9.65 The setup is with $Y_1, \dots, Y_n \sim$ Bernoulli (p). iid
We are seeking an MVUE of $p(1-p)$ Let $W = \sum Y_i$.

a) $T = \begin{cases} 1 & \text{if } Y_1 = 1 \text{ and } Y_2 = 0 \\ 0 & \text{otherwise.} \end{cases}$

$$\mathbb{E}T = \sum T(x)\mathbb{P}(x) = \mathbb{P}(Y_1 = 1 \text{ and } Y_2 = 0) \overset{\text{independence}}{=} \mathbb{P}(Y_1 = 1)\cdot\mathbb{P}(Y_2 = 0) = p\cdot(1-p)$$

b) $\mathbb{P}(T = 1 \,/\, W = w)$
We have w ones to distribute on n values in total.
This can be done in $\binom{n}{w}$ different ways.

For $T=1$ we distribute $w-1$ ones on $n-2$ values: $\binom{n-2}{w-1}$.

$$\text{So } \mathbb{P}(T=1 \,/\, W=w) = \binom{n-2}{w-1} \Big/ \binom{n}{w} = \frac{(n-2)!}{(w-1)!(n-\cancel{0}-1)!} \Big/ \frac{n!}{w!(n-w)!} =$$

$$= \frac{(n-2)!}{n!} \cdot \frac{w!}{(w-1)!} \cdot \frac{(n-w)!}{(n-2-(w-1))!} = \frac{1}{n\cdot(n-1)} \cdot w \cdot (n-w) = \frac{w(n-w)}{n(n-1)}.$$

c) $$\mathbb{E}(T/W) = \sum T(x)\mathbb{P}(x/W) = \frac{w(n-w)}{n(n-1)} = \frac{1}{n-1}\cdot\frac{w}{n}\cdot\frac{(n-w)}{n}\cdot n$$

$$= \frac{n}{n-1}\cdot\frac{w}{n}\cdot\left(1-\frac{w}{n}\right) = \frac{n}{n-1}\cdot\bar{Y}\cdot(1-\bar{Y}).$$

This is the Rao-Blackwell estimator; produced from the sufficient statistic of a full rank exponential family.
Hence MVUE.

9.67 $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$.

Joint likelihood is $\exp\left[-\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum(Y_i-\mu)^2\right] =$

$= \exp\left[-\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(\sum Y_i^2 - 2\mu\sum Y_i + n\mu^2\right)\right]$.

So Lehmann-Scheffé's likelihood ratio for samples $x_1, \ldots, x_n$; $y_1, \ldots, y_n$ will cancel all terms that don't depend on the sample:

$$\frac{L(x)}{L(y)} = \exp\left(\frac{-1}{2\sigma^2}\left(\sum X_i^2 - \sum Y_i^2\right) + \frac{\mu}{\sigma^2}\left(\sum X_i - \sum Y_i\right)\right)$$

Unless $\sum X_i^2 = \sum Y_i^2$ and $\sum X_i = \sum Y_i$, this depends on σ^2 and possibly on μ. But if and only if equality, the ratio is independent of μ and σ^2.

Hence $\left(\sum Y_i^2, \sum Y_i\right)$ is a minimal sufficient statistic.

9.97 $Y_1, \ldots, Y_n \sim \text{Geometric}(p)$

a) The mean of the geometric distribution is $\frac{1}{p}$.

So pick $\bar{Y} = \frac{1}{\hat{p}}$ or $\hat{p} = \frac{1}{\bar{Y}}$.

b) The joint likelihood is $p^n (1-p)^{\sum Y_i - n}$ Write $\Sigma = \sum Y_i$

Critical point:

$$\frac{d}{dp} L = \frac{d}{dp} p^n (1-p)^{\Sigma - n} = n p^{n-1} (1-p)^{\Sigma - n} + p^n (\Sigma - n) \cdot (1-p)^{\Sigma - n - 1} \cdot (-1)$$

$$= 0 \quad \text{when} \quad n p^{n-1} (1-p)^{\Sigma - n} = p^n (\Sigma - n)(1-p)^{\Sigma - n - 1}$$

$$n(1-\hat{p}) = \hat{p}(\Sigma - n)$$

$$n - \hat{p}n = \Sigma\hat{p} - \hat{p}n$$

$$n = \Sigma\hat{p}$$

$$\hat{p} = \frac{n}{\sum Y_i} = \frac{1}{\bar{Y}}.$$

Wald's Test

Suppose $\hat{\vartheta}$ is an estimator of ϑ with an approximately normal distribution

We will construct a hypothesis test for $H_0: \vartheta = \vartheta_0$ vs. $H_A: \vartheta > \vartheta_0$.

If ϑ is close to ϑ_0 we would expect $\hat{\vartheta}$ close to ϑ_0 — whereas if $\vartheta > \vartheta_0$ we would expect $\hat{\vartheta} > \vartheta_0$.

So large values for $\hat{\vartheta}$ would lead us to reject H_0.
We get a test on the structure:

$H_0: \vartheta = \vartheta_0$
$H_A: \vartheta > \vartheta_0$
Test statistic: $\hat{\vartheta}$
Rejection Region: $\{\hat{\vartheta} > k\}$ for some k.

We choose the threshold k according to the level we want for the test: $k = \vartheta_0 + z_\alpha \sigma_{\hat{\vartheta}}$ where $F^{-1}_{N(0,1)}(1-\alpha)$.

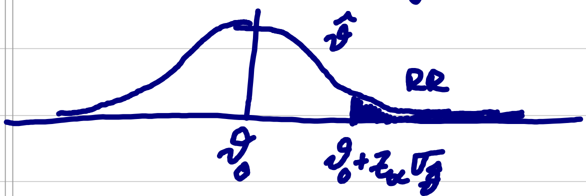
We can rewrite the test statistic;

$\hat{\vartheta} > \vartheta_0 + z_\alpha \sigma_{\hat{\vartheta}}$ is equivalent to

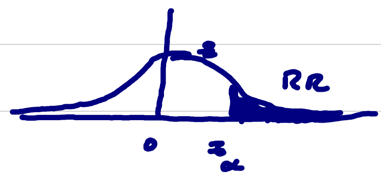
$\dfrac{\hat{\vartheta} - \vartheta_0}{\sigma_{\hat{\vartheta}}} > z_\alpha$ yielding: Test statistic: $z = \dfrac{\hat{\vartheta} - \vartheta_0}{\sigma_{\hat{\vartheta}}}$

Rejection Region: $\{z > z_\alpha\}$

Graphically, we can illustrate the construction using the distribution of $\hat{\vartheta}$ under H_0:



or using z:



The transformation comes from going from (under H_0)

$$\hat{\vartheta} \sim N(\vartheta_0, \sigma_{\hat{\vartheta}}^2) \quad \text{to} \quad \frac{\hat{\vartheta} - \vartheta_0}{\sigma_{\hat{\vartheta}}} \sim N(0,1)$$

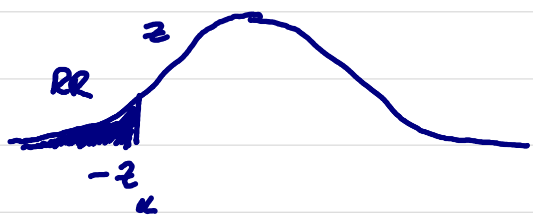
With the same transformation we can construct a lower tail test:

$H_0: \vartheta = \vartheta_0$

$H_A: \vartheta < \vartheta_0$

Test statistic: $\dfrac{\hat{\vartheta} - \vartheta_0}{\sigma_{\hat{\vartheta}}} =: z$

Rejection Region: $\{z < -z_\alpha\}$



For a two-tailed hypothesis $H_A: \vartheta \neq \vartheta_0$, we get to divide the probability mass between the tails.

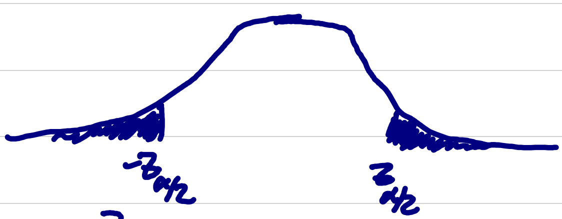
Since we are assuming a normal distribution the easy way to do this is to assign $\alpha/2$ to each tail:

$H_0: \vartheta = \vartheta_0$

$H_A: \vartheta \neq \vartheta_0$

Test statistic: $z = \dfrac{\hat{\vartheta} - \vartheta_0}{\sigma_{\hat{\vartheta}}}$

Rejection Region: $\{ |z| < z_{\alpha/2} \}$



False Acceptance and power calculation

Definition The power of a test at some parameter value ϑ is

$$\text{power}(\vartheta) = P(\text{rejection} \mid \vartheta)$$

Hence, $\alpha = \text{power}(\vartheta_0)$ and $\beta(\vartheta) = 1 - \text{power}(\vartheta)$.

In general, power can be quite complicated to calculate; but in the Wald case it's easier.

Let's focus on the upper tail case:

$$RR = \{ \hat{\vartheta} : \hat{\vartheta} > k)$$

$$\beta(\vartheta_A) = \mathbb{P}(\text{acceptance} \mid \vartheta_A) = \mathbb{P}(\hat{\vartheta} \le k \mid \vartheta_A) =$$

$$= \mathbb{P}\left(\frac{\hat{\vartheta} - \vartheta_A}{\sigma_{\hat{\vartheta}}} \le \frac{k - \vartheta_A}{\sigma_{\hat{\vartheta}}} \,\middle|\, \vartheta_A \right)$$

When we apply this to some of our known estimators, we can get very concrete connections between sample sizes and power.

Say, for $\bar{X}$ estimating μ, the Wald test yields

$$\beta = \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} \le \frac{k - \mu_A}{\sigma/\sqrt{n}} \right) = \mathbb{P}(Z \le -z_\beta)$$

where k is chosen for

$$\alpha = \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{k - \mu_0}{\sigma/\sqrt{n}} \right) = \mathbb{P}(Z > z_\alpha)$$

So we get: $\dfrac{k - \mu_0}{\sigma/\sqrt{n}} = z_\alpha \qquad \dfrac{k - \mu_A}{\sigma/\sqrt{n}} = -z_\beta$

Solve for $k = \mu_0 + z_\alpha\left(\frac{\sigma}{\sqrt{n}}\right) = \mu_A - z_\beta\left(\frac{\sigma}{\sqrt{n}}\right)$

Hence $(z_\alpha + z_\beta)\left(\frac{\sigma}{\sqrt{n}}\right) = \mu_A - \mu_0$

So $$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_A - \mu_0)^2}$$

Hence, out of
- power
- level
- sample size
- variance
- detection distance

choose any 5 to compute the 6th.

Duality

The expressions we are writing down now look familiar — we had something very similar with confidence intervals earlier.

This is a general pattern: a level α acceptance region corresponds to a $1-\alpha$ confidence region

If C is a $1-\alpha$ confidence region for ϑ, we can build a test that rejects exactly when $\vartheta_0 \in C$

$$\alpha = \mathbb{P}(\text{rejection} \mid H_0) = \mathbb{P}(\vartheta_0 \notin C \mid \vartheta = \vartheta_0) = 1 - \mathbb{P}(\vartheta \in C) = 1-(1-\alpha).$$

If R is a level α rejection region parametrized by ϑ_0 for some $H_0 : \vartheta = \vartheta_0$, then

$C(\hat{\vartheta}) = \{ \vartheta_0 : \hat{\vartheta} \text{ is not rejected by } R(\vartheta_0) \}$ is a $1-\alpha$ confidence region.

$$\mathbb{P}(\vartheta \in C(\hat{\vartheta}) \mid H_0) = \mathbb{P}(\hat{\vartheta} \text{ is not rejected by } \vartheta_0 \mid \vartheta = \vartheta_0) = \mathbb{P}(\hat{\vartheta} \text{ is not rejected by } \vartheta)$$
$$= 1 - \mathbb{P}(\hat{\vartheta} \text{ is rejected by } \vartheta \mid H_0) = 1 - \alpha.$$