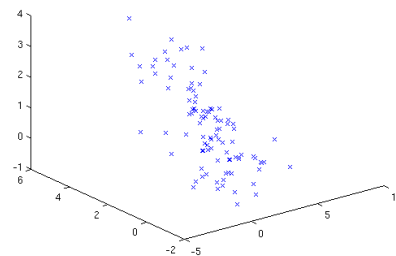


```
plot3(x(1,:),x(2,:),x(3,:),'x');
```

2) Sample size

A recurrent theme in statistical data analysis is that large numbers of observations are required in order to obtain reliable estimates. The present problem aims at illustrating this phenomenon, using a Gaussian distribution as an example data source. We will draw n sample points from a one-dimensional Gaussian, sort them into a histogram, and see how stable the result is with respect to different samples. The basic procedure is the following:

- Choose a number n of sample points and a number N_{bins} of histogram bins.

- Use the matlab function `randn` to draw n samples from a one-dimensional Gaussian distribution ($\mu = 0, \sigma = 1$).

- Turn the complete data sample into a histogram using the function `hist`.

Please complete the following steps:

a) Produce four histograms with $n = 100$ and $N_{\text{bins}} = 10$. Plot the histograms using the `plot` function.
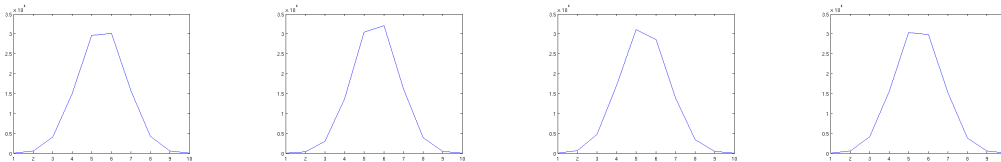
Solution:

```
n = 100; nbins = 10;
x=randn(n,1);
h = hist(x,nbins);
figure; plot(h);
```
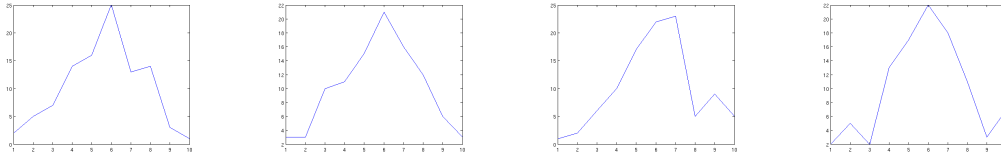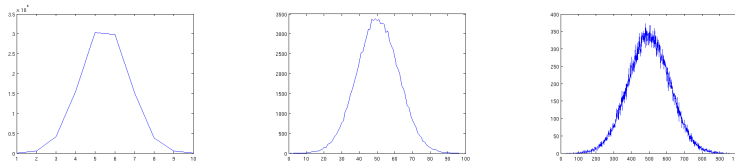

b) Repeat the procedure with $n = 100000$.
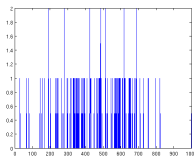
Solution:

c) For $n = 100000$, plot one histogram each for $N_{\text{bins}} \in \{10, 100, 1000\}$.

Solution:

d) Finally, choose $n = 100$ and $N_{\text{bins}} = 1000$.

Solution:

e) Give a brief discussion of the results. Remember, this is about sample size and reliability of estimates.

Solution: a+b) Good reproduction of the distribution requires large samples.

c) Too many bins produce noisy distributions.

d) Insufficient regularization: the ratio n/N_{bins} is too small (histogram bins are too small / too few samples for appropriate approximation of the distribution).

3

3) Analytic MLE

In this exercise, we analytically derive maximum likelihood estimators for the parameters of an example model distribution. While most textbooks discuss the Gaussian example, we consider the *gamma distribution* here.

The gamma distribution is univariate (one-dimensional) and continuous. We consider a parameterization that is controlled by two parameters, the *location parameter* μ and the *shape parameter* ν. For a gamma-distributed random variable X, we write $X \sim \mathcal{G}(\mu, \nu)$. $\mathcal{G}$ is defined by the following density function:

$$p(x|\mu, \nu) := \left(\frac{\nu}{\mu}\right)^{\nu} \frac{x^{\nu-1}}{\Gamma(\nu)} \exp\left(-\frac{\nu x}{\mu}\right) \; ,$$

where $x \geq 0$ and $\mu, \nu > 0$. The symbol Γ denotes the *gamma function* (that's where the distribution's name comes from), which is a generalization of the factorial $n!$ to arbitrary values from the real line. Fortunately, we will not need to evaluate the gamma function explicitly here.

Whenever $\nu > 1$, the gamma density has a single peak, much like a Gaussian. Unlike the Gaussian, it is not symmetric. The first two moment statistics of the gamma distribution are given by

$$\mathsf{E}[X] = \mu \qquad \text{and} \qquad \mathsf{Var}[X] = \frac{\mu^2}{\nu} \tag{1}$$

for $X \sim \mathcal{G}(\mu, \nu)$. The following plots should give you a rough idea of how the gamma density looks like for different parameter values:
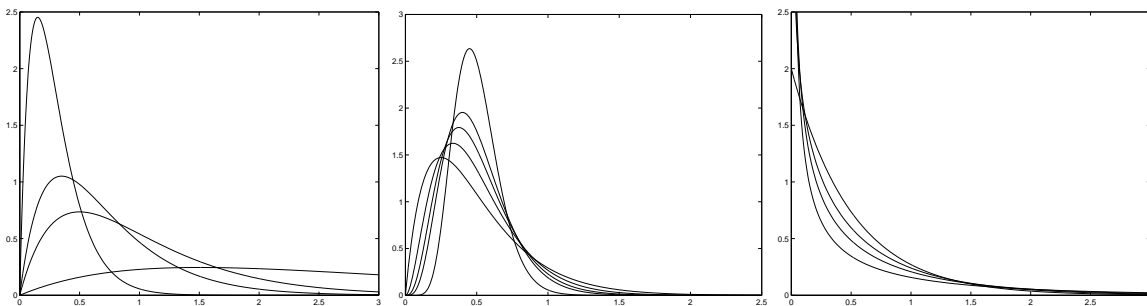
Left: The plot shows the density for different values of the location parameter ($\mu = 0.3, 0.5, 1.0, 3.0$), with the shape parameter fixed to $\nu = 2$. Since $\nu > 1$, the densities peak. As we increase μ, the peak moves to the right, and the curve flattens. *Middle:* For $\mu = 0.5$ fixed, we look at different values of the shape parameter ($\nu = 2, 3, 4, 5, 19$). Again, all the densities peak, with the peak shifting to the right as we increase ν. *Right:* If $\nu < 1$, the density turns into a monotonously decreasing function. The smaller the value of ν, the sharper the curve dips towards the origin.

a) Write the general analytic procedure to obtain the maximum likelihood estimator (including logarithmic transformation) in the form of a short algorithm or recipe. A few words are enough, but be precise: Write all important mathematical operations as formulae. Assume that data is given as an i.i.d. sample $x_1, \ldots, x_n$. Denote the conditional density in question by $p(x|\theta)$, and the likelihood by $l(\theta)$. Make sure both symbols show up somewhere in your list, as well as a logarithm turning a product into a sum.

Solution:

- *data: i.i.d. sample* $x_1, \ldots, x_n \sim p(x|\theta)$
- *i.i.d.* $\Rightarrow p(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta) =: l(\theta)$
- $\log$ *is monotonous* $\Rightarrow \operatorname{argmax}_\theta l(\theta) = \operatorname{argmax}_\theta \log(l(\theta))$

- *necessary condition for maximum:*

$$\frac{\partial}{\partial \theta} l(\theta) = 0 \quad \text{or} \quad \frac{\partial}{\partial \theta} \log\left(l(\theta)\right) = 0$$

- *check second derivative:*

$$\frac{\partial^2}{\partial \theta^2} \log\left(l(\theta)\right) < 0$$

b) Derive the ML estimator for the location parameter μ, given data values $x_1, \ldots, x_n$. Conventionally, the ML estimator for a parameter is denoted by adding a hat symbol: $\hat{\mu}$. Given both the statistics of the gamma distribution (cf. (1)) and what you know about MLE for Gaussians, the result should not come as a surprise.

Solution: Partial derivative for the location parameter μ:

$$
\begin{aligned}
\frac{\partial}{\partial \mu} \log\left(l(\theta)\right) &= \frac{\partial}{\partial \mu} \log\left(\prod_i p\left(x_i|\mu, \nu\right)\right) \\
&= \frac{\partial}{\partial \mu} \sum_i \log\left(\left(\frac{\nu}{\mu}\right)^\nu \frac{x_i^{\nu-1}}{\Gamma(\nu)} \exp\left(-\frac{\nu x_i}{\mu}\right)\right) \\
&= \sum_i \left(\underbrace{\frac{\partial}{\partial \mu} \log\left(\nu^\nu \frac{x_i^{\nu-1}}{\Gamma(\nu)}\right)}_{=0} + \frac{\partial}{\partial \mu} \log\left(\mu^{-\nu} \exp\left(-\frac{\nu x_i}{\mu}\right)\right)\right) \\
&= \sum_i \left(\frac{\partial}{\partial \mu}\left(-\nu \log \mu\right) + \frac{\partial}{\partial \mu}\left(-\frac{\nu x_i}{\mu}\right)\right) \\
&= \sum_i \left(-\frac{\nu}{\mu} + \frac{\nu x_i}{\mu^2}\right)
\end{aligned}
$$

Setting to zero:

$$
\begin{aligned}
\sum_i \left(-\frac{\nu}{\mu} + \frac{\nu x_i}{\mu^2}\right) = 0 \quad &\Rightarrow \quad \sum_i \left(-\nu\mu + \nu x_i\right) = 0 \\
&\Rightarrow \quad -n\nu\mu + \nu \sum_i x_i = 0 \\
&\Rightarrow \quad \mu = \frac{1}{n}\sum_i x_i
\end{aligned}
$$

c) By now you should have some proficiency at deriving ML estimators, so a look at the gamma density will tell you that things get more complicated for the shape parameter: ν appears inside the gamma function, and both inside and outside the exponential. Thus, instead of deriving an explicit formula of the form $\hat{\nu} := \ldots$, please show the following: Given an i.i.d. data sample $x_1, \ldots, x_n$ and the value of μ, the ML estimator $\hat{\nu}$ for the gamma distribution shape parameter solves the equation

$$\sum_{i=1}^n \left(\log\left(\frac{x_i \hat{\nu}}{\mu}\right) + 1 - \frac{x_i}{\mu} - \phi\left(\hat{\nu}\right)\right) = 0,$$

where $\phi(\nu) = \frac{\partial \Gamma(\nu)}{\partial \nu}/\Gamma(\nu)$ denotes the so-called *digamma function*.

Solution: Partial derivative for the shape parameter ν:

$$\frac{\partial}{\partial \nu} \log\left(l(\theta)\right) = \frac{\partial}{\partial \nu} \sum_i \log\left(\left(\frac{\nu}{\mu}\right)^{\nu} \frac{x_i^{\nu-1}}{\Gamma(\nu)} \exp\left(-\frac{\nu x_i}{\mu}\right)\right)$$

$$= \sum_i \frac{\partial}{\partial \nu} \left(\nu \log \nu - \nu \log \mu + (\nu - 1)\log x_i - \log(\Gamma(\nu)) - \frac{\nu x_i}{\mu}\right)$$

$$= \sum_i \left(1 + \log \nu - \log \mu + \log x_i - \frac{\frac{\partial \Gamma(\nu)}{\partial \mu}}{\Gamma(\nu)} - \frac{x_i}{\mu}\right)$$

$$= \sum_i \left(\log\left(\frac{\nu x_i}{\mu}\right) + 1 - \frac{x_i}{\mu} - \phi(\nu)\right)$$

Setting this to zero gives the desired implicit formula for $\hat{\nu}$. A solution can be found iteratively e.g. by using Newton's method.