Lecturenotes Statistics I – Contents

Uniform and General Distributions

$$u(x) = \begin{cases} 1 & \text{for} \quad 0 \le x < 1; \\ 0 & \text{elsewhere.} \end{cases}$$

The corresponding distribution function is

$$U(x) = \int_{-\infty}^{x} u(x')\, dx' = \begin{cases} 0 & \text{for} \quad x < 0; \\ x & \text{for} \quad 0 \le x \le 1; \\ 1 & \text{for} \quad x > 1. \end{cases}$$

It allows for the construction of general probability distributions. Let

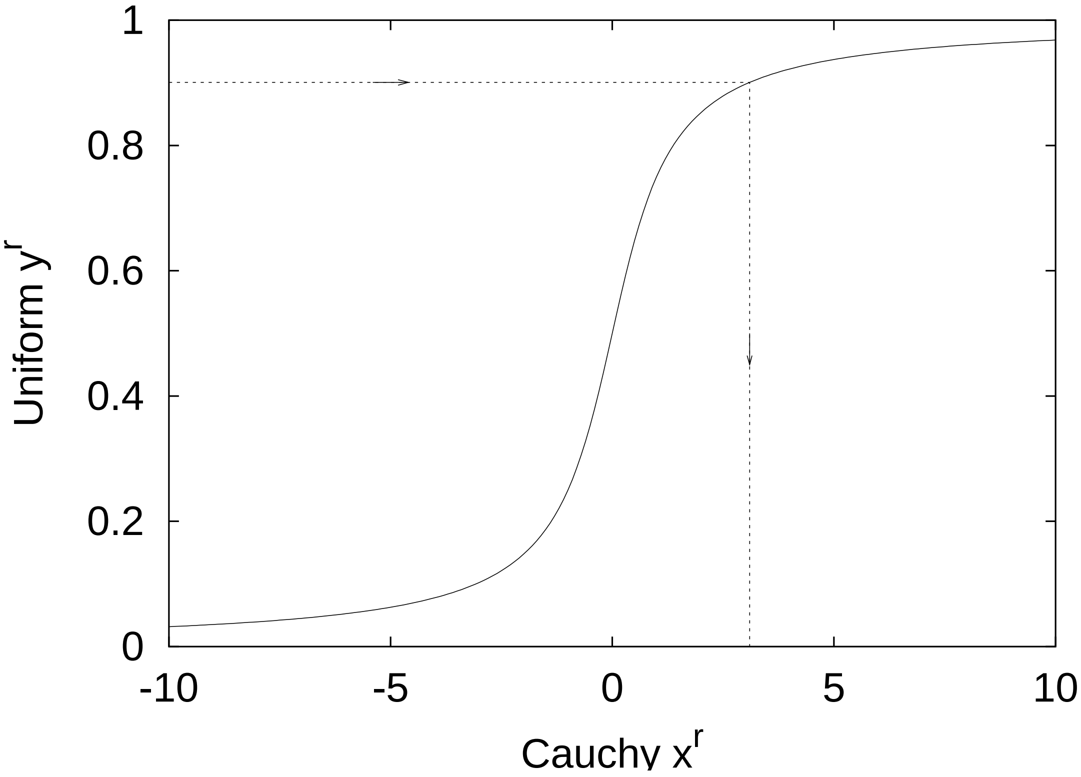
$$y = F(x) = \int_{-\infty}^{x} f(x')\, dx' \; .$$

For y^r being a uniformly distributed random variable in $[0, 1)$

$x^r = F^{-1}(y^r)$ is then distributed according to the probability density $f(x)$.

Example

Mapping of the uniform to the Cauchy distribution.



Confidence Intervals and Sorting

One defines **q-tiles** (also **quantiles** or **fractiles**) x_q of a distribution function by

$$F(x_q) \ = \ q \ .$$

An example is the **median** $x_{\frac{1}{2}}$. The probability content of the **confidence interval**

$$[x_q, x_{1-q}] \ \text{ is } \ p = 1 - 2q \ .$$

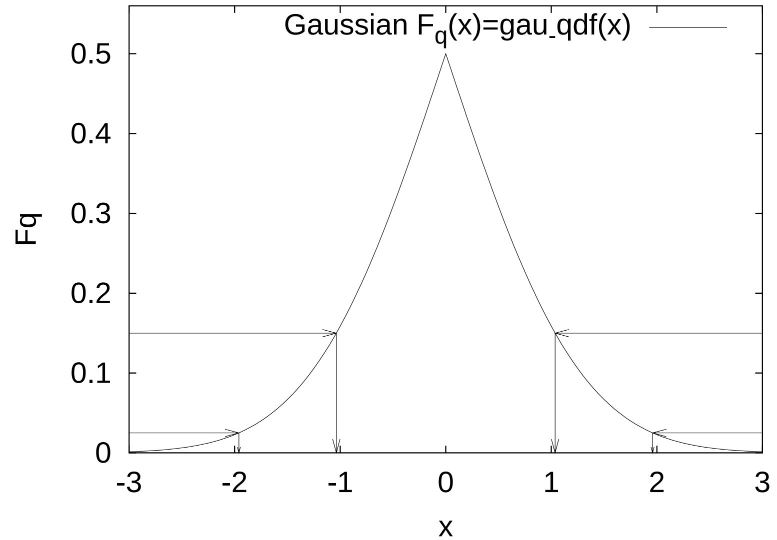
Example: Gaussian or normal distribution of variance σ^2 :

$$[-n\sigma, +n\sigma] \ \Rightarrow \ p = 0.6827 \text{ for } n = 1, \ \ p = 0.9545 \text{ for } n = 2 \ .$$

The **peaked distribution function**

$$F_q(x) = \begin{cases} F(x) \text{ for } F(x) \le \frac{1}{2}, \\ 1 - F(x) \text{ for } F(x) > \frac{1}{2}. \end{cases}$$

provides a graphical visualization of probability intervals of such a distribution:



Sorting allows for an empirical estimate. Assume we generate n random number $x_1, ..., x_n$. We may re-arrange the x_i in increasing order:

$$x_{\pi_1} \leq x_{\pi_2} \leq \ldots \leq x_{\pi_n}$$

where $\pi_1, \ldots, \pi_n$ is a permutation of $1, \ldots, n$.

An estimator for the distribution function $F(x)$ is then the **empirical distribution function**

$$\overline{F}(x) = \frac{i}{n} \quad \text{for} \quad x_{\pi_i} \leq x < x_{\pi_{i+1}},\ i = 0, 1, \ldots, n-1, n$$

with the definitions $x_{\pi_0} = -\infty$ and $x_{\pi_{n+1}} = +\infty$. To calculate $\overline{F}(x)$ one needs an efficient way to **sort** n data values in ascending (or descending) order. In the STMC package this is provided by a **heapsort** routine, which arrives at the results in $O(n \log_2 n)$ steps.

Example: Gaussian distribution in assignment a0106_02 (200 and 20,000 data).