Advanced statistics and modelling

6. week

Parameter inference in statistics
Hypothesis testing, MLE, Bootstrap

Calculating c and T in general

Recap: Calculate parameters for a model (3 approaches):

- assume normal distribution,
- assume all possible values to be observed
- assume the observation is the most probable realization

.

Calculating c and T in general

Recap: Calculate parameters for a model (3 approaches):

- assume normal distribution,
- assume all possible values to be observed
- assume the observation is the most probable realization

Approach No. 1:
The Wald test assumes normal distributions: $N(\mu, \sigma^2)$ needs the expected value and the variance of the quantity.

.

Calculating c and T in general

Recap: Calculate parameters for a model (3 approaches):

- assume normal distribution,
- assume all possible values to be observed
- assume the observation is the most probable realization

Approach No. 1:
The Wald test assumes normal distributions: $N(\mu, \sigma^2)$ needs the expected value and the variance of the quantity.

.

How to calculate $\overline{()}, \hat{se}$ and how to derive $T(x)$ functions in general?

Calculating c and T in general

Recap: Calculate parameters for a model (3 approaches):

- assume normal distribution,
- assume all possible values to be observed
- assume the observation is the most probable realization

Approach No. 1:
The Wald test assumes normal distributions: $N(\mu, \sigma^2)$ needs the expected value and the variance of the quantity.

.

How to calculate $\overline{()}, \hat{se}$ and how to derive $T(x)$ functions in general?
Three methods:

- Bootstrap – all observed
- Method of Moments – all observed weakly
- Maximum Likelihood Estimation – most probable observed

Bootstrap

Theory behind this method:

$$\mathbb{E}(Y) = \lim_{B\to\infty} \frac{1}{B} \sum_{i=1}^{B} Y_i$$

For each function h with a finite mean

$$\mathbb{E}(h(Y)) = \lim_{B\to\infty} \frac{1}{B} \sum_{i=1}^{B} h(Y_i) = \langle h \rangle_B$$

Here B is the number of independent measurements (sampling with replacement) on the original dataset.

.

Bootstrap

Theory behind this method:

$$\mathbb{E}(Y) = \lim_{B\to\infty} \frac{1}{B} \sum_{i=1}^{B} Y_i$$

For each function h with a finite mean

$$\mathbb{E}(h(Y)) = \lim_{B\to\infty} \frac{1}{B} \sum_{i=1}^{B} h(Y_i) = \langle h \rangle_B$$

Here B is the number of independent measurements (sampling with replacement) on the original dataset.

.

In case the function h has $k > 1$ arguments, we simply draw k data with replacement (due to independence) from the original dataset.

.

This delivers the expectation value as $\langle . \rangle_B$

The standard error is just another function: $se^2 = \left\langle (. - \langle . \rangle_B)^2 \right\rangle_B$

Bootstrap

Formal definition:
Bootstrap method is used for approximating the expected value and the standard error of any function from a measured dataset. E.g. if the $T(X)$ test statistics is approximated, then X is distributed according to a fixed, but unknown F distribution.
Steps for any $T(X)$ function:

- Draw k points from the measured dataset: this follows the F distribution, since the measured dataset has values according to F.
- Compute $T(X)$ where X is k dimensional vector.
- Repeat B times the above steps
- $E_{bootstrap} = \frac{1}{B}\sum_b^B T_b$
- $SE_{bootstrap} = \frac{1}{B}\sum_b^B (T_b - \frac{1}{B}\sum_r^B T_r)^2$

Note: Jackknife was a similar, replica based method.
Note2: if $T(X)$ is the test function for the mean,
then $T(X) = \sum_i^n X_i/n$ is the average, using all values of the sample.

Bootstrap for confidence intervals

Using bootstrap for confidence intervals $C(a, b)$:

Bootstrap for confidence intervals

Using bootstrap for confidence intervals $C(a, b)$:

- Normal method: $\hat{T} \pm z_{\alpha/2}\hat{se}_{bootstrap}$, so

  $a = \hat{T} - z_{\alpha/2}\hat{se}_{bootstrap}$, $b = \hat{T} + z_{\alpha/2}\hat{se}_{bootstrap}$

Bootstrap for confidence intervals

Using bootstrap for confidence intervals $C(a,b)$:

- Normal method: $\hat{T} \pm z_{\alpha/2}\hat{se}_{bootstrap}$, so

  $a = \hat{T} - z_{\alpha/2}\hat{se}_{bootstrap}$, $b = \hat{T} + z_{\alpha/2}\hat{se}_{bootstrap}$
- Pivotal method: $C = (a,b)$

  $a = \hat{T} + H^{-1}(1-\alpha/2)$ and $b = \hat{T} + H^{-1}(\alpha/2)$

  where $H(r) = \mathbb{P}_F(\hat{T} - T \leq r)$

Bootstrap for confidence intervals

Using bootstrap for confidence intervals $C(a, b)$:

- Normal method: $\hat{T} \pm z_{\alpha/2}\hat{se}_{bootstrap}$, so

  $a = \hat{T} - z_{\alpha/2}\hat{se}_{bootstrap}$, $b = \hat{T} + z_{\alpha/2}\hat{se}_{bootstrap}$
- Pivotal method: $C = (a, b)$

  $a = \hat{T} + H^{-1}(1 - \alpha/2)$ and $b = \hat{T} + H^{-1}(\alpha/2)$

  where $H(r) = \mathbb{P}_F(\hat{T} - T \leq r)$

- good approximation: $a = 2\hat{T} - T^*_{1-\alpha/2}$, $b = 2\hat{T} - T^*_{\alpha/2}$,

  where T^*_β is β-sample-quantile of $T_1, T_2, \ldots, T_B$ replicas.

Bootstrap for confidence intervals

Using bootstrap for confidence intervals $C(a, b)$:

- Normal method: $\hat{T} \pm z_{\alpha/2}\hat{se}_{bootstrap}$, so

  $a = \hat{T} - z_{\alpha/2}\hat{se}_{bootstrap}$, $b = \hat{T} + z_{\alpha/2}\hat{se}_{bootstrap}$
- Pivotal method: $C = (a, b)$

  $a = \hat{T} + H^{-1}(1 - \alpha/2)$ and $b = \hat{T} + H^{-1}(\alpha/2)$

  where $H(r) = \mathbb{P}_F(\hat{T} - T \leq r)$
  - good approximation: $a = 2\hat{T} - T^*_{1-\alpha/2}$, $b = 2\hat{T} - T^*_{\alpha/2}$,

    where T^*_β is β-sample-quantile of $T_1, T_2, \ldots, T_B$ replicas.
- Percentile interval: $(T^*_{\alpha/2}, T^*_{1-\alpha/2})$

Parameter inference

How can we find $T = T(t_1, t_2, \ldots)$ functions?

Parameter inference

How can we find $T = T(t_1, t_2, \ldots)$ functions?
E.g. by estimating the parameters $t_1, t_2, \ldots$..
Two methods:

- moments (MME)
- max. likelihood (MLE)

Method of moments (MME)

Given a measured dataset, any moments of the distribution can be approximated:

$$\alpha_k = \int x^k \, dF = \frac{1}{n} \sum x_i^k$$

Method of moments (MME)

Given a measured dataset, any moments of the distribution can be approximated:

$$\alpha_k = \int x^k \, dF = \frac{1}{n} \sum x_i^k$$

If the distribution F depends on the parameter t,
the left hand side provides a formula for t based on the m-th moment.

$$\alpha_m(t) = \frac{1}{n} \sum x_i^m$$

Method of moments (MME)

Given a measured dataset, any moments of the distribution can be approximated:

$$\alpha_k = \int x^k \, dF = \frac{1}{n} \sum x_i^k$$

If the distribution F depends on the parameter t, the left hand side provides a formula for t based on the m-th moment.

$$\alpha_m(t) = \frac{1}{n} \sum x_i^m$$

Expressing t at the left and substituting the data values at the right, one gets an approximation for the parameter t.

Method of moments (MME)

Given a measured dataset, any moments of the distribution can be approximated:

$$\alpha_k = \int x^k \, dF = \frac{1}{n} \sum x_i^k$$

If the distribution F depends on the parameter t,
the left hand side provides a formula for t based on the m-th moment.

$$\alpha_m(t) = \frac{1}{n} \sum x_i^m$$

Expressing t at the left and substituting the data values at the right, one gets an approximation for the parameter t.
In case of more parameters, use so many moments as many unknown parameters.

Method of moments (MME)

Given a measured dataset, any moments of the distribution can be approximated:

$$\alpha_k = \int x^k \, dF = \frac{1}{n} \sum x_i^k$$

If the distribution F depends on the parameter t,
the left hand side provides a formula for t based on the m-th moment.

$$\alpha_m(t) = \frac{1}{n} \sum x_i^m$$

Expressing t at the left and substituting the data values at the right, one gets an approximation for the parameter t.
In case of more parameters, use so many moments as many unknown parameters.
Some properties of the method:

- An estimate for $\hat{t}$ exists with $\mathbb{P} = 1$

Method of moments (MME)

Given a measured dataset, any moments of the distribution can be approximated:

$$\alpha_k = \int x^k \, dF = \frac{1}{n} \sum x_i^k$$

If the distribution F depends on the parameter t,
the left hand side provides a formula for t based on the m-th moment.

$$\alpha_m(t) = \frac{1}{n} \sum x_i^m$$

Expressing t at the left and substituting the data values at the right, one gets an approximation for the parameter t.
In case of more parameters, use so many moments as many unknown parameters.
Some properties of the method:

- An estimate for $\hat{t}$ exists with $\mathbb{P} = 1$
- The estimate is consistent: $\hat{t} \to t$

Method of moments (MME)

Given a measured dataset, any moments of the distribution can be approximated:

$$\alpha_k = \int x^k \, dF = \frac{1}{n} \sum x_i^k$$

If the distribution F depends on the parameter t,
the left hand side provides a formula for t based on the m-th moment.

$$\alpha_m(t) = \frac{1}{n} \sum x_i^m$$

Expressing t at the left and substituting the data values at the right, one gets an approximation for the parameter t.
In case of more parameters, use so many moments as many unknown parameters.
Some properties of the method:

- An estimate for $\hat{t}$ exists with $\mathbb{P} = 1$
- The estimate is consistent: $\hat{t} \to t$
- The estimate is asymptically normal: $\sqrt{n}(\hat{t} - t) \to N(0, \sigma^2)$ (allows Wald test)

Maximum Likelihood Estimate (MLE)

We measure n values: $X_1, X_2, \ldots, X_n$
These are IID random variables with PDF $f(X_i, \theta)$

- The **likelihood function** is $\mathcal{L}(\theta) = \prod_i f(X_i, \theta)$
- The **log-likelihood function** is $\ell(\theta) = \sum_i \log f(X_i, \theta)$
- The **Maximum Likelihood Estimator** is $\hat{\theta}$ that maximizes $\mathcal{L}$

Properties of MLE

- The $\mathcal{L}(\theta) \in [0, \infty)$: it is the join density of data, but not a probability: $\int d\theta \mathcal{L}(\theta) \neq 1$
- consistent: $\hat{\theta} \to \theta$
- equivariant:if $\hat{\theta}$ MLE of θ then $g(\hat{\theta})$ MLE of $g(\theta)$
- asymptotically normal:$\sqrt{n}(\hat{\theta} - \theta) \to N(0, \hat{se})$
- asymptotically optimal: MLE has smallest variance for large samples

Working with sheets

- Open weblink
  - Open file: sh1 in your browser
- create an empty sheet under your google drive
- Copy and paste data from first sheet to your sheet
- Now you are ready to work with your sheet: add some formulas, data etc

Note: if your browser is set to a default language other than English, you will see translated version of functions and menus of the instructing screen.
Do not worry, in most cases you can give function names in English in this case as well.

Method of moments, MME 1.

- Find the appropriate distribution of inter arrival times!
  - Which distribution?
  - Express the first moment with the parameter of the distribution!
- Check the fitted parameter: compare distribution from data and fitted PDF! (Use QQ-plot!)
- Calculate the residuals as well!

Note:

$$\begin{aligned} F(x) &= 1 - e^{-\lambda x} \\ f(x) &= \lambda e^{-\lambda x} \\ \langle x \rangle &= 1/\lambda \end{aligned}$$

Method of Momements, MME 2.

- Set up a dataset with uniformly distributed random numbers! You can choose to copy numbers from your "master" sheet, or from the "sh1" file (sheet: Fit_Uniform), or generate your self with the function "rand()".
- Try to estimate the parameters of the uniform distribution for your dataset!
  - Note: you have to find two parameters now, so you need to calculate two moments.

How are the parameters, the moments and the QQ-plot changing?

Method of Momements, MME 2.

- Set up a dataset with uniformly distributed random numbers! You can choose to copy numbers from your "master" sheet, or from the "sh1" file (sheet: Fit_Uniform), or generate your self with the function "rand()".
- Try to estimate the parameters of the uniform distribution for your dataset!
  - Note: you have to find two parameters now, so you need to calculate two moments.
  - Hint: $\langle Uniform(a,b) \rangle = \frac{a+b}{2}$ and $\sigma^2(Uniform(a,b)) = \frac{(b-a)^2}{12}$
- Prepare a QQ-plot for testing the fitted values!
  - Fit the best line on the QQ-plot!

How are the parameters, the moments and the QQ-plot changing?

Method of Momements, MME 2.

- Set up a dataset with uniformly distributed random numbers! You can choose to copy numbers from your "master" sheet, or from the "sh1" file (sheet: Fit_Uniform), or generate your self with the function "rand()".
- Try to estimate the parameters of the uniform distribution for your dataset!
  - Note: you have to find two parameters now, so you need to calculate two moments.
  - Hint: $\langle Uniform(a,b)\rangle = \frac{a+b}{2}$ and $\sigma^2(Uniform(a,b)) = \frac{(b-a)^2}{12}$
- Prepare a QQ-plot for testing the fitted values!
  - Fit the best line on the QQ-plot!
- Experiment with the parameters: try to
  - shift the raw dataset,

How are the parameters, the moments and the QQ-plot changing?

Method of Momements, MME 2.

- Set up a dataset with uniformly distributed random numbers! You can choose to copy numbers from your "master" sheet, or from the "sh1" file (sheet: Fit_Uniform), or generate your self with the function "rand()".
- Try to estimate the parameters of the uniform distribution for your dataset!
  - Note: you have to find two parameters now, so you need to calculate two moments.
  - Hint: $\langle Uniform(a,b)\rangle = \frac{a+b}{2}$ and $\sigma^2(Uniform(a,b)) = \frac{(b-a)^2}{12}$
- Prepare a QQ-plot for testing the fitted values!
  - Fit the best line on the QQ-plot!
- Experiment with the parameters: try to
  - shift the raw dataset,
  - rescale the dataset,

How are the parameters, the moments and the QQ-plot changing?

Method of Momements, MME 2.

- Set up a dataset with uniformly distributed random numbers! You can choose to copy numbers from your "master" sheet, or from the "sh1" file (sheet: Fit_Uniform), or generate your self with the function "rand()".
- Try to estimate the parameters of the uniform distribution for your dataset!
  - Note: you have to find two parameters now, so you need to calculate two moments.
  - Hint: $\langle Uniform(a,b)\rangle = \frac{a+b}{2}$ and $\sigma^2(Uniform(a,b)) = \frac{(b-a)^2}{12}$
- Prepare a QQ-plot for testing the fitted values!
  - Fit the best line on the QQ-plot!
- Experiment with the parameters: try to
  - shift the raw dataset,
  - rescale the dataset,
  - insert outliers into the data

How are the parameters, the moments and the QQ-plot changing?

Method of Momements, MME 2.

- Try to estimate the parameters of the uniform distribution for your dataset!
  - Note: you have to find two parameters now, so you need to calculate two moments.
  - Hint: $\langle Uniform(a, b)\rangle = \frac{a+b}{2}$ and $\sigma^2(Uniform(a, b)) = \frac{(b-a)^2}{12}$
- Prepare a QQ-plot for testing the fitted values!
  - Fit the best line on the QQ-plot!
- Experiment with the parameters: try to
  - shift the raw dataset,
  - rescale the dataset,
  - insert outliers into the data

How are the parameters, the moments and the QQ-plot changing?
Note: in real scenarios you will meet similar biased or transformed data usually.

Method of Moments, MME 3

- We have a dataset with counts of incoming calls at the secretary of the dean for each working hours in a week. Assuming constant calling rate, try to fit an appropriate distribution for the number of calls!
  Which distribution would you try first?

Method of Moments, MME 3

- We have a dataset with counts of incoming calls at the secretary of the dean for each working hours in a week. Assuming constant calling rate, try to fit an appropriate distribution for the number of calls!
  Which distribution would you try first?
- Calls can be modeled by a stationary Poisson process. How many parameters do you need to fit?

Method of Moments, MME 3

- We have a dataset with counts of incoming calls at the secretary of the dean for each working hours in a week. Assuming constant calling rate, try to fit an appropriate distribution for the number of calls!
  Which distribution would you try first?
- Calls can be modeled by a stationary Poisson process. How many parameters do you need to fit?
- Prepare a QQ-plot for testing the fitted values! Take into account, that you work with a discrete distribution!

Poisson:

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Method of Moments, MME 3

- We have a dataset with counts of incoming calls at the secretary of the dean for each working hours in a week. Assuming constant calling rate, try to fit an appropriate distribution for the number of calls!
  Which distribution would you try first?
- Calls can be modeled by a stationary Poisson process. How many parameters do you need to fit?
- Prepare a QQ-plot for testing the fitted values! Take into account, that you work with a discrete distribution!
- Fit the best line on the QQ-plot! Check the slope and the abscissa of the line! Correspond these numbers to your expectations?

Poisson:

$$P(k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

Method of Moments, MME 3

- We have a dataset with counts of incoming calls at the secretary of the dean for each working hours in a week. Assuming constant calling rate, try to fit an appropriate distribution for the number of calls!
  Which distribution would you try first?
- Calls can be modeled by a stationary Poisson process. How many parameters do you need to fit?
- Prepare a QQ-plot for testing the fitted values! Take into account, that you work with a discrete distribution!
- Fit the best line on the QQ-plot! Check the slope and the abscissa of the line! Correspond these numbers to your expectations?
- Plot the residuals and plot the cumulative distribution functions! Can you retain the Poissonian assumption?

Poisson:

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Method of Moments, MME 3

- We have a dataset with counts of incoming calls at the secretary of the dean for each working hours in a week. Assuming constant calling rate, try to fit an appropriate distribution for the number of calls!
  Which distribution would you try first?
- Calls can be modeled by a stationary Poisson process. How many parameters do you need to fit?
- Prepare a QQ-plot for testing the fitted values! Take into account, that you work with a discrete distribution!
- Fit the best line on the QQ-plot! Check the slope and the abscissa of the line! Correspond these numbers to your expectations?
- Plot the residuals and plot the cumulative distribution functions! Can you retain the Poissonian assumption?

Conclusions:QQ-plot indicates, that this is the good distribution family. But further parameters are to be fitted: inhomogeneous Poisson process could be better.
Poisson:

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Maximum Likelihood Estimation, MLE

- Find the MLE for the uniform distribution using the data from the MME example!

Maximum Likelihood Estimation, MLE

- Find the MLE for the uniform distribution using the data from the MME example!
- Recall: PDF of uniform

$$f(x) = \frac{1}{b-a} \quad x \in [a, b]$$

- Can you solve it? How do you interpret the result?

Maximum Likelihood Estimation, MLE

- Find the MLE for the uniform distribution using the data from the MME example!
- Recall: PDF of uniform

$$f(x) = \frac{1}{b-a} \quad x \in [a,b]$$

- MLE:

$$\mathcal{L}(a,b) = \prod_i f(x_i; a, b) = \frac{1}{(b-a)^n}$$

where x_i are fixed numbers from the n data.

- Can you solve it? How do you interpret the result?

Maximum Likelihood Estimation, MLE

- Find the MLE for the uniform distribution using the data from the MME example!
- Recall: PDF of uniform

$$f(x) = \frac{1}{b-a} \quad x \in [a, b]$$

- MLE:

$$\mathcal{L}(a, b) = \prod_i f(x_i; a, b) = \frac{1}{(b-a)^n}$$

  where x_i are fixed numbers from the n data.
- $\mathcal{L}$ seems to be good behaving function, find the maximum with usual analysis:

$$\partial_a \mathcal{L} = 0$$

$$\partial_b \mathcal{L} = 0$$

- Can you solve it? How do you interpret the result?

Maximum Likelihood Estimation, MLE

- Because $f(x) = 0$ if $x \notin [a, b]$, we have

$$a \leq x_1 \leq x_2 \leq \ldots \leq x_n \leq b$$

where data (x_i) are in increasing order.

Maximum Likelihood Estimation, MLE

- Because $f(x) = 0$ if $x \notin [a, b]$, we have

$$a \leq x_1 \leq x_2 \leq \ldots \leq x_n \leq b$$

  where data (x_i) are in increasing order.
- $\mathcal{L}$ is not differentiable in x_1 and x_n!

Maximum Likelihood Estimation, MLE

- Because $f(x) = 0$ if $x \notin [a, b]$, we have

  $$a \leq x_1 \leq x_2 \leq \ldots \leq x_n \leq b$$

  where data (x_i) are in increasing order.
- $\mathcal{L}$ is not differentiable in x_1 and x_n!
- MLE:

  $$(a, b) = (min(x), max(x))$$

  But: this is a biased estimation.
  (Recall: what is a biased estimation?)

Maximum Likelihood Estimation, MLE

- Because $f(x) = 0$ if $x \notin [a, b]$, we have

$$a \leq x_1 \leq x_2 \leq \ldots \leq x_n \leq b$$

  where data (x_i) are in increasing order.
- $\mathcal{L}$ is not differentiable in x_1 and x_n!
- MLE:

$$(a, b) = (min(x), max(x))$$

  But: this is a biased estimation.
  (Recall: what is a biased estimation?)
- Try to correct the results to have an unbiased estimation with estimation values:

$$(a, b) = (\langle min(x) \rangle, \langle max(x) \rangle)$$

Maximum Likelihood Estimation, MLE

- Calculate $\langle max(x) \rangle$ from the CDF!

Maximum Likelihood Estimation, MLE

- Calculate $\langle max(x) \rangle$ from the CDF!
- Utilize the independence of data points!

Maximum Likelihood Estimation, MLE

- Calculate $\langle max(x) \rangle$ from the CDF!
  - Utilize the independence of data points!

$$
\begin{aligned}
F_{max}(y) &= \mathbb{P}((x_1 < y) \cap (x_2 < y) \cap \ldots \cap (x_n < y)) \\
. &= \prod_i F(y) = F^n(y) = \frac{(y-a)^n}{(b-a)^n}
\end{aligned}
$$

Maximum Likelihood Estimation, MLE

- Calculate $\langle max(x) \rangle$ from the CDF!
- Utilize the independence of data points!

$$\begin{aligned} F_{max}(y) &= \mathbb{P}((x_1 < y) \cap (x_2 < y) \cap \ldots \cap (x_n < y)) \\ . &= \prod_i F(y) = F^n(y) = \frac{(y-a)^n}{(b-a)^n} \end{aligned}$$

- Expectation value for max(x):

$$\langle max(x) \rangle = \int_a^b dy\ yn \frac{(y-a)^{n-1}}{(b-a)^n} = \frac{n}{n+1}(b-a) + a$$

Maximum Likelihood Estimation, MLE

- Calculate $\langle max(x) \rangle$ from the CDF!
- Utilize the independence of data points!

$$\begin{aligned} F_{max}(y) &= \mathbb{P}((x_1 < y) \cap (x_2 < y) \cap \ldots \cap (x_n < y)) \\ . &= \prod_i F(y) = F^n(y) = \frac{(y-a)^n}{(b-a)^n} \end{aligned}$$

- Expectation value for max(x):

$$\langle max(x) \rangle = \int_a^b dy\ yn \frac{(y-a)^{n-1}}{(b-a)^n} = \frac{n}{n+1}(b-a) + a$$

- From this we have an estimate for b, assuming a is known:

$$b = \frac{n+1}{n}(max(x) - a) + a$$

Maximum Likelihood Estimation, MLE

- Calculate $\langle max(x) \rangle$ from the CDF! Utilize the independence of data points!

$$\begin{aligned} F_{max}(y) &= \mathbb{P}((x_1 < y) \cap (x_2 < y) \cap \ldots \cap (x_n < y)) \\ . &= \prod_i F(y) = F^n(y) = \frac{(y-a)^n}{(b-a)^n} \end{aligned}$$

- Expectation value for max(x):

$$\langle max(x) \rangle = \int_a^b dy\ yn \frac{(y-a)^{n-1}}{(b-a)^n} = \frac{n}{n+1}(b-a) + a$$

- From this we have an estimate for b, assuming a is known:

$$b = \frac{n+1}{n}(max(x) - a) + a$$

- Because we do not know a at the beginning, we use the MLE value for a. Then (using a similar formula for the minimum), the estimated b will be used for a. Iterate this procedure until convergence!

Bootstrap

- Calculate the 95% confidence interval for the median of experimental data! Apply bootstrap method with percentile calculation!

Bootstrap

- Calculate the 95% confidence interval for the median of experimental data! Apply bootstrap method with percentile calculation!
- Create a sample with replacement, and calculate the median for the sample! Repeat 2000 times!

Bootstrap

- Calculate the 95% confidence interval for the median of experimental data! Apply bootstrap method with percentile calculation!
- Create a sample with replacement, and calculate the median for the sample! Repeat 2000 times!
- Create PDF for the median values from the bootstrapped sample,calculate the percentile values. Remind: confidence intervals are two sided by default.

Multiple testing

- Problem:
  - We have 25 disease indicators for a child. We want to conduct an experiment, where we need a healthy subject, none of the indicators are allowed to show a disease. Our lab is able to test **each indicator** at 0.95 confidence against the disease. We found some indicators, where the test for being healthy was rejected. Should we discard this subject, if our threshold is 95% confidence level for conducting the experiment?

Multiple testing

- Problem:
  - We have 25 disease indicators for a child. We want to conduct an experiment, where we need a healthy subject, none of the indicators are allowed to show a disease. Our lab is able to test **each indicator** at 0.95 confidence against the disease. We found some indicators, where the test for being healthy was rejected. Should we discard this subject, if our threshold is 95% confidence level for conducting the experiment?
- Solution:

Multiple testing

- Problem:
  - We have 25 disease indicators for a child. We want to conduct an experiment, where we need a healthy subject, none of the indicators are allowed to show a disease. Our lab is able to test **each indicator** at 0.95 confidence against the disease. We found some indicators, where the test for being healthy was rejected. Should we discard this subject, if our threshold is 95% confidence level for conducting the experiment?
- Solution:
  - Bonferroni method

Multiple testing

- Problem:
  - We have 25 disease indicators for a child. We want to conduct an experiment, where we need a healthy subject, none of the indicators are allowed to show a disease. Our lab is able to test **each indicator** at 0.95 confidence against the disease. We found some indicators, where the test for being healthy was rejected. Should we discard this subject, if our threshold is 95% confidence level for conducting the experiment?
- Solution:
  - Bonferroni method
  - Benjamini-Hochenberg method

Multiple testing

Multiple testing: Bonferroni correction

- How to set the confidence level for **each test** if we want to ensure the probability of false rejection of **any** null hypothesis is less than α?

Multiple testing: Bonferroni correction

- How to set the confidence level for **each test** if we want to ensure the probability of false rejection of **any** null hypothesis is less than α?
- We have m separate hypothesis tests

$$H_0^i \leftrightarrow H_1^i \text{ with } p - \text{value } P_i$$

Multiple testing: Bonferroni correction

- How to set the confidence level for **each test** if we want to ensure the probability of false rejection of **any** null hypothesis is less than α?
- We have m separate hypothesis tests

$$H_0^i \leftrightarrow H_1^i \text{ with } p - \text{value } P_i$$

- Bonferroni correction:

$$\text{reject } H_0^i \text{ if } P_i < \alpha/m$$

Multiple testing: Bonferroni correction

- How to set the confidence level for **each test** if we want to ensure the probability of false rejection of **any** null hypothesis is less than α?
- We have m separate hypothesis tests

$$H_0^i \leftrightarrow H_1^i \text{ with } p - \text{value } P_i$$

- Bonferroni correction:

$$\text{reject } H_0^i \text{ if } P_i < \alpha/m$$

Proof: $\mathbb{P}$(any test A_i falsely rejected)
$= \mathbb{P}(\cup_i^m A_i) \leq \sum_i^m \mathbb{P}(A_i) = \sum_i^m \alpha/m = \alpha$

Multiple testing: Bonferroni correction

- How to set the confidence level for **each test** if we want to ensure the probability of false rejection of **any** null hypothesis is less than α?
- We have m separate hypothesis tests

$$H_0^i \leftrightarrow H_1^i \text{ with } p - \text{value } P_i$$

- Bonferroni correction:

$$\text{reject } H_0^i \text{ if } P_i < \alpha/m$$

Proof: $\mathbb{P}$(any test A_i falsely rejected)
$= \mathbb{P}(\cup_i^m A_i) \leq \sum_i^m \mathbb{P}(A_i) = \sum_i^m \alpha/m = \alpha$

- This correction is very conservative.

Multiple testing: Bonferroni correction

- How to set the confidence level for **each test** if we want to ensure the probability of false rejection of **any** null hypothesis is less than α?
- We have m separate hypothesis tests

$$H_0^i \leftrightarrow H_1^i \text{ with } p - \text{value } P_i$$

- Bonferroni correction:

$$\text{reject } H_0^i \text{ if } P_i < \alpha/m$$

Proof: $\mathbb{P}$(any test A_i falsely rejected)
$= \mathbb{P}(\cup_i^m A_i) \leq \sum_i^m \mathbb{P}(A_i) = \sum_i^m \alpha/m = \alpha$

- This correction is very conservative.
- The power of this correction is quite low.

Multiple testing: B-H correction

Benjamini-Hochberg correction

- Allow some more false rejection, but control the False Discovery Rate (FDR):

$$FDR = \left\langle \frac{\text{number of false rejections}}{\text{number of all rejections}} \right\rangle \leq \alpha$$

Multiple testing: B-H correction

Benjamini-Hochberg correction

- Allow some more false rejection, but control the False Discovery Rate (FDR):

$$FDR = \left\langle \frac{\text{number of false rejections}}{\text{number of all rejections}} \right\rangle \leq \alpha$$

- The correction procedure:
  - order the tests by increasing p-values: $P_i \leq P_j$ for $i < j$

Multiple testing: B-H correction

Hypothesis testing, MLE, Bootstrap

Introduction

Bootstrap theory

Method of moments theory

Maximum Likelihood Estimate theory

Technical

MME

MME2

MME3

MLE

Bootstrap example

Independence testing

Benjamini-Hochberg correction

- Allow some more false rejection, but control the False Discovery Rate (FDR):

$$FDR = \left\langle \frac{\text{number of false rejections}}{\text{number of all rejections}} \right\rangle \leq \alpha$$

- The correction procedure:
  * order the tests by increasing p-values: $P_i \leq P_j$ for $i < j$
  * set temporary threshold $t_i = i \frac{1}{C_m} \frac{\alpha}{m}$ for all tests, where $C_m = \sum_{i=1}^{m} 1/i$

Multiple testing: B-H correction

Benjamini-Hochberg correction

- Allow some more false rejection, but control the False Discovery Rate (FDR):

$$FDR = \left\langle \frac{\text{number of false rejections}}{\text{number of all rejections}} \right\rangle \leq \alpha$$

- The correction procedure:
  * order the tests by increasing p-values: $P_i \leq P_j$ for $i < j$
  * set temporary threshold $t_i = i \frac{1}{C_m} \frac{\alpha}{m}$ for all tests, where $C_m = \sum_{i=1}^{m} 1/i$
  * Starting from the largest P_i and going down, find the first $P_j < t_j$

Multiple testing: B-H correction

Benjamini-Hochberg correction

- Allow some more false rejection, but control the False Discovery Rate (FDR):

$$FDR = \left\langle \frac{\text{number of false rejections}}{\text{number of all rejections}} \right\rangle \leq \alpha$$

- The correction procedure:
  * order the tests by increasing p-values: $P_i \leq P_j$ for $i < j$
  * set temporary threshold $t_i = i \frac{1}{C_m} \frac{\alpha}{m}$ for all tests, where $C_m = \sum_{i=1}^{m} 1/i$
  * Starting from the largest P_i and going down, find the first $P_j < t_j$
  * reject all H_0 with lower p-value.

Multiple testing: B-H correction

Benjamini-Hochberg correction

- Allow some more false rejection, but control the False Discovery Rate (FDR):

$$FDR = \left\langle \frac{\text{number of false rejections}}{\text{number of all rejections}} \right\rangle \leq \alpha$$

- The correction procedure:
  * order the tests by increasing p-values: $P_i \leq P_j$ for $i < j$
  * set temporary threshold $t_i = i \frac{1}{C_m} \frac{\alpha}{m}$ for all tests, where $C_m = \sum_{i=1}^{m} 1/i$
  * Starting from the largest P_i and going down, find the first $P_j < t_j$
  * reject all H_0 with lower p-value.
- The B-H correction is more popular in case of testing a large numberof hypotheses.

Multiple testing: B-H correction

Benjamini-Hochberg correction

- Allow some more false rejection, but control the False Discovery Rate (FDR):

$$FDR = \left\langle \frac{\text{number of false rejections}}{\text{number of all rejections}} \right\rangle \leq \alpha$$

- The correction procedure:
  - * order the tests by increasing p-values: $P_i \leq P_j$ for $i < j$
  - * set temporary threshold $t_i = i \frac{1}{C_m} \frac{\alpha}{m}$ for all tests, where $C_m = \sum_{i=1}^{m} 1/i$
  - * Starting from the largest P_i and going down, find the first $P_j < t_j$
  - * reject all H_0 with lower p-value.
- The B-H correction is more popular in case of testing a large numberof hypotheses.
- The correction factor for independent tests is smaller: $C_m \leq 1$

Independence

Main concept:

- two by two table of outcomes

.	$a=0$	$a=1$	.
$b=0$	p_{00}	p_{01}	$p_{0.}$
$b=1$	p_{10}	p_{11}	$p_{1.}$
.	$p_{.0}$	$p_{.1}$	$p_{..}=1$

Independence

Main concept:

- two by two table of outcomes

.	$a=0$	$a=1$	.
$b=0$	p_{00}	p_{01}	$p_{0.}$
$b=1$	p_{10}	p_{11}	$p_{1.}$
.	$p_{.0}$	$p_{.1}$	$p_{..}=1$

- Odds ratio:

$$OR = \frac{p_{00}p_{11}}{p_{01}p_{10}}$$

Independence

Main concept:

- two by two table of outcomes

.	$a = 0$	$a = 1$	.
$b = 0$	p_{00}	p_{01}	$p_{0.}$
$b = 1$	p_{10}	p_{11}	$p_{1.}$
.	$p_{.0}$	$p_{.1}$	$p_{..} = 1$

- Odds ratio:

$$OR = \frac{p_{00}p_{11}}{p_{01}p_{10}}$$

- a and b are independent iff:

$$OR = 1$$

or equivalently

$$p_{ij} = p_{i.}p_{.j}$$

Independence

Main concept:

- two by two table of outcomes

.	$a=0$	$a=1$	.
$b=0$	p_{00}	p_{01}	$p_{0.}$
$b=1$	p_{10}	p_{11}	$p_{1.}$
.	$p_{.0}$	$p_{.1}$	$p_{..}=1$

- Odds ratio:

$$OR = \frac{p_{00}p_{11}}{p_{01}p_{10}}$$

- a and b are independent iff:

$$OR = 1$$

or equivalently

$$p_{ij} = p_{i.}p_{.j}$$

- Main tests:
- χ^2 sheet: "chi2"
- Fisher-exact, sheet: "fish"