Deliverable 3.1: Sampling protocols for inorganic ENP from at least 3 matrices (meat, soup, olive oil)

Ersbøll, Bjarne Kjær; Dahl, Vedrana Andersen; Löschner, Katrin; Larsen, Erik Huusfeldt

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

NanoLyse

Nanoparticles in Food: Analytical methods for detection and characterisation

245162: Collaborative Project

Seventh Framework Programme,
Theme 2: Food, Agriculture and Fisheries, and Biotechnology

Deliverable: 3.1

Title: Sampling protocols for inorganic ENP from at least 3 matrices (meat, soup, olive oil)

Author(s)	: Bjarne Kjær Ersbøll, Vedrana Andersen, Katrin Löschner, and Erik Huusfeldt Larsen
Beneficiary(s)	: DTU
Date of preparation	: 31.12.2010
Periodic covered	: 1.1.2010-31.12.2010
Status	: Completed

Dissemination level		
PU	Public	
PP	Restricted to other participants	
RE	Restricted to a group specified by the consortium	
CO	Confidential, only members of the consortium	

Deliverable 3.1: Sampling protocols for inorganic ENP from at least 3 matrices (meat, soup, olive oil)

Bjarne Kjær Ersbøll, Vedrana Andersen,
Katrin Löschner, & Erik Huusfeldt Larsen

January 12, 2011

Contents

1 Introduction **3**

2 T3.1 Establishing a strategy and methodology for representative sampling of solid or liquid foods for inorganic ENP analysis **4**

3 Representative sampling **4**

3.1 From bulk sample to representative sample 5

3.2 Representative sampling - possible ISO standards 7

3.2.1 Representative sampling. A possible ISO standard for olive oil 7

3.2.2 Representative sampling. A possible ISO standard for olive oil and soup 8

3.2.3 Representative sampling. A possible ISO standard for soup (using milk as proxy) 8

3.2.4 Representative sampling. Possible ISO standards for meat . . 10

3.2.5 Representative sampling. NMKL Procedure No. 12 (2002): Guide on sampling for analysis of foods 10

3.2.6 Representative sampling. Most important issues 10

4 Sources of variation **11**

5 Some distributions **13**

5.1 Count data: Poisson 14

5.2 Count data: Negative binomial 15

5.3 Continuous data: Normal 16

5.4 Continuous data: Lognormal 16

6 Statistical considerations regarding sample size **17**

6.1 Confidence intervals 17

6.1.1 An Excel spreadsheet for estimation of confidence limits . . . 19

6.2 Estimating a suitable sample size 19

6.3 Estimating a suitable sample size - an approximation 22

6.3.1 An Excel spreadsheet for estimation of sample size 23

7 Summary and conclusion **24**

1 Introduction

The NanoLyse project aims at detection and analysis of nanoparticles in the food matrix. Any chemical or physical analysis begins with considerations on sampling of the food to be tested, including sub-sampling of the portion taken for instrumental analysis. In chemical analysis of atoms or molecules in food, even at the nM concentration level, sub-sampling often allows for representative sampling because the distribution of the very large number of analyte atoms or molecules per unit volume of the food is uniform.

In contrast, the situation may be different for the same analytes if they are contained in nanoparticles in the food matrix. Each nanoparticle may contain a large number of atoms or molecules. This means that the corresponding number of particles per unit volume is much lower causing a risk of non-representative sampling, or sampling with a larger standard deviation, than was the case for sampling of food containing atoms or molecules. In the NanoLyse project it is foreseen that the instrumental analysis of nanoparticles contained in food or in a food extract will be carried out using a hyphenated system comprising a separation device for size fractionation of the distribution of sizes of nanoparticles, followed by an on-line and real-time detector for quantitative measurement of the particles eluting from the separation device.

Especially at low average concentrations of nanoparticles in the food matrix, understanding "low" as a concentration approaching the instrumental limit of detection, relatively few particles of a given size will exist in the food sub-sample taken for analysis. Hereby the assumption made for atoms or molecules that a very large number of atoms/molecules were contained in the sub-sample taken cannot any longer be made for their nanoparticle counterparts. Therefore the risk of non-representative sampling increases, meaning that the sub-sample no longer represents the composition of the original food sample from which the sub-sample was taken.

Following sub-sampling and any possible sample pre-treatment, the hyphenated instrumentation samples events and therefore a discreet distribution is formally appropriate. One of the most common discreet distributions is the Poisson distribution which is frequently used to describe count data. For the Poisson assumption to hold it is important that the particles are distributed randomly in the (sub-)sample being analysed. If this is not the case then adjustments have to be made on the samples size calculations.

2 T3.1 Establishing a strategy and methodology for representative sampling of solid or liquid foods for inorganic ENP analysis

The milestone (M3.1) and the deliverable (D3.1) for month 12 are both titled: "Sampling protocols for inorganic ENP from at least 3 matrices (e.g. meat, soup, olive oil)". In order to understand the problem consider Figure 1 which indicates some of the steps involved in analysing a (sub-)sample of a matrix for nanoparticles.

- First we have the matrix containing the nanoparticles. In the domain of analysing nanoparticles it is common to call this "the sample". Here this is illustrated by a bowl of soup.
- From the sample a so-called *representative sub-sample* is taken. It is not easy task to ensure representativeness of the sub-sample as this depends heavily on the matrix considered. It is important to have domain knowledge of how nanoparticles behave for each matrix considered. Such domain knowledge often relies on the experience of the laboratory involved and good laboratory practice would be to write down standard operating procedures for dealing with combinations of food-matrices and nanoparticles. As a starting point we refer to several ISO-standards and a Nordic guideline which all consider representative sampling of food-stuffs of different kinds.
- The subsample is processed in an extraction step which removes the food matrix and renders the nanoparticles open to analysis.
- The subsample enters the injector (or nebuliser).
- The subsample enters the Field Flow Fractionation unit (FFF).
- The subsample enters the Inductively Coupled Plasma-Mass Spectrometry+Light Scattering (ICPMS+LS) unit.

Each of these steps introduces a further sub-sampling of the sample from the step before. This implies that the final count in the ICPMS+LS unit is a factor 10,000-50,000 less than that of the original subsample. Again it is important that each of the steps is controlled well enough that the subsample is representative at all steps.

3 Representative sampling

This document is concerned with analysing the uncertainty of content of nanoparticles in samples of food. Based on a subsample taken from a suitable food-matrix (here meat, soup, and olive-oil are considered) an analysis is performed using FFF-ICPMS+LS.

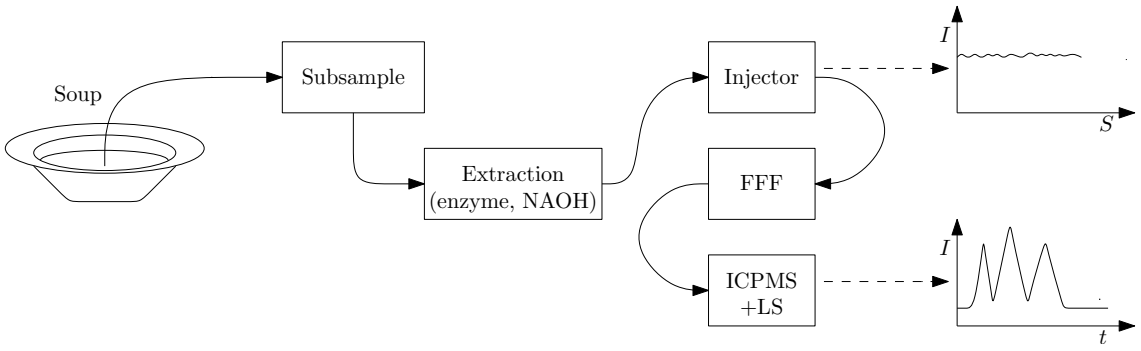


Figure 1: Some of the steps involved in sampling the matrix. At each step in the hyphenated system the sample is further sub-sampled.

It is important to note that the result of this analysis cannot be used to extrapolate back to the original food-matrix, unless the subsample can be considered representative. In this context representative sampling simply means that the subsample contains the same fraction(s) of nanoparticles as the original food-matrix. Given suitable knowledge of the food-matrix, one could in principle apply systematic sampling. Assume for instance that it is known that nanoparticles are always distributed evenly throughout the fat-phase (in one concentration) and the meat-phase (in another concentration) of a beef-steak. In that case one could carefully choose a meat sample and a fat sample of the steak. Each of the two samples is now representative of each of the two phases. Further, given they are proportional in size to the phases in the original sample, then the two subsamples together constitute a representative subsample of the whole beef-steak.

Unfortunately, such thorough knowledge of the sample is rarely available. One is therefore forced to rely on the only reliable alternative, namely random sampling. Random sampling means giving each possible sample location the same probability of entering the subsample. Usually this is achieved by some sort of homogenization of the sample, as described in a number of different standardisation documents.

3.1 From bulk sample to representative sample

Figure 2 shows a schematic overview of achieving a representative sample from two different types of bulk material. This figure and the wording to the left at each subsampling step are typical of the jargon encountered in national and international standards. In the NanoLyse project the jargon differs from this. For instance: the bowl of soup is called a sample and the (representative) sample taken from that is called a subsample. However, the figure still illustrates the central issue, namely the importance carefully assuring the representativeness of the sample.

Securing representativity depends on the nature of the matrix. In the NanoLyse project it has been chosen to investigate three different food matrices: oil, soup, and meat. These might be considered as:

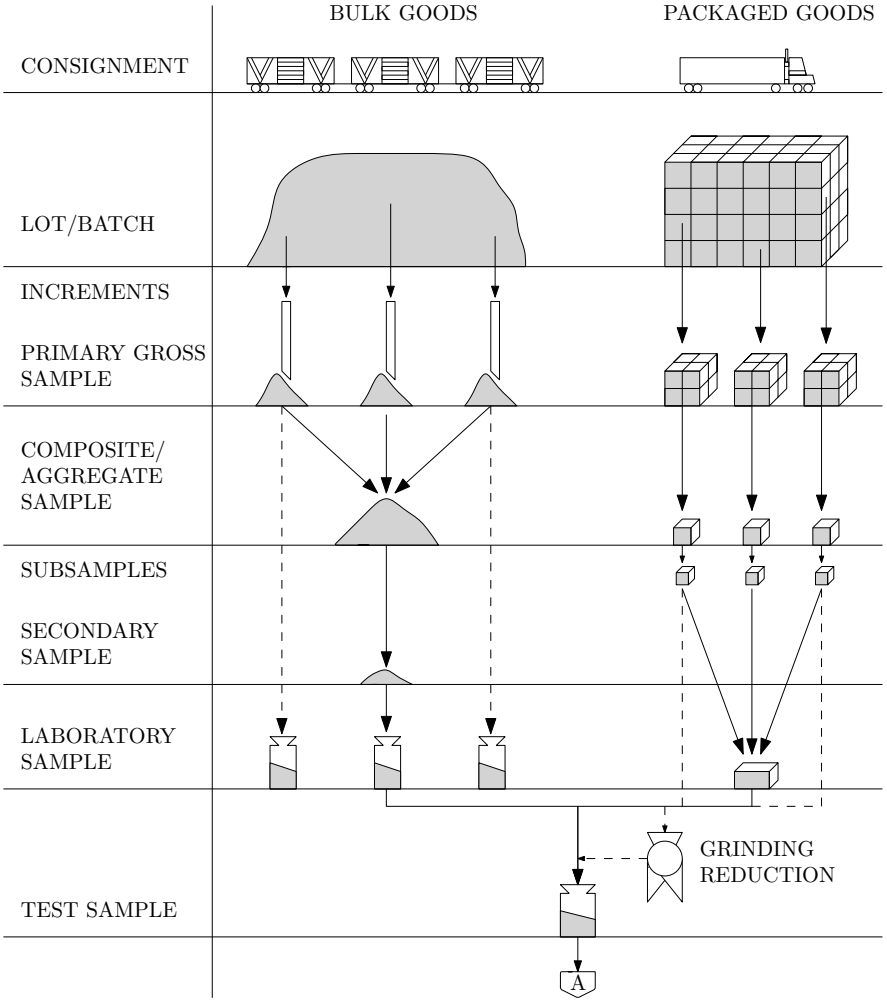


Figure 2: Illustration of acquisition of a representative laboratory sample. Adapted from: NMKL Procedure No. 12 (2002): Guide on sampling for analysis of foods.

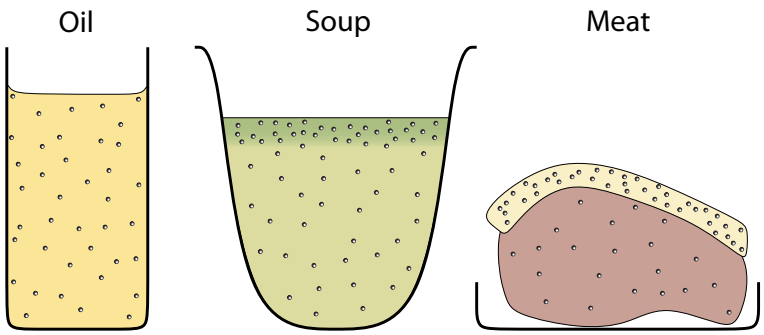


Figure 3: Sampling different matrices where the distribution of ENP might be different in different phases.

- Oil: Fluid considered to consist of one phase, relatively easy to homogenise.
- Soup: Fluid, considered to consist of two or more phases.
- Meat: Solid, considered to consist of two or more phases, probably even heterogeneous.

The distribution of nanoparticles might depend on phase as illustrated in Figure 3. Here the distribution of nanoparticles is seen to depend on which phase is considered.

3.2 Representative sampling - possible ISO standards

As a guideline to representative sampling one may consider different ISO standards on the subject. A number of such ISO standards exist and we will refer to a few here. Basically the standards are just common sense written down formally. Each standard is based on the consensus of the experience of a group of experts within the area. They can be viewed as a practical guidance on what to do depending on the nature of the bulk. In the case of the NanoLyse project the standards may be more or less related to the matrices. Nevertheless, they are a good starting point for the work which should be done within the international community in order to provide sufficient and relevant guidelines for the problem of representative sampling.

3.2.1 Representative sampling. A possible ISO standard for olive oil

The NanoLyse project considers olive oil as one of the three matrices. One relevant ISO-standard for sampling oil is:

- ISO 212 Essential oils Sampling, 2nd ed. 2007

In this standard the recommendation is to shake the sample and then to sample three increments, at 20%, 50%, and 95% of the container height. It is not mentioned why

the sampling increments are not taken at symmetric heights. Also, often standards mention examples of sampling apparatus. However, in this standard, no apparatus is mentioned.

3.2.2 Representative sampling. A possible ISO standard for olive oil and soup

Another standard which may be relevant for olive oil and which also is relevant for certain types of soup is:

- ISO 5555 Animal and vegetable fats and oils – Sampling, 3rd ed. 2001.

In this standard the recommendation is to homogenize the sample if possible. Then one should sample increments at the bottom (10% height), middle (50% height), and top (90% height) of the container. It is noted that in this case the sampling heights are symmetric.

Should the sample be inhomogeneous then the recommendation is to sample increments at depths for each 300 mm (a tank is being considered in the standard). Around the layer(s) between different compositions one should sample more densely e.g. for each 100 mm. Then one should mix appropriate increments proportional to the thickness of the layers.

In this standard several sampling apparatus are mentioned. A couple of the more relevant ones are shown in Figure 4.

3.2.3 Representative sampling. A possible ISO standard for soup (using milk as proxy)

Another standard which may be relevant for soup is a standard on milk and milk products:

- ISO 707 Milk and milk products – Guidance on sampling, 3rd ed. 2008.

Here it is recommended to thoroughly mix all liquids, by inverting, stirring, by pouring to and from one product container to another of the same volume, until sufficient homogeneity is obtained while avoiding foaming. Take the sample immediately after mixing. In certain cases, it will be necessary to take a number of samples to produce a composite of corresponding minimum sample size. The recommended minimum recombined sample size is given as approx. 100ml.

In this standard several sampling apparatus are mentioned. A couple of the more relevant ones are shown in Figure 5.

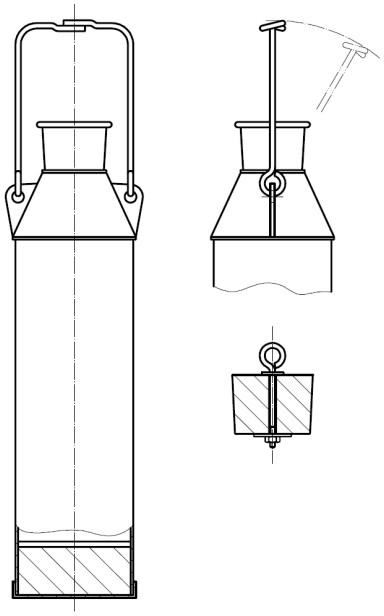


Figure B.1 — Simple weighted sample can

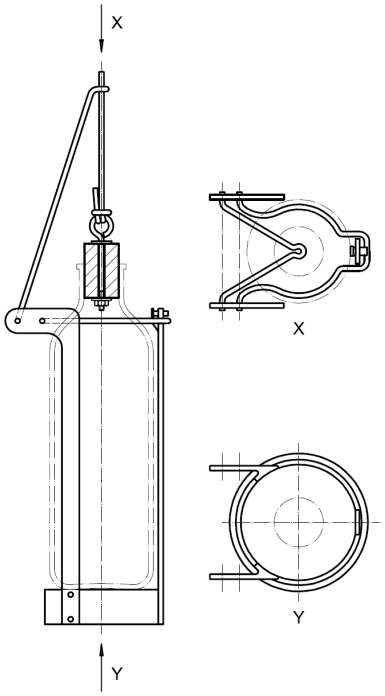
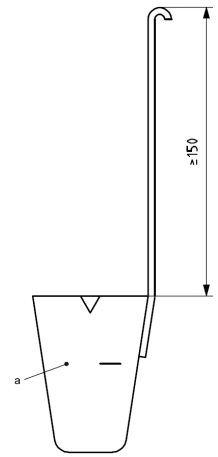


Figure B.2 — Weighted cage for sample bottle

Figure 4: Examples of sampling equipment. Taken from International Standard ISO 5555, Animal and vegetable fats and oils – Sampling, 3rd ed. 2001.

Dimensions in millimetres



a ⩾ 50 ml

Figure A.3 — Suitable dipper for liquids

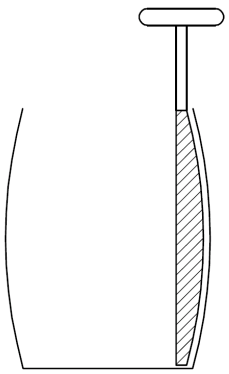


Figure A.4 — Suitable stirrer for mixing sweetened condensed milk in barrels

Figure 5: Examples of sampling equipment. Taken from International Standard ISO 707 IDF 50, Milk and milk products – Guidance on sampling, 3rd ed. 2008.

3.2.4 Representative sampling. Possible ISO standards for meat

Two standards which consider sampling of meat products have been identified:

- ISO 6887-2 Microbiology of food and animal feeding stuffs Preparation of test samples, initial suspension and decimal dilutions for microbiological examination Part 2: Specific rules for the preparation of meat and meat products, 1st ed. 2003.
- ISO 17604 Microbiology of food and animal feeding stuffs. Carcass sampling for microbiological analysis, 1st ed. 2003.

The first of the two standards considers a range of different states the meat sample might be in:

- Frozen products: Products stored frozen should be brought to a consistency that allows sampling.
- Hard and dry products: Rotary homogenizer for max 2.5 min. Mince or grind for max 1 min.
- Liquid and non-viscous products: Test sample should be taken after shaking by hand.
- Heterogeneous products: Sampling by taking aliquots of each component proportionally. Homogenize by mincing or grinding.

The second of the two standards is probably not as relevant to the NanoLyse project as the previous one. Nevertheless it is mentioned for sake of completeness.

3.2.5 Representative sampling. NMKL Procedure No. 12 (2002): Guide on sampling for analysis of foods

This guide is actually a collection of and interpretation of (ISO) standards which has been developed by the Nordic Committee on Food Analysis. It is not based on the most recent ISO standards, nevertheless it gives a good comprehensive overview and serves as a good starting point.

3.2.6 Representative sampling. Most important issues

In the NMKL guide as in all of the previously mentioned ISO standards the keywords towards representative sampling are:

- If possible homogenise your sample then take the necessary subsample.

- If it is not possible to homogenize your sample then take aliquots of the different phases. Recombine the aliquots proportionally to the phases in the sample. This then constitutes the subsample.

The guide and the standards give examples of sampling apparatus which can be used as guidelines for choosing apparatus relevant for the NanoLyse project. It is important that the participating laboratories agree on the sampling procedure and apparatus for each of the matrices involved in the project. In this way the results from the different laboratories will become more comparable.

4 Sources of variation

The final variation which should be reported depends more or less complicatedly on the different steps necessary in order to analyse the sample. As a starting point consider Figure 1 again. The steps starting with the (bulk) sample of soup are roughly as follows:

- Take a representative sub-sample. Due to unavoidable sampling error this induces a variance: σ_1^2
- Perform the extraction step (enzymatic, NaOH,...). This introduces a further variance: σ_2^2
- The injector step also introduces a variance: σ_3^2
- as does the FFF (Field Flow Fractionation) unit: σ_4^2
- the ICP-MS+LS (Inductively Coupled Plasma-Mass Spectrometry+Light Scattering) unit: σ_5^2
- the calibration curve: σ_6^2
- the Xx/Rh ratio, both of which have error: σ_7^2
- multiplication by flow rate: σ_8^2

Most of the steps above include a further sub-sampling or dilution of the original sub-sample. Some researchers have suggest that the dilution factor might be in the order of magnitude: 50,000. Furthermore, one could also consider the influence of different laboratories, different technicians, different equipment, etc. These sources will of course also contribute to the final variation. The resulting error variance σ^2 is therefore a (more or less complicated) combination of all the above mentioned variances. Figure 6 visualises the combination of variances.

It is clearly not obvious which "theoretically correct" probability model applies. It is expected that the size of the different variance components can be estimated or

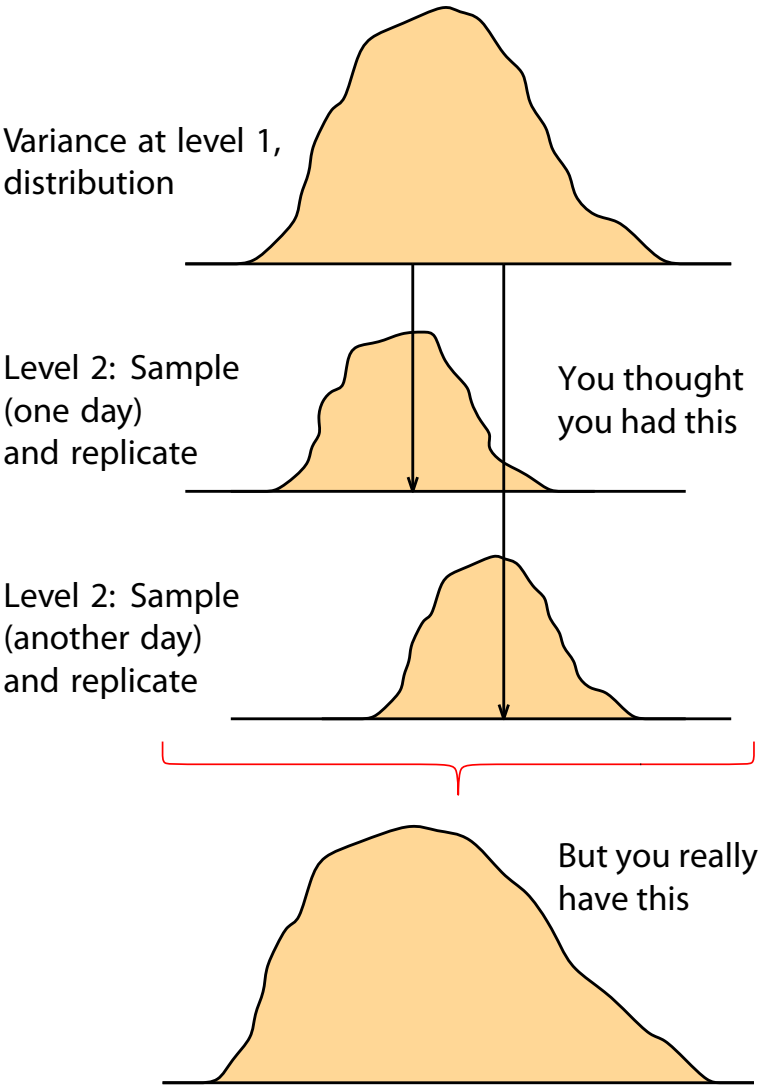


Figure 6: Illustration of variation issues. The actual variance may be larger than first anticipated.

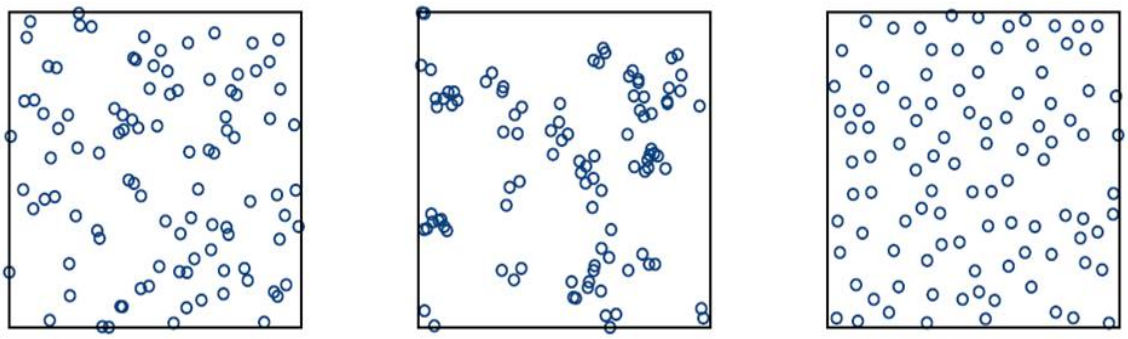


Figure 7: Left: Complete spatial randomness. Middle: Clustering (attractive). Right: Regularity (repulsive).

derived by performing a so-called (fractional-) factorial experiment and analysing this by means of analysis of variance (ANOVA) techniques. It is recommended that this issue is investigated sometime during the remaining course of the NanoLyse project.

5 Some distributions

In this section we will consider a few distributions which are assumed to be relevant for data from ENP experiments. These data are assumed intrinsically to be count data. Before performing any statistical analysis it is important to bring the data on a form, which represents the actual counts in the ICP-MS+LS. E.g. if the read-out is in ng it has to be transformed into counts by dividing by the mass of a typical particle. We will term this number the *raw count*.

Since the data are intrinsically count data it is natural to consider a couple of the most common distributions:

- The Poisson distribution is the most common choice and results from the assumption of complete spatial (volumetric) randomness of in principle infinitely small particles as seen in the leftmost sub-figure in Figure 7.
- The negative binomial distribution is one possible choice if the assumption of complete spatial randomness is violated towards that of spatially attractive clustering as seen in the middle sub-figure in Figure 7.

If and when a factorial experiment to assess the different sources of variation as mentioned above has been performed, it might be advantageous to consider distributions which are typical of continuous data:

- The most common continuous distribution is surely the normal or Gaussian distribution. It is very flexible and can often be used even with quite serious

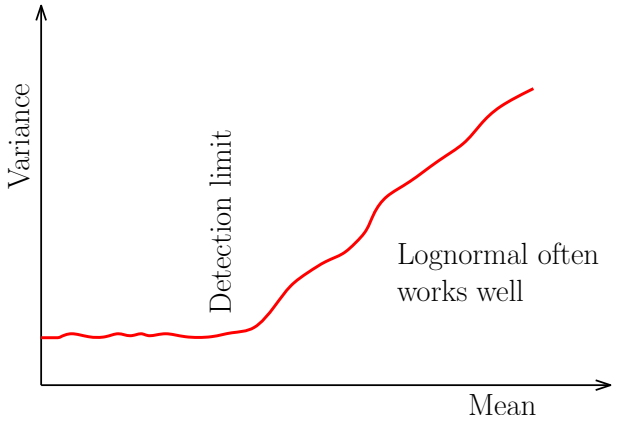


Figure 8: Typical mean and variance relation for chemical laboratory experiments. After applying a log-transformation to such data a normal model may be applicable.

deviations from the normal assumptions. Poisson distributed data can for instance be analysed approximately in a model using the normal distribution. Sometimes the approximation becomes better after applying a square-root transformation. The normal distribution lends itself naturally towards the so-called additive error model where error variances can be combined additively. Quite complicated statistical models can be handled in this setup.

- Another common choice is the lognormal distribution. This handles cases where the error structure is multiplicative in nature. By simply transforming the original (lognormal) data by the log-transformation the error structure turns into an additive one and the normal distribution and models known from that apply. Often in chemical laboratory experiments the mean-variance relationship may be as depicted in Figure 8. In such cases a log-transformation is often applied

In the following we give some formulae related to the different distributions.

5.1 Count data: Poisson

As mentioned above this is a very common choice when considering count data.

Notation

$$\text{Pois}(\lambda)$$

Density

$$\frac{\lambda^k}{k!}e^{-\lambda}$$

Mean

$$\lambda$$

Variance

$$\lambda$$

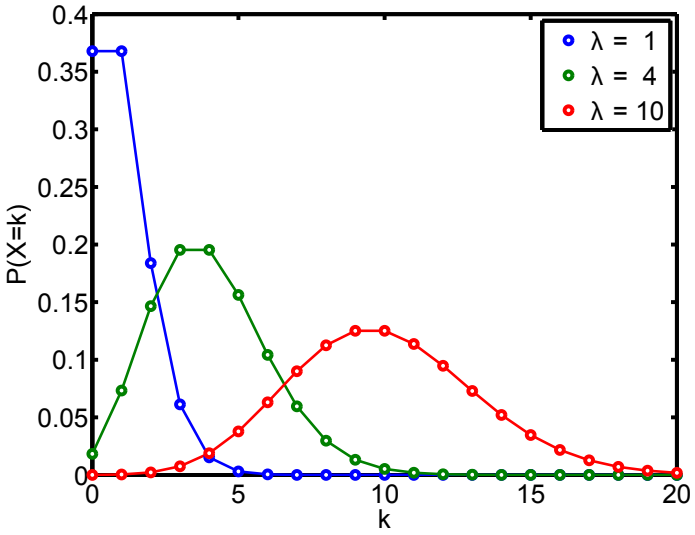


Figure 9: Poisson distribution.

It is noted that the distribution is governed by a single parameter: λ and that the mean equals the variance. Graphs for different values of λ are shown in Figure 9. (Note that the distribution is discrete. The lines connecting the dots are merely added as a visual aid.)

5.2 Count data: Negative binomial

As mentioned, this is one possible choice when considering (attractively) clustered particles.

Notation

$$\mathrm{NB}(r, p)$$

Density

$$\begin{pmatrix} k+r-1 \\ r-1 \end{pmatrix} (1-p)^r p^k$$

Mean

$$r\frac{p}{1-p}$$

Variance

$$r\frac{p}{(1-p)^2}$$

In this case the distribution is governed by two parameters, n and p allowing for a more flexible distribution. Note that the variance is larger than the mean. This is often called over-dispersion. It accounts for the fact that if we sub-sample in the middle sub-figure in Figure 7 we will very often get either many particles or no particles, hence the larger variation as compared to the Poisson case. Graphs for

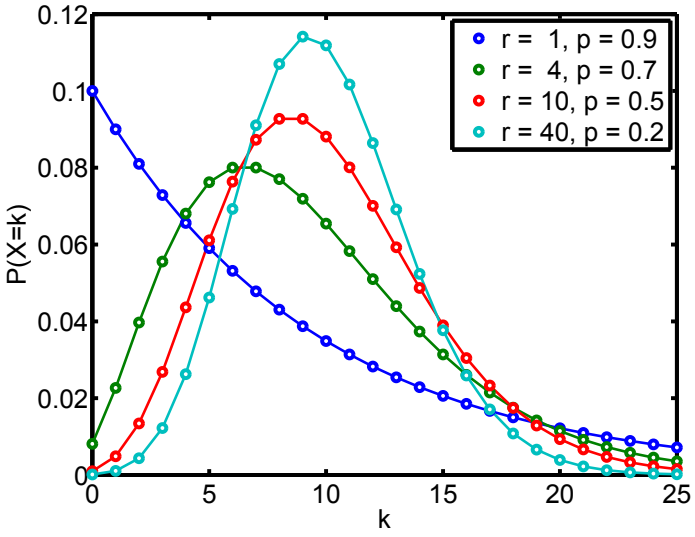


Figure 10: Negative binomial distribution.

different values of n and p are shown in Figure 10. (Again note that the distribution is discrete. The lines connecting the dots are merely added as a visual aid.)

5.3 Continuous data: Normal

The notation for the normal distribution is given below. In Figure 11 different distributional curves for different values of the parameters μ and σ are shown.

Notation

$$\mathcal{N}(\mu, \sigma^2)$$

Density

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

Mean

$$\mu$$

Variance

$$\sigma^2$$

5.4 Continuous data: Lognormal

The notation for the lognormal distribution is given below. In Figure 12 different distributional curves for different values of the parameters α and β are shown.

Notation

$$\ln \mathcal{N}(\mu, \sigma^2)$$

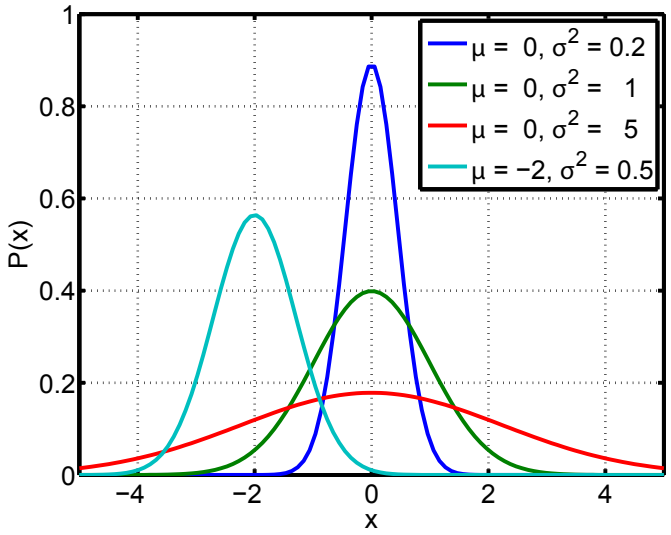


Figure 11: Normal distribution.

Density

$$\frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$$

Mean

$$e^{\mu+\frac{\sigma^2}{2}}$$

Variance

$$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$$

6 Statistical considerations regarding sample size

In the NanoLyse project it has not yet been possible to perform an experiment to assess the order of magnitude of all sources of variation. However, even in that case the natural sampling variance is unavoidable. If complete spatial randomness of the particles is a reasonable assumption and other sources of variation are small compared to the variation from random sampling, then the Poisson distribution can be applied. This gives us access to estimating confidence intervals and sample size considerations.

6.1 Confidence intervals

Assume an experiment has been performed and the number of particles actually counted by the system is x. Then the maximum likelihood estimate $\hat{\lambda}$ of the parameter λ in the Poisson distribution is given as:

$$\hat{\lambda} = x$$

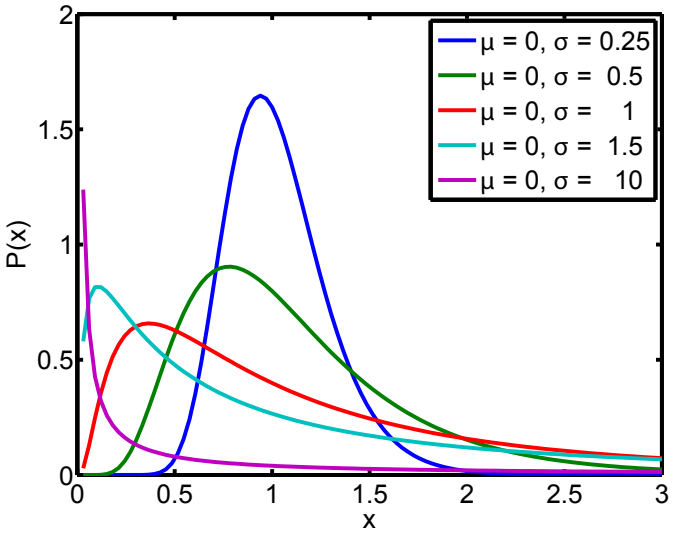


Figure 12: Lognormal distribution.

and the corresponding $1-\alpha$ confidence interval is given by

$$[\frac{1}{2}\chi^2(2x)_{\alpha/2}; \frac{1}{2}\chi^2(2x+2)_{1-\alpha/2}]$$

where: $1-\alpha$ is the width of the confidence interval (typical values are 0.90, 0.95, and 0.99) and $\chi^2(\nu)_\gamma$ is the γ quantile in a χ^2 distribution on ν degrees of freedom.

As an example consider an experiment where 10 particles were detected. In that case parameter λ for the part of the sample actually analysed has a maximum likelihood estimate:

$$\hat{\lambda} = 10$$

The corresponding 95% confidence interval is given by:

$$[0.5 \cdot \chi^2(20)_{0.025}; 0.5 \cdot \chi^2(22)_{0.975}] = [0.5 \cdot 9.591; 0.5 \cdot 36.781] = [4.80; 18.39]$$

Assuming a dilution factor from (sub-)sample to actual counts of 50,000 the estimated number of nanoparticles becomes:

$$50,000 \cdot 10 = 500,000.$$

Similarly, the confidence interval transforms into:

$$50,000 \cdot [4.80; 18.39] = [240,000; 919,500]$$

Note that in order to get a narrower confidence interval the actually analysed part of the sample has to be larger. If a twice as large part of the sample is analysed then the count would be twice as large, namely 20. The dilution factor would then only be 25,000 and the confidence interval becomes:

$$25,000 \cdot [0.5 \cdot \chi^2(40)_{0.025}; 0.5 \cdot \chi^2(42)_{0.975}] = [305,000; 772,000]$$

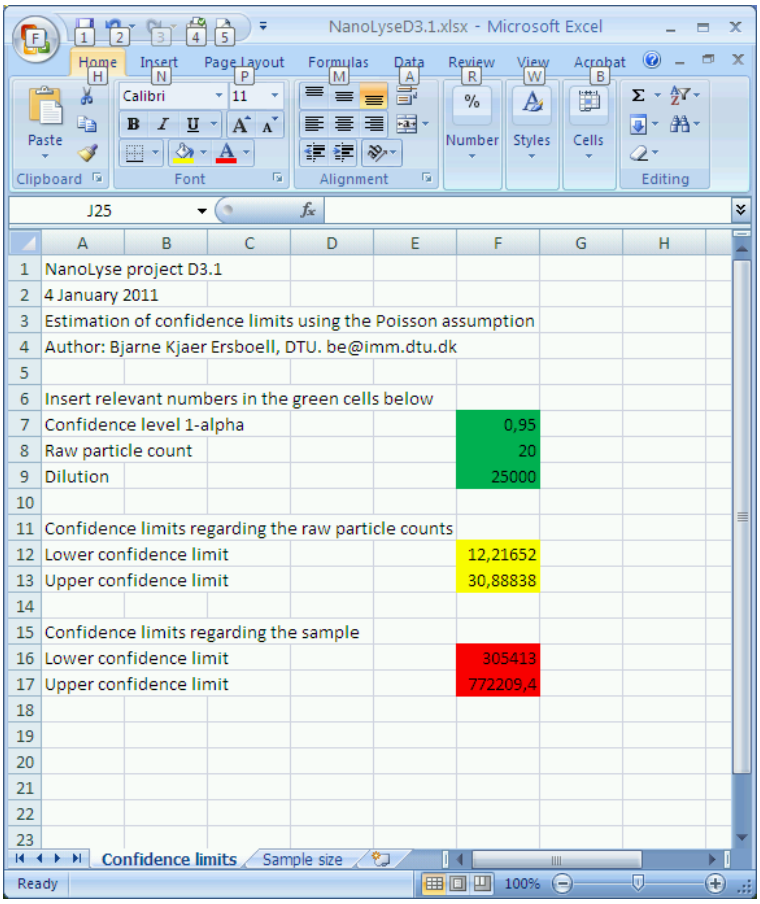


Figure 13: Screenshot of Excel spreadsheet for estimation of confidence limits.

6.1.1 An Excel spreadsheet for estimation of confidence limits

An Excel spreadsheet has been developed to help in estimating the confidence limits given above. A screenshot is shown in Figure 13. Enter relevant parameter values in the green cells. The yellow cells then give the confidence limits on the raw counts. Similarly the red cells give the corresponding confidence limits at the sub-sample or sample level, depending on the dilution factor assumed.

6.2 Estimating a suitable sample size

Assume a qualified guess on the expected number of counts λ_0 actually measured for a certain sub-sample size exists. Furthermore, assume a certain width of the $1-\alpha$ confidence interval is requested. Since the confidence interval itself is a statistic it will of course vary in width. However, we can demand that the width is less than or equal to the requested width a certain fraction β of the time. The parameter β is called the power. Often β is set at 80%. Other common values of β are 90% or 95%.

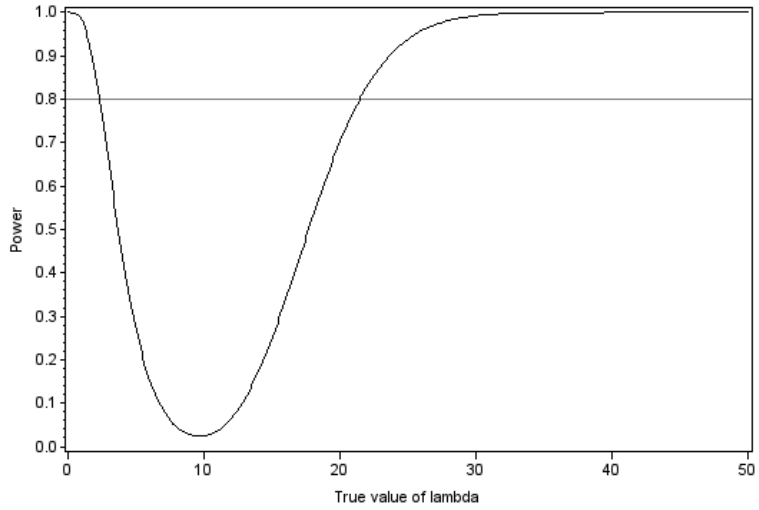


Figure 14: Power curve for $n\lambda_0 = 10$.

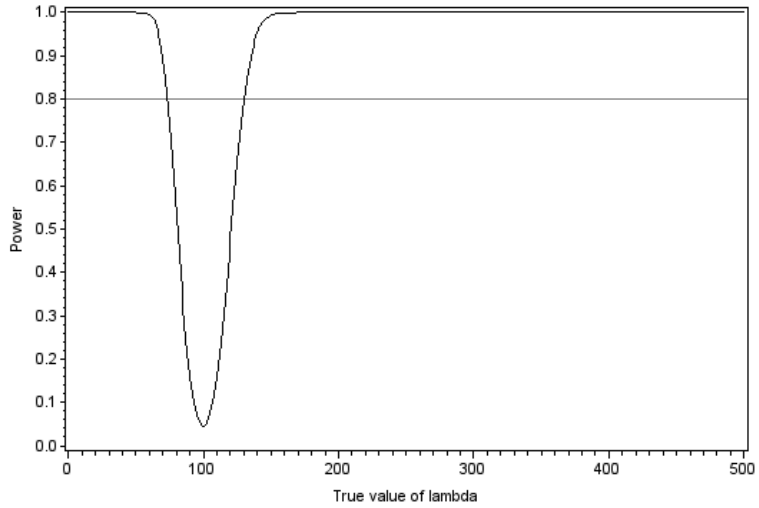


Figure 15: Power curve for $n\lambda_0 = 100$.

The power in $n\lambda$ is given by the expression:

$$P\{\text{Pois}(n\lambda) < \text{Pois}(n\lambda_0)_{\alpha/2}\} + P\{\text{Pois}(n\lambda) > \text{Pois}(n\lambda_0)_{1-\alpha/2}\}$$

Figures 14, 15, 16, and 17 show examples of power curves for different values of $n\lambda_0$. The intersections between the power curves and the horizontal line drawn at 0.8 shows the requested maximum width of the confidence interval. Note how the interval becomes relatively more narrow as $n\lambda_0$ increases.

Solving the above equation for n gives how many times larger (or smaller) the actual sub-sample *at least* should be. Consider a case, where we expect the raw count for a certain volume of the sub-sample to be 10. We want the width of the 95% confidence interval to be less than or equal to 4 (roughly corresponding to $[\lambda_{0.025}; \lambda_{0.975}] = [8; 12]$) 80% of the time. (Note: from e.g. Figure 14 $\lambda_{0.025}$ and $\lambda_{0.975}$ are clearly not symmetric around λ, so the interval is not exactly $[8; 12]$ but a bit

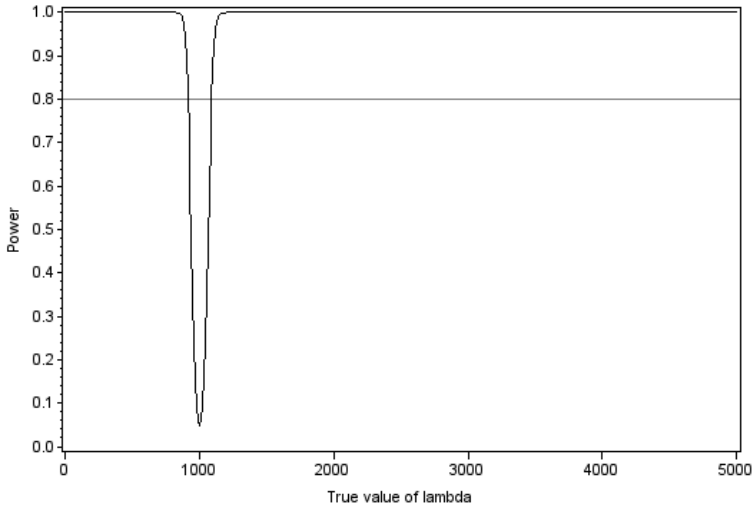


Figure 16: Power curve for $n\lambda_0 = 1000$.

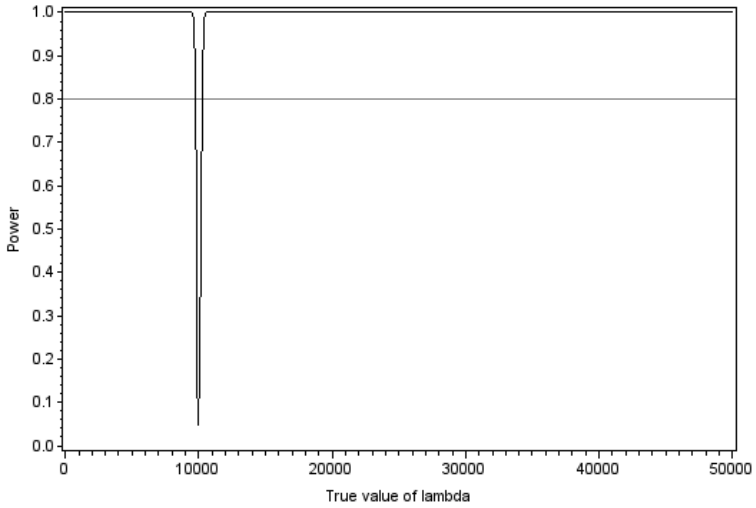


Figure 17: Power curve for $n\lambda_0 = 10000$.

skewed to the right.)

In summary we have:
$\lambda_0 = 10$ (expected count),
$w = \lambda_{0.975} - \lambda_{0.025} = 4$ (width),
$\alpha = 1 - 0.95 = 0.05$ (1-confidence level), and
$\beta = 0.80$ (power).

We then need to solve:

$$P\{\text{Pois}(n\lambda_{0.025}) < \text{Pois}(n \cdot 10)_{0.025}\} + P\{\text{Pois}(n\lambda_{0.975}) > \text{Pois}(n \cdot 10)_{0.975}\}$$

or approximately:

$$P\{\text{Pois}(n \cdot 8) < \text{Pois}(n \cdot 10)_{0.025}\} + P\{\text{Pois}(n \cdot 12) > \text{Pois}(n \cdot 10)_{0.975}\}$$

for n. Unfortunately this has to be done by trial and error.

6.3 Estimating a suitable sample size - an approximation

For reasonably large counts (10 or more as a rule of thumb) the sample size can be found by an approximation to the normal distribution. The mean and the variance of the Poisson distribution $\text{P}(\lambda_0)$ are both λ_0. We may then use the normal distribution $\text{N}(\lambda_0, \lambda_0)$ as an approximation.

In that case n can be found as:

$$n = \frac{\lambda_0}{(w/2)^2} \cdot (z_{1-\alpha/2} + z_\beta)^2$$

Where z_ν is the ν quantile in the normal distribution. For the example above we get:

$$n = \frac{10}{(4/2)^2} \cdot (z_{1-\alpha/2} + z_\beta)^2 = \frac{10}{4} \cdot (1.96 + 0.8416)^2 = 2.5 \cdot 7.85 = 19.6.$$

In other words one should analyse a sample about 20 times larger in order to end up with an expected count of around $20 \cdot 10 = 200$. The corresponding lower and upper limits with 80% power will then be about $20 \cdot [8; 12] = [160; 240]$. As a check consider:

$$P\{\text{Pois}(160) < \text{Pois}(200)_{0.025}\} + P\{\text{Pois}(160) > \text{Pois}(200)_{0.975}\}$$

or

$$P\{\text{Pois}(160) < 172\} + P\{\text{Pois}(160) > 228\} = 0.82$$

So the power at $n \cdot \lambda = 160$ is about 82%. At $n \cdot \lambda = 240$ the power is about:

$$P\{\text{Pois}(240) < \text{Pois}(200)_{0.025}\} + P\{\text{Pois}(240) > \text{Pois}(200)_{0.975}\}$$

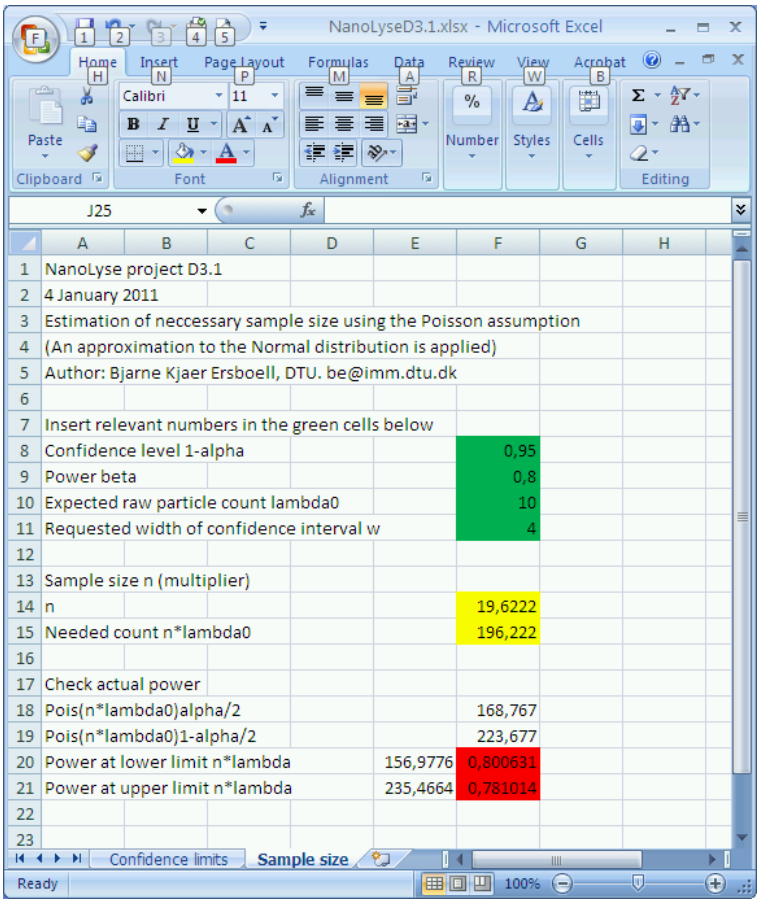


Figure 18: Screenshot of Excel spreadsheet for estimation of sample size.

or

$$P\{\text{Pois}(240) < 172\} + P\{\text{Pois}(240) > 228\} = 0.77$$

or about 77%. In both cases this is considered sufficiently close to the requested power.

6.3.1 An Excel spreadsheet for estimation of sample size

An Excel spreadsheet has been developed to help in estimating the sample sizes as given above. A screenshot is shown in Figure 18. Enter relevant values in the green cells. The yellow cells show how many times larger the sub-sample needs to be in order to achieve the needed number of raw counts to fulfil the requirements. The red cells are a check of the power achieved at the two points corresponding to the intersections between the power curve and the horizontal line drawn at 0.8 in e.g. Figure 14. This shows the requested maximal width of the confidence interval. Note that the spreadsheet uses an approximation to the Normal distribution in a couple of places. Furthermore, for simplicity the confidence interval is assumed to be symmetric which it is known not to be. As mentioned above this should not be an issue for reasonably large (raw) counts.

In summary: what is important is the expected *raw* particle count at the ICP-MS+LS instrument. This number is expected to be thousandsof times less than the actual count in the soup bowl. It is the number of atoms actually measured divided by number of atoms in a particle.

Considering Figure 18 one notes the researcher expects a raw count of around 10 particles for a unit volume (say 1μl). She would then like to end up with a 95% confidence interval with maximal width 4 (so the interval is about [8;12], or loosely: 10$\pm$2). The power for this should be 80%. The first yellow field then says she needs to analyse 19.6 times more or about 20μl. This will then end up with a raw count of about 200 (196), which is sufficient to secure the width ratio of 4/10.

If there are different particle sizes involved in the sample, then the particle size with lowest count should be used for dimensioning.

Again, 20μl is a lower limit, since many other sources of variation are bound to exist.

7 Summary and conclusion

In this report we consider different issues related to sampling matrices containing nano particles. First the issue of preparing a representative sub-sample is addressed by referring to different ISO standards and a Nordic guideline on the matter. Then different relevant distributions and different sources of variation are considered. Since no data is yet available from actual so-called factorial experiments it has not been possible to assess the order of magnitude of these possible sources of variation. However, the intrinsic sampling variation is unavoidable. Assuming the nanoparticles are Poisson distributed both confidence intervals and sample size estimates can be provided. Formulas and an Excel spreadsheet for both are given.

Once again it should be stressed that experiments should be conducted in order to reveal the main sources of variation for the hyphenated system. The formulas and spreadsheet given depend on the Poisson assumption to be at least approximately true. When more is known about the main sources of error then other formulae for confidence limits and sample size calculation may have to be developed.

References

[1] *ISO 212 Essential oils Sampling*, 2nd edition, 2007. ISO copyright office, Case postale 56, CH-1211 Geneva 20, Web www.iso.org, Published in Switzerland

[2] *ISO 5555 Animal and vegetable fats and oils Sampling*, 3rd edition, 2001. ISO copyright office, Case postale 56, CH-1211 Geneva 20, Web www.iso.org, Published in Switzerland

[3] *ISO 212 Essential oils Sampling*, 2nd edition, 2007. ISO copyright office, Case postale 56, CH-1211 Geneva 20, Web www.iso.org, Published in Switzerland

[4] *ISO 6887-2 Microbiology of food and animal feeding stuffs Preparation of test samples, initial suspension and decimal dilutions for microbiological examination Part 2: Specific rules for the preparation of meat and meat products*, 1st edition, 2003. ISO copyright office, Case postale 56, CH-1211 Geneva 20, Web www.iso.org, Published in Switzerland

[5] *ISO 17604 Microbiology of food and animal feeding stuffs Carcass sampling for microbiological analysis*, 1st edition, 2003. ISO copyright office, Case postale 56, CH-1211 Geneva 20, Web www.iso.org, Published in Switzerland

[6] *NMKL Procedure No. 12: Guide on sampling for analysis of foods*, 2002. The General Secretariate of NMKL, c/o National Veterinary Institute, PB 8156 Dep, N-0033 Oslo, Norway