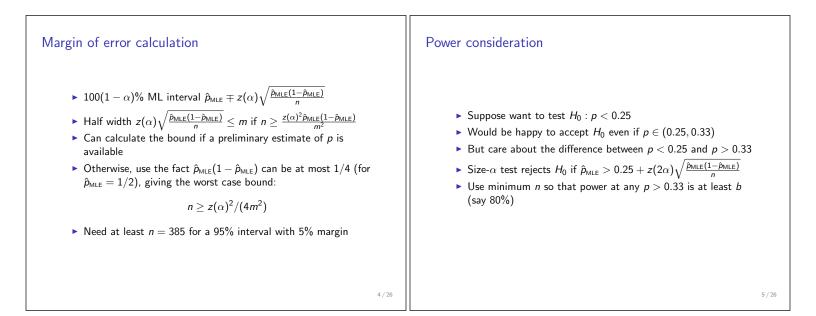
	Uses of frequentist calculations
Design of Studies and Comparing Procedures Surya Tokdar	<list-item><list-item><list-item><list-item><list-item><list-item><list-item><list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item></list-item>

Sample size determination	Margin of error consideration
 To estimate the prevalence of child malnutrition in a country Quantity of interest: p =proportion of malnourished children under age 10 Survey n children under 10 years, data X = number of malnourished What n to use? [Surveying is expensive] 	 Model X ~ Bin(n, p), p ∈ (0, 1) Want a 95% confidence interval for p with 5% margin of error Means interval half-width is no larger than .05 Use smallest n to meet margin of error
2 / 26	3 / 26



Power calculation

• Fix $p^* > 0.33$ • Power at p^* equals $P_{[X|p^*]}\left(\hat{p}_{MLE} > 0.25 + z(2\alpha)\sqrt{\frac{\hat{p}_{MLE}(1-\hat{p}_{MLE})}{n}}\right)$ $= P_{[X|p^*]}\left(T > \hat{\delta}(p^*) + z(2\alpha)\right)$ where • $T = \frac{\hat{p}_{MLE} - p^*}{\sqrt{\hat{p}_{MLE}(1-\hat{p}_{MLE})/\sqrt{n}}} \sim AN(0,1)$ • $\hat{\delta}(p^*) = \frac{0.25 - p^*}{\sqrt{\hat{p}_{MLE}(1-\hat{p}_{MLE})/\sqrt{n}}} \approx \sqrt{n} \frac{0.25 - p^*}{\sqrt{p^*(1-p^*)}}$ • So power at p^* is approximately $1 - \Phi(\hat{\delta}(p^*) + z(2\alpha))$

From power to sample size

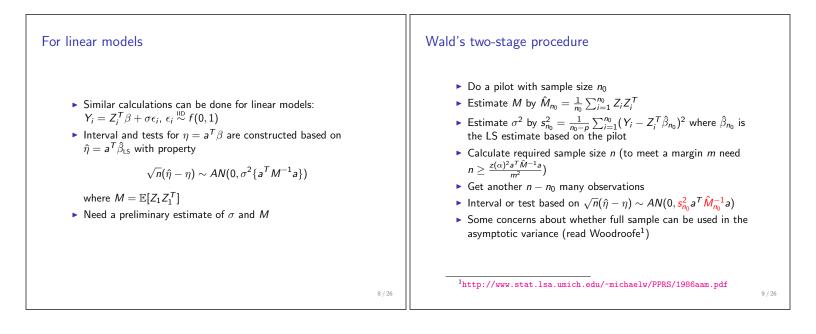
 \blacktriangleright Minimum power is at $p^*=0.33$ (among all $p^*>0.33)$ with $\hat{\delta}(p^*)\approx -0.17\sqrt{n}$

▶ Need $1 - \Phi(-0.17\sqrt{n} + z(2\alpha)) \ge b$ which happens when

$$n \ge \left\{ \frac{z(2\alpha) - \Phi^{-1}(1-b)}{0.17} \right\}^2$$

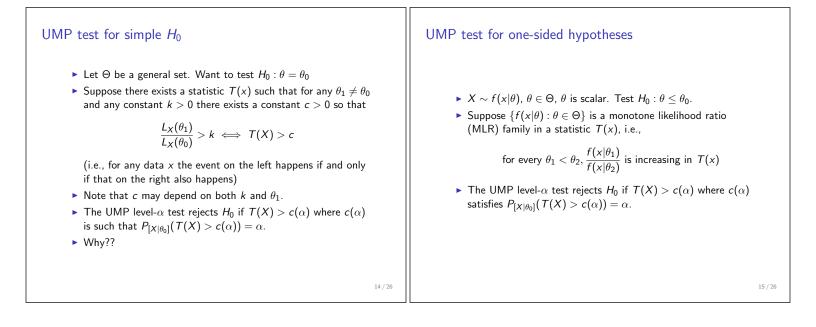
For a size-5% test to have at least 80% power at all p > 0.33 we need at least n = 214.

7 / 26

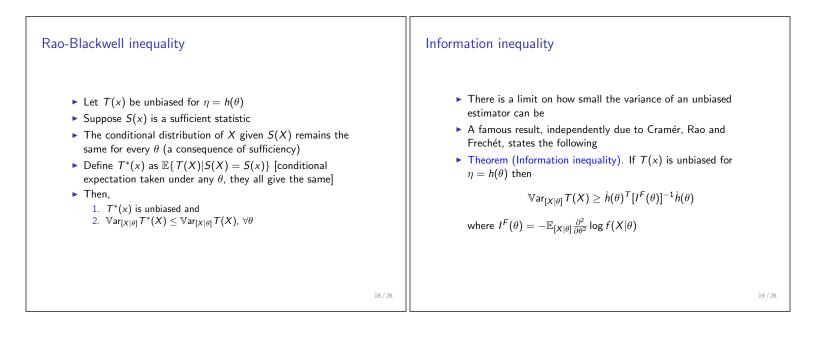


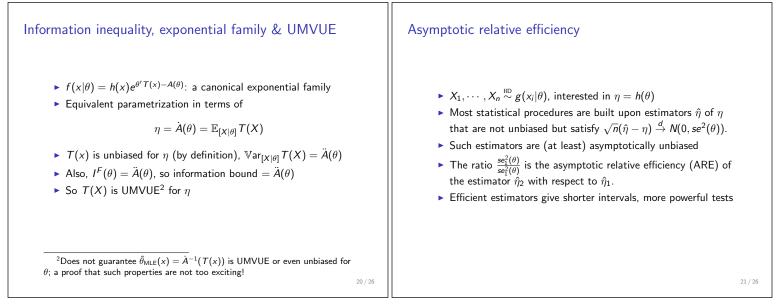
6 / 26

Most powerful tests	The Neyman-Pearson lemma
 X ~ f(x θ), θ ∈ Θ Want to test H₀ : θ ∈ Θ₀ A level-α test δ is most powerful at a θ₁ ∈ Θ \ Θ₀ if there is no level-α test δ' with β(θ₁, δ') > β(θ₁, δ). A test δ is said to be the uniformly most powerful level-α test if δ is a most powerful level-α test at every θ₁ ∈ Θ \ Θ₀. 	 Consider Θ = {θ₀, θ₁} and we want to test H₀ : θ = θ₀ The Neyman-Pearson lemma says that the most power level α test is given by Reject H₀ if Λ(X) = L_X(θ₁)/L_X(θ₀) > k where k is such that P_[X θ₀](Λ(X) > k) ≤ α₀ but P_[X θ₀](Λ(X) > k') > α₀ for all k' < k. i.e., k is the smallest threshold satisfying the size condition
12 / 26	13 / 26



UMP unbiased	Optimal estimation: UMVUE
 Not too many families are MLR in any statistic Consequently, not too many UMP tests are known A slightly less demanding criterion is to be UMP among all tests that are unbiased, i.e., whose power function at any θ ∉ Θ₀ is at least as much as the size α In most examples, ML tests are UMP among this class of tests 	 In general, tests and intervals are based on estimators How do we compare estimators? An old favorite of classical statisticians' is the uniformly minimum variance unbiased estimator (UMVUE). An estimator T(x) of η = h(θ) is said to be UMVUE if T(x) is unbiased, i.e., E_[X θ]T(X) = h(θ) ∀θ and for any unbiased T(X), Var_[X θ]T(X) ≤ Var_[X θ]T(X), ∀θ





Efficiency of MLE	Efficiency of MLE (contd)
 Usually √n(η̂_{MLE} - η) → N(0, h(θ) TI₁^F(θ)⁻¹h(θ)). That is, asymptotically, MLE is unbiased and meets the information bound (note, I^F(θ) = nI₁^F(θ)) So information inequality suggests that η̂_{MLE} should have smaller variance than any other estimate that is asymptotically normal with mean η. A precise statement is given below. 	 If θ̃ is a minimum contrast estimate of θ then η̃ = h(θ̃) satisfies √n(η̃ - η) → N(0, s̃e²(θ)) with s̃e²(θ) ≥ μ(θ)^T I₁^F(θ)⁻¹μ(θ) That is, the MLE is efficient among all minimum contrast estimators that are asymptotically normal.
22 / 26	23 / 26

Robust procedures

Median is more robust than mean

- For non-parametric models, optimal procedures are difficult to find
- A non-parametric model contains many regular parametric sub-models, their corresponding ML procedures are optimal within that sub-model
- However, it might be possible to find a procedure that remains competitive across all sub-models
- Such procedures are called robust (there are other definitions of robustness, but the essence is this)
- Two examples

• $X_i = \mu + \epsilon_i, \ \epsilon_i \stackrel{\text{IID}}{\sim} f \in \mathcal{F}_{sym}, \ \mu \in (-\infty, \infty)$

- \mathcal{F}_{sym} contains all pdfs that are symmetric around 0
- Two estimators of μ : \bar{x} and x_{med}
- ARE of x_{med} w.r.t \bar{x} is $4f_0(0)\sigma_f^2$, where $\sigma_f^2 = \int x^2 f(x) dx$
- Consider $f = (1 \epsilon)N(0, \sigma^2) + \epsilon N(0, k\sigma^2)$
- ARE = 64% for ϵ = 0 (best case for mean, being the MLE for that sub-model)
- $\blacktriangleright \ \, {\rm For} \ \epsilon > {\rm 0}, \ {\rm ARE} \rightarrow \infty \ {\rm as} \ k \rightarrow \infty$
- Regression counterpart: least-squares vs median (quantile) regression

25 / 26

24 / 26

Wilcoxon's rank-sum test vs t-test

- $X_i \stackrel{\text{\tiny IID}}{\sim} g(x), \ Y_j \stackrel{\text{\tiny IID}}{\sim} g(x-\theta), \ g \ \text{arbitrary}$
- To test $H_0: \theta = 0$
- Let δ_1 be size- α t-test, δ_2 be size- α W's test
- \blacktriangleright Fixed desired power $\beta > \alpha$
- Pitman efficiency: Consider a sequence (θ_k)_{k≥1} → 0. Let n_{k,1} and n_{k,2} be sample sizes needed by δ₁ and δ₂ to have power β at θ_k. Relative Pitman efficiency of δ₂ to δ₁ is lim_{k→∞} n_{k,2}/n_{k,2}
- \blacktriangleright Limit does not depend on β
- Relative P efficiency of W's test w.r.t t-test is
 - 1. 0.95 when g is normal (the best case scenario in favor of t-test)
 - 2. Can be arbitrarily large (and approach $\infty)$ for non-normal (heavy tailed) g

26 / 26