Probabilistic Graphical Models
Christoph Lampert <chl@ist.ac.at>
TA: all of us <pgm_2016@lists.ist.ac.at>
Exercise Sheet 3 v1.0

1 Maximum Likelihood Parameter Estimation: Gaussians

(Reminder: an estimator $\hat{E}$ is called *unbiased*, if it's expected valued is the true target value E, *i.e.* $\mathbb{E}\hat{E} = E$)

In the lecture we saw: the maximum likelihood parameter estimates for a Gaussian random variable are (for samples $x_1, \dots, x_n$):

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad\qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

a) Show that the estimator $\hat{\mu}$ is *unbiased.*

b) Show that the estimator $\hat{\sigma}^2$ is *not unbiased.*

c) Can you construct an *unbiased* estimator of σ^2?

2 Maximum Likelihood Parameter Estimation: Coin Toss

In the lecture we showed how to derive the maximum-likelihood parameter estimation rules for a coin toss using constrained optimization with Lagrangian multipliers.

a) Derive the same result but using the parameterization $p(\texttt{head}) = \theta$ and $p(\texttt{tail}) = 1 - \theta$ (which makes things much easier).

3 Conditionals

Two research labs work independently on the relationship between discrete variables X and Y. Lab A proudly announces that they have ascertained the distribution $p(x|y)$ from data (let's call it $p_A(x|y)$). Lab B proudly announces that they have ascertained $p(y|x)$ from data (called $p_B(y|x)$).

a) Is it always possible to find a joint distribution $p(x, y)$ consistent with the results of both labs?

b) Is it possible to define consistent marginals $p(x)$ and $p(y)$, in the sense that $p(x) = \sum_y p_A(x|y)p(y)$ and $p(y) = \sum_x p_B(y|x)p(x)$? If so, explain how to find such marginals. If not, explain why not.

4 Estimating Entropy

(Reminder: the *entropy* of a discrete random variable $X \sim p(x)$ is $H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$.)

For a set of samples $x_1, x_2, \dots,$ study the following *plug-in estimator* of the entropy

$$\hat{H}_n(X) = -\sum_{x\in\mathcal{X}} \hat{p}_n(x)\log\hat{p}_n(x)$$

where $\hat{p}_n(x) = \frac{1}{n}\sum_{i=1}^{n} [\![x_i = x]\!]$ is the maximum likelihood estimate of the probability distribution.

4.1 Simulation

Consider a random variable X with $\mathcal{X} = \{1, 2, 3, 4\}$ and $p(x) = (\frac{1}{4}, \frac{1}{3}, \frac{1}{4}, \frac{1}{6})$. Write a program/routine that for given $k \in \mathbb{N}$:

a) produces k samples i.i.d. from p

b) from the samples, computes the maximum likelihood estimate of $\hat{p}(x)$

c) computes the entropy of the estimated distribution (make sure that it handles $0 \log 0$ correctly)

For each $k \in \{1, 5, 10, 20, 50, 100, 200, 500, 1000\}$ run the program 100 times.

d) plot the *average estimated entropy* for each k and their *standard error of the mean* (=standard deviation divided by the square root of the number of repeats). Make sure to choose a reasonable parametrization of the axes.

e) plot the true entropy as a constant line in the same figure

f) interpret your results

4.2 Analysis

g) Show that $\hat{H}_n(X)$ is *biased* as an estimator of the true entropy.

h) Show that it *underestimates* the true entropy, *i.e.* $\mathbb{E}\hat{H}_n(X) - H(X) \geq 0$.

i) Is $\hat{H}_n(X)$ *consistent*?

Hint: You should be able to solve g) yourself. If you cannot solve h) and/or i), feel free to consult the literature, e.g. [G. P. Basharin, *On a Statistical Estimate for the Entropy of a Sequence of Independent Random Variables*, Theory of Probability & Its Applications 1959 4:3, 333-336].

4.3 Alternatives

For continuous random variables, the plug-in estimator cannot be used directly, since we cannot estimate $\hat{p}$ that easily.

j) Search the literature to find at least different ways to estimate the entropy of a continuous random variables.