BIOSTATISTICS 667
Sample Size

Sample size calculations are typically done at the study design stage. They can be based on several considerations such as: 1) Hypothesis testing, Type-I and Type-II error control; 2) Width of confidence intervals; 3) Balancing cost and precision (e.g. minmizing the variance subject to cost constraints).

Sample size calculations can be based on a scalar parameter or a vector of parameters. Here we focus on scalar parameters.

The following structure arises in many problems. There is a scalar parameter θ and estimator $\hat{\theta}$ such that

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0,\lambda^2) \text{ as } n \to \infty,$$

where n quantifies the "sample size" (more on this below). Convergence in distribution allows us to use the cdf of the $N(0,\lambda^2)$ to approximate the cdf of $\sqrt{n}(\hat{\theta}-\theta)$.

In many problems λ is unknown. However, since plugging in a consistent estimator of it does not affect the approximation we are using here, we will proceed as if λ is known.

Suppose we want to test the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1 > \theta_0$ using a 2-sided test with type-I error α and type-II error γ (hence power $1-\gamma$).

Now we define specific values of λ. Under H_0 (i.e. when H_0 is true),

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,\lambda_0^2) \text{ as } n \to \infty,$$

and under H_1,

$$\sqrt{n}(\hat{\theta}-\theta_1) \xrightarrow{d} N(0,\lambda_1^2) \text{ as } n \to \infty.$$

The examples below will make it clear that whether λ_0 and λ_1 are the same depends on the specific problem.

Example (The one-sample normal problem): $Y_1, \cdots, Y_n$ iid $\sim N(\theta,\sigma^2)$, $\hat{\theta} = \bar{Y}$ is the sample mean, $\sqrt{n}(\bar{Y}-\theta) \sim N(0,\sigma^2)$. So $\lambda^2 = \sigma^2$.

Example (The one-sample Bernoulli problem): $Y_1, \cdots, Y_n$ iid Bernoulli(θ), $\hat{\theta} = \bar{Y}$ is the sample mean,

$$T = \sqrt{n}(\bar{Y}-\theta) \xrightarrow{d} N(0,\theta(1-\theta)) \quad \text{ as } n \to \infty.$$

So $\lambda^2 = \theta(1-\theta)$.

Example (The two-sample normal problem): Two independent samples, $Y_1, \cdots, Y_n$ iid $\sim N(\mu_1,\sigma^2)$, $X_1, \cdots, X_n$ iid $\sim N(\mu_2,\sigma^2)$, $\bar{Y}$ and $\bar{X}$ are the sample means, $\theta := \mu_1 - \mu_2$, $\hat{\theta} = \bar{Y} - \bar{X}$,

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0,2\sigma^2) \quad \text{ as } n \to \infty.$$

So $\lambda^2 = 2\sigma^2$. Note that the total sample size is $2n$. But it is trivial to rewrite the above as $\sqrt{2n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, 4\sigma^2)$, but then we have to define λ differently, $\lambda^2 = 4\sigma^2$.

Example (The two-sample Bernoulli problem): Two independent samples, $Y_1, \cdots, Y_n$ iid Bernoulli(μ_1), $X_1, \cdots, X_n$ iid Bernoulli(μ_2), $\bar{Y}$ and $\bar{X}$ are the sample means, $\theta := \mu_1 - \mu_2$, $\hat{\theta} = \bar{Y} - \bar{X}$,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mu_1(1-\mu_1) + \mu_2(1-\mu_2)) \quad \text{as } n \to \infty.$$

So $\lambda^2 = \mu_1(1-\mu_1) + \mu_2(1-\mu_2)$.

For $p \in (0,1)$, define the standard normal quantiles $Z_p = \Phi^{-1}(p)$. For example, $Z_{1-0.05/2} \approx 1.96, Z_{0.9} \approx 1.282, Z_{0.8} \approx 0.8416$.

Define

$$T := \sqrt{n}(\hat{\theta} - \theta_0)/\lambda_0.$$

Then the 2-sided test that rejects H_0 if $|T| > Z_{1-\alpha/2}$ has size α asymptotically.

Now we compute the power of this test,

$$\text{power } = P(T < -Z_{1-\alpha/2}|H_1) + P(T > Z_{1-\alpha/2}|H_1).$$

Note that under H_1 the distribution of $\hat{\theta}$, and hence of T, shifts to the right (as compared to under H_0), the area in the left tail ($T < -Z_{1-\alpha/2}$) becomes negligible and power comes mainly from the right tail (The opposite happens if $\theta_1 < \theta_0$). That is

$$\begin{aligned}
\text{power} &\approx P(T > Z_{1-\alpha/2}|H_1) \\
&= P(\hat{\theta} > \theta_0 + Z_{1-\alpha/2}\lambda_0/\sqrt{n}|H_1) \\
&= P(\sqrt{n}(\hat{\theta} - \theta_1)/\lambda_1 > \sqrt{n}(\theta_0 - \theta_1 + Z_{1-\alpha/2}\lambda_0/\sqrt{n})/\lambda_1|H_1) \\
&\approx 1 - \Phi(\sqrt{n}(\theta_0 - \theta_1)/\lambda_1 + Z_{1-\alpha/2}\lambda_0/\lambda_1).
\end{aligned}$$

If the power is $1 - \gamma$ then

$$\sqrt{n}(\theta_0 - \theta_1)/\lambda_1 + Z_{1-\alpha/2}\lambda_0/\lambda_1 = -Z_{1-\gamma}$$

and

$$\sqrt{n} \;=\; \frac{Z_{1-\alpha/2}\lambda_0/\lambda_1 + Z_{1-\gamma}}{(\theta_1 - \theta_0)/\lambda_1} = \frac{Z_{1-\alpha/2} + Z_{1-\gamma}\lambda_1/\lambda_0}{(\theta_1 - \theta_0)/\lambda_0} = \frac{Z_{1-\alpha/2}\lambda_0 + Z_{1-\gamma}\lambda_1}{\theta_1 - \theta_0}$$

The formula given in some textbooks,

$$\sqrt{n} \;=\; \frac{Z_{1-\alpha/2} + Z_{1-\gamma}}{(\theta_1 - \theta_0)/\lambda},$$

assumes implicitly that $\lambda_0 = \lambda_1$. This assumption is not always valid, as seen in the Bernoulli examples above.

Example (Logistic model for a binary response, width of a confidence interval): A confidence interval for an odds ratio will be computed as $\exp\{\hat{\theta} \pm 1.96se\}$. Suppose that with $n = 63$

subjects, the asymptotic standard error (already divied by $\sqrt{n}$) is $se = 1.5181$ (see ss01.sas). Suppose also that we want the upper end of the confidence interval to be no more than double the point estimate. This means that $\exp(\hat{\theta} + 1.96se) < 2\exp(\hat{\theta})$, i.e. $\exp(1.96se) < 2$, or $1.96se < \log 2$, which implies that $se < (\log 2)/1.96 = 0.35365$. In order to reduce the standard error from 1.5181 to 0.35365, $\sqrt{n}$ must be multiplied by 1.5181/0.35365, i.e. n must be multiplied by $(1.5181/0.35365)^2$. Hence the required sample size is $63(1.5181/0.35365)^2 \approx 1161$.

Example (Longitudinal study, logistic model for a binary response, power of a test): Suppose $\theta_1 - \theta_0 = 0.4437$, type-I error $\alpha = 0.05$ and type-II error $\gamma = 0.2$ (hence power is 0.8), $\lambda_0 = 1.1952$, $\lambda_1 = 1.3581$ (these are based on one subject per group, see ss06.sas). Then

$$\sqrt{n} \quad = \quad \frac{1.96(1.1952) + 0.8416(1.3581)}{0.4437} = 7.856$$

So $n \approx 62$ subjects per group.

An example of balancing cost and precision. K subjects, n observations each, $\mathrm{E}[Y_{ij}] = \mu$, $\mathrm{var}(Y_{ij}) = \sigma^2$, pairwise correlation ρ, $\hat{\mu} = \sum_{i=1}^{K} \bar{Y}_i/K$, $\bar{Y}_i := \sum_{j=1}^{n} Y_{ij}/n$, and

$$\mathrm{var}(\hat{\mu}) = \{1 + (n-1)\rho\}\sigma^2/(nK).$$

Total cost $T = K(A + n)$ dollars is fixed, \$A to recruit, \$1 per observation. Optimal (smallest variance) n is $n^* = \sqrt{A(1-\rho)/\rho}$. If $n^* < 1$, we take $n = 1$. If n^* is between integers m and $m+1$, then $\mathrm{var}(\hat{\mu})$ is evaluated at $n = m$ and $n = m+1$, and the value with the smaller variance is chosen. For each n, the value of K must be as large as possible subject to $K(A + n) \leq T$. Examples:

A	ρ	n^*
≤ 0.25	0.2	1
1	0.2	2
4	0.2	4
16	0.2	8

Some normal quantiles:

p	$\Phi^{-1}(p)$
0.7	0.5244
0.75	0.6745
0.8	0.8416
0.85	1.0364
0.9	1.2816
0.95	1.6449
0.975	1.9600
0.99	2.3263
0.995	2.5758
0.999	3.0902