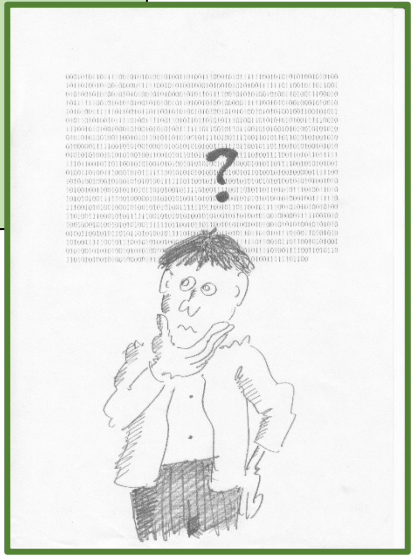
Random Sampling and Confidence Intervals

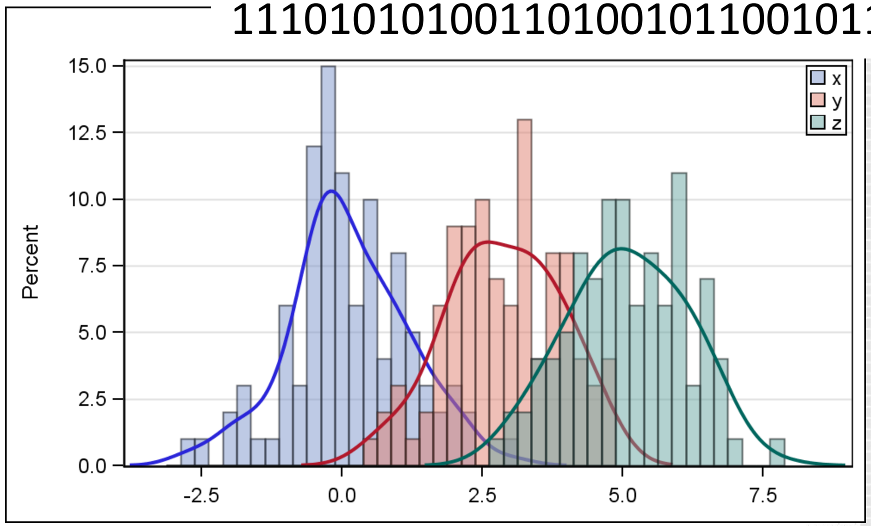
Statistics and data analysis

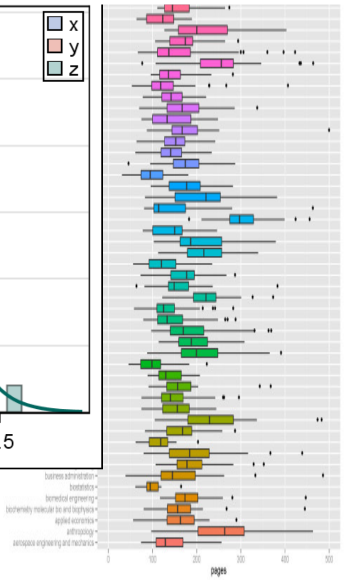
Zohar Yakhini

IDC, Herzeliya



00100111010101001010101001001000010
10101000101011111010110100110010001
11101010100110100101100101101100010





Outline

- Confidence intervals
- Bayes credibility intervals
- Likelihood intervals

Confidence Intervals

Confidence Interval: An interval of values computed from the observed data, that, with some quantifiable (typically high) confidence covers the true population (full distribution) value for an estimated parameter.

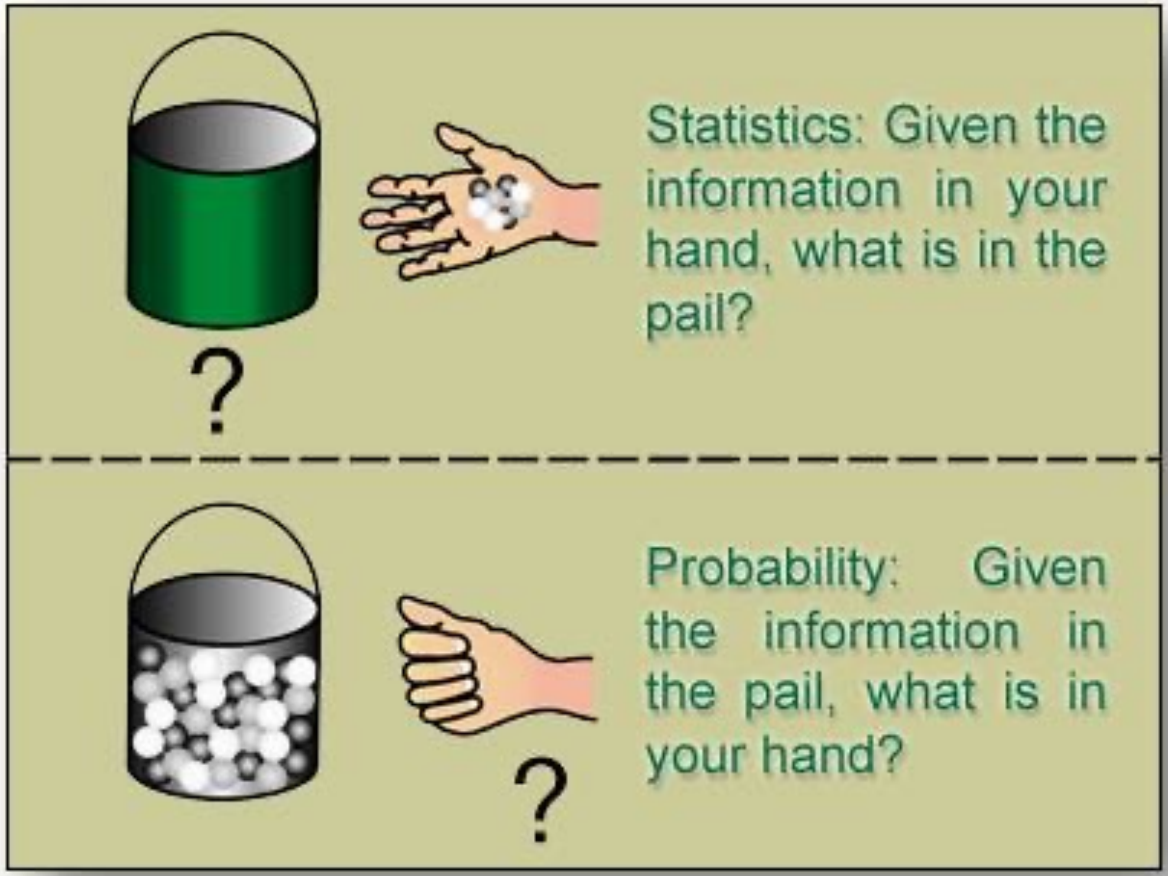
We compute confidence intervals using values computed from the sample, NOT the known values from the population

We then say that we are 95% confident that the true population parameter (e.g proportion) is inside the interval

Probability theory and statistics

Statistics – given observations, what can we say about the underlying mechanism/system that gave rise to these observations?

Probability – assuming a model - what is the expected behavior of observations from the model?



Pfizer/BioNTech announcement

A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6). Similar vaccine efficacy (generally 90 to 100%) was observed across subgroups defined by age, sex, race, ethnicity, baseline body-mass index, and the presence of coexisting conditions. Among 10 cases of severe Covid-19 with onset after the first dose, 9 occurred in placebo recipients and 1 in a BNT162b2 recipient. The safety profile of BNT162b2 was characterized by short-term, mild-to-moderate pain at the injection site, fatigue, and headache. The incidence of serious adverse events was low and was similar in the vaccine and placebo groups.

HG p-value for the BNT162b2 results

$$N = 43448, B = 21720, n = 162 + 8 = 170$$

$$X \sim HG(N, B, n) = HG(43448{,}21780{,}170)$$

$$X \cong Y \sim Binomial\left(n, p = {}^{B}/_{N}\right)$$

```
from scipy.stats import binom, norm
X = hg(M = 43448, n = 170, N = 21720) # Note: N->M, B->N
print(f'{X.cdf(8):.3e}')
Y = binom(n=170,p=21720/43448)
print(f'{Y.cdf(8):.3e}')
```

```
8.056e-39
1.058e-38
```

Quantification?

The hypergeometric analysis only addresses the p-value of the observed low numbers of vaccine recipients amongst the (confirmed) infected population.
They also report efficacy numbers.
How were these computed?

Pfizer/BioNTech announcement

A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; **BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6).** Similar vaccine efficacy (generally 90 to 100%) was observed across subgroups defined by age, sex, race, ethnicity, baseline body-mass index, and the presence of coexisting conditions. Among 10 cases of severe Covid-19 with onset after the first dose, 9 occurred in placebo recipients and 1 in a BNT162b2 recipient. The safety profile of BNT162b2 was characterized by short-term, mild-to-moderate pain at the injection site, fatigue, and headache. The incidence of serious adverse events was low and was similar in the vaccine and placebo groups.

Proportion estimate

$$(162 - 8)/162 \approx 0.95$$

Confidence interval?
Credible interval?

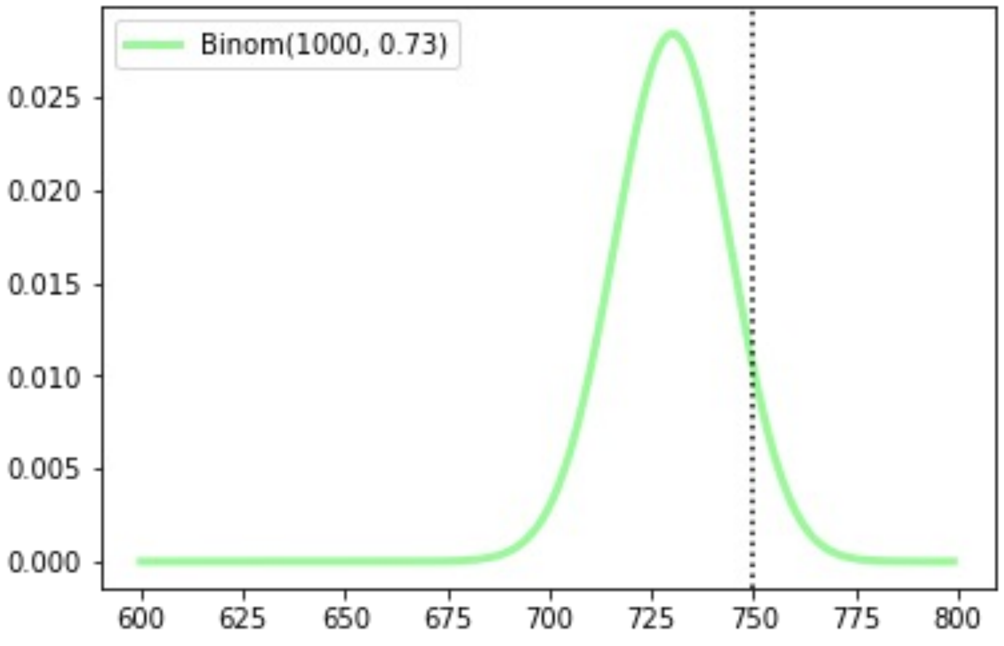
A coin tossing story …

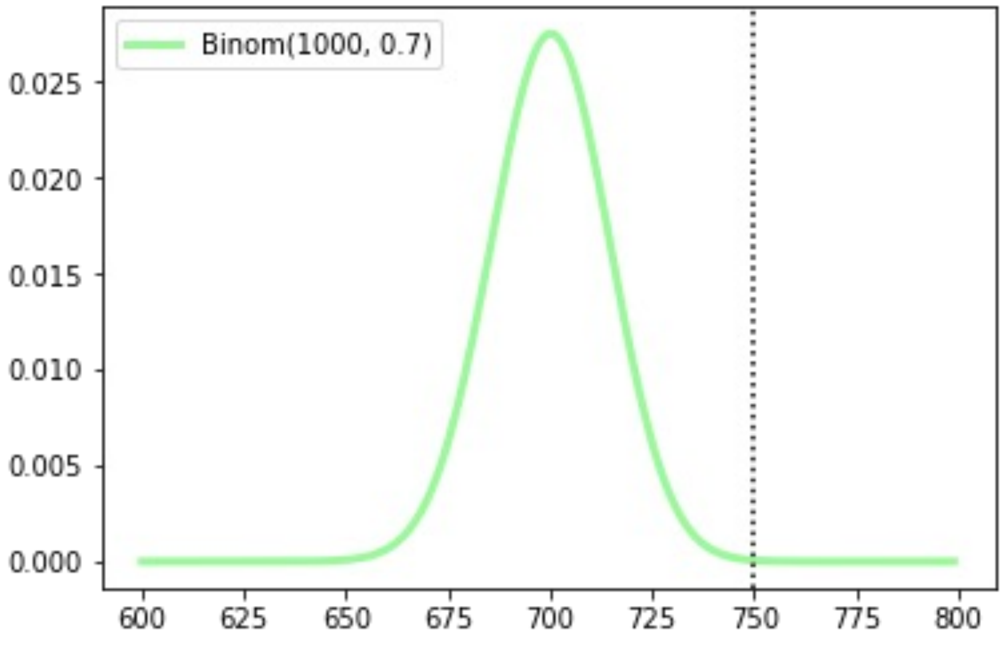
- Toss a p coin 1000 times. You see 750 Hs
- What is the MLE for p? What more can we say?
- Consider the interval $[0.71, 0.79]$. Is (the real) p in this interval?
- Assume $p < 0.71$.
  Let Y be the binomial random variable that describes the count of Hs.
- What is $Prob(Y \geq 750)$?
- Use the Binomial distribution:
  `1-binom.cdf(750)` ≈ 0.002
- Now lets construct an interval $[p_l , p_u]$ so that for any p outside this interval the p-value of our observation will be < 0.01.

A Likelihood Interval for our coin

$$p_l = max\{p: 1 - Binom(1000, p).cdf(749) < 0.01\}$$

$$p_u = min\{p: Binom(1000, p).cdf(750) < 0.01\}$$





Likelihood Int: [0.72,0.78]

Confidence Intervals

- A random interval in which we expect to find some parameter of some given distribution, with a designated confidence.
- Drawing n independent samples from some distribution, π, with a parameter θ we will produce an estimate $\hat{\theta}$ (a random variable) and then compute an interval around it so that the probability, under π^n, of the interval to include θ is $\geq 1 - \alpha$.
- **The calculation will be a closed form formula.**

CLT, reminder

Recall that for i.i.d sampling we have, for any $z \in \mathbb{R}$:

$$P\left(\frac{\bar{X}_n - \mu}{\sigma \ / \ \sqrt{n}} \leq z\right) \rightarrow \Phi(z) \quad \text{as } n \rightarrow \infty$$

CTRs

- Performance criterion for recommendation systems – the Click Through Rate - CTR
- The online store Randoporium is considering a new advertisement policy, Policy A, as a candidate for replacing the production advertising policy, Policy B.
- In the experiment we observe the following empirical CTRs:

$$\hat{p}(A) = 30K/1M \quad \text{and} \quad \hat{p}(B) = 25K/1M$$

- What conclusions can we draw from this?

CTRs

- Consider the sampling of the A group. Each potential click through is a Bernoulli event X_i with some $p = p(A)$ (this is what we want to estimate) and we observe the (empirical) proportion:

$$\hat{p} = \hat{p}(A) = \frac{30K}{1M} = \frac{1}{1M} \sum_{i=1}^{1M} X_i$$

- By CLT (assuming that the samples are close enough to independent):

$$\forall \gamma > 0 \quad P\left(\left| \frac{\sqrt{n}\,(\hat{p} - p)}{\sqrt{p(1-p)}} \right| \leq \gamma \right) \approx 2\Phi(\gamma) - 1$$

CTRs

- After some algebra and using $\gamma = 1.96$ we get:

$$P\left(\hat{p} - 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq p \ \leq \hat{p} + 1.96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx 0.95$$

- This almost gives us an interval where we expect to find p
- But not quite … why?
- Another question:
  the probability here, is under which model?

CTRs

- Using a stronger CLT (and assuming a large n), we get that for independent coin tossing:

$$\forall \gamma > 0 \quad P\left(\left| \frac{\sqrt{n}\,(\hat{p} - p)}{\sqrt{\hat{p}(1-\hat{p})}} \right| \leq \gamma \right) \approx 2\Phi(\gamma) - 1$$

Which, upon setting $\hat{\sigma} = \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$ yields:

$$P(\hat{p} - 1.96 \cdot \hat{\sigma} \leq p \leq \hat{p} + 1.96 \cdot \hat{\sigma}) \approx 0.95$$

Confidence intervals for proportions

If we sample independent Bernoulli events with probability of success p then (for sufficiently large values of n) we have:

$\forall \gamma > 0$

$$P(\hat{p} - \gamma \cdot \hat{\sigma} \leq p \ \leq \hat{p} + \gamma \cdot \hat{\sigma}) \approx 2\Phi(\gamma) \ - 1$$

where $\hat{p}(A)$ is the empirical proportion and $\hat{\sigma} = \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$

All quantities here can be calculated from the OBSERVED data

$(1-\alpha)$ CIs for proportions:

If we sample independent Bernoulli events with probability of success p then (for sufficiently large values of n) the $(1-\alpha)$ confidence interval for p is obtained from finding γ so that:

$$P(\hat{p}-\gamma\cdot\hat{\sigma}\leq p\ \leq\hat{p}+\gamma\cdot\hat{\sigma})\approx$$

$$2\Phi(\gamma)\ -1=1-\alpha$$

which yields: $\gamma=\gamma(\alpha)=\ \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$

Often denoted $z\left(\frac{\alpha}{2}\right)$

CTRs

- Back to Randoporium, we have

$$P\big(p(A) \in \hat{p}(A) \pm 1.96 \cdot \hat{\sigma}(A)\big) \approx 0.95$$
$$P\big(p(B) \in \hat{p}(B) \pm 1.96 \cdot \hat{\sigma}(B)\big) \approx 0.95$$

- Which means that w 95% confidence we have:

$$p(A) \in 0.03 \pm 1.96 \cdot \underbrace{0.001}_{sqrt(10^{-6})} \cdot 0.17 = 0.03 \pm 0.00034$$
$$p(B) \in 0.025 \pm 1.96 \cdot \underbrace{0.001 \cdot 0.15}_{\hat{\sigma}(B)} = 0.025 \pm 0.0003$$

Back to our coin

Likelihood Int for $\alpha = 0.01$: $[0.72, 0.78]$

Confidence Int for $\alpha = 0.01$: $[0.71, 0.79]$

A different coin

Now assume that we observed $950\ H$s in our coin tossing experiment

Likelihood Int for $\alpha = 0.01$: $[0.93, 0.96]$

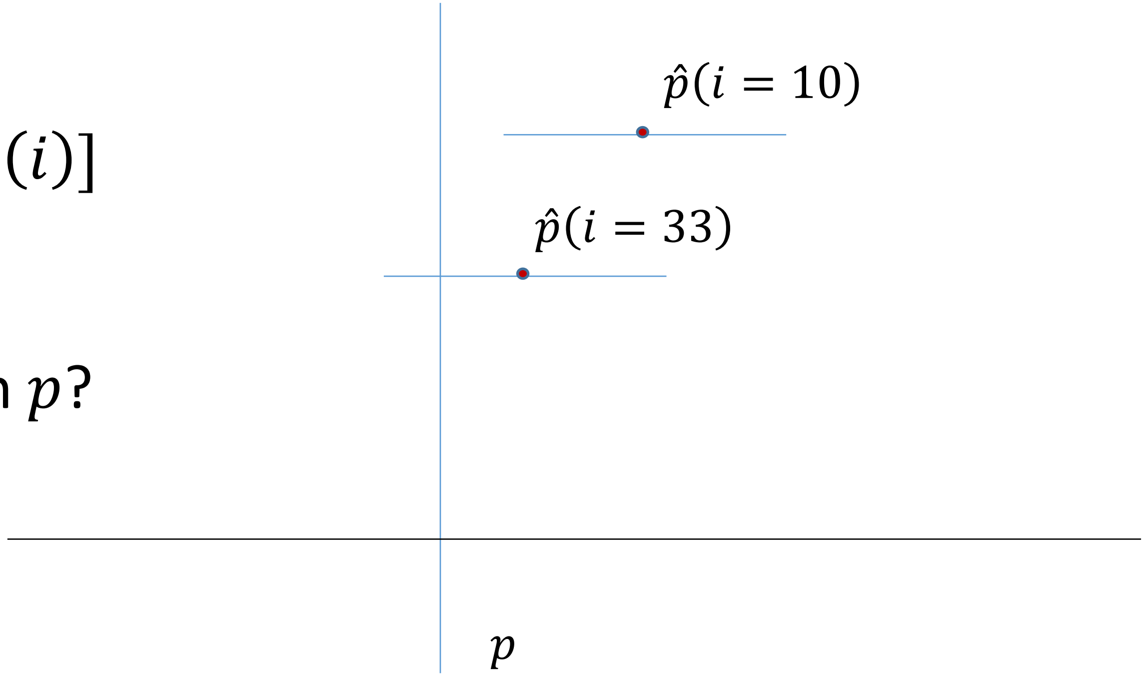
Confidence Int for $\alpha = 0.01$: $[0.93, 0.97]$

A $\hat{p}$ experiment

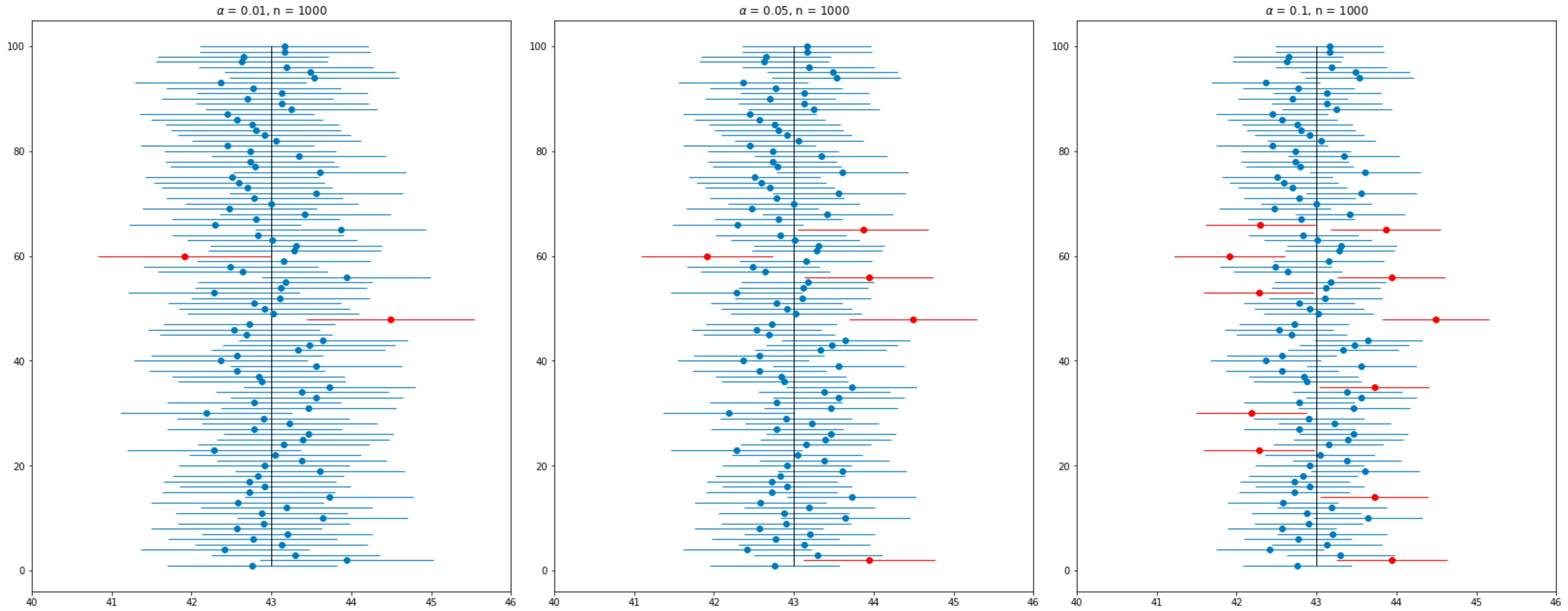
$$P(\hat{p} - \gamma \cdot \hat{\sigma} \leq p \leq \hat{p} + \gamma \cdot \hat{\sigma}) \approx 2\Phi(\gamma) - 1 = 1 - \alpha$$

$$\gamma = \gamma(\alpha) = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

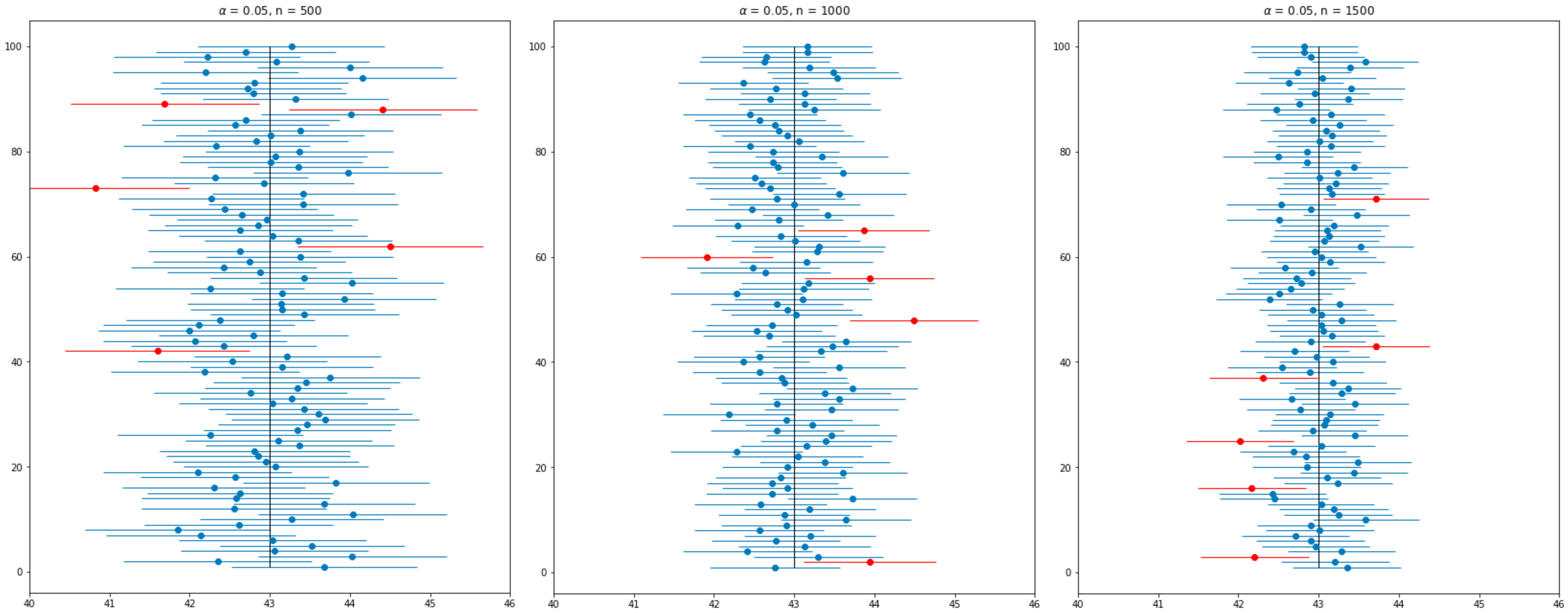
- For $i = 1..100$, draw an instance of $Binom(n, p)$
- For each instance compute $\hat{p}(i)$ by MLE
- Fix α and let $\gamma = \gamma(\alpha)$
- Now form the interval
$$[\hat{p}(i) - \gamma \cdot \hat{\sigma}(i), \hat{p}(i) + \gamma \cdot \hat{\sigma}(i)]$$
- Note that this is a random interval
- Does it or does it not contain p?
- In how many of your instances will it contain p?



The effect of α and of n



The effect of α and of n



The effect of α and of n

Partial code:

```
z_critical = stats.norm.ppf(q = 1-alpha/2)  # Get the z-critical value (two tails)

for i in range(n_samples):
    sample = np.random.choice(a= population_ages, size = sample_size)
    sample_mean = sample.mean()
    sample_stdev = sample.std()  # Get the sample standard deviation
    margin_of_error = z_critical * (sample_stdev/math.sqrt(sample_size-1))

    bot = sample_mean - margin_of_error
    top = sample_mean + margin_of_error
    confidence_interval = (bot, top)
```

CTRs

- Performance criterion for recommendation systems – the Click Through Rate - CTR
- The online store Randoporium is considering a new advertisement policy, Policy A, as a candidate for replacing the production advertising policy, Policy B.
- In the experiment we observe the following empirical CTRs:

$$\hat{p}(A) = 30K/1M \quad \text{and} \quad \hat{p}(B) = 25K/1M$$

- What conclusions can we draw from this?

Difference in proportions

$$\Delta \in \; (\hat{p}_1 \; - \hat{p}_2) \; \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \left(\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}\right)^{\frac{1}{2}}$$

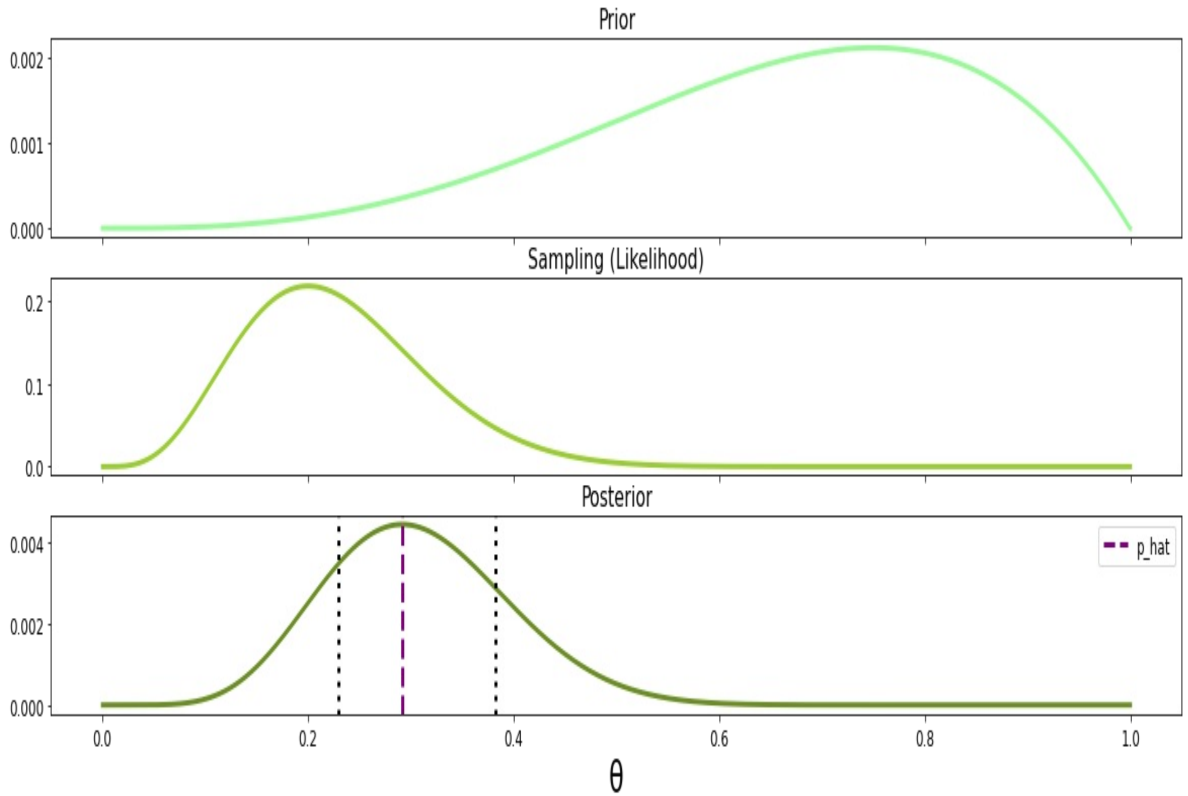
The data support, with 0.99 confidence, that by moving from Policy B to Policy A we will improve our CTR by XXX%

Pfizer/BioNTech announcement

A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6). Similar vaccine efficacy (generally 90 to 100%) was observed across subgroups defined by age, sex, race, ethnicity, baseline body-mass index, and the presence of coexisting conditions. Among 10 cases of severe Covid-19 with onset after the first dose, 9 occurred in placebo recipients and 1 in a BNT162b2 recipient. The safety profile of BNT162b2 was characterized by short-term, mild-to-moderate pain at the injection site, fatigue, and headache. The incidence of serious adverse events was low and was similar in the vaccine and placebo groups.

The Bayesean approach: Credible Intervals

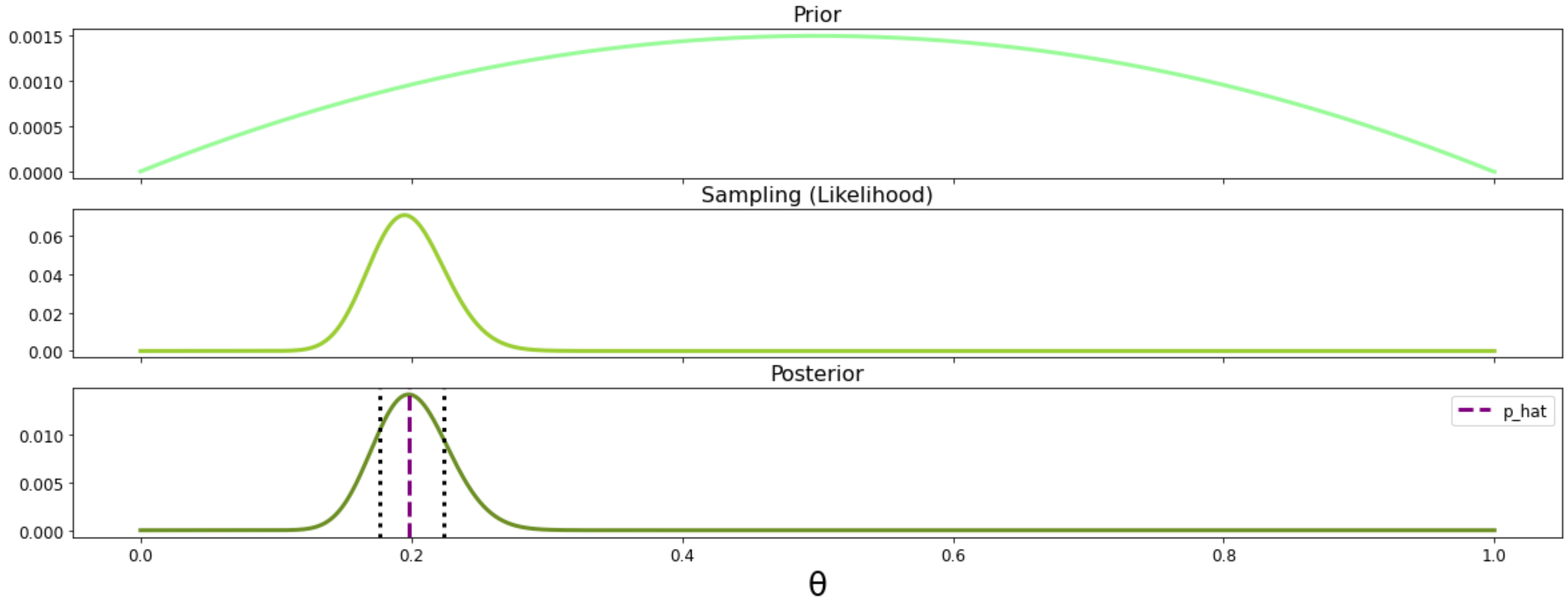
- Prior: $\pi(\theta)$
- Likelihood – having observed the data we compute
$$L(\theta) = Prob(D;\ \theta)$$
- Note that the MLE approach will return $\hat{\theta} = \operatorname{argmax} L(\theta)$
- In the Bayes approach we will now compute a posterior:
$$\psi(\theta) = \pi(\theta)L(\theta)$$
- We will return $\hat{\theta} = \operatorname{argmax} \psi(\theta)$ and compute credibility intervals for the actual θ



```
coin_prob = 0.2
n = 20

# iid (independent and identically distributed) assumption; generate the data
flip_data = np.random.binomial(n=1, p=coin_prob, size=n)
# Beta parameters
a = 4
b = 6 - a
# the domain of \theta
theta_range = np.linspace(0, 1, 1000)

# compute the three curves
# use discrete pdf to compute the pdf of the prior adjusted to the domain resolution
theta_range_e = theta_range + 0.001
prior = stats.beta.cdf(x = theta_range_e, a=a, b=b) - stats.beta.cdf(x = theta_range, a=a, b=b)
# prior = stats.beta.pdf(x = theta_range, a=a, b=b)
likelihood = stats.binom.pmf(k = np.sum(flip_data), n = len(flip_data), p = theta_range)
posterior = likelihood * prior   # element-wise multiplication
normalized_posterior = posterior / np.sum(posterior)
```



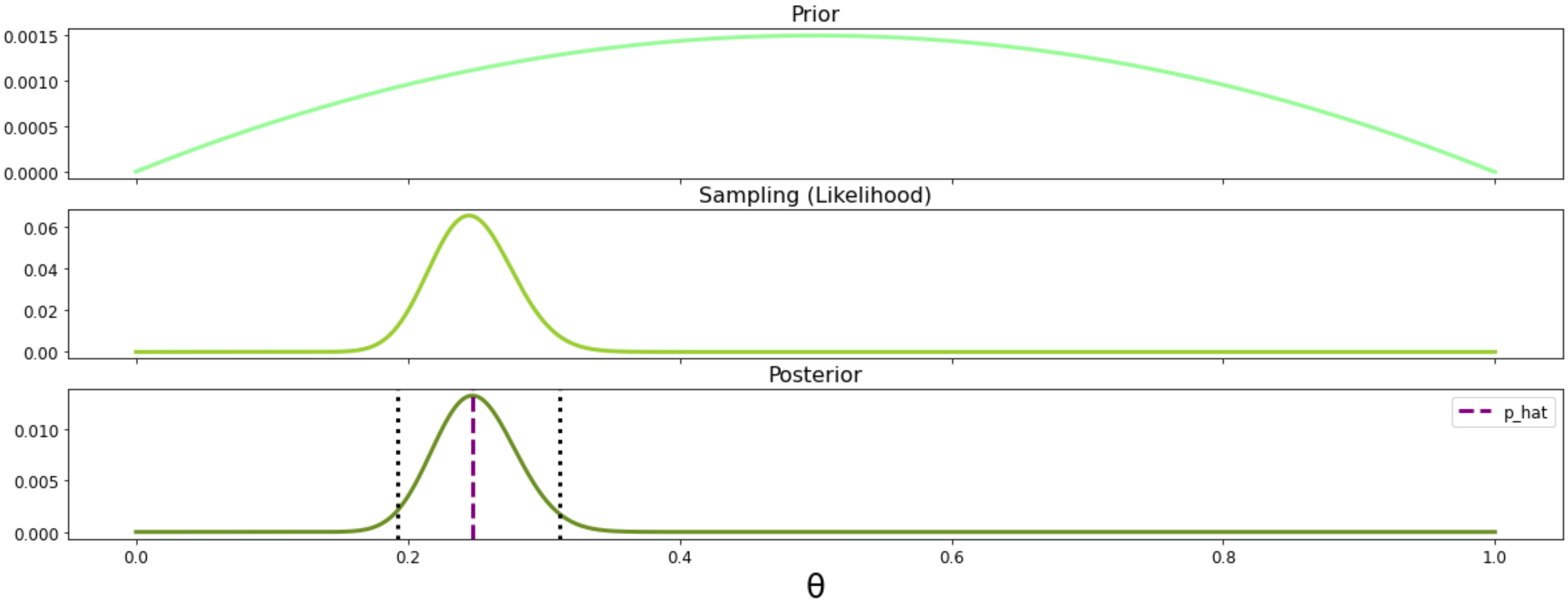
p = 0.2

a = 2, b = 2

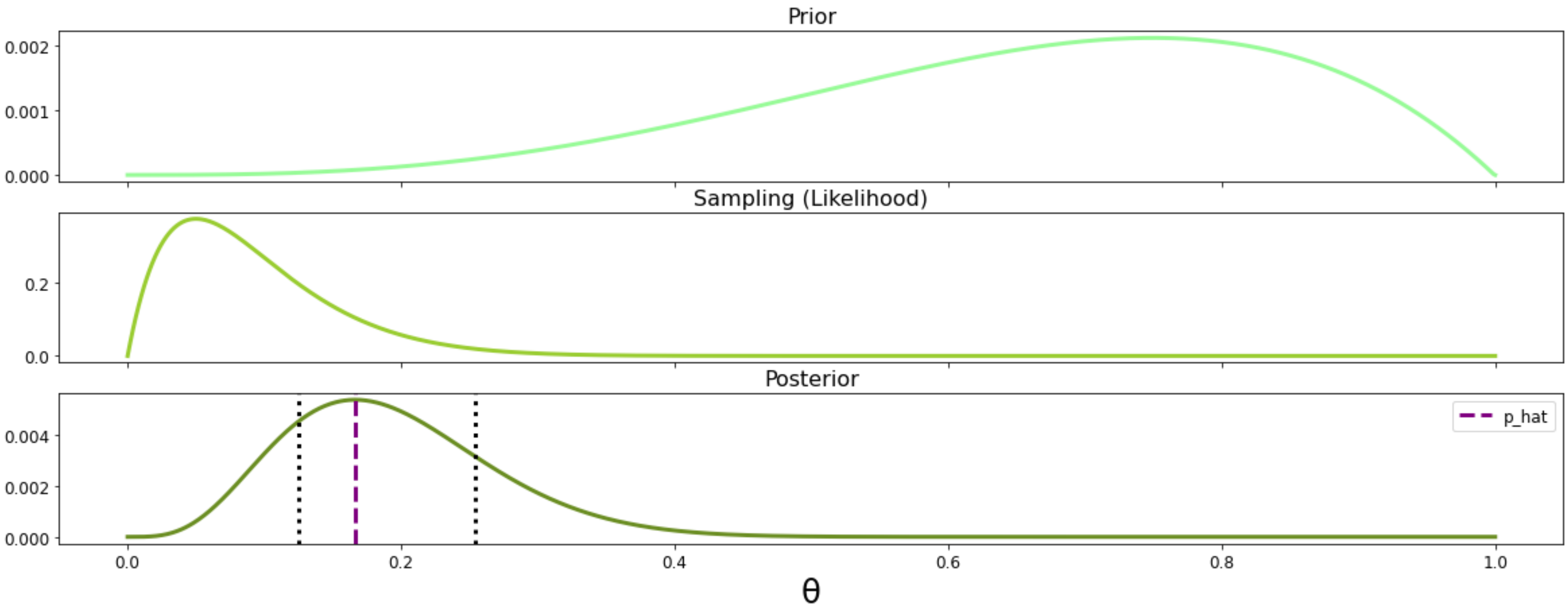
n = 200

theta_hat = 0.198

alpha = 0.4



p = 0.2
a = 2, b = 2
n = 200
theta_hat = 0.24
alpha = 0.05



p = 0.2
a = 4, b = 2
n = 20
theta_hat = 0.17
alpha = 0.4

Pfizer/BioNTech announcement

A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6). Similar vaccine efficacy (generally 90 to 100%) was observed across subgroups defined by age, sex, race, ethnicity, baseline body-mass index, and the presence of coexisting conditions. Among 10 cases of severe Covid-19 with onset after the first dose, 9 occurred in placebo recipients and 1 in a BNT162b2 recipient. The safety profile of BNT162b2 was characterized by short-term, mild-to-moderate pain at the injection site, fatigue, and headache. The incidence of serious adverse events was low and was similar in the vaccine and placebo groups.

Summary

- Confidence interval for the population mean, based on the sample mean and the sample std, can be derived from the CLT
- In the Bernoulli case we can directly compute likelihood intervals.
- The more acceptable/routine/standard frequentist/classical approach is to use Confidence Intervals
- Working under the p model the $1-\alpha$ CI (a random interval) will include the true parameter p with probability $1-\alpha$
- CIs for other quantities (difference in proportions)
- A CI for the median
- The Bayesean approach: Bayes credibility intervals