tail bounds + MLE

LECTURE 20

CHERNOFF BOUND

Let $X_1, X_2, \dots, X_n$ be *independent* Bernoulli random variables. Let $X = \sum X_i$, and $\mu = \mathbb{E}[X]$. For any $0 \le \delta \le 1$

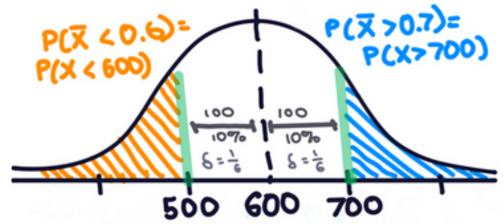
$\mathbb{P}(X \le (1-\delta)\mu) \le e^{\left(-\frac{\delta^2 \mu}{2}\right)}$ and $\mathbb{P}(X \ge (1+\delta)\mu) \le e^{\left(-\frac{\delta^2 \mu}{3}\right)}$

LEFT TAIL — RIGHT TAIL

Requirements:
1. X is a sum of independent Bernoulli random variables.
2. We know $\mathbb{E}[X]$

*You run a poll of 1000 people where 60% of true population supports you. What is the probability that the poll is **not** within 10% of the true value?*

Goal: bound $\mathbb{P}(|\bar{X} - 0.6| \ge 0.1) = \mathbb{P}(\bar{X} \le 0.5) + \mathbb{P}(\bar{X} \ge 0.7)$



1. bound the left tail

$\mathbb{P}(\bar{X} \le 0.5)$

2. bound the right tail

$\mathbb{P}(\bar{X} \ge 0.7)$

3. putting it together

UNION BOUND

For any events E, F
$\mathbb{P}(E \cup F) \le \mathbb{P}(E) + \mathbb{P}(F)$

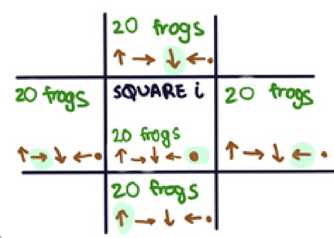
Sometimes we don't't' have enough information to compute this probability exactly, so we use the union bound to bound that probability

There are 20 frogs on each location in a 5x5 grid. Each frog independently jumps L, R, U, D, or neither with equal probability. Bound the probability at least one square ends up with at least 36 frogs.

1. Apply Union Bound

2. Apply Chernoff Bound to bound each of P(Ai)

3. Put it all together



MAXIMUM LIKELIHOOD ESTIMATION

Goal: derive an estimate θ^ for the parameter θ based on observed data

1. Run the experiment a bunch of times (i.e., collect many samples from the distribution) -> *data*
   e.g., we flip a coin that follows Ber(p) *5 times and write down the results - HTTTH*
2. Estimate the missing rules (unknown parameter(s)) *based on the data*
   › *Guess rules to maximize probability of the events we saw* (relative to other choices of the rules).
   e.g., what is the value of p *that makes the probability of seeing HTTTH the highest?*

1. DEFINE LIKELIHOOD FUNCTION

$\mathcal{L}(E;\theta)$ is $\mathbb{P}(E)$ when the experiment is run with θ

"what is probability of seeing the event E (in our case, the set of data), if the experiment is run with the parameter θ?"

Likelihood Function: Likelihood of n observations (from discrete distribution)

$$\mathcal{L}(x_1, x_2, \ldots, x_n;\theta) = \Pi_i^n \mathbb{P}(x_i;\theta)$$

Coin example

We ran the experiment 10 times independently. The result was HTTTHHTHHH

$\mathcal{L}(\textbf{HTTTHHTHHH};\boldsymbol{\theta}) =$

"Probability of observing HTTTHHTHHH if $\boldsymbol{\theta}$ is probability of heads on a single flip"

2. TAKE LOG OF THE LIKELIHOOD (LOG-LIKELIHOOD

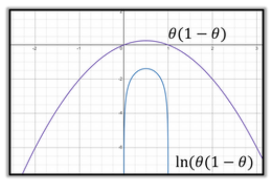
The product rule is not fun!! So, take the log of the likelihood function before taking the derivative!

*ln(a*b)=ln(a)+ln(b)*

Can we still take the max? Yes! ln() is an increasing function, so

$\text{argmax}_\theta \ln(\mathcal{L}(E;\theta)) = \text{argmax}_\theta \mathcal{L}(E;\theta)$

"the log of the likelihood will increase as the likelihood increases and vice verse, so, the value of θ that maximizes the log likelihood also maximized the likelihood"



Coin Example: $\ln(\mathcal{L}(\text{HTTTHHTHHH};\theta) =$

Now, we want to find the value of theta, that will maximize the likelihood (maximize the probability of the data)

The MLE of the parameter θ is: $\hat{\theta} = \text{argmax}_\theta \ \mathcal{L}(E;\theta)$

3. TAKE DERIVATIVE OF THE LOG-LIKELIHOOD

$\frac{d}{d\theta}\ln(\mathcal{L}(\cdot)) =$

4. SET DERIVATIVE TO 0, SOLVE FOR THE MLE

$\frac{6}{\hat{\theta}} - \frac{4}{1-\hat{\theta}} = 0 \Rightarrow$

5. 2ND DERIVATIVE TEST TO VERIFY MAX

$\frac{d^2}{d\theta^2} =$

GENERAL STEPS

We're given that there's a distribution with some unknown parameter(s) θ.

There are independent observations x1, x2,..., xn from this distribution.

To find the MLE for this unknown parameter(s) θ....

1. Write the likelihood function

› multiply (not add) probabilities of seeing each of the observations based on θ

2. Take the log ln(..) of the likelihood function (makes the math easier)

3. Take the derivative of the log-likelihood function

4. Set the derivative to 0, and solve for the MLE $\hat{\theta}$

› remember to switch from θ to $\hat{\theta}$ in this step because we're now solving for the MLE

5. Verify it is a maximum with second derivative test (not required for 312)