4. Probability Theory, Distributions and all that

See: Bishop Chapters 1 and 2.
Experiment
(Bishop 1.2)

- Have two boxes with fruit
- Randomly pick a box
- Pick fruit from box and note its type
- Put fruit back
Experiment (Bishop 1.2)

- Random variables
  - B: box being red r or blue b
  - F: fruit being apple a or orange o

Let $P(B=b)$ denote the probability for picking the blue box.

Probabilities: positive, normalized and additive
Generalize

Consider two random variables $X$ and $Y$

For now: use relative frequencies for probabilities
Joint and marginal probabilities

Joint probability\[ p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} \]

Marginal probability\[ p(X = x_i) = \frac{c_i}{N} \]

Note: \[ c_i = \sum_j n_{ij} \]

Therefore\[ p(X = x_i) = \sum_j p(X = x_i, Y = y_j) \]
Illustration for joint and marginal probability
Conditional probability

Suppose we fix $X = x_i$ and call this a conditional probability

$$p(Y = y_j \mid X = x_i) = \frac{n_{ij}}{c_i}$$

Relation to joint probability (product rule)

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

$$= \frac{n_{ij}}{c_i} \frac{c_i}{N}$$

$$= p(Y = y_j \mid X = x_i) p(X = x_i)$$
Illustration for conditional probability

We know it is $Y=1$
The rules of probability

Sum rule

\[ p(X) = \sum_Y p(X, Y) \]

Product rule

\[ p(X, Y) = p(Y | X) p(X) \]

Bayes theorem

\[ p(X | Y) p(Y) = p(Y | X) p(X) \]
Probability densities

• Continues events \(x\)
  – (e.g. weight of an apple)

• \(p(x)\) is called probability density

• It is

  \[\text{positive: } p(x) \geq 0\]

  \[\text{normalised: } \int p(x)dx = 1\]
Expectation values

• Expectation value

\[ \mathbb{E}[f] = \int f(x) p(x) \, dx \]

• Can be approximated by

\[ \mathbb{E}[f] \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i) \]

Where \( x_i \) is sampled from the distribution \( p(x) \)
Basic Expectation Values

- **Normalization**
  \[ 1 = E[1] = \int p(x)dx \]

- **Mean**
  \[ \mu = E[x] = \int x p(x)dx \]

- **Variance**
  \[ \sigma^2 = E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx \]
  (\(\sigma\) is called standard deviation)

- **Entropy**
  \[ H = E[-\ln p(x)] = \int_{-\infty}^{\infty} [-\ln p(x)] p(x)dx = -\int_{-\infty}^{\infty} p(x) \ln p(x) dx \]
Expectation values for several random variables

- Partial expectation value \( \mathbb{E}_x[f(x, y)] = \sum_x f(x, y) p(x, y) \)
  
  (subscript denotes the variable that is summed over)

- Conditional expectation \( \mathbb{E}_x[f \mid y] = \sum_x f(x, y) p(x \mid y) \)

- Covariance
  
  \[
  \text{cov}[x, y] = \mathbb{E}_{x,y}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])] = \mathbb{E}_{x,y}[xy] - \mathbb{E}[x] \mathbb{E}[y]
  \]
One Dimensional Gaussian Density (Univariate Density)

\[ N(x \mid \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2} \]

Terminology: “Normal Distribution” is a different term for Gaussian Densities

\[ \maple \]
Maximum Likelihood Estimation
Maximum Likelihood Estimation

- A popular statistical method used to estimate parameters of the underlying probability distribution of given samples
- It finds the most likely value of the parameter (the likelihood of the sample is the highest when the estimate parameter is used)
- Developed solely by R. Fisher between 1912 and 1922.
Bernoulli Experiment

- Draw ball from a box
- Fraction of green balls is $p$
- Return the ball afterwards

Length of sequence: $N=8$

$P_{\text{green}} = 0.6$

Probability of that sequence?
Bernoulli Experiment

Probability of that sequence

\[ P_{\text{Sequence}} = (1 - p_{\text{green}})(1 - p_{\text{green}})(1 - p_{\text{green}})p_{\text{green}}p_{\text{green}}(1 - p_{\text{green}})(1 - p_{\text{green}})p_{\text{green}} \]

\[ P_{\text{Sequence}} = (1 - p_{\text{green}})^5 p_{\text{green}}^3 \]
Generalization to sequence with specific order of balls

\( p \): probability for a green ball
\( N \): length of sequence
\( x \): number of green balls in a specific sequence
Probability

\[
P_{Sequence} = (1 - p)^{N-x} p^x
\]
Bernoulli Experiment

Probability of any sequence with $x$ green balls (ignoring order of drawing)

\[
P_{\text{Sequence}} = \binom{N}{x} (1 - p)^{N-x} p^x
\]

\[
= \frac{N!}{(N - x)!x!} (1 - p)^{N-x} p^x
\]
Estimation problem for Bernoulli Experiment

Result $x$ of drawing $N$ balls is known
What could be $p$?

$p=?$

Estimate $p$
Maximum Likelihood (ML) Estimator

Maximize the probability (likelihood) of the known sequence:

\[
\hat{p} = \arg \max_p P_{\text{Sequence}} = \arg \max_p \left[ \binom{N}{x}(1 - p)^{N-x} p^x \right]
\]

\[
L(p) = \binom{N}{x}(1 - p)^{N-x} p^x
\]

\[
\frac{dL(p)}{dp} = \binom{N}{x}(1 - p)^{N-x} p^x \left( \frac{x}{p} - \frac{N-x}{1-p} \right) = 0
\]

Relative frequency

\[\hat{p} = \frac{x}{N}\]

Warning: ML-estimate is not always relative frequencies (see e.g. Duda+Hart exercise 24)
Summary

- Probabilities
  - Joint, marginal, conditional
- Probability densities
- Expectation values
- Parameter estimation