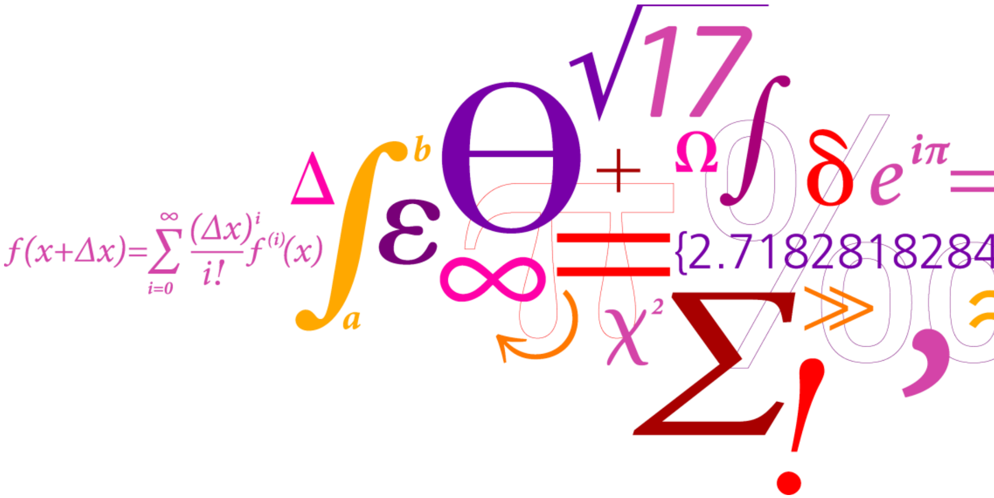
Group testing – pool sizes, effects and group sensitivity

Anders Stockmarr

DTU Statistics and Data Analysis

DTU, 7/5 2020



DTU Compute
Institut for Matematik og Computer Science

Why Group testing

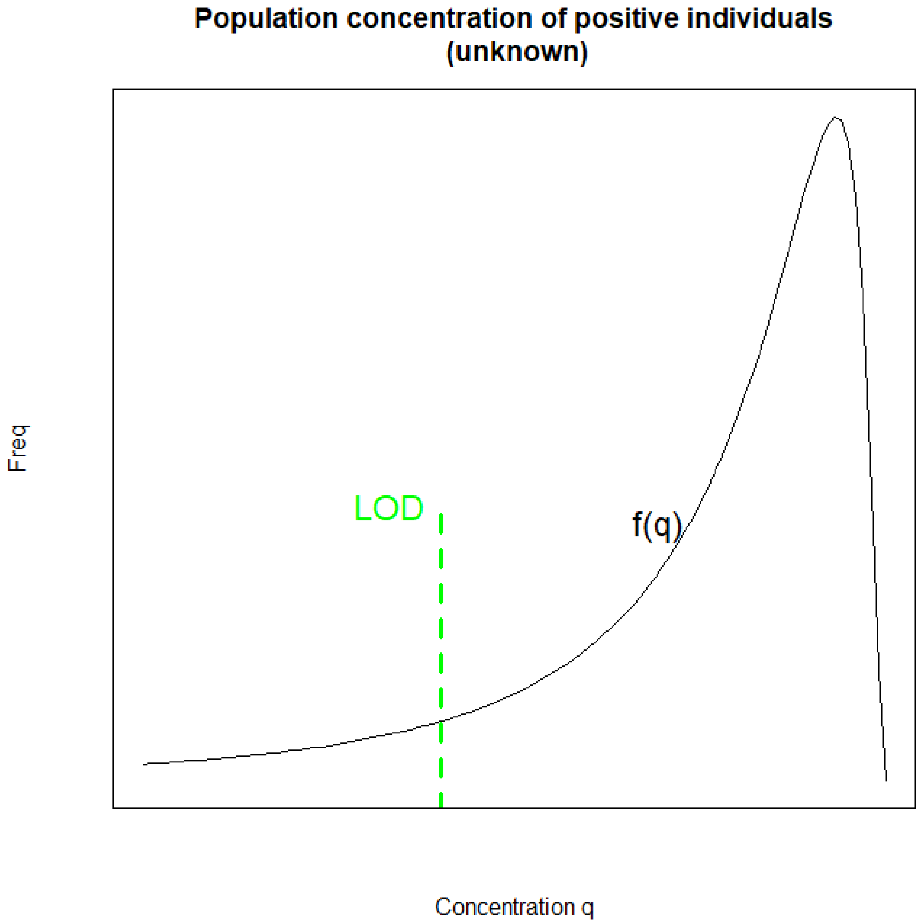
- Fewer tests save:

Equipment

Time

Sars-Cov-2 testing

- RT PCR, result is a score in copies/µL.



Imperfect tests

- Even if a person is infected, the sample may not contain virus (throat swab in the case of a lung infection)

- Even if a sample from an individual contains virus, the test may not detects it. Reasons:

  - The concentration is below the Limit of Detection (LOD);

  - The test fail to accurately measure the concentration.

The probability of a positive sample given infection is expressed as the sensitivity λ.

Probabilistic approach to tests

Assume that the test is being conducted for a disease with prevalence p. The test has a sensitivity of λ, ie.

$$P(Test\ positive|Disease\ positive) = P(T+|D+) = \lambda$$

For simplicity, perfect specificity is assumed, ie.

$$P(Test\ negative|Disease\ negative) = P(T-|D-) = 1$$

Probabilistic approach to tests

The **test result** $\mathbb{X}$ depends on both the **concentration** q (with density f for positive individuals), and on **measurement uncertainty**.

Will assume a multiplicative conditional model so that

$$\log(\mathbb{X}|q) \sim N(\log(q), \sigma^2)$$

Ie. Conditionally on q, $\mathbb{X}$ is log-normal:

$$log(\mathbb{X}) = \log(q) + \sigma\varepsilon$$

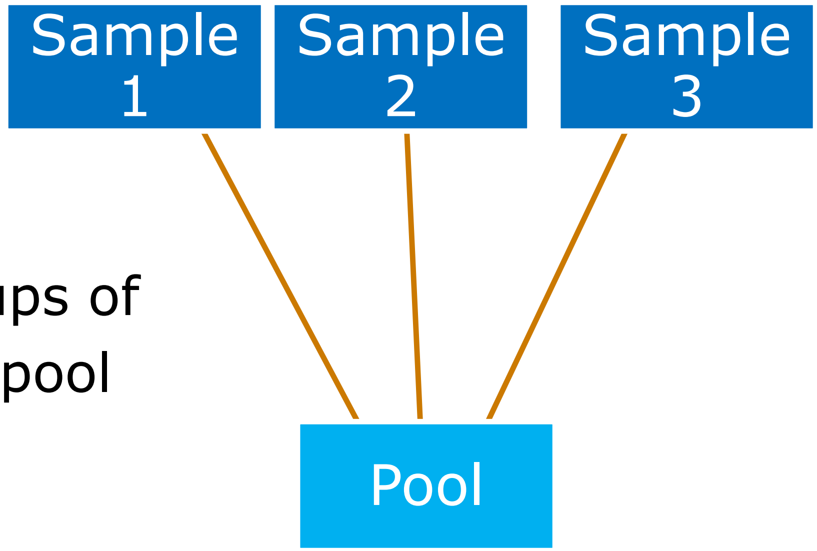
concentration

measurement uncertainty

(this assumption only plays a part for the last part of the talk)

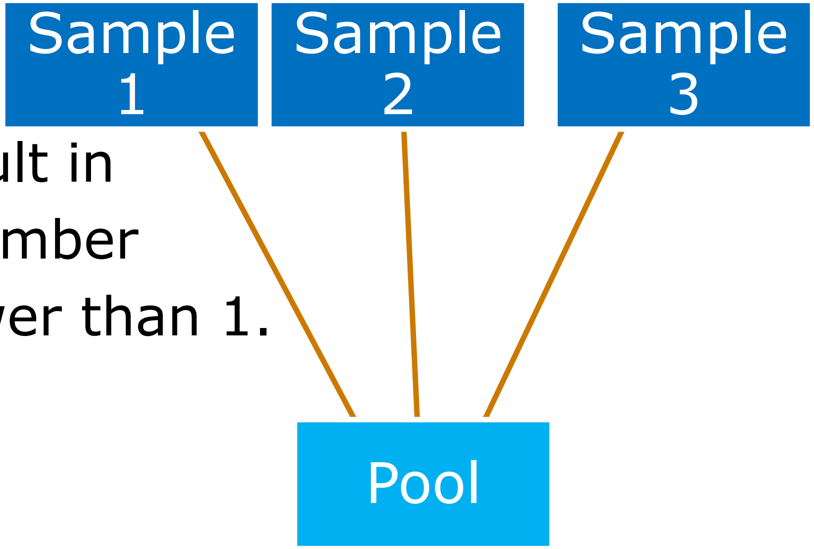
The test is **positive** if $\mathbb{X} > LOD$ (another value could apply)

Pool tests



- Simple pooling scheme:
  - **Divide** the subjects into groups of size k for some integer, and pool their samples.
  - **Test** the pooled sample;
  - If the pooled sample is positive, **retest** every individual in the group.

Pooled tests



- Well chosen group sizes will result in a lesser **effort**; the expected number of tests per individual will be lower than 1.

- The effort is (k>1)

$$\lambda_k\left(1-(1-p)^k\right)+\frac{1}{k}$$

Ex. $k=1$: Effort is 1 ☹ .
Ex. $k=n$: Effort may be >1.

Somewhere in between is better…

Minimizing the effort

For a fixed sensitivity, the group size with minimum effort can determined analytically as

$$k = 2Argmin\left(W_0\left(-\sqrt{-\frac{\log(1-p)}{4\lambda}}\right), \left(W_{-1}\left(-\sqrt{-\frac{\log(1-p)}{4\lambda}}\right)\right)\right)/\log(1-p)$$

Where W_0, W_{-1} are the two branches of the Lambert function, and 'Argmin' chooses the one with the lowest effort (for realistic values of p it will be W_0).

	$p = 0.1$		$p = 0.01$		$p = 0.001$		$p = 0.0001$		$p = 0.00001$	
	k	Effort	k	Effort	k	Effort	k	Effort	k	Effort
$\lambda = 1$	4	0.59	11	0.20	32	0.06	101	0.02	317	0.006
$\lambda = 0.9$	4	0.56	11	0.19	34	0.06	106	0.02	334	0.006
$\lambda = 0.8$	4	0.53	12	0.17	36	0.06	112	0.02	354	0.006
$\lambda = 0.7$	5	0.49	13	0.16	39	0.05	120	0.02	379	0.005

What if the the pool test is positive?

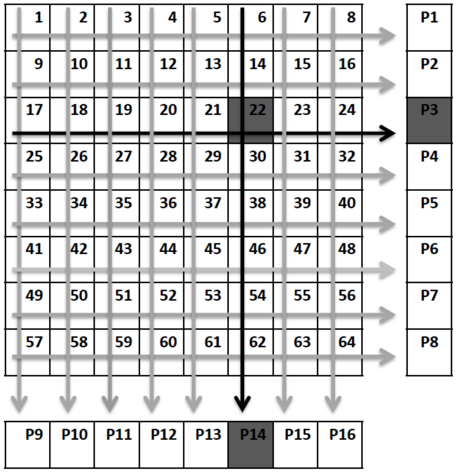
- The same sample appears in multiple pools: Results are stochastically **dependent** through the concentration q;

Consider

$$P(\mathbb{X}_2\ positive, \mathbb{X}_{pool}\ positive) \\ = \lambda_k P(\mathbb{X}_2\ positive | \mathbb{X}_{pool}\ positive)$$

- If $\mathbb{X}_2$ is a single test result, it is the probability that **simple pooling** catches a case;

- If $\mathbb{X}_2$ is a pooled test containing the same (case) individual, it is the probability that Kaare Græsbøls's **2nd order non-hierachical testing** catches a case.



What if the the pool test is positive? - Initial thoughts

- With knowledge on f and σ^2, one may derive (approximate) expressions for the probability that a case will reulst in a detection.

- Disregard two positive samples in the pool (low prevalence), assume that the sample **dilutes the concentration** k-fold and consider the event

$$\left\{\mathbb{X}_{pool} > LOD\right\}; \text{ ie. } \left\{\log\left(\mathbb{X}_{pool}\right) = \log(q/k) + \sigma\varepsilon >\right.$$

What if the the pool test is positive?
- Initial thoughts – loss of sensitivity

$\{\mathbb{X}_{pool} > LOD\}$; ie. $\{\log(\mathbb{X}_{pool}) = \log(q/k) + \sigma\varepsilon >$

Conditional probability of a single test being positive

Thus:

$$P\left(\mathbb{X}_{single}\ positive \middle| \mathbb{X}_{pool}\ positive\right) \approx$$

$$P\left(\mathbb{X}_{single}\ positive \middle| q > kLODe^{-1.64\sigma}\right) =$$

$$\int_{\log(kLOD)-1.64\sigma}^{\infty} \int_{LOD}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\log(q))^2} f(q)dx\, dq$$

- For known values of f and σ^2 this integral can be calculated, and provides the approximate probability that an individual is tested positive, given the pooled sample is positive.

- Together with λ_k, this is a measure of the loss in sensitivity.

Research

- More thorough expressions of conditionally positive tests.
- Detect analytical expressions for optimal pool sizes with varying sensitivity;

- Derive optimal pool size analytically under more advanced pooling schemes than simple pooling;

- **How to get the parameter values?**

\- Need f and σ^2; can be derived from (at least) double measurements on the same sample. Will be dependent on lab technician, kit, lab, etc. Should be found from a random effects model. Looking into possibilities with Johan, Kaare and others.