The Power

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Hypothesis testing

1. Choose a conjecture
2. Choose null hypothesis
3. Choose test statistic
4. Decide significance level α
5. Gather data and compute test statistic
6. Compute p value
7. Reject the null hypothesis if p value $\leq \alpha$

$\mathrm{P}\left(\text{False positive}\right) \leq \alpha$

1. Choose a conjecture
2. Choose null hypothesis
3. Choose test statistic
4. Decide significance level α
5. Gather data and compute test statistic
6. Compute p value
7. Reject the null hypothesis if p value $\leq \alpha$

Is it enough to control false positives?

No, what about false negatives?

The power is the probability of a true positive

The power function

Parametric testing:

Distribution of test statistic depends on parameters θ

The power function

$$\text{pow}(\theta) := \mathrm{P}\left(\text{pv}\left(\tilde{t}_\theta\right) \leq \alpha\right)$$

maps θ to probability of rejecting the null hypothesis

Power function

Null hypothesis: $\theta \in \Theta_{\text{null}}$

For $\theta \in \Theta_{\text{null}}$, do we want $\text{pow}(\theta)$ large or small?

$$\text{pow}(\theta) = \mathrm{P}\left(\text{False positive}\right) \leq \alpha$$

Power function

Parameters associated to alternative hypothesis: Θ_{alt}

For $\theta \in \Theta_{\text{alt}}$, do we want $\text{pow}(\theta)$ large or small?

$$\text{pow}(\theta) = \mathrm{P}\left(\text{True positive}\right)$$

This is the power!

Die rolls

Conjecture: Probability of rolling $3 > 1/6$

Null hypothesis: Probability of rolling $3 \leq 1/6$

Test statistic: Number of 3s (out of n)

Distribution?

Binomial with parameters n and θ

Rejection region

$$\mathcal{R} := \{\tau : \mathrm{P}\left(\tilde{t}_{\text{null}} \geq \tau\right) \leq \alpha\}$$

Rejection threshold

$$\tau_{\text{thresh}} := \min_{1 \leq \tau \leq n} \{\tau : \mathrm{P}\left(\tilde{t}_{\text{null}} \geq \tau\right) \leq \alpha\}$$

We reject if and only if test statistic $\geq \tau_{\text{thresh}}$

If $\tau < \tau_{\text{thresh}} \implies \tau \notin \mathcal{R}$

If $\tau \geq \tau_{\text{thresh}} \implies \tau \in \mathcal{R}$

$$\mathrm{P}\left(\tilde{t}_{\text{null}} \geq \tau\right) \leq \mathrm{P}\left(\tilde{t}_{\text{null}} \geq \tau_{\text{thresh}}\right) \leq \alpha$$

Rejection threshold

$$\begin{aligned}\tau_{\text{thresh}} &:= \min_{1 \le \tau \le n} \left\{\tau : \mathrm{P}\left(\tilde{t}_{\text{null}} \geq \tau\right) \leq \alpha\right\} \\ &= \min_{1 \le \tau \le n} \left\{\tau : \sum_{i=\tau}^{n} \binom{n}{i} \theta_{\text{null}}^{i} \left(1 - \theta_{\text{null}}\right)^{n-i} \leq \alpha\right\}\end{aligned}$$

Dependence on significance level α?

$\alpha := 0.01 \implies \tau_{\text{thresh}} = 27$

$\alpha := 0.05 \implies \tau_{\text{thresh}} = 24$

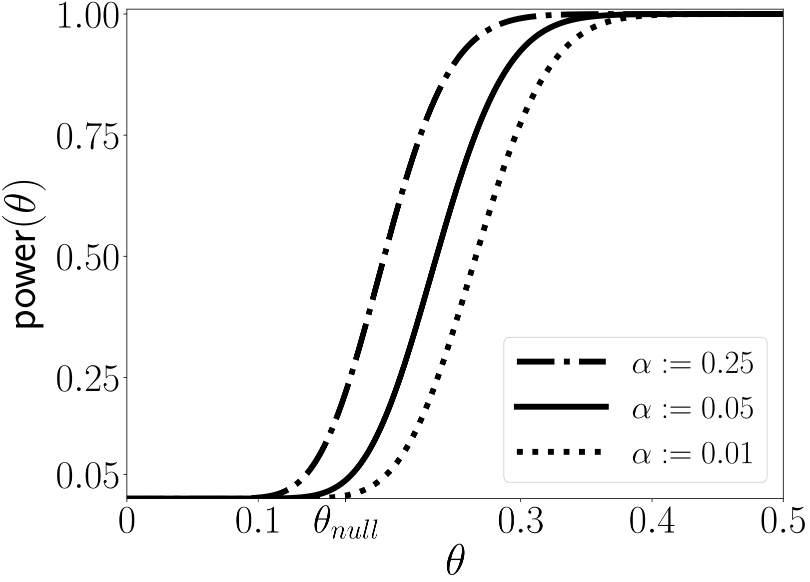
$\alpha := 0.25 \implies \tau_{\text{thresh}} = 20$

Power function

$$
\begin{aligned}
\mathsf{pow}(\theta) &:= \mathrm{P}\left(\mathsf{pv}\left(\tilde{t}_\theta\right) \leq \alpha\right) \\
&= \mathrm{P}\left(\tilde{t}_\theta \geq \tau_{\mathsf{thresh}}\right) \\
&= \sum_{i=\tau_{\mathsf{thresh}}}^{n} \binom{n}{i} \theta^i \left(1-\theta\right)^{n-i}
\end{aligned}
$$

Dependence on θ?

$n := 100$



1.00
0.75
0.50
0.25
0.05
power(θ)
0
0.1
θ_{null}
0.3
0.4
0.5
θ
$\alpha := 0.25$
$\alpha := 0.05$
$\alpha := 0.01$

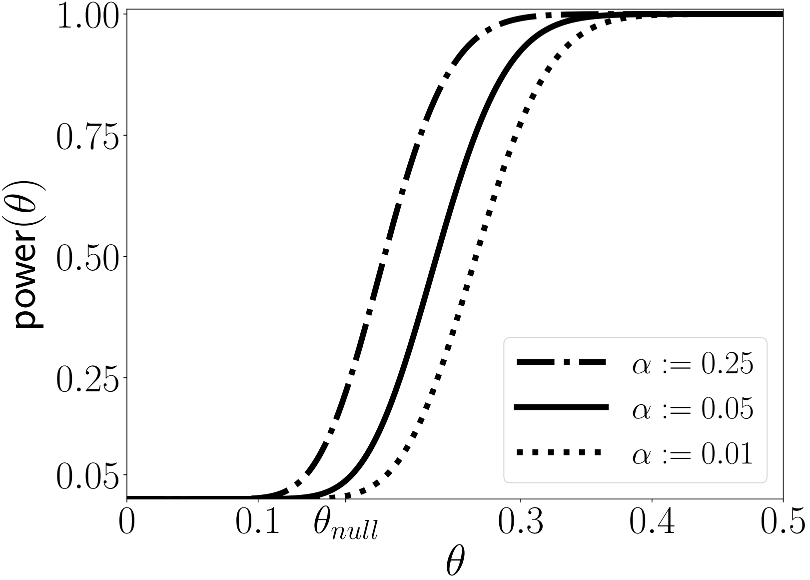
Dependence on α?

For any $\alpha_1 \leq \alpha_2$

$$\mathrm{P}\left(\mathsf{pv}\left(\tilde{t}_\theta\right) \leq \alpha_1\right) \leq \mathrm{P}\left(\mathsf{pv}\left(\tilde{t}_\theta\right) \leq \alpha_2\right)$$

Great, so let's just increase α, right?

$n := 100$

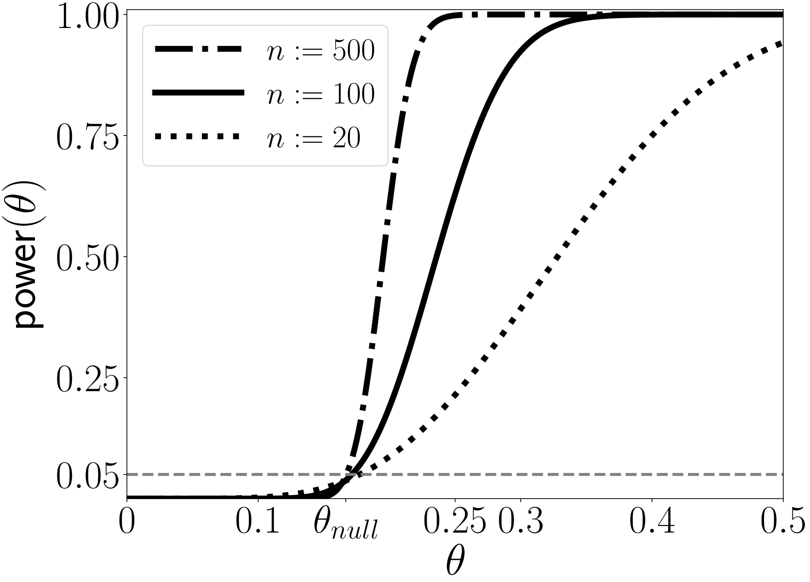


power(θ)
1.00
0.75
0.50
0.25
0.05
0
0.1
θ_{null}
0.3
0.4
0.5
θ
$\alpha := 0.25$
$\alpha := 0.05$
$\alpha := 0.01$

Key question

How can we increase power while keeping α fixed? More data!

$\alpha := 0.05$



$n := 500$
$n := 100$
$n := 20$
1.00
0.75
0.50
0.25
0.05
power(θ)
0
0.1
θ_{null}
0.25
0.3
0.4
0.5
θ

Antetokounmpo's free throws

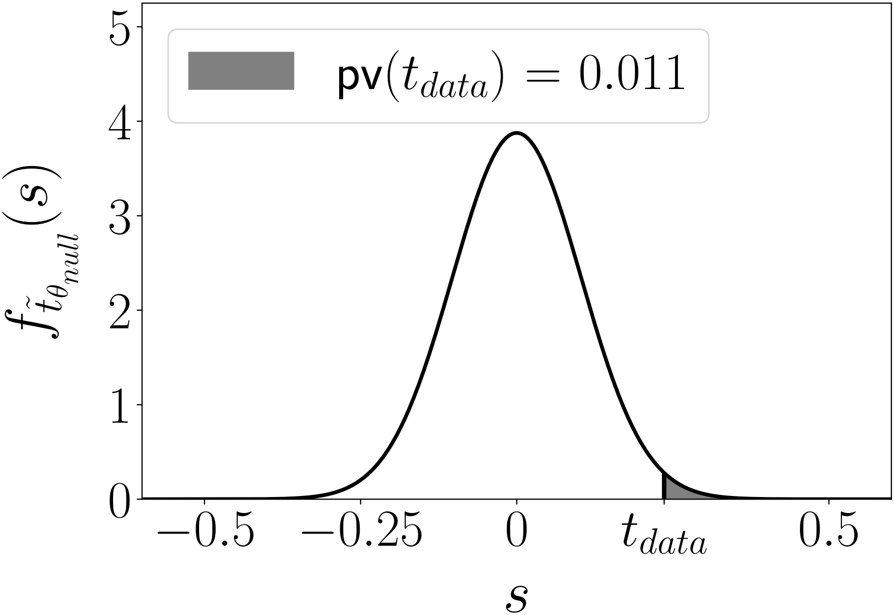
Conjecture: Free throw percentage is higher at home than away

Null hypothesis: Percentage is the same

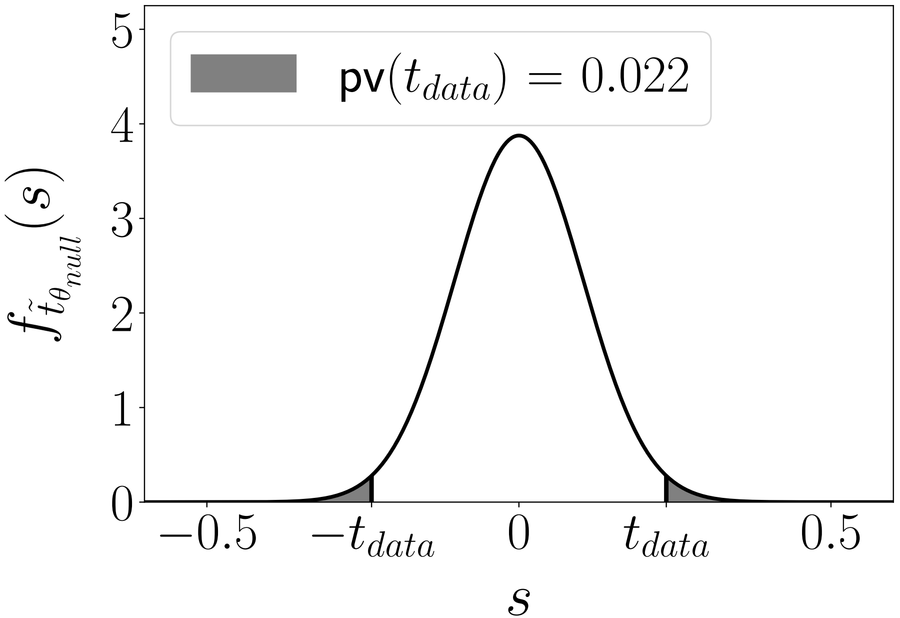
Test statistic:

$$\frac{\text{Made at home}}{\text{Attempted at home}} - \frac{\text{Made away}}{\text{Attempted away}}$$

One-tailed test



Two-tailed test



Power function

Parametric model:

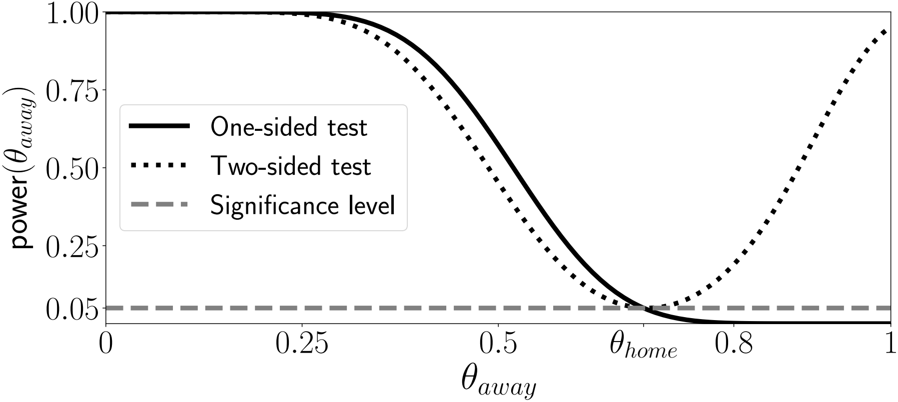
Home free-throw %: θ_{home}

Away free-throw %: θ_{away}

We fix $\theta_{\text{home}} := 0.685$ (season %)

Power function as a function of θ_{away}

Power function for fixed θ_{home}



Monte Carlo estimation

Goal: Estimate power via Monte Carlo simulations

Choose parameter θ

1. Simulate k independent samples of the test statistic $\tilde{t}_\theta$

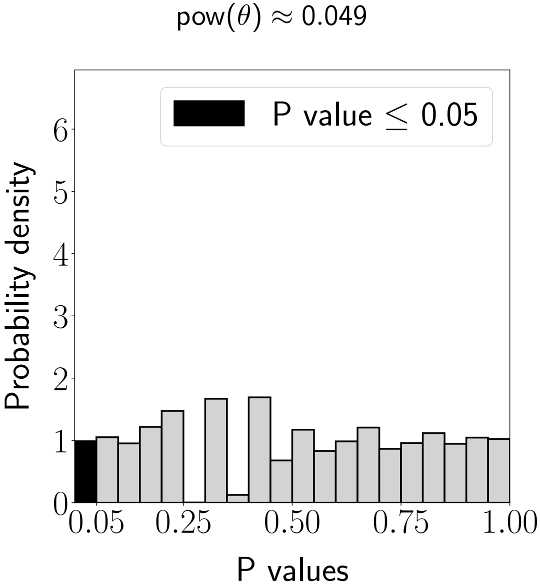
$$t_1,\ t_2,\ t_3, \ldots,\ t_k$$

2. Power $\approx$ fraction of null-hypothesis rejections

$$\text{pow}(\theta) := \mathrm{P}\left(\text{pv}\left(\tilde{t}_\theta\right) \leq \alpha\right) \approx \frac{\sum_{i=1}^{k} 1(\text{pv}\left(t_i\right) \leq \alpha)}{k}$$

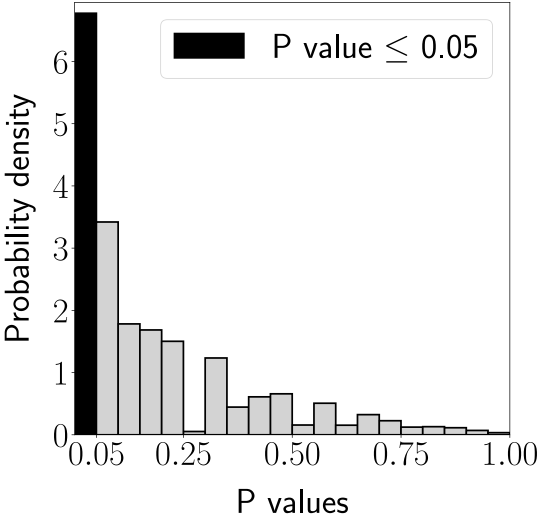
where $1(\text{pv}\left(t_i\right) \leq \alpha)$ is 1 if $\text{pv}\left(t_i\right) \leq \alpha$ and 0 otherwise

Antetokounmpo's free throws: $\theta_{\text{away}} := \theta_{\text{home}}$

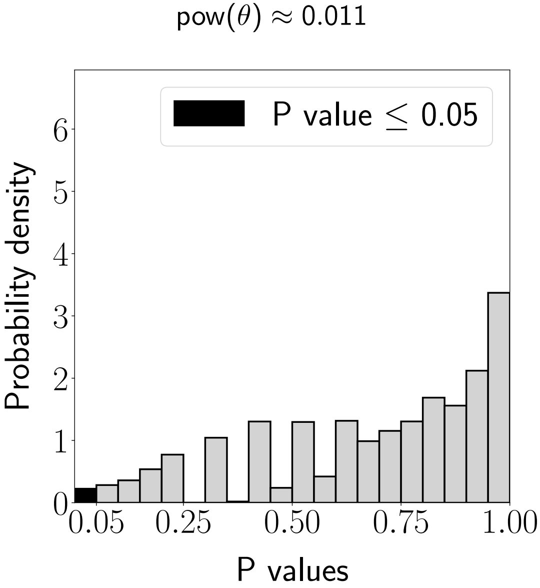


$\theta_{\text{away}} := 0.55 < \theta_{\text{home}}$

$\text{pow}(\theta) \approx 0.339$



$\theta_{\text{away}} := 0.75 > \theta_{\text{home}}$



What have we learned

Definition of power

Definition of power function

Derivation for parametric tests

Monte Carlo power estimation