Oliver Kohlbacher, Sven Nahnsen, Knut Reinert

3. Basic statistics for computational MS

Outline

- Probability distributions
  - Discrete probability distributions
  - Continuous probability distribution
- p-values and false discovery rates
- Mixture modeling
- Expectation-Maximization algorithm

Random variables

- A ***random variable***, usually written *X*, is a variable whose possible values are numerical outcomes of a random phenomenon. These values can be interpreted as probabilities. There are two types of random variables, ***discrete*** and ***continuous.***
  - ***Discrete*** random variables have a countable number of outcomes, e.g., dice
  - ***Continuous*** random variables have an infinite continuum of possible values, e.g., blood pressure

Probability functions/distribution

- A probability distribution is a function that describes the probability of a random variable taking certain values
- A probability function maps the possible values of x against their respective probabilities of occurrence, $p(x)$
- $p(x)$ is a number from 0 to 1.0.
- The area under a probability function is always 1.

Mean and variance

- If we understand the underlying probability distribution of a certain phenomenon, then we can make informed decisions based on how we expect *x* to behave on-average.
- The expected value is just the weighted average or mean (μ) of random variable *x*.
    - A random variable X takes values x_1 with a probability p_1, x_2 with p_2,… and x_n with p_n, the expected value or mean is then given by

$$E[X] = \mu = \sum_{i=1}^{n} x_i p_i$$

For example: see the average weight and isotope distribution (lecture 2)

Mean and variance

- The variance is a measure that describes how far the numbers are from the mean
  - A random variable *X* has the expected value (mean) $\mu=E(X)$, then the variance is given by

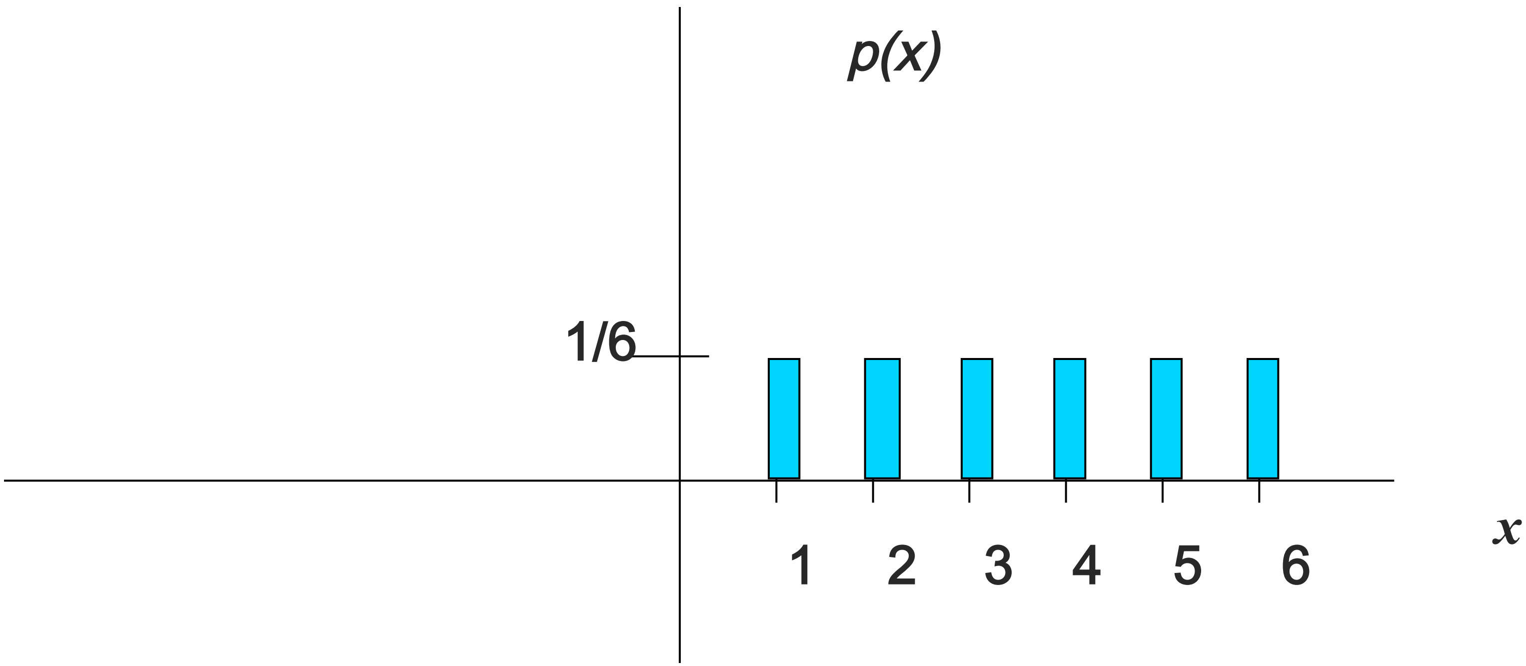
$$Var\left[X\right] = E\left[(X-\mu)^2\right] = E\left[X^2\right] - \left(E\left[X\right]\right)^2$$

- The variance is often also denoted as σ^2, where is σ is defined as the **standard deviation** of the random variable *X*

$$\sigma = \sqrt{Var\left[X\right]}$$

- How can you relate the concepts of accuracy and precision to the measure of variance (see slides from last week)?
- What about reproducibility

Discrete example: roll of a die



$$\sum_x P(x) = 1$$

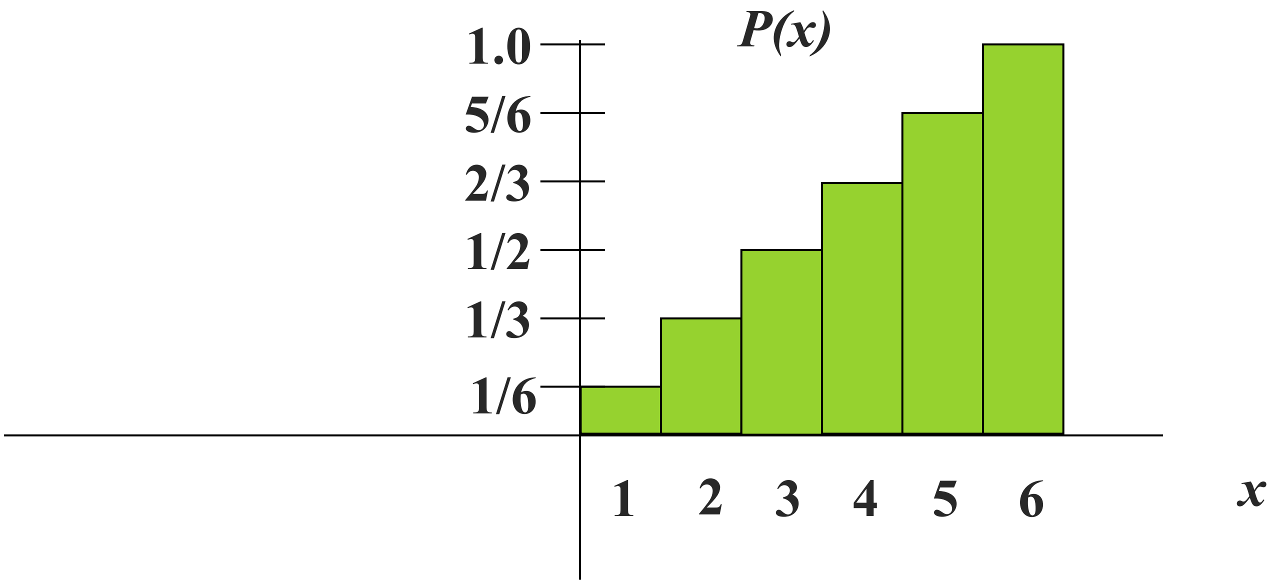
Discrete example: roll of a die

x	$p(x)$
1	$p(x = 1) = 1/6$
2	$p(x = 2) = 1/6$
3	$p(x = 3) = 1/6$
4	$p(x = 4) = 1/6$
5	$p(x = 5) = 1/6$
6	$p(x = 6) = 1/6$

$$p(x \leq 6) = 1$$

Cumulative distribution function (CDF)

…also called probability density function…



Cumulative distribution function

x	$P(x \leq A)$
1	$P(x \leq 1) = 1/6$
2	$P(x \leq 2) = 2/6$
3	$P(x \leq 3) = 3/6$
4	$P(x \leq 4) = 4/6$
5	$P(x \leq 5) = 5/6$
6	$P(x \leq 6) = 6/6$

Examples

1. What's the probability that you roll a 3 or less?

$$P(x \leq 3) = \frac{1}{2}$$

2. What's the probability that you roll a 5 or higher?

$$P(x \geq 5) = 1 - P(x \leq 4) = 1 - \frac{2}{3} = \frac{1}{3}$$

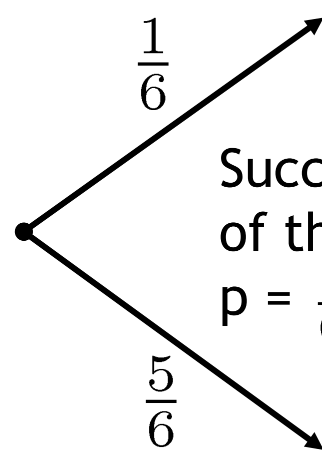
Important discrete distributions…

…for computational mass spectrometry..

- Binomial distribution
  - E.g., isotope distribution of a single atom and one additional isotope peak (lecture 2)
- Poisson distribution
  - Peptide identification, peptide quantification

Bernoulli experiment

- Jakob Bernoulli
- A Bernoulli experiment is a random experiment where the random variable can take only two values
  - success
  - 1-success (no success)

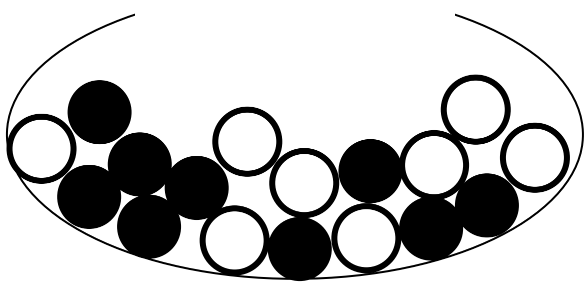


Success probability p
of the Bernoulli experiment:
p = $\frac{1}{6}$

- Example: role a die
  - 6: success
  - 1,2,3,4,5: no success

Binomial distribution

- Independent Bernoulli experiments build the basis for binomial distributions
  - Trials with two possible outcomes (e.g., flipping a coin)
  - n independent (repeated) trials are performed
  - p, the probability of success, is the same in every experiment



- N marbles in a jar
- r black and N-r white
- What is the probability to have k black marbles, if n are drawn with replacement ?

Important notations

- *x*: the number of successes that result from the binomial experiment
- *n*: the number of trials in the binomial experiment
- *p*: the probability of success in an individual trial
- *q*: the probability o failure (=(1-p)).
- B(x;n,p): Binomial probability – the probability than an *n*-trial binomial experiment results in exactly *x* successes with a success probability is *p*.
- $\binom{n}{r}$ "n choose r" the number of different ways to choose *r* things out of *n*.

Mini example

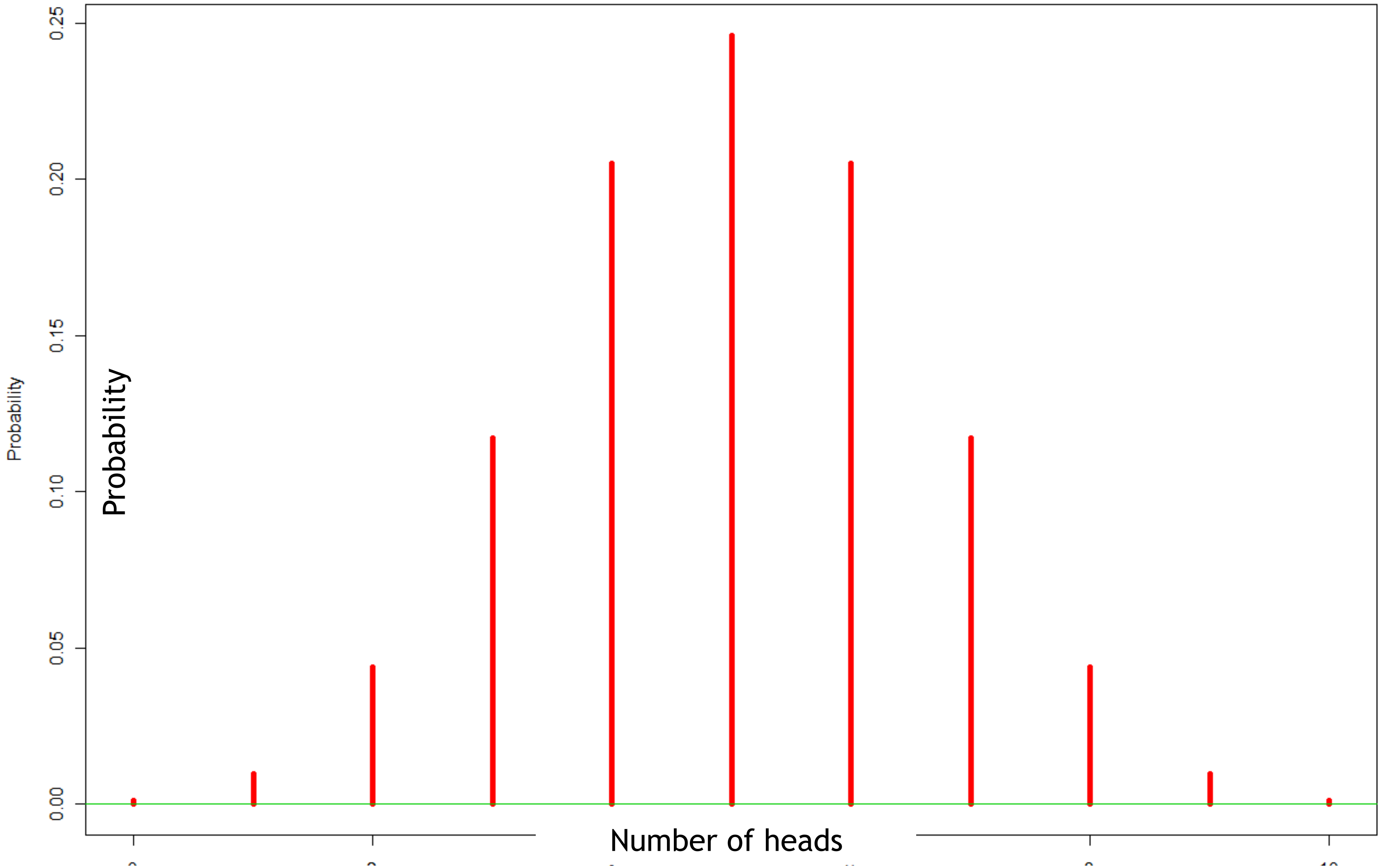
- Draw twice a single marble from a jar containing 10 black and 10 white marbles (with replacement)
- The probability of having k black marbles is:

$$B(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

# of black marbles	probability
0	0.25
1	0.5
2	0.25

- The mean of the prob. distribution is $\mu = n \cdot p$
- The variance σ^2 is $n \cdot (1-p) \cdot p$

Throwing a coin 10 times



Poisson approximation of the binomial distribution

Lets P(x=k) denote the binomial distribution and set p= λ/n. Then it holds in the limit:

$$\lim_{n\to\infty} P(X=k) = \lim_{n\to\infty} \frac{n!}{(n-k)!k!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$\Leftrightarrow \lim_{n\to\infty} P(X=k) = \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(\frac{n(n-1)(n-2)...(n-k+1)}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n}$$

$$\Leftrightarrow \lim_{n\to\infty} P(X=k) = \lim_{n\to\infty}\frac{\lambda^k}{k!}\left(\frac{n(n-1)(n-2)...(n-k+1)}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n}$$

$$\lim_{n\to\infty}\left(\frac{n(n-1)(n-2)...(n-k+1)}{n^k}\right)=1 \qquad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}=1 \qquad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=e^{-\lambda}$$

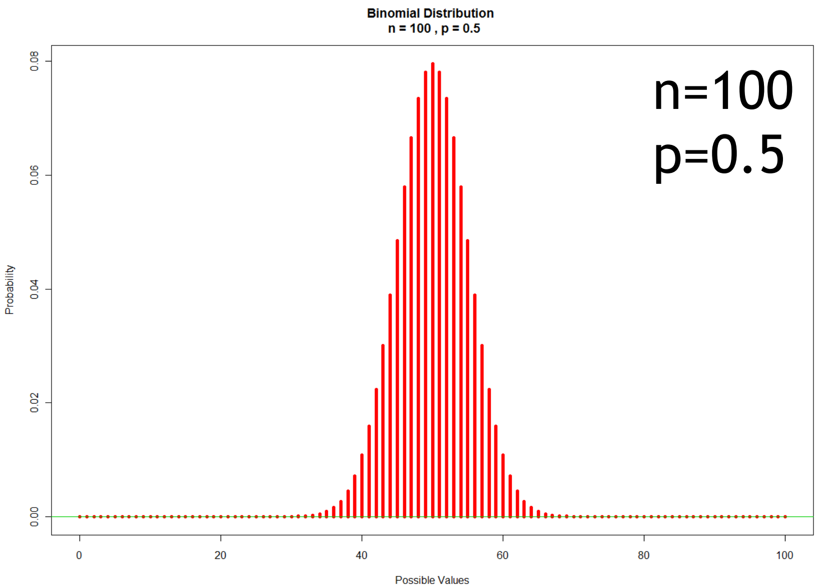
Binomial approximation of the Poisson distribution

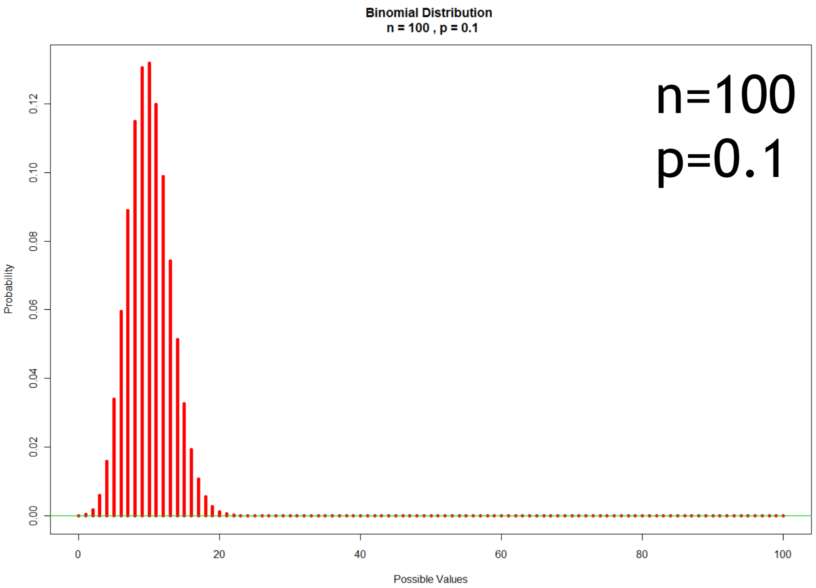
And we end with: $\lim_{n\to\infty} P(X=k) = \left(\frac{\lambda^k e^{-\lambda}}{k!}\right)$

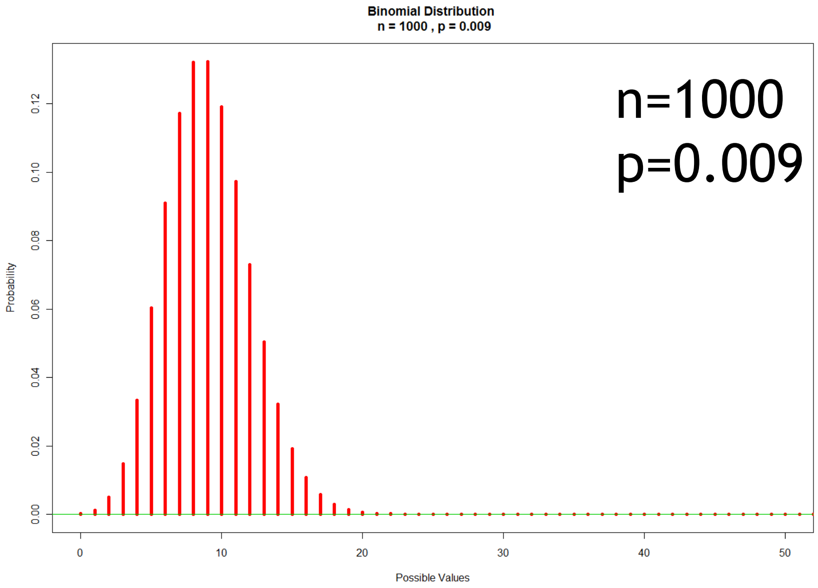
This is the Poisson distribution function. The Poisson distribution approximates a Bernoulli experiment with a high number of repeats and low success probability. Therefore it is also called **Poisson law of small numbers.**

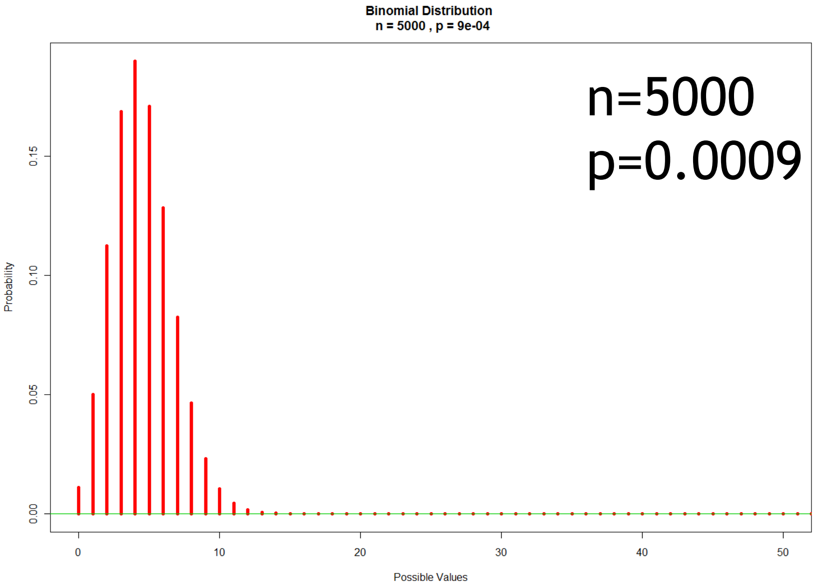
- The mean of a Poisson distribution is λ
- The standard deviation is given by $\sqrt{\lambda}$

Binomial distributions

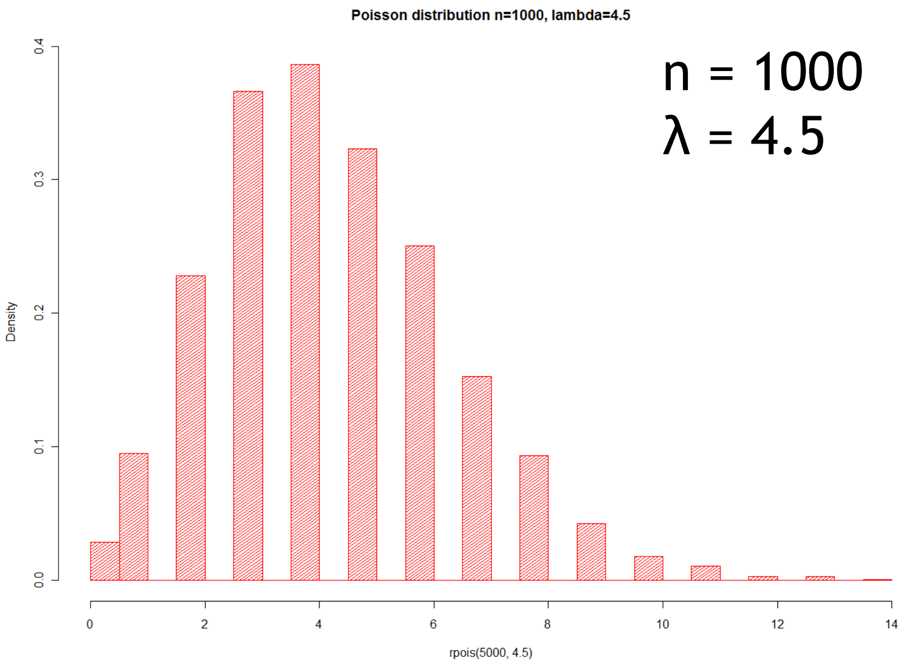


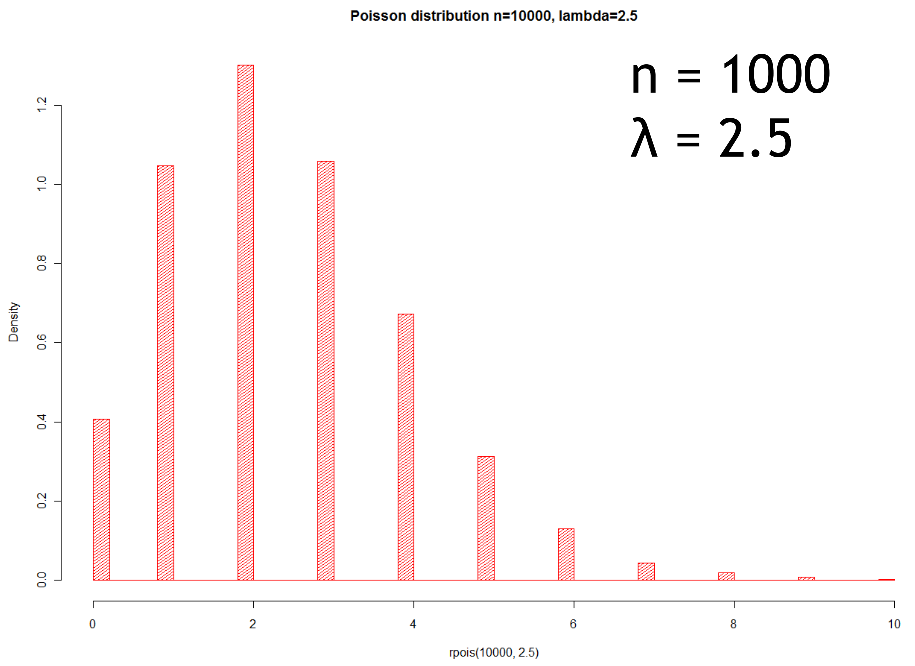






Poisson distributions





Continuous random variables

- The probability function that accompanies a continuous random variable is a continuous mathematical function that integrates to 1.
- The probabilities associated with continuous functions are just areas under the curve (integrals!).

Important continuous distribution:

Gaussian distribution

The Gaussian Distribution

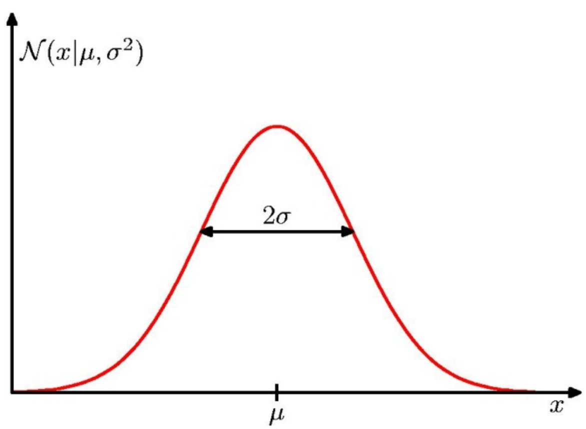
- The probability function is given by

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

- Per definition we have

$$N(x,\mu,\sigma^2) \geq 0 \text{ and } \int_{-\infty}^{\infty} N(x,\mu,\sigma^2)dx = 1$$

- The probability function results in the well-known bell-shape projection



Gaussian Mean and Variance

- The expectation value is calculated as follows,

$$E\left[x\right] = \int_{-\infty}^{\infty} \mathcal{N}\left(x, \mu, \sigma^2\right) x dx = \mu$$

- Furthermore,

$$E\left[x^2\right] = \int_{-\infty}^{\infty} \mathcal{N}\left(x, \mu, \sigma^2\right) x^2 dx = \mu^2 + \sigma^2$$

- Resulting in the general variance of Gaussian distributions

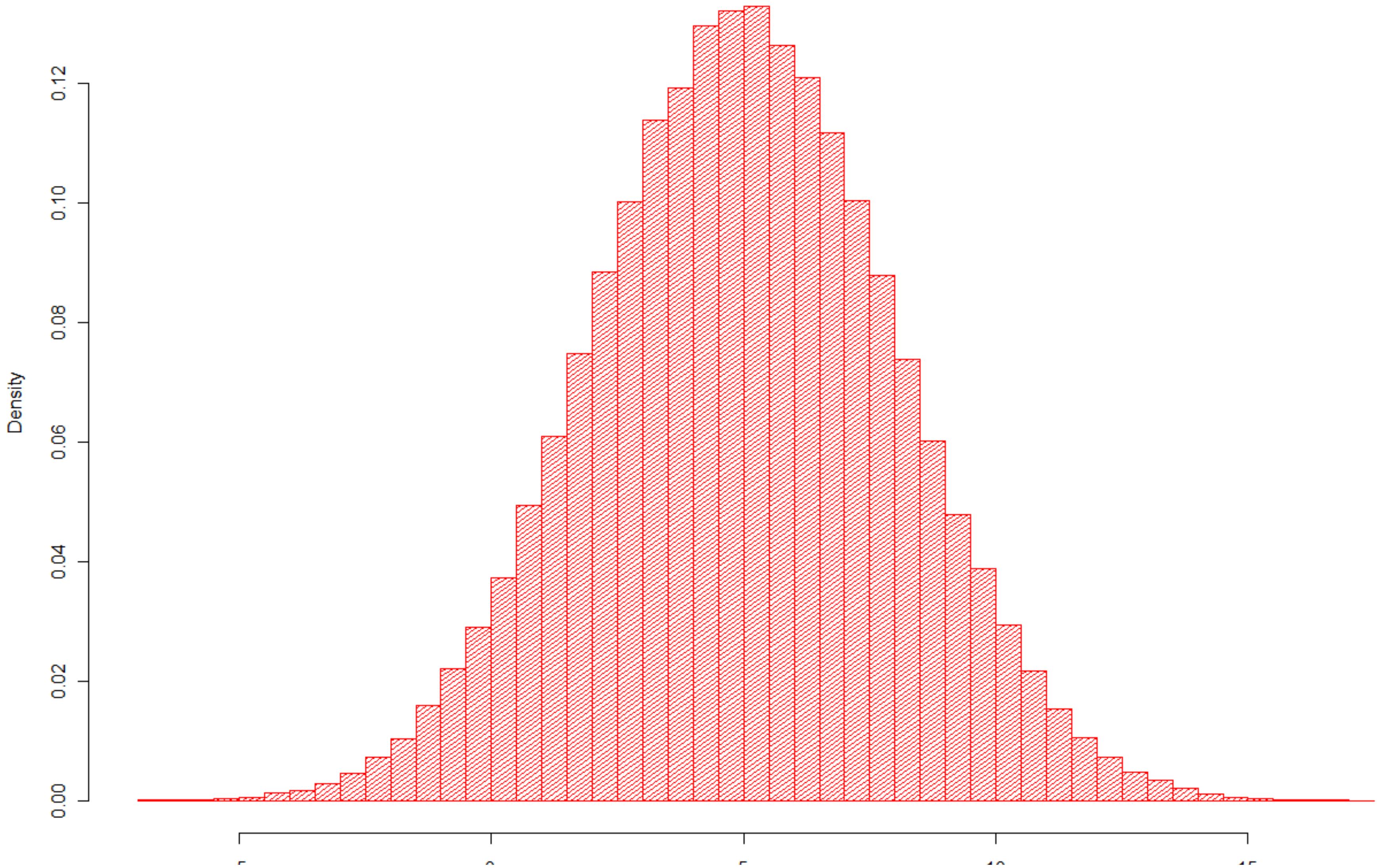
$$Var\left[x\right] = E\left[x^2\right] - E\left[x\right]^2 = \sigma^2$$

Standard normal distribution

- The standard normal distribution corresponds to the general form of the Gaussian distribution with $\mu=0$ and $\sigma^2=1$
- An arbitrary normal distribution can be converted to a standard normal distribution via *Z-transformation*
  $Z(X)=X-\mu/\sigma$
  resulting in

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

Gaussian distribution



Error function

A Gaussian distribution can also be estimated with an error distribution:

Given a real number $r \in \mathbb{R}$

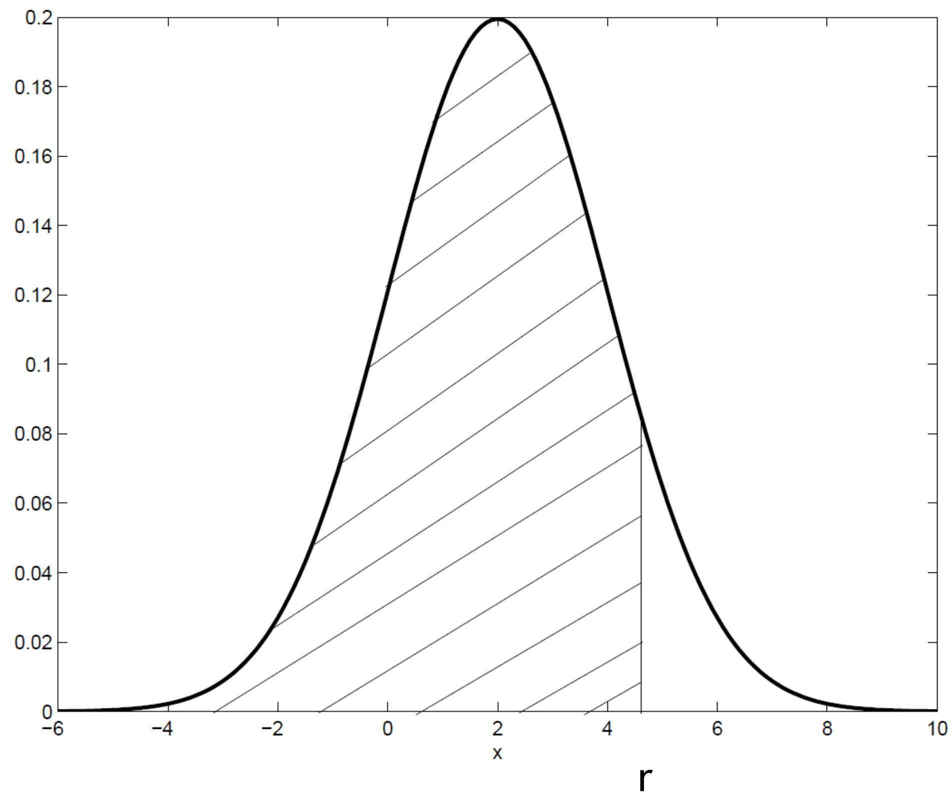
The probability that a random variable $X \sim N(\mu, \sigma^2)$ takes values is given by

$$P\{X \leq r\} = \int_{-\infty}^{r} f(x)dx = \int_{-\infty}^{r} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Error function

The Gaussian error function is denoted as

$$erf(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{\frac{-y^2}{2}} dy$$



With the Gaussian error function $P\{X \leq r\}$ can be expressed as

$$P(X \leq r) = \begin{cases} 0.5 - erf(\frac{\mu - r}{\sigma})), & \text{for } r \leq \mu \\ 0.5 + erf(\frac{r-\mu}{\sigma})), & \text{for } r \geq \mu \end{cases}$$

This allows the evaluation of the probability that a random variable *Y* lies in an interval around the mean value *μ*

Error function

The probability that a Gaussian random variable lies in the interval $[\mu - 2\sigma, \mu + 2\sigma]$ is equal to 0.95452.

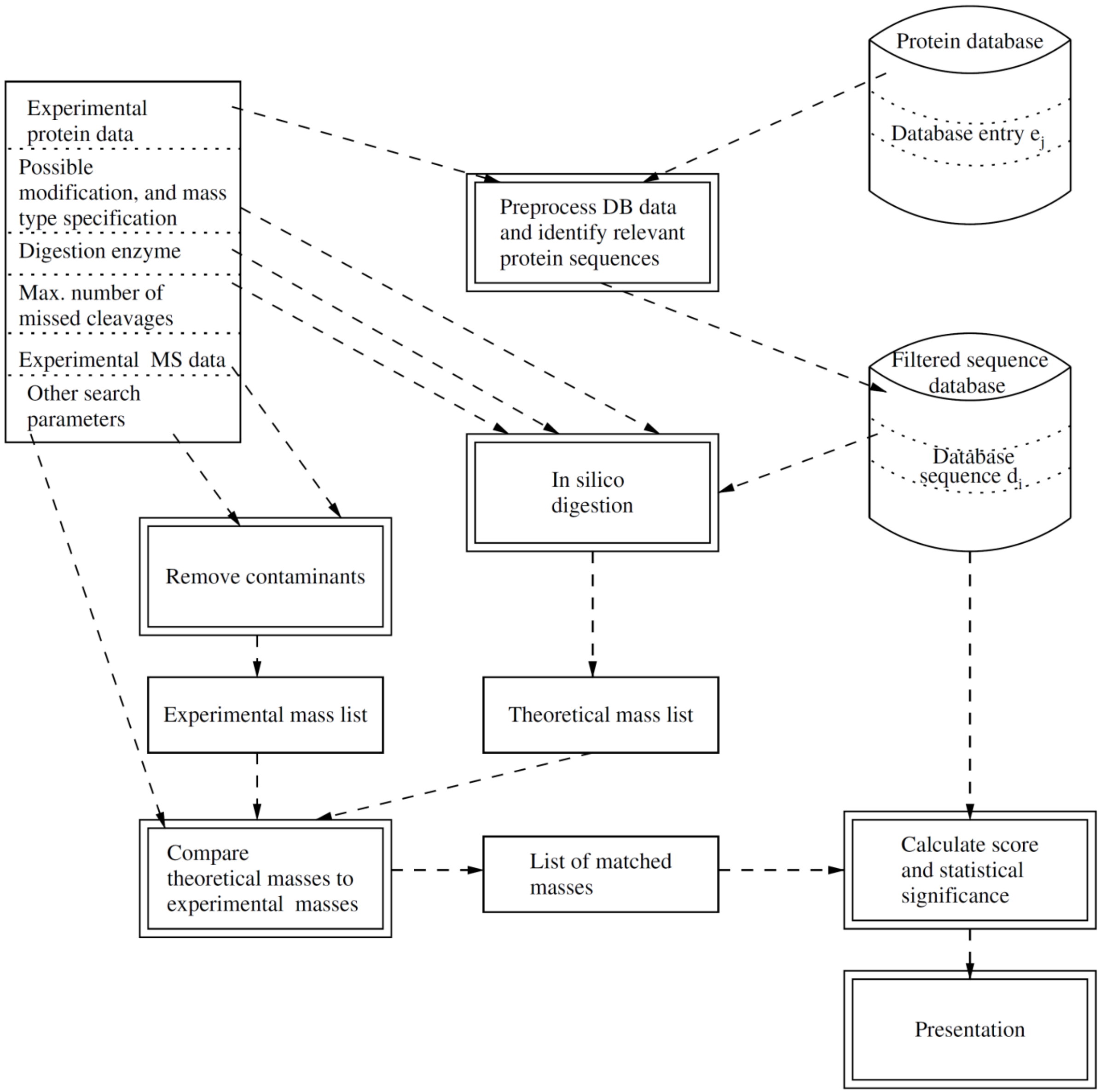
$$erf(2) = \frac{1}{\sqrt{2\pi}} \int_0^2 e^{\frac{-y^2}{2}} dy = 0.47726$$

$$P(|X - \mu| \leq 2\sigma) = 2erf(2)$$

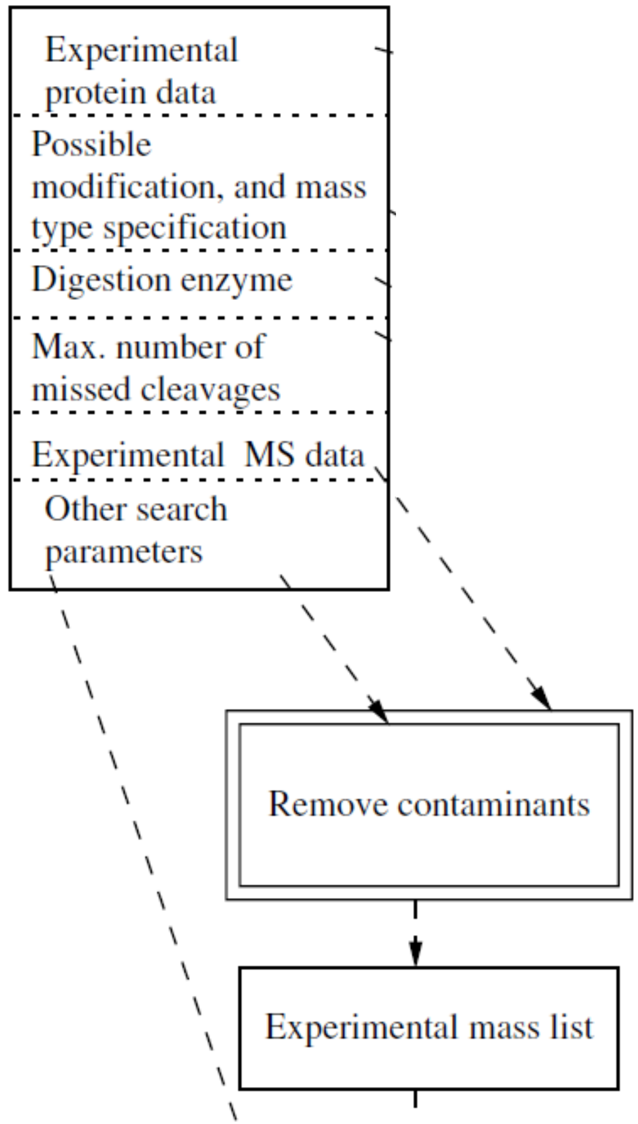
Applications

- Probabilities in proteomics
  - Post-search-processing of peptide identification results

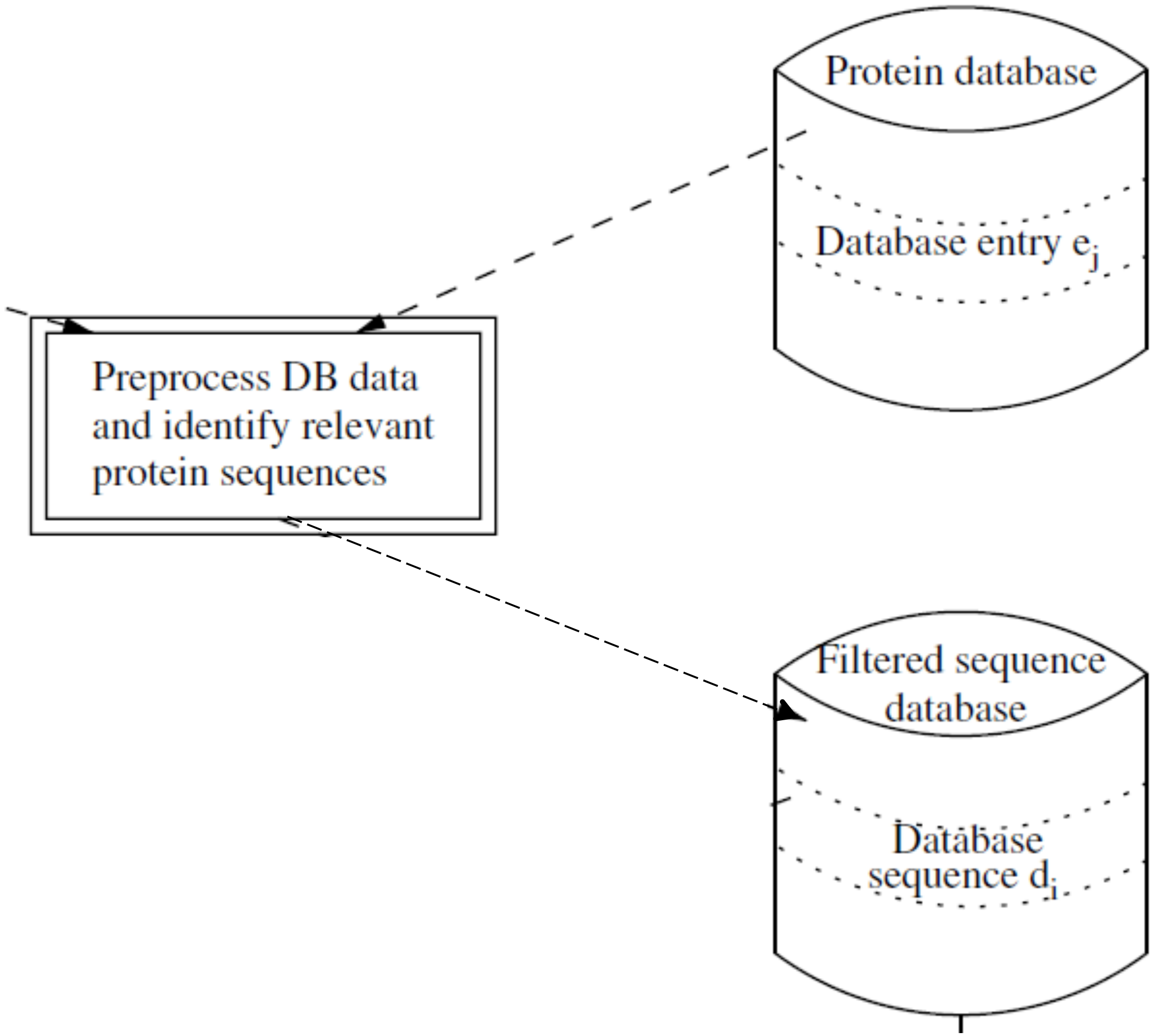
Identification workflow



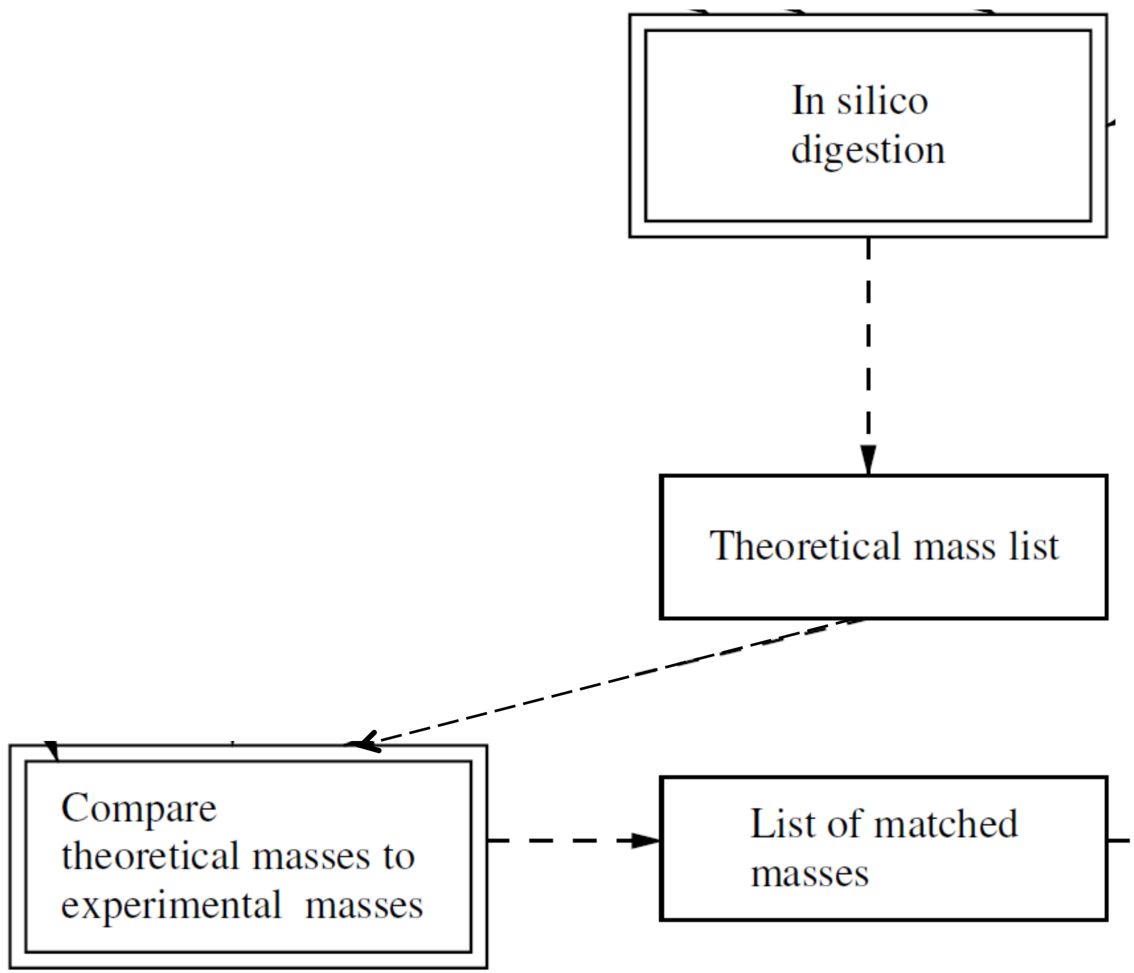
Experimental parameters



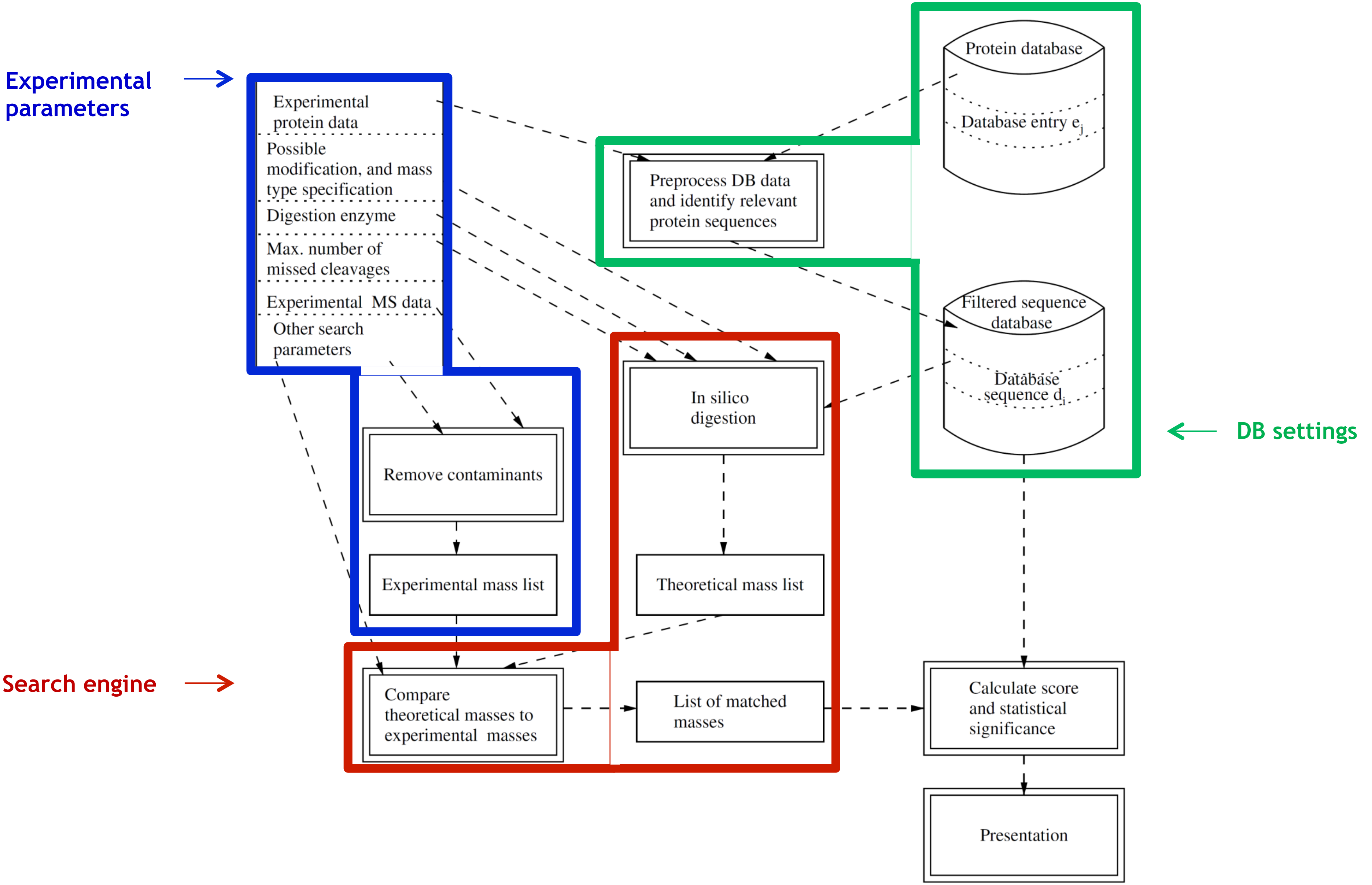
Database settings



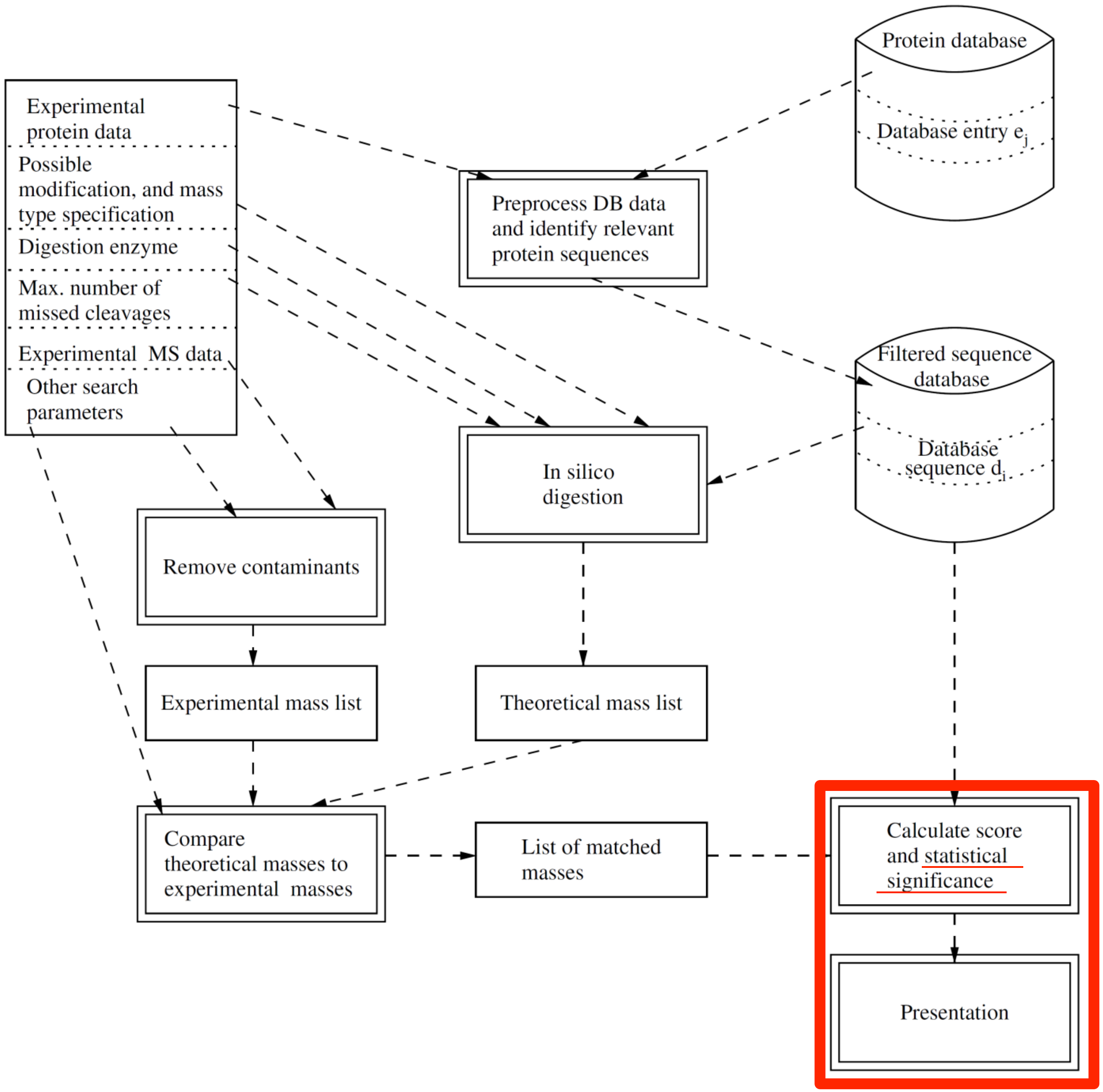
Search engine



Identification workflow



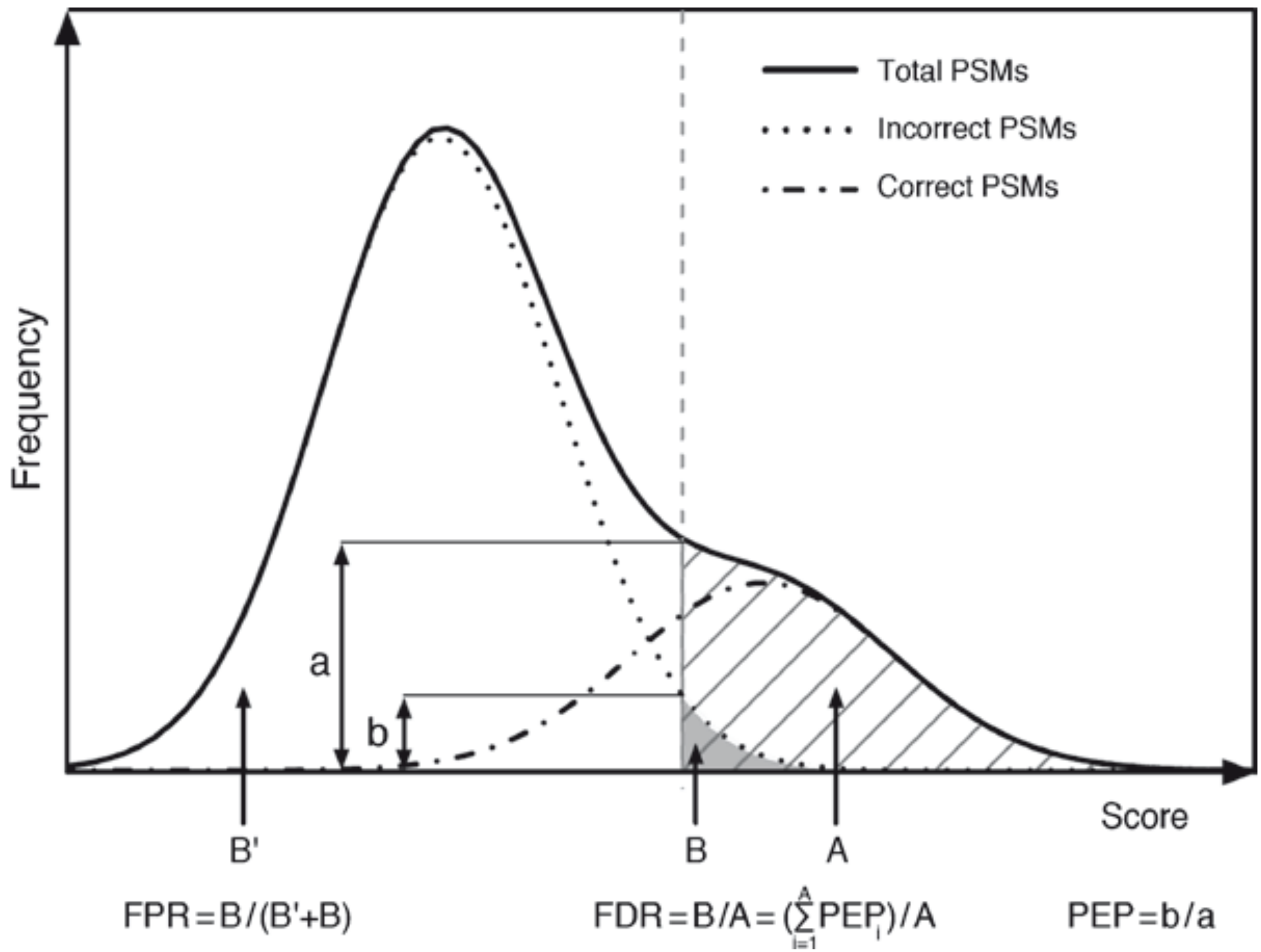
Identification workflow



Use statistics to assess significance of IDs

Peptide spectrum matchings (PSMs)

- Search engines assign scores to each peptide sequence that matches the theoretical spectrum
  - Most common search engines:
    - Sequest, Mascot, OMSSA, X!Tandem
- These peptide spectrum matchings are called *PSMs*
- All peptide candidates are ranked according to their PSM score
- Usually the top hit is reported
- However not all top scoring peptide identifications are correct (e.g., if the correct sequence is not in the DB, there might still be a PSM which is wrong)



- Brosch and Choudhary, Methods in Molecular Biology, 2010, Volume 604, 43-5

Problems with judging PSMs

- Heuristic score cut-offs are used
- Low score thresholds will accept more PSMs, but at the cost of more false positives (FP)
- High score thresholds reduce the error rate, but decrease identification rates as well
- The main problem is that the actual error remains unknown
- If heuristic methods are used, the results between two different approaches can vary by as much as 50 % (using the very same data set)

From PSMs to meaningful values

- p-values
- False discovery rates and q-values
- Posterior error probabilities

p-values

- Widely used statistical significance measure
- p-values in the database search context:=
  The probability of observing an incorrect PSM with a given score or better
- Hence, a low p-value indicates a low probability that the observed PSM is incorrect
- The p-value can be derived from the false positive rate (FPR), the fraction of incorrect PSMs above a certain score threshold over all PSMs
- Problems associated with p-value calculations
  - The FPR is usually unknown
  - p-values should be corrected for multiple hypothesis testing

- Note: There are also scoring algorithms that directly calculate p-values based on the theoretical and experimental spectrum comparison, but this works only for very simple scoring schemes and for rather small datasets

p-values in statistical testing

- p-values are used to judge the significance of a test for the null-hypothesis
- Null-hypothesis:= corresponds to the default position, e.g., *random chance peptide identification* or mean values of two independent measurements are *not* different
- Alternative-hypothesis:=the opposite positions, e.g., *non random peptide identification*
- Usually, the null hypothesis can not be formally proven, but statistical testing can accept or reject the null-hypothesis
- The null-hypothesis is rejected if the p-value is less then a significance level α (e.g., 0.05 or 0.01)

p-value example

- Given 10^4 PSMs with p-value cut-off of 0.05
- We then expect $0.05\times 104=500$ incorrect PSMs simply by chance
- Needs to be corrected for multiple hypothesis testing (10^4 tests are performed)
- Bonferroni correction would lead to $p-value/\# tests=0.05/104=0.000005$…new p-value cut-off
- Very stringent!
- Another way to account for multiple hypothesis testing: False discovery rates

False discovery rates (FDRs)

- Another approach to control for multiple hypothesis
- FDR:= expected proportion of incorrect predictions amongst a selected set of predictions
- For our MS problem this can be interpreted as a fraction of incorrect PSMs within a selected set of PSMs above a certain score threshold

False discovery rates (FDRs)

Peptide identification	Search engine score	True/false
LCEVEEGDKEDVDK	s_1	T
YTAQVDAEEKEDVK	s_2	T
IVADKDYSVTANSK	s_3	T
TGIEIIKK	s_4	T
DLGEEHFK	s_5	T
TASSDTSEELNSQDSPK	s_6	F
GAGGENEPPAAAPEPR	s_7	T
IKDPDAAKPEDWDDR	s_8	T
VDEVGGEALGR	s_9	T
SEEQLKEEGIEYK	s_{10}	F
LHVDPENFK	s_{11}	T
FSTVAGESGSADTVRDPR	s_{12}	T
AEEDEILNR	s_{13}	F

- The FDR of the entire list is calculated as

$$FDR = \frac{FP}{FP+TP}$$

- Here: 3/13: 10 PSM are considered identified at a FDR of 23 %

q-values

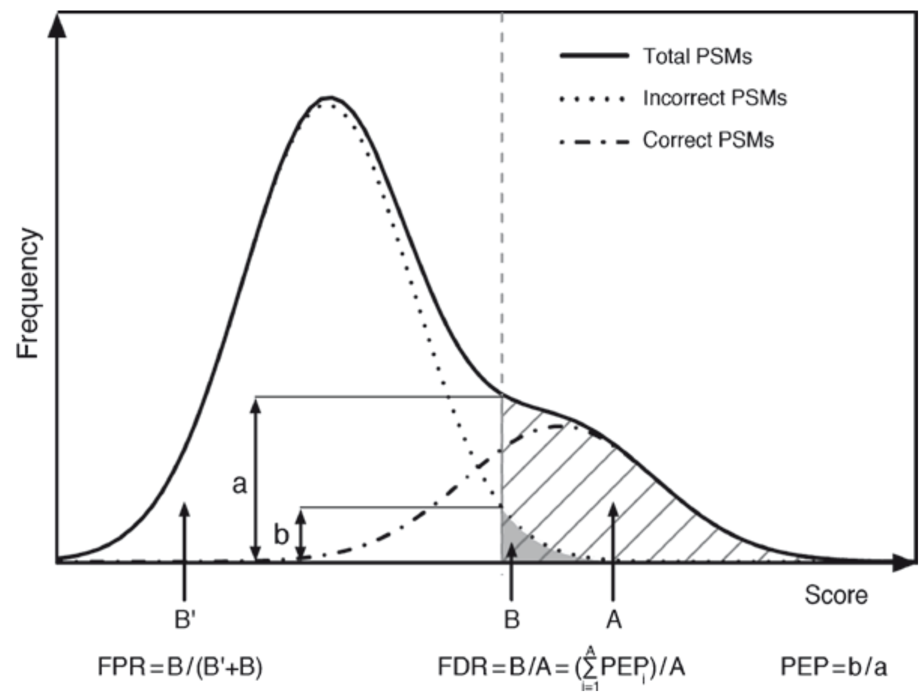
Peptide identification	q-value score	True/false
LCEVEEGDKEDVDK	0.00	T
YTAQVDAEEKEDVK	0.00	T
IVADKDYSVTANSK	0.00	T
TGIEIIKK	0.00	T
DLGEEHFK	0.00	T
TASSDTSEELNSQDSPK	-	F
GAGGENEPPAAAPEPR	0.11	T
IKDPDAAKPEDWDDR	0.11	T
VDEVGGEALGR	0.11	T
SEEQLKEEGIEYK	-	F
LHVDPENFK	0.16	T
FSTVAGESGSADTVRDPR	0.16	T
AEEDEILNR	-	F

- The q-value can be understood as the minimal FDR level at which a PSM can be accepted

Posterior error probabilities

- Assuming a bimodal distribution; this can also be considered as two distinct distributions
- One distribution describing the incorrect peptide assignments
- One distributions describing the correct peptide assignments
- The **posterior error probability** denotes the probability that a given peptide assignment score lies in the first distribution

  The probability of being incorrect



- **PEPs** can be inferred via mixture modeling and the expectation-maximization algorithm

What is false?

- A general problem for any statistical assessment is the missing knowledge on what is false and what is true
- All presented methods need to make assumptions on false positive assignments

- Target-decoy database searches
- Mixture modeling and expectation-maximization algorithm

Mixture model

A *statistical law* explains a phenomenon in terms of the probability of occurrence of its underlying relationships.

A *k*-component mixture model is a weighted sum of laws, the likelihood of a sample x being given by

$$f(x) = \sum_{i=1}^{K} \pi_i f_i(x)$$

With the constraint that:

$$\sum_{i=1}^{K} \pi_i = 1$$

If the laws f_i are probability distributions *f* is also a probability distribution (a mixture of probability distributions)

Joint density

- Consider the joint function $f(x, y)$ with,

$$f(x, y) \geq 0 \ \forall x, y \text{ and } x, y \in \left]-\infty, \infty\right[$$

- If $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \ dx \ dy = 1$

- Then, $f(x, y)$ is called a joint density function over *x* and *y*

Marginal density

- Consider the joint density $f(x,y)$, with

$$P(a \le x \le b \wedge c \le y \le d) = \int_a^b \int_c^d f(x,y)\ dx\ dy$$

- To calculate the probability for $a \le x \le b$ we need to look at

$$P(a \le x \le b \wedge -\infty \le y \le \infty) = \int_a^b \int_{-\infty}^{\infty} f(x,y)\ dx\ dy$$

- Furthermore, we define

$$f_x(x) = \int_{-\infty}^{\infty} f(x,y)\ dy$$

- With this, we have

$$P(a \le x \le b \wedge -\infty \le y \le \infty) = \int_a^b f_x(x)\ dx$$

- And $f_x(x)$ is called the marginal density function of the random variable *x*

Conditional density

- The conditional density of a random variable y for known occurrences of $x \in X$ is defined as follows,

$$f(y|x = X) = \frac{f(x,y)}{f_x(x)}$$

- Where $f(x, y)$ is the joint distribution of *x* and *y* and $f_x(x)$ is the marginal distribution of *x*

- The conditional mean is then given as

$$E(y|x = X) = \sum_{y \in Y} y f(y|x = X)$$

Mixture model

Mixture modeling and Expectation-Maximization (EM) algorithm

EM slides are based on a lecture given by Jörg Rahnenführer at TU Dortmund in 2007

The EM Algorithm

A. Dempster et al., Maximum likelihood from incomplete data via the EM algorithm (with discussion), J. R. Statist. Soc. B. 39 (1977) 1-38. Also Hastie, Tibshirani, Friedman, pages 238ff)

- Two-component mixture model
- Example: 20 data points
- Distribution apparently bi-modal
- Fit two Gaussians

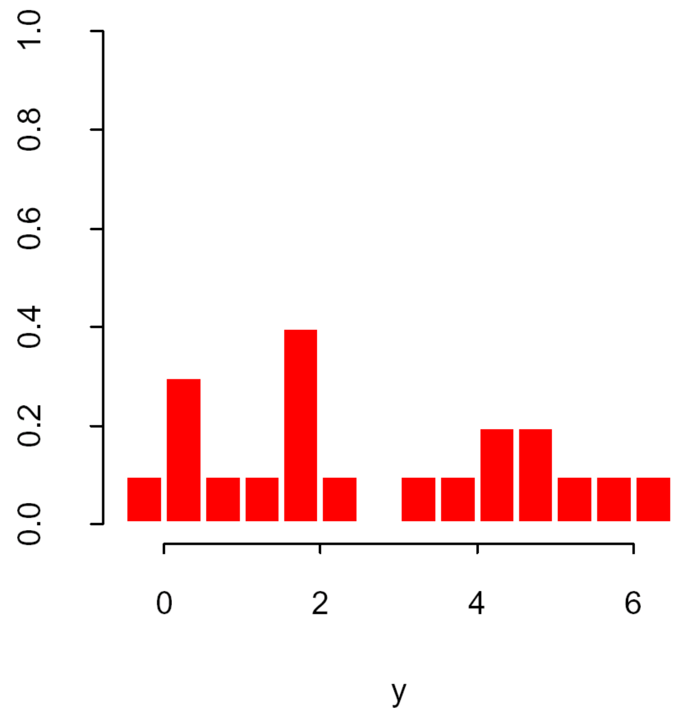
$$Y_1 \sim N(\mu_1, \sigma_1)$$

$$Y_2 \sim N(\mu_2, \sigma_2)$$

$$Y = (1 - \Delta) Y_1 + \Delta Y_2$$

$$\Delta \in \{0, 1\}$$

$$P(\Delta = 1) = \pi$$



20 data points

-0.39	0.12	0.94	1.67	1.76	2.44	3.72	4.28	4.92	5.53
0.06	0.48	1.01	1.68	1.80	3.25	4.12	4.60	5.28	6.22

The EM Algorithm

- Two-component mixture model
- Example: 20 data points
- Distribution apparently bi-modal
- Fit two Gaussians

$$Y_1 \sim N(\mu_1, \sigma_1)$$

$$Y_2 \sim N(\mu_2, \sigma_2)$$

$$Y = (1-\Delta)Y_1 + \Delta Y_2$$

$$\Delta \in \{0, 1\}$$

$$P(\Delta = 1) = \pi$$

- This is a generative representation
- Let $\phi_\theta(x) = N(\mu, \sigma^2), \quad \theta = (\mu, \sigma^2)$
- Then the density of Y is

$$g_Y(y) = (1-\pi)\phi_{\theta_1}(y) + \pi\phi_{\theta_2}(y)$$

- Fit with max-likelihood
- Parameters

$$\theta = (\pi, \theta_1, \theta_2) = (\pi, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$$

- Log-likelihood function based on N training cases

$$l(\theta; \mathbf{Z}) = \sum_{i=1}^{N} \log[(1-\pi)\phi_{\theta_1}(y_i) + \pi\phi_{\theta_2}(y_i)]$$

The EM Algorithm

- Direct optimization is difficult for the sum under the log
- Thus, let us assume that we know the Δ_i for all training inputs
- Joint density is $\phi(\Delta, y) = [(1-\Delta)\phi_{\theta_1}(y)]\ [\Delta\phi_{\theta_2}(y)]$

- The log-likelihood for the complete data is

$$\ell_0(\theta; \Delta, y) = \sum_{i=1}^{N} [(1-\Delta_i)\log\phi_{\theta_1}(y_i) + \Delta_i \log\phi_{\theta_2}(y_i)]$$

- Max-likelihood estimates are the sample mean and standard deviation of the respective subclasses of the training data for $\Delta_i = 0,1$

The EM Algorithm

- Since the Δ_i are actually unknown we proceed iteratively
- Step 1 (Expectation): Substitute for each Δ_i its expected value (responsibility of model 2 for observation i) as derived from the present model.

$$\gamma_i(\theta) = \mathrm{E}(\Delta_i \mid \theta, \mathbf{Z}) = \Pr(\Delta_i = 1 \mid \theta, \mathbf{Z})$$

  This is done by computing the relative densities of the training points under each model.
- Step 2 (Maximization): Compute new max-likelihood parameters

The EM Algorithm

- **The EM algorithm for two-component Gaussian mixtures**

1. **Take initial guesses** $\hat{\pi}, \hat{\mu}_1, \hat{\sigma}_1^2, \hat{\mu}_2, \hat{\sigma}_2^2$ **for the parameters**
2. **Expectation Step: Compute the responsibilities**

$$\hat{\gamma}_i = \frac{\hat{\pi}\phi_{\hat{\theta}_2}(y_i)}{(1-\hat{\pi})\phi_{\hat{\theta}_1}(y_i) + \hat{\pi}\phi_{\hat{\theta}_2}(y_i)}, \quad i = 1, \ldots, N$$

3. **Maximization Step: Compute the weighted means and variances**

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{N}(1-\hat{\gamma}_i)y_i}{\sum_{i=1}^{N}(1-\hat{\gamma}_i)}, \qquad \hat{\sigma}_1^2 = \frac{\sum_{i=1}^{N}(1-\hat{\gamma}_i)(y_i-\hat{\mu}_1)^2}{\sum_{i=1}^{N}(1-\hat{\gamma}_i)},$$

$$\hat{\mu}_2 = \frac{\sum_{i=1}^{N}\hat{\gamma}_i y_i}{\sum_{i=1}^{N}\hat{\gamma}_i}, \qquad \hat{\sigma}_2^2 = \frac{\sum_{i=1}^{N}\hat{\gamma}_i(y_i-\hat{\mu}_1)^2}{\sum_{i=1}^{N}\hat{\gamma}_i},$$

$$\hat{\pi} = \sum_{i=1}^{N}\hat{\gamma}_i / N$$

4. **Iterate 2 and 3 until convergence**

The EM Algorithm

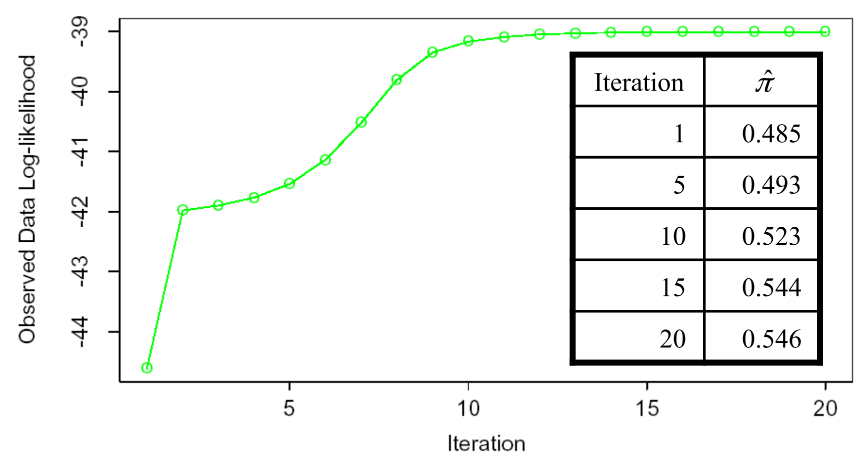
- How to choose the start values?
- For $\hat{\mu}_1$ and $\hat{\mu}_2$ choose two of the y_i at random. Set both $\hat{\sigma}_1$ and $\hat{\sigma}_2$ to the overall sample variance. Set $\hat{\pi} = 0.5$
- Global maxima of the log-likelihood function

$$\hat{\mu}_1 = y_i, \quad \text{for any } i \in (1, \ldots, n)$$

$$\hat{\sigma}_1^2 = 0$$

- Makes the log-likelihood function infinite
- Not a useful maximum

- Thus, we are looking for local maxima, for which $\hat{\sigma}_1, \hat{\sigma}_2 > 0$
- There can be many such maxima $\hat{\sigma}_1, \hat{\sigma}_2 > 0$
- Thus start with many random start solutions with $\hat{\sigma}_1^2, \hat{\sigma}_2^2 > 0.5$ and pick the outcome with the largest log-likelihood value

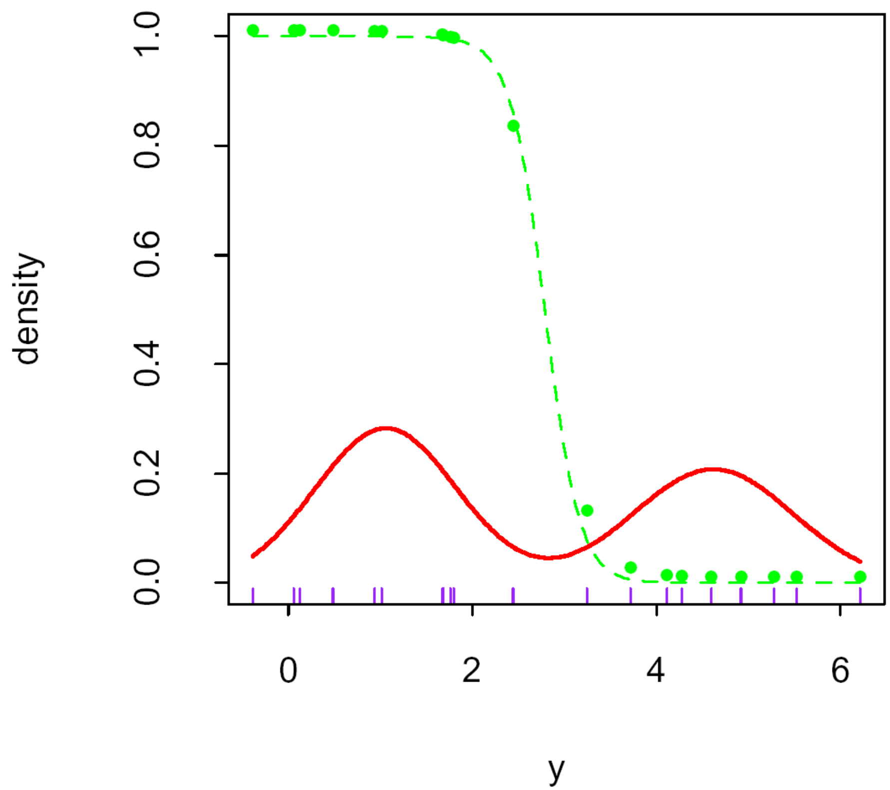


The EM Algorithm

- Final estimates

$$\hat{\mu}_1 = 4.62, \qquad \hat{\sigma}_1^2 = 0.87$$
$$\hat{\mu}_2 = 1.06, \qquad \hat{\sigma}_2^2 = 0.77$$
$$\hat{\pi} = 0.546$$

Gaussian mixture density
Responsibility for left class
(**Posterior error probabilities**)



Sources

- Eidhammer et al., Computational Methods for Mass Spectrometry Proteomics. Wiley. 2007.
- Kristin L. Sainani, Stanford University
  - www.stanford.edu/~kcobb/hrp259/lecture4.ppt
  - www.stanford.edu/~kcobb/hrp259/lecture5.ppt
- Christopher M. Bishop, Pattern Recognition and Machine Learning. 2006
- Brosch and Choudhary, Scoring and Validation of Tandem MS Peptide Identification Methods. Methods in Molecular Biology, 2010, Volume 604, 43-5
- Elias et al., Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nature Methods, Vol.2, No.9, 2005
- http://www.colorado.edu/economics/morey/6818/jointdensity.pdf

Materials

- Learning Units 3A and 3B