Estimation and hypothesis testing for rare binary traits:
Controlling size and increasing power by binomial group testing

7th ISTA Seminar on Statistics in Seed Testing

Frank Schaarschmidt, Bioinformatics Unit, Faculty of Natural Sciences, University of Hannover
schaarschmidt@bioinf.uni-hannover.de
Outline

• Introduction: Objective of group testing
• Confidence intervals for group testing
• Experimental design for hypothesis testing: based on power
• Alternative Intervals
• Experimental design for estimation: MSE and interval width
Objective of binomial group testing

Estimation (confidence intervals)
Hypothesis testing (testing plans, p-values or confidence intervals)
for small binomial proportions

Here, focus on testing the hypotheses:
\[ H_0 : \pi \geq \pi_0 \text{ vs. } H_1 : \pi < \pi_0 \]
Respectively estimation of upper \((1-\alpha)\times100\%\) confidence limits for \(\pi\)
Idea of binomial group testing

Binomial testing:

\[
\begin{array}{cccccccccc}
\odot & \odot & \odot & \odot & \bullet & \odot & \odot & \odot & \odot & \odot \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

10 seeds
10 assays
\[\hat{p} = \frac{1}{10} = 0.1\]

→ Observation of individuals

Binomial group testing:

\[
\begin{array}{cccccccccc}
\bullet & \odot & \bullet & \odot & \odot & \bullet & \odot & \odot & \odot & \odot \\
\odot & \bullet & \odot & \odot & \odot & \bullet & \odot & \odot & \odot & \odot \\
\odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot \\
\odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot \\
\odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot \\
\odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot & \odot \\
\end{array}
\]

50 seeds assigned
to 10 groups
10 assays
\[\hat{p} = 1 - \left(1 - \frac{4}{10}\right)^{\frac{1}{5}} = 0.0971\]

→ Observation of groups
Assumptions

1. Trait of interest *i.i.d. Bernoulli* ($\pi$) for each unit in the population.
2. Individuals *randomly assigned to groups*.
3. All groups contain same number of units.
4. Group size does not influence the probability of a single unit to show the trait of interest. *Sensitivity* and *Specificity* of the assay must be *sufficient to detect a group as positive* if and only if at least one *single member is positive*. 
Why confidence intervals?

More general methodology, useful for:
1) Hypothesis testing against a priori known threshold proportion
2) Estimation including the uncertainty of sampling
   → Explorative stage of research
   → Decisions without a priori defined thresholds (plant breeding)

The most important risk can be pre-specified via $\alpha$.

CI methods for binomial proportion can be easily applied for group testing (Tebbs and Bilder, 2004).
Upper confidence limits for testing impurity

- Null hypothesis $H_0: \pi \geq \pi_0$ can be rejected if upper confidence limit excludes the threshold $\pi_0$ (LQL)
- Exact upper $(1-\alpha)*100\%$-Clopper-Pearson CI guarantees a consumers risk of maximally size $\alpha$
- Producers risk $\beta$ depends on the true proportion $\pi$, threshold $\pi_0$ (LQL), number of groups $n$ and size of groups $s$
CI for a-posteriori decisions

For many internal decisions there are **no a priori defined thresholds**, f.e.:

- select inbred lines with a **low** proportion of **susceptible** individuals
- **Internal quality control** for seed borne diseases, germination,…

**If aim is only estimation:** other methods than the exact Clopper-Pearson CI are available, producing intervals of **shorter width** (Brown et al. 2001, Tebbs and Bilder 2004)
Experimental design for group testing

Testing the hypotheses: $H_0 : \pi \geq \pi_0$ vs. $H_1 : \pi < \pi_0$

A problem with more than two dimensions:

- Number of groups $n$
- Group size $s$
- True proportion $\pi$ (AQL)
- Pre-specified threshold $\pi_0$ (LQL)
- Consumers risk ($\alpha$ of upper limit for impurity)
- Producers risk ($\beta$)
The importance of experimental design

1) **Group size too small:**
insensitive design: the event Y=0 mostly favouring rejection of H₀ has very high probability
→ H₀ can never be rejected

2) **Group size appropriate:**

3) **Group size too large:** high probability to observe only positive groups Y=n=20,
→ For these cases estimator p=1 although the true proportion π=0.005 \(\rightarrow\) bias
Experimental design for testing plans (SeedCalc 6.1.4)

Testing plans consider producers risk for $\pi$ (AQL) and consumers risk for $\pi_0$ (LQL) for a certain critical event $c$ (Remund et al. 2001) and fixed number of groups $n$ and group size $s$.

- Try different combinations $\{n, s, c\}$ to find a testing plan with appropriate consumers and producers risk.
Experimental design for **group testing confidence intervals**

Confidence level \((1-\alpha)\) is **pre-specified**

**Consumers risk:** maximally \(\alpha\) if upper Clopper Pearson confidence limit for impurity excludes \(\pi_0\)

**Producers risk:** considered for **all possible events**, including the problematic event \(Y=n\), resulting in **bias** of the estimator

Producers risk is a **non-monotone** function of \(n\) and \(s\)
Closed calculation of power for group testing confidence intervals

Calculate the expectation of rejecting the null hypotheses from all possible outcomes $y=0,\ldots,n$ and its probabilities:

$$\text{power}(n, s, \pi, \pi_0) = \sum_{y=0}^{n} I(y, n, s, \pi_0) \binom{n}{y} (1-(1-\pi)^{s})^{y} (1-\pi)^{s(n-y)}$$

where $I(y, n, s, \pi_0) = 1$ if CI excludes $\pi_0$, and 0 otherwise

- Similar calculation: coverage probability, bias of estimator, expected interval width
- Increasing intensity of computation for $n > 300$
Find **optimal group size** \( s \) for a fixed number of groups \( n \)

\[ n=10 \text{ groups, } \]
\[ s= 1,\ldots,500 \]

Power to reject \( H_0: \pi \geq 0.01 \) \( \pi = 0,\ldots,0.01 \)
Find **optimal group size** $s$ for different values of $\pi$

Reject $H_0$: $\pi \geq 0.01$ in favor of $H_A$: $\pi < 0.01$

For a fixed number of groups $n=10$, find optimal group size in $s=1, \ldots, 500$:

Assume different values of true proportion $\pi=0.003, 0.004, 0.005$

**Local maxima of power do not depend on $\pi$!**
Local maxima of power $\rightarrow$ local minima of coverage probability

**Threshold** is known a priori: $\pi_0 = 0.01$
true proportion $\pi$ is unknown, for power calculation $\pi = 0.003$

**Group sizes** $s$ with local maximal power have $\alpha$ closest to the nominal level ($0.05$)
Optimal group sizes $s$ depend on $n$, $\alpha$, $\pi_0$
Conclusions: hypothesis tests using upper Clopper-Pearson limits

Select designs \( \{n,s\} \) with (local) maximum of power, then:

→ **Consumers risk** (local) close to \( \alpha \) for any value of \( \pi \)
→ **Producers risk** (local) minimal for given parameters
→ **Producers risk** also has local minimum for all smaller \( \pi \)

decrease of power due to bias \( (Y=n) \) explicitly considered
Example: Testing a seed lot for presence of GMO not approved in the EU

Threshold proportion: $\pi_0 = 0.005$
Assay method is sensitive to find 1 GMO in bulk samples of size $s=3000$
Budget for about $n=20$ assays per seed lot

Consumers risk: $\alpha=0.05$
Producers risk: $\beta=0.10$

$\pi=0.002$, and $\pi=0.0025$ for comparison

For which group size $s$ this can be achieved?
n=20, power to reject $H_0$: $\pi \geq 0.005$ if $\pi = 0.002, 0.0025$ (AQL)

Power

$s=452$

$s=239$

group size $s$

bias of the estimator

bias($\hat{p}$)

$\pi = 0.002$

$\pi = 0.0025$

group size $s$
Objective: Estimation, no a-priori defined threshold

Alternative, asymptotic intervals:

**Wilson Score, Second-Order-Corrected** Interval:

- Shorter in width,
- **Mean** coverage probability closer to \((1-\alpha)\),
- But: no guarantee of pre-specified \(\alpha\) for all situations

See discussion in Brown et al. 2001, Cai 2005
Experimental design if objective is only estimation

1) Chose \(\{n, s\}\) so that \(\text{MSE}(p)\) is minimal
   (Swallow, 1985)

2) Chose \(\{n, s\}\) so that expected interval width is minimal
Software for this problem: **R-package binGroup**

Freeware for R under Windows:
Confidence intervals for group testing (exact and asymptotic methods): `bgtCI()`
Tests (→ p-values) for group testing (exact and asymptotic): `bgtTest()`
Power calculation for a given design: `bgtPower()`
Find the power-optimal number of groups or group size: `nDesign()` and `sDesign()`
Expected interval width: `bgtWidth()`
References


