ii. $X_1, \ldots, X_k \sim \mathcal{NB}\text{in}(\theta, m)$

- Number of trials it takes to get m successes, if each has success probability θ
- Likelihood for one observation $L(\theta) = \binom{X-1}{m-1}\theta^m(1-\theta)^{X-m}$
- Log likelihood for one observation $\ln(\binom{X-1}{m-1}) + m\ln(\theta) + (X-m)\ln(1-\theta)$
- Overall log likelihood $\Sigma_{j=1}^n[\ln(\binom{X_j-1}{m-1}) + m\ln(\theta) + (X_j - m)\ln(1-\theta)] = \Sigma_{j=1}^n \ln(\binom{X_j-1}{m-1}) + nm\ln(\theta) + (\Sigma_{j=1}^n X_j - nm)\ln(1-\theta)$
- $l'(\theta) = nm/\theta - (\Sigma_{j=1}^n X_j - nm)/(1-\theta)$
- MLE satisfies $nm\hat{\theta} - (\Sigma_{j=1}^n X_j - nm)/(1-\hat{\theta}) = 0$, $nm - nm\hat{\theta} - (\Sigma_{j=1}^n X_j - nm)\hat{\theta} = 0$, $nm - \Sigma_{j=1}^n X_j\hat{\theta} = 0$, $\hat{\theta} = m/\bar{X}$.
- By invariance, $\hat{\mu} = \bar{X}$, unbiased.

iii. Bivariate Normal

- $X_i \sim \mathcal{N}(0,1)$, $Y_i|X_i \sim \mathcal{N}(\rho X_i, 1-\rho^2)$
- Log likelihood is

$$l(\rho) = \sum_{j=1}^n \left[-\frac{X_j^2}{2} - \frac{(Y_j - \rho X_j)^2}{2(1-\rho^2)} - \ln(2\pi) - \tfrac{1}{2}\ln(1-\rho^2)\right]$$

- $l'(\rho) =$

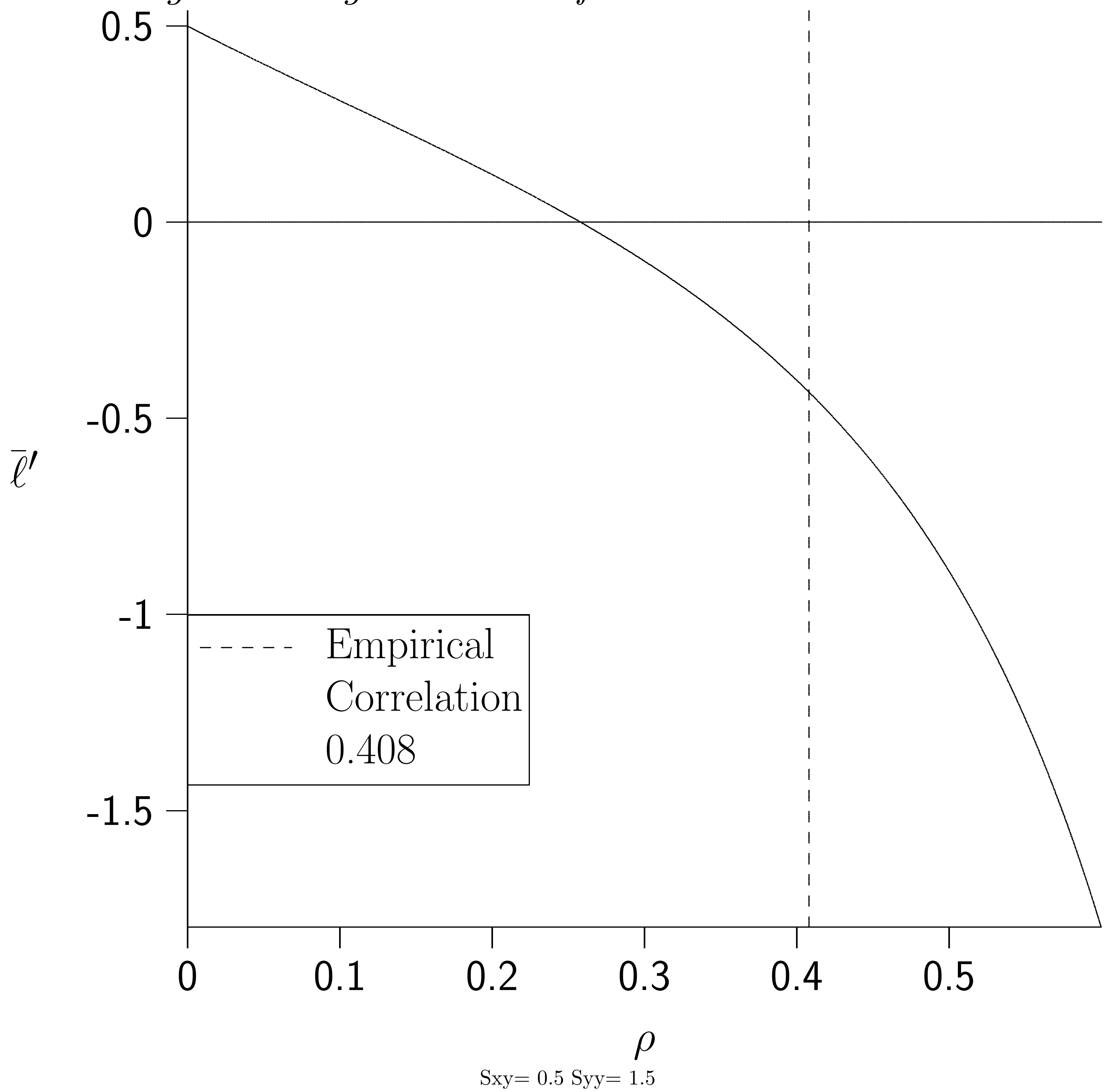
$$= \sum_{j=1}^{n} [\frac{-(2(1-\rho^2)(Y_j-\rho X_j)2(-X_j)-(Y_j-\rho X_j)^2 4\rho)}{4(1-\rho^2)^2}$$
$$+ \rho/(1-\rho^2)]$$
$$= \sum_{j=1}^{n} [\frac{(1-\rho^2)(Y_j-\rho X_j)X_j-(Y_j^2-2\rho X_j Y_j+\rho^2 X_j^2)\rho)}{(1-\rho^2)^2}$$
$$+ \rho/(1-\rho^2)]$$
$$= n\frac{\rho-\rho^3+S_{xy}+\rho^2 S_{xy}-2\rho S_{yy}}{(1-\rho^2)^2}$$

- $S_{xx} = \Sigma_{j=1}^n X_j^2/n$, $S_{yy} = \Sigma_{j=1}^n Y_j^2/n$, $S_{xy} = \Sigma_{j=1}^n X_j Y_j/n$.
- Gives cubic equation in ρ
- Has factor $S_{xy}-\rho$ if $S_{yy}=1$.
- See Fig 10.

iv. Weibull

- $X_1,\ldots,X_k \sim \exp(-x^\alpha/\theta)\theta^{-1}\alpha x^{\alpha-1}$, α known
- $X_j^\alpha \sim \mathcal{E}(\theta)$
- MLE $\hat\theta = \Sigma_{j=1}^n X_j^\alpha/n$
- $\mathrm{E}[X_j] = \Gamma(1+1/\alpha)\theta^{1/\alpha}$
- m.o.m.e. satisfies $\bar X = \Gamma(1+1/\alpha)\theta^{1/\alpha}$ if and only if $\hat\theta = (\bar X/\Gamma(1+1/\alpha))^\alpha$

Fig. 10: Log Likelihood for Normal Correlation



c. Rules for function maximization:

i. Differentiate the log likelihood function.

ii. Equate these to 0.

iii. Solve

iv. Ensure that the solution is a local maximum,

- possibly by checking to see that the second derivative < 0 at the proposed maximum,

v. Ensuring that our local max is a global max, possibly by checking

- that either $L''(\boldsymbol{\theta}) < 0 \, \forall \boldsymbol{\theta}$

d. Properties of Maximum Likelihood estimates?

i. Are they unbiased? No, but almost...

ii. Are they sufficient? No, but almost...

iii. Are they efficient? No, but almost...

Aa

11. See Table 3 for a summary.

WMS: 10.1-10.2

III. Hypothesis Testing:

A. What if instead of asking what values of θ are reasonable, we ask the question: Is the value 0 (or any other number of interest a $priori$) reasonable?

B. Formal statement of problem:

Table 3: Summary of Estimators

Distribution	Density	Estimator m.o.m.e.	Estimator m.l.e.
$\mathcal{B}\text{in}(m,\theta)$	$\binom{m}{x}\theta^x(1-\theta)^m$		$\hat{\theta} = \bar{X}/m$
$\mathcal{NB}\text{in}(m,\theta)$	$\binom{x-1}{m-1}\theta^x(1-\theta)^m$		$\hat{\theta} = m/\bar{X}$
$\mathcal{P}(\theta)$	$\exp(-\theta)\theta^x/x!$		$\hat{\theta} = \bar{X}$
$\mathcal{N}(\mu,\sigma^2)$	$\frac{\exp(-(x-\mu)^2/2)}{\sigma\sqrt{2\pi}}$	$\hat{\mu} = \bar{X}$, $\hat{\sigma} = \sqrt{\frac{\Sigma_{j=1}^n (X_j-\bar{X})^2}{n}}$	
$\mathcal{U}(0,\theta)$	$\begin{cases} 1 & \text{if } x \in [0,\theta] \\ 0 & \text{ow} \end{cases}$	$\hat{\theta} = 2\bar{X}$	$\hat{\theta} = \max(X_j)$
$\mathcal{C}(\theta)$	$\frac{1}{\pi(1+(x-\theta)^2)}$	Doesn't exist	Exists; no closed form expression
$\mathcal{W}(\alpha,\theta)$	$\exp\left(-\frac{x^\alpha}{\theta}\right)\left(\frac{\alpha}{\theta}\right) \times x^{\alpha-1}$	$\hat{\theta} = \frac{\Sigma_{j=1}^n X_j^{1/\alpha}}{n}$	$\hat{\theta} = \left(\frac{\bar{X}}{\Gamma(1+\frac{1}{\alpha})}\right)^\alpha$

1. Given data $X_1, \cdots, X_n$ from a model

$$f_{X_1,\cdots,X_n}(x_1, \cdots, x_n; \theta),$$

2. wish to differentiate between two hypothesis
   a. the $null\,hypothesis$, that θ takes on a value in some set, vs.
   b. $alternative\,hypothesis$ that θ is in some other set.
3. Each of these hypotheses is called
   a. $simple$ if the set consists of one element
   b. $composite$ otherwise

4. A *test* is a rule that decides on the basis of the data reject null or don't reject null.

   a. Arrange results into Table 4.

Table 4: Classification of Hypothesis Test Results

	Test Rejects	Test Doesn't Reject
Null False	Correct	Type II Error
Null True	Type I Error	Correct

   b. Type I error rate called *level* or *size* .

C. What makes a good test? Among test of a fixed size,

1. Why is the alternative of randomly rejecting with the same probability, without regard to data, a bad test?

2. Want the Type II error rate small.

   a. Want the *power* , or probability of correct decision under the alternative, high.

D. General Construction

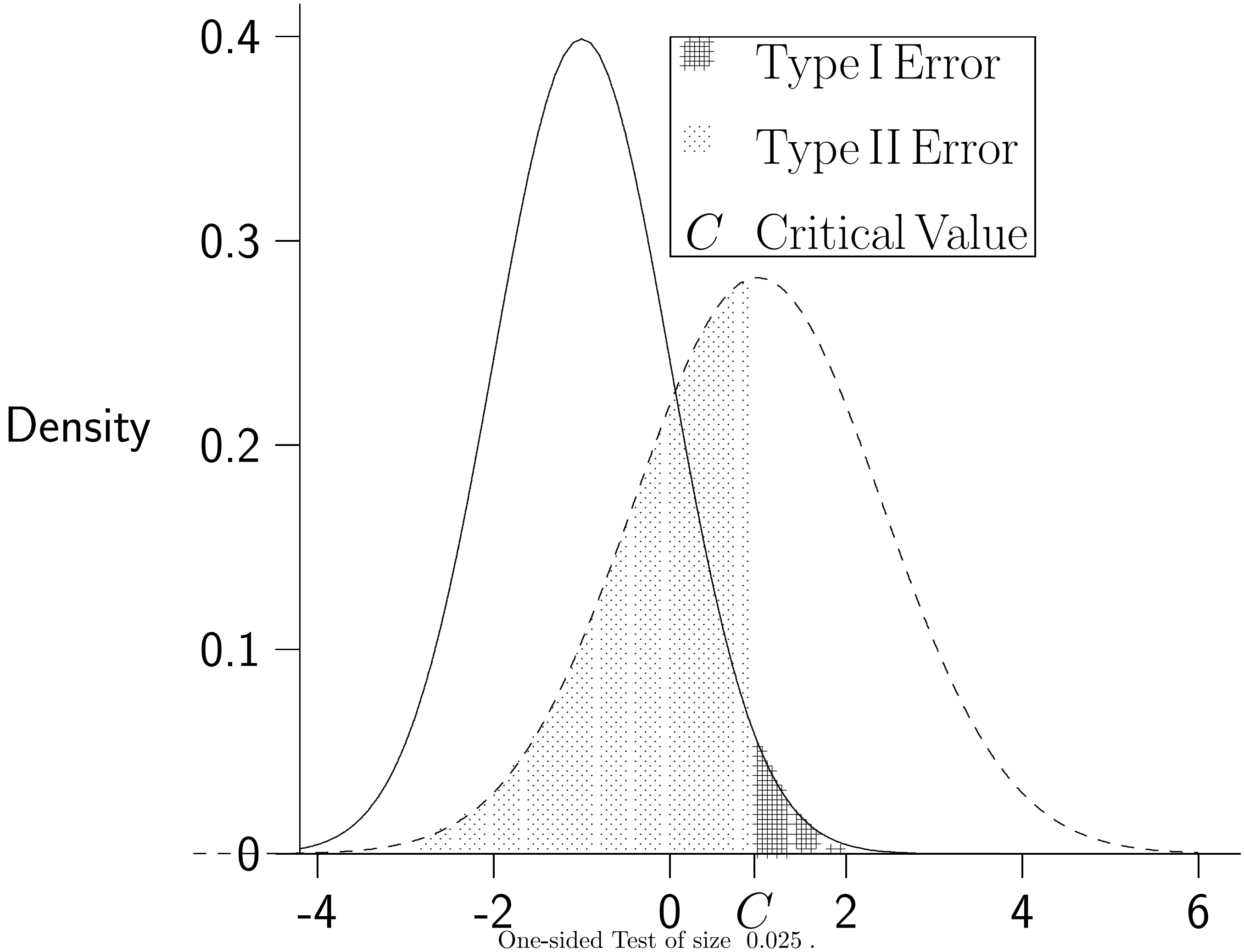
1. Create a *test statistic*

   a. that gives more evidence against H_0 the bigger it is,

2. Select a *critical value* to which test statistic is compared.

   a. reject H_0 if the statistic is equal to or larger than critical value.

Fig. 11: Error Types in Normal Testing



3. See Fig 11 for a diagram.

E. Examples:

1. Normal Case:

   a. $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$

   b. $H_0 : \mu = \mu_0 (= 0)$, $H_A : \mu = \mu_A = 1$.

   c. Pick type I error rate α

   d. Use as test statistic $\bar{X}$,

i. find the critical value C to make the test "Reject if $\bar{X} \geq c$" have size α.

ii. Equivalently, ask for the value of C such that under H_0, $\mathrm{P}[Q \geq C] = \alpha$.

- $\mathrm{P}_0\left[\bar{X} > C\right] = \bar{\Phi}((C - \mu_0)/(\sigma/\sqrt{n}))$
- $\bar{\Phi}((C - \mu_0)/(\sigma/\sqrt{n})) = \alpha$
- $(C - \mu_0)/(\sigma/\sqrt{n})) = z_\alpha$ implies $C = \mu_0 + \sigma z_\alpha/\sqrt{n}$.
- Here you can choose C and get α, or choose α to get C.

iii. Rejection region is $\{\bar{X} \geq C\}$.

2. Binomial Case:

a. Problem:

i. $X \sim \mathcal{B}\text{in}(m, \pi)$

ii. $H_0 : \pi = \pi_0 (= .65)$, $H_A : \pi > \pi_0$.

iii. type I error rate α

b. Use as test statistic observed defective count X,

c. For what value of C does $\mathrm{P}_0[X \geq C] = \alpha$?

d. Rejection region is then $\{X \geq C\}$.

i. Via approximation, need C such that $\mathrm{P}[Q \geq C] \approx \alpha$.

- Since $(X - m\pi)/\sqrt{m\pi(1 - \pi)} \sim N(0, 1)$,

$\mathrm{P}\left[(X - m\pi)/\sqrt{m\pi(1-\pi)} \geq z_\alpha\right] \approx \alpha$,

- and $\mathrm{P}\left[\geq m\pi + z_\alpha\sqrt{m\pi(1-\pi)}\right] \approx \alpha$;
- hence $m\pi + z_\alpha\sqrt{m\pi(1-\pi)/m}$ is the approximate critical value.

ii. If ex. $m = 10$, $\alpha = 0/025$, rejection region is $\{X \geq 6.5 + \sqrt{.65 \times .35 \times 10}1.96 = 9.456$.

- Rejection region is $\{X \geq 9.456\}$.
- Level is $\mathrm{P}_0\,[X \geq 9.456] = \mathrm{P}_0\,[X = 10] = .65^{10} = .0134$: Level too low
  - ▷ Try lower C: $\mathrm{P}_0\,[X \geq 9.2] = \mathrm{P}_0\,[X = 10] = .65^{10} = .0134$: No effect
  - ▷ Try lower C: $\mathrm{P}_0\,[X \geq 8.8] = 10 * .65^9 * .35 + .65^{10} = 0.0860$: Too high
  - ▷ Take away: For discrete examples, set of available levels is restricted.
  - ▷ Null hypothesis that has $\alpha = 0.025$ for $C = 9.5$ is $\pi_0^{10} = \alpha$, or $\pi_0 = (0.025)^{1/10} = .691$.

WMS: 10.7

F. From Simple to Composite Hypotheses, when T is generally larger for larger values of θ .

1. Our test of a null hypothesis vs. a simple alternative generally used the alternative only to indicate whether rejection region is of the form $\{T \geq C\}$ or $\{T \leq C\}$
   a. and also used θ_0 only to set C , and not to determine the shape of the region.
2. Hence such a test of $H_0 : \theta = \theta_0$ vs. $H_A : \theta = \theta_1$ for some $\theta_1 > \theta_0$ is also adequate for $H_A : \theta > \theta_0$.
   a. Under these circumstances, test is the same for all of the alternatives.
   b. Ex., For $X \sim \mathcal{B}\text{in}(\pi, m)$, $H_0 : \pi = .691$, $\alpha = 0.025$, $H_A : \pi = \pi_A$ has rejection region $\{X = 10\}$, for any $\pi_A > .691$.
   c. So we can use this test for the composite alternative $\pi > .691$.
   d. Called a *one-sided test* .
3. Furthermore, if we were to reduce θ_0 , this would generally have the effect of reducing the test level.
4. Hence the same test can be used for $H_0 : \theta = \theta_0$ vs. $H_A : \theta > \theta_0$.

WMS: 10,5

G. Sometimes alternatives are on both sides of the null, level α .

1. In this case, usually construct two one-sided tests in opposite directions, levels α_L and α_U , such that $\alpha_L + \alpha_u = \alpha$.
   a. Often, but not always, $\alpha_L = \alpha_U$.
   b. Separate rejection regions $\{T \leq C_l\}$ and $\{T \geq C_U\}$.
   c. Resulting rejection region is $\{T \leq C_l\} \cup \{T \geq C_U\}$.
   d. Such a procedure is called a *two-sided test* .

H. Solution using c.i.s

1. Null hypothesis θ_0 vs alternative that $\theta = \theta_1 > \theta_0$.
2. Construct test
   a. Make a $1 - \alpha$ c.i.
      i. One-sided of form $(L(\text{data}), \infty]$
   b. Reject H_0 if θ_0 is outside the c.i.; don't reject otherwise
   c. Type I error α .