# Probability and likelihood

Miroslav Kulich

Intelligent and Mobile Robotics Group
Czech Institute of Informatics, Robotics and Cybernetics
Czech Technical University in Prague

Winter semester 2018/2019

# Probability

- Probability is the measure of the likelihood that an event will occur.
  [Probability". Webster's Revised Unabridged Dictionary. G & C Merriam, 1913]

- The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible. [Pierre-Simon Laplace, A Philosophical Essay on Probabilities, 1814]

$$P(A) = \frac{N_A}{N}$$

  - Limitations:
    - a finite number of outcomes expected
    - all possible outcomes are equally likely

# Probability

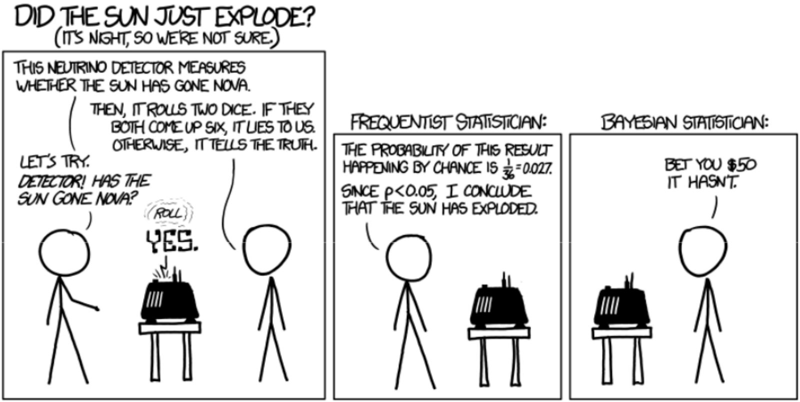
- Frequentism: (Frequency) probability of an event is its relative frequency over time,

$$P(A) = \lim_{n \to \infty} \frac{n_a}{n}$$

  where $n_a$ is the number of occurrences of an event $A$ in $n$ trials

- Kolmogorov (axiomatic) probability:
  - Non-negativity $P(E) \in \mathcal{R}, P(E) \geq 0 \quad \forall E \in F$
  - Unitarity: $P(\Omega) = 1$
  - $\sigma-$additivity: $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$

- Subjectivism (Bayesians): Probability is a measure of the 'degree of belief' of the individual assessing the uncertainty of a particular situation.




Randall Munroe, xkcd: *Frequentists vs. Bayesians* (http://xkcd.com/1132)

# Likelihood

- A likelihood function (likelihood) is a function of the parameters of a statistical model, given specific observed data.

$$\mathcal{L}(\theta|O) = P(O|\theta)$$

- The likelihood function is conditioned on the observed $O$ and that it is a function of the unknown parameters $\theta$.

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{\textit{likelihood} \cdot \textit{prior}}{\textit{evidence}}$$

$x \ldots$ state $\quad y \ldots$ data

# Maximum likelihood estimation

- A method of estimating the parameters of a statistical model, given observations
- Goal: find the parameter values that maximize the likelihood function, given the observations and a statistical model:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta; x)$$

- log-likelihood:

$$\ell(\theta; x) = \ln \mathcal{L}(\theta; x)$$

- Closed-form solution exists in many cases
- Use numerical methods otherwise (e.g. Expectation-Maximization)

# Maximum likelihood estimation - Example

- Assume a normal distribution $\mathcal{N}(\mu, \sigma^2)$:

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

given i.i.d samples $X_i$, $\sigma^2$ is known, $\mu$ to be estimated

$$\ln \mathcal{L}\left(\theta | X_1, X_2, \ldots X_n\right) = \ln \left( \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i-\theta)^2}{2\sigma^2}} \right)$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(X_i - \theta)^2$$

$$\frac{\partial \ln \mathcal{L}(\theta | X_1, X_2, \ldots, X_n)}{\partial \theta} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(X_i - \theta) = 0$$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{N} X_i = \bar{X}_n$$ $\leftarrow$ **sample mean**