Statistical Measurement of Information Leakage

Tom Chothia

Univ. of Birmingham

Joint work with Kostas Chatzikokolakis and Apratim Guha

Overview

- Estimate information leakage statistically from trail runs of a real system.
  - Automatic tool to calculate information leakage.
- We work out bounds on the possible error.
- We present an ***if, and only if***, test for ***zero*** leakage.
- For actuate results you need more samples than the (no. of inputs) x (no. of observations).

Information Theory

Entropy: $H(X) = - \sum_x p(x).\log(p(x))$

the amount of uncertainty in X.

Conditional Entropy: $H(Y|X) = \sum_x p(x).H(Y|X=x)$

the amount of uncertain in Y if you know X

Mutual Information: $I(X;Y) = H(X) - H(X|Y)$

the reduce of uncertainty you get in X if you know Y.

Relative Entropy: $D(p||q) = \sum_x p(x).\log(p(x) / q(x))$

“distance” from one distribution to another.

Information Theory

- A **Channel,** has inputs X, outputs Y, and a probability transition matrix W(x|y).

- Information sent across the channel = I(X;Y),
  - We define I(Q,W) = I(Q;Y)

- Maximum rate is the **Channel Capacity**:

$$C(W) = \text{Max}_Q\ I(Q,W)$$

the most information you can send across a channel no matter how data is sent.

Information Leakage = Capacity(System)

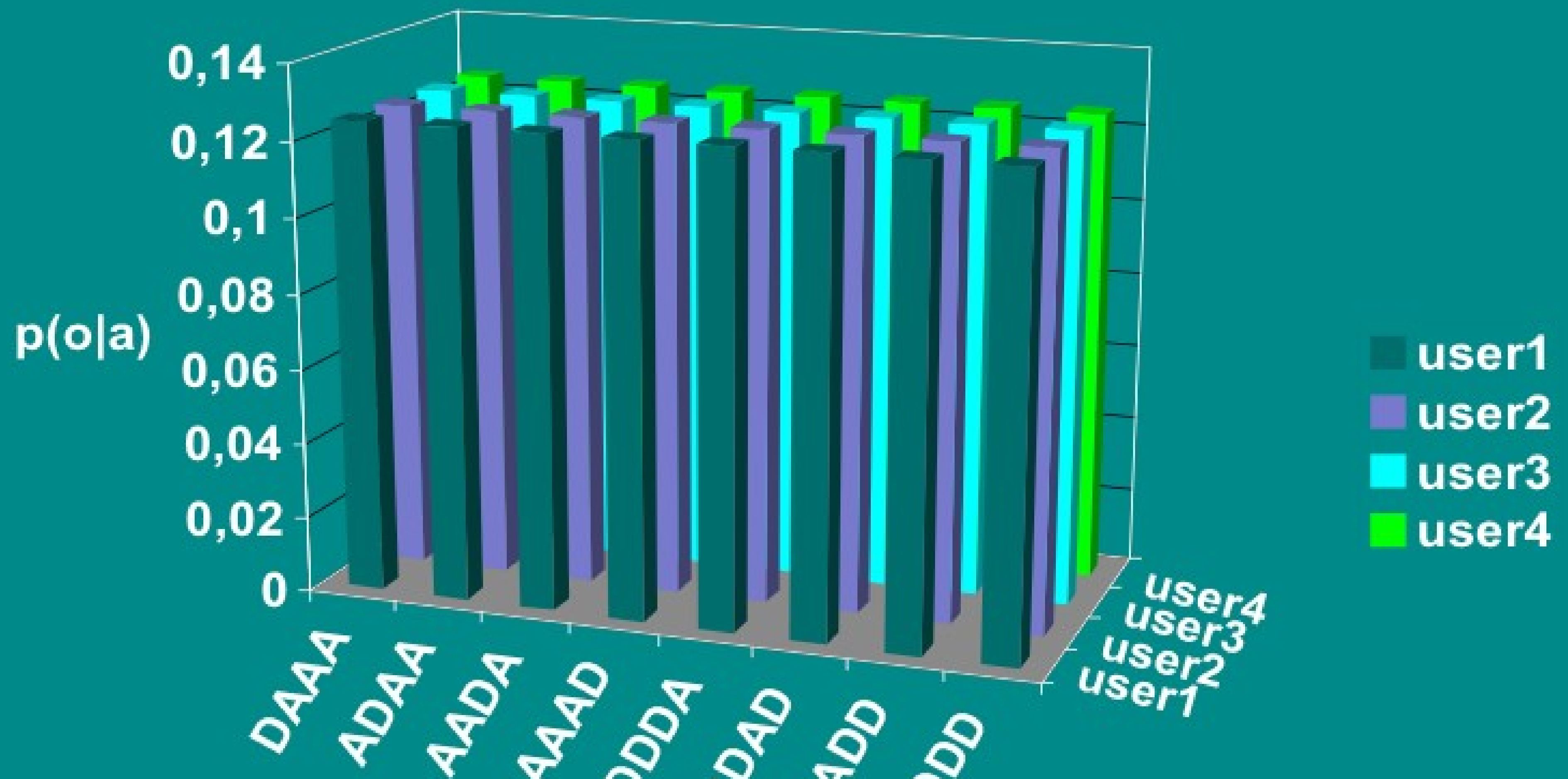
Following Chatzikokolakis et al.

- Think of the whole system as a channel.
  - The guilty user is the input to the “channel”.
  - The observable actions are the outputs from the “channel”.

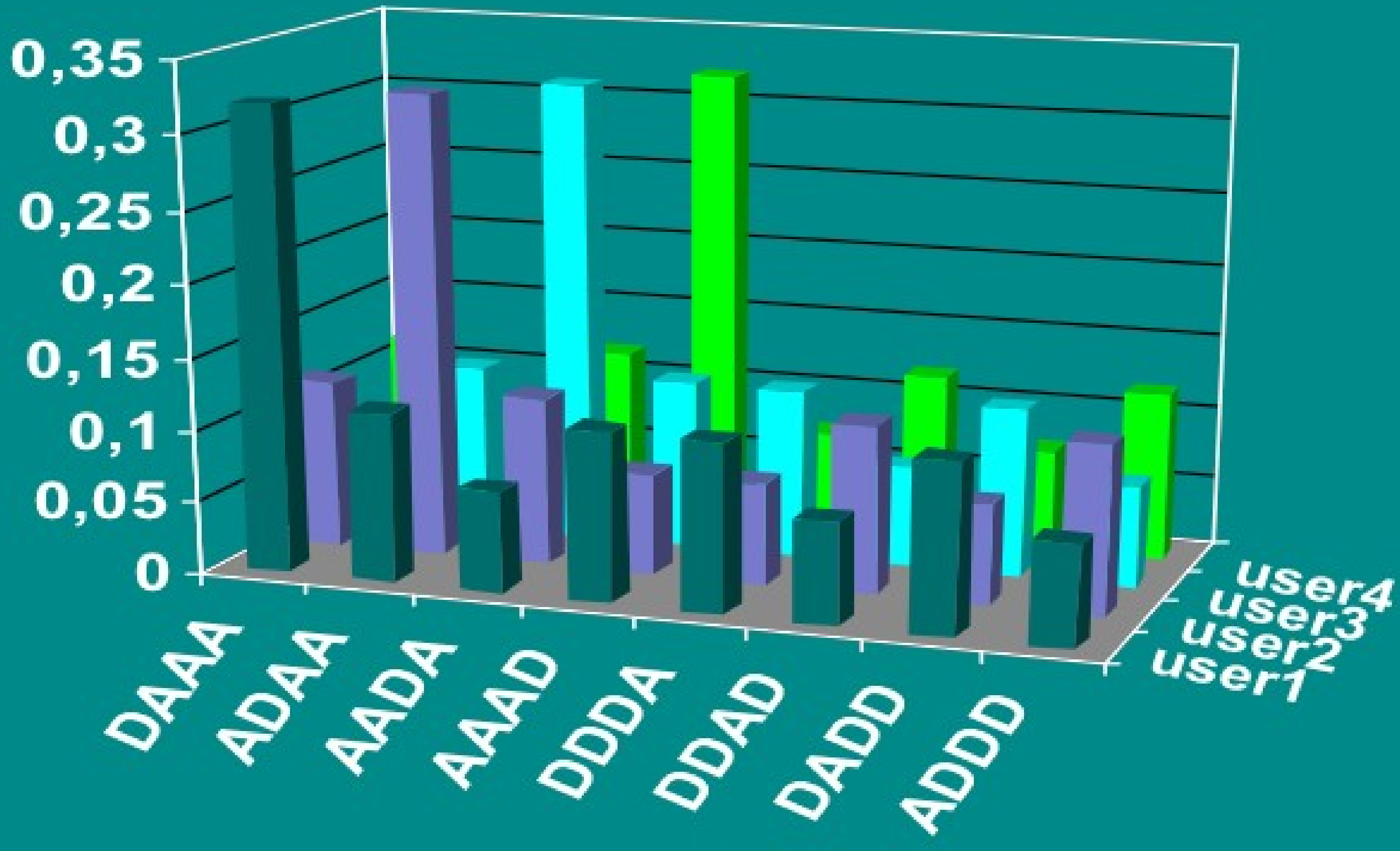
- Capacity tells us what we can learn about the users from the observable actions.

- Similar approach can be taken for Information Flow e.g. Millen, Clark et al. etc.

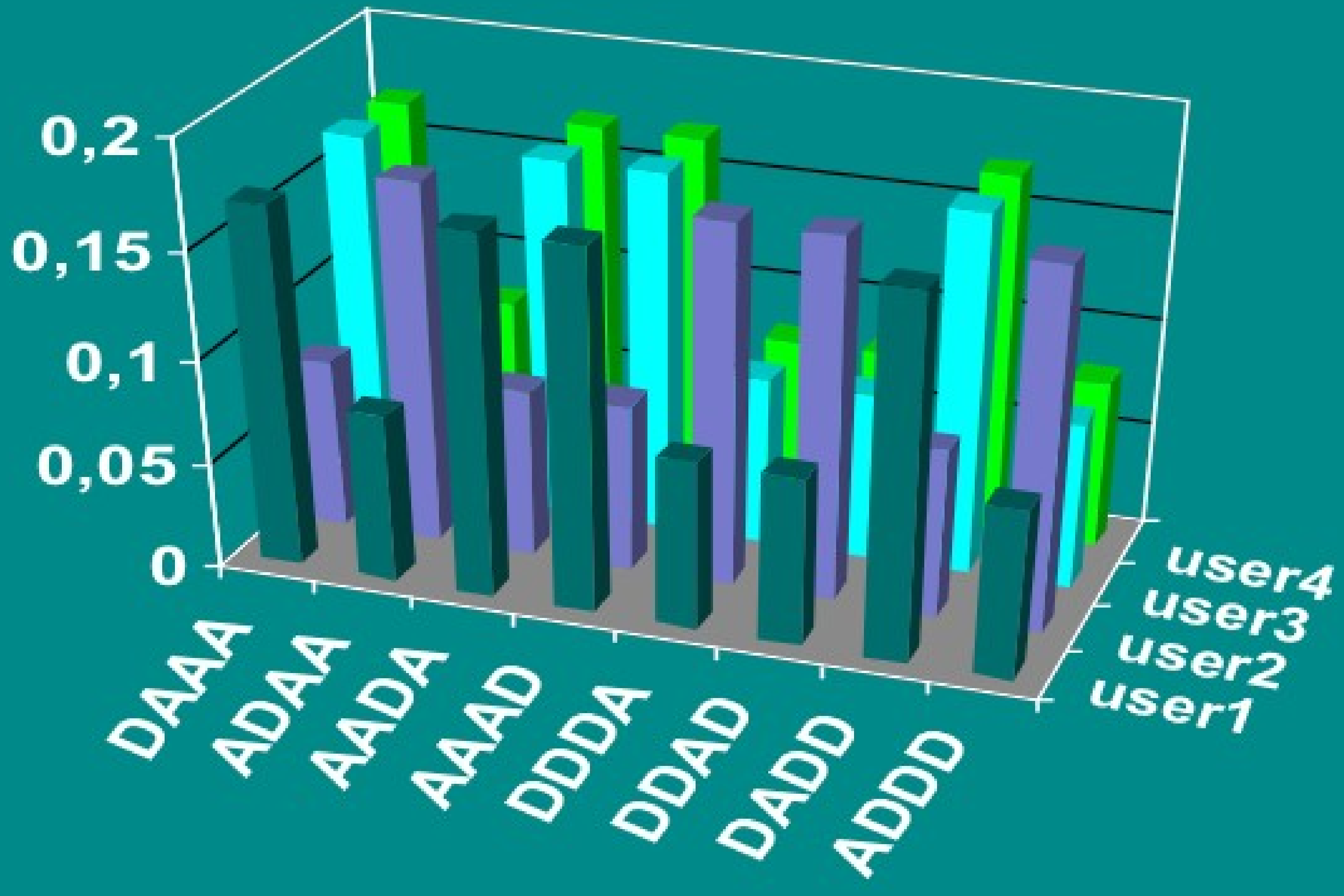
The Conditional Probabilities of the D.C. Protocol



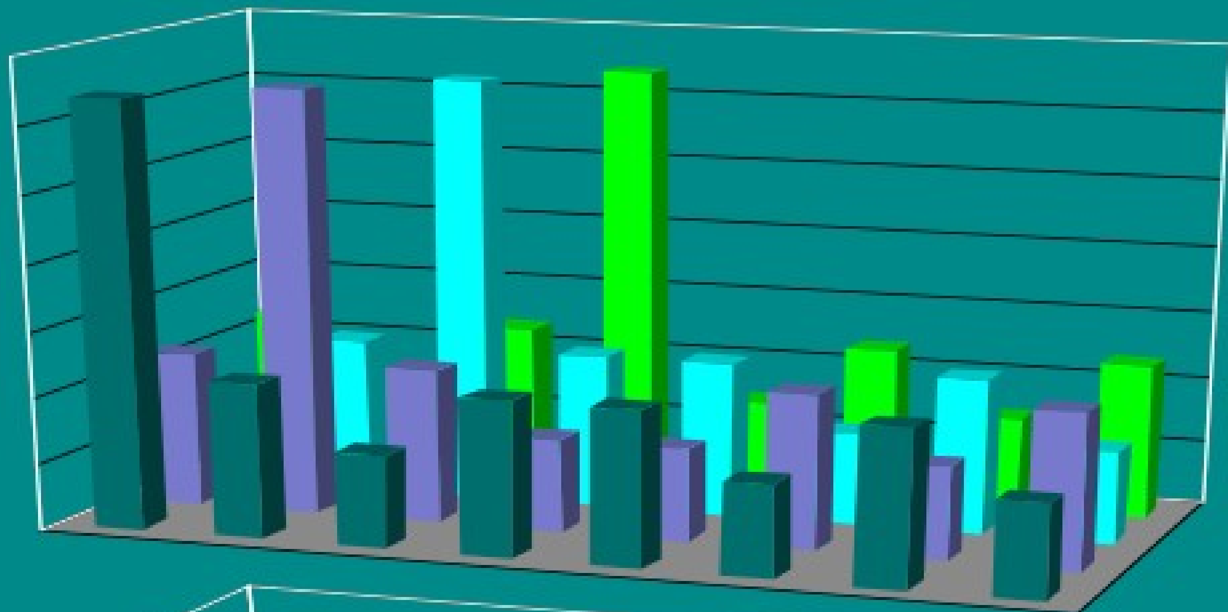
Biased Coins: if p(heads) = 0.25



2 Out of 4 p(heads) = 0.8

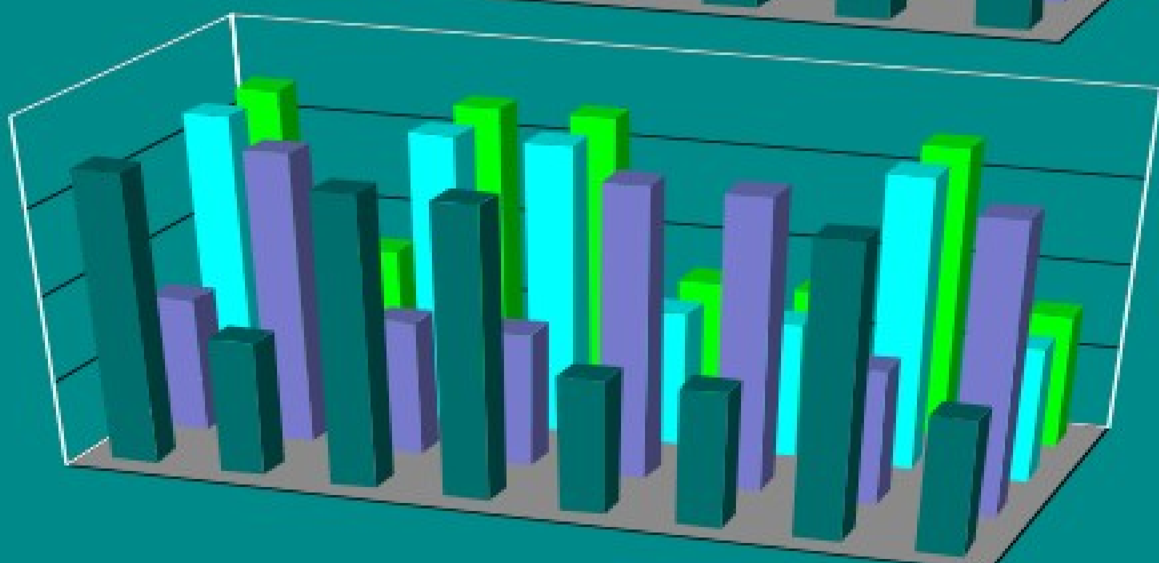


Which is Better?



All coins
P(heads) = 0.75

Anonymity
= 0.170536



2 coins
P(heads) = 0.8

Anonymity
= 0.095619

Applying this to Real Systems

How do we apply information theoretic measures of information leakage to real systems?

- Source code may be too complex to analyse directly or model / behave probabilistically.
- Leak may be caused by the implementation:
  - Time based attack on RSA, (Paul Kocher)
  - Bandwidth attack on Tor (Murdoch & Danezis)
  - CPU Heat attack on Tor (Murdoch)

Demo

- Examples of the kinds of sampled data we want to analyse

Blahut-Arimoto Algorithm

How do we find the maximising input distribution?

$I(X;Y) = H(X) - H(X|Y)$

$= \sum_{x,y} p(x)\, W(y|x) \log \left(W(y|x) \;/\; \sum_{x'} p(x')W(y|x)\right.$

$= \sum_x p(x).D(\; W(\_|x) \;\|\; \sum_{x'} p(x')W(\_|x') \;)$

$= \sum_x p(x).D_x(W\|pW)$

Distribution of y given x

Distribution of y

$\sum_x p(x).D_x(W\|pW) \leq C(W) \leq \mathrm{Max}_x\, D_x(W\|pW)$

Blahut-Arimoto Algorithm.

1) Guess an input distribution $p^0(a)$ e.g., uniform

2) Improve the guess, for all x:

$$p^{n+1}(x) = \frac{\exp(\ D_x(W \,||\, p^n W)\ )}{\sum_{x'} \exp(\ D_{x'}(W \,||\, p^n W)\ )}$$

3) Repeat until $I(p^n, W) - \mathrm{Max}_x\, D_x(W||p^n W) < e$

Can be tweaked for super linear convergence, conditional mutual information etc.

Other Ways to Find Capacity

- Special cases can be calculated directly
- A “gradient climb” e.g. Frank-Wolfe algorithm
- Kuhn-Tucker Theorem/Lagrange multipliers.

Method of Analysing Anonymity

- To analyse a system we define the inputs and outputs.
  - Some abstraction might be needed to make the number of observations manageable

- We run tests of the system for each input.

- From these tests we estimate a matrix.

- We estimate capacity, from the matrix.

Terms

W : the matrix of the true system.

$\hat{W}_n$: a matrix estimated from n samples.

Q : the input dist. that maximise M.I.

$\hat{Q}_m(\hat{W}_n)$: the B.A. algorithm applied to W_n.

$C = I(Q,W)$: the true system capacity.

$\hat{C}_{n,m} = I(\hat{Q}_m(\hat{W}_n),\hat{W}_n)$: estimate of capacity ??

Convergence

Theorem: $\hat{C}_{n,m}$ almost surely convergences to C as $n,m \rightarrow \infty$

i.e., for any p_e and error e there exists n‘ & m’ such that for n > n’ and m > m’:

$$p(|\, C - \hat{C}_{n,m} \,| > e\,) < p_e$$

The Distribution of Anonymity

We can get bounds on the error by ask what distribution $\hat{C}_{n,m}$ comes from.

Adapting a statistical method from Rao:

- We find the Taylor expansion of the $\hat{C}_{n,m}$
- We drop the terms smaller than sampleSize^{-2}
- We then calculate the mean and variance.
- We find the distribution using the central limit theory.

Estimated Value

As we can’t find the distribution for the maximising distribution we relate our estimate to $I(\hat{Q}_m(\hat{W}_n),W)$

Lemma: The estimate

- is less than or equal to the capacity,
- equals zero if, and only if, the capacity equals zero.

Expectation and Variance

To find a distribution we need to find the expectation:

E(X): the average value

And the variance:

$$Var\ (X) = E(mean - x)^2$$

What We Know

K_{ij} is the number of times the pair (i,j) shows up in our test.

Let the true prob: $p(i,j) = {}^{h}K_{ij}/n$

Then maximum likehood tells us that

- $E(K_{ij} - {}^{h}K_{ij}) = 0$
- $E((K_{ij} - {}^{h}K_{ij})^2) = p(i).\ W(j|i)(1-W(j|i))$
- $E((K_{ij} - {}^{h}K_{ij})^3) = K_{ij}(2W(j|i)^2-3W(j|i)+1)$ ...

Taylor's Theorem

To find the value of a function at x (near a):

$$f(x) = f(a) + \frac{f'(a)(x-a)}{1!} + \frac{f''(a)(x-a)^2}{2!} + \frac{f''(a)(x-a)^3}{3!} + ...$$

We take I(X,_) as "f", W_n as "x" and W as "a" to give get an expression for the estimate in terms of the true value.

Taylor Expansion of Entropy

$I_n(X,Y) = H(X) + H_n(Y) - H_n(X,Y)$

$E(I_n(X,Y)) = E(H(X)) + E(H_n(Y)) - E(H_n(X,Y))$

$H(X,Y) \quad = - \sum_{x,y} p(x,y)\log(p(x,y))$

$H_n(X,Y) = - \sum_{x,y} K_{ij}/n.\log(K_{ij}/n)$

$H_n(X,Y) \quad = -\sum_{x,y} {}^hK_{ij}/n - 1/n.\sum_{x,y}(1+{}^hK_{ij}/n)$

$\quad -\sum_{x,y}(K_{ij}-{}^hK_{ij})^2/n.\,{}^hK_{ij})$

$\quad + \sum_{x,y}(K_{ij}+{}^hK_{ij})^3/6n.{}^hK_{ij}{}^2) + O(n^{-2})$

$E(H_n(X,Y)) \; = \; H(X,Y) - I(J-1)/2n + O(n^{-2})$

For Non-Zero Mutual Information

When the true value is not 0, an estimation of capacity is drawn from a normal distribution with:

Mean: $I(\hat{Q}_m(\hat{W}_n),W) + \frac{(I-1)(J-1)}{2n} + O(n^{-2})$

Variance: ...

Variance

$$\frac{1}{n} \sum_x Q(x).\left(\sum_y W(y|x).\left(\log\left(\frac{Q(x).W(y|x)}{\sum_{x'} Q(x')W(y|x')}\right)\right)^2 - \left(\sum_y W(y|x).\log\left(\frac{Q(x).W(y|x)}{\sum_{x'} Q(x')W(y|x')}\right)\right)^2\right)$$

$$+O(n^{-2})$$

When I = 0

- The $O(n^{-1})$ term disappears with X and Y are independent.

- In which case we need to find the $O(n^{-2})$ term.

- Following Rao, we observe when I = 0 the

$$\sum_{ij} \left((K_{ij} - E(K_{ij}))^2 / E(K_{ij}) \right) \sim \chi^2$$

and that this approximates mutual infromation.

Results for I = 0

When the true value is 0, an estimation of capacity (or mutual information) is drawn from the distribution:

$$2n.I \sim \chi^2((noOfInputs-1)(noOfOutputs-1))$$

Mean: (noOfInputs-1)(noOfOutputs-1)/2

Variance: (noOfInputs-1)(noOfOutputs-1)/2n^2

Upper Bound on the Variance

- In both cases var(C(W)) < I.J / n

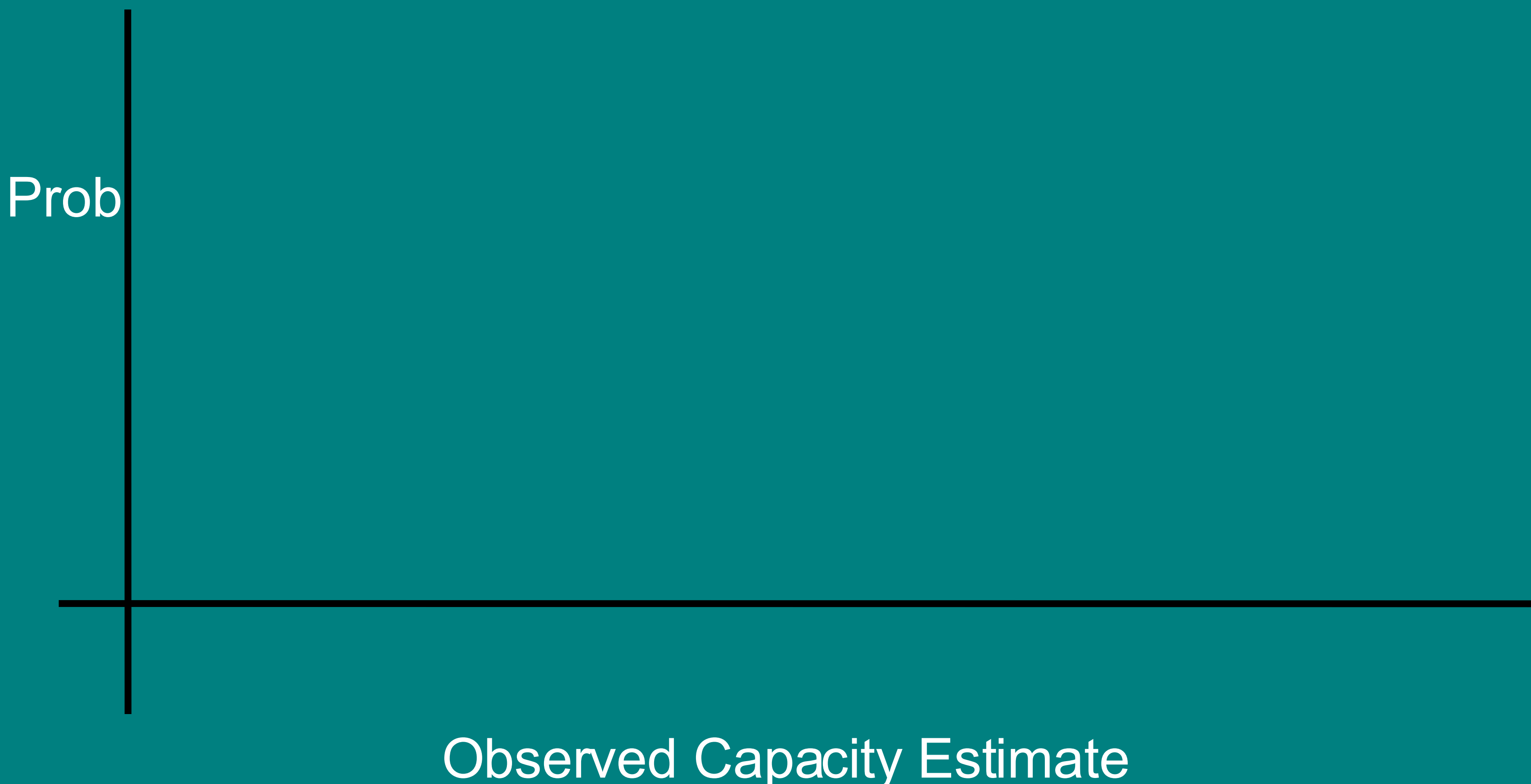
- Rule of thumb:
  - if I.J >> n the variance will be low and the results actuate.
  - If you can get this many samples then statically analysis is useful, otherwise not.

To Analyse a System.

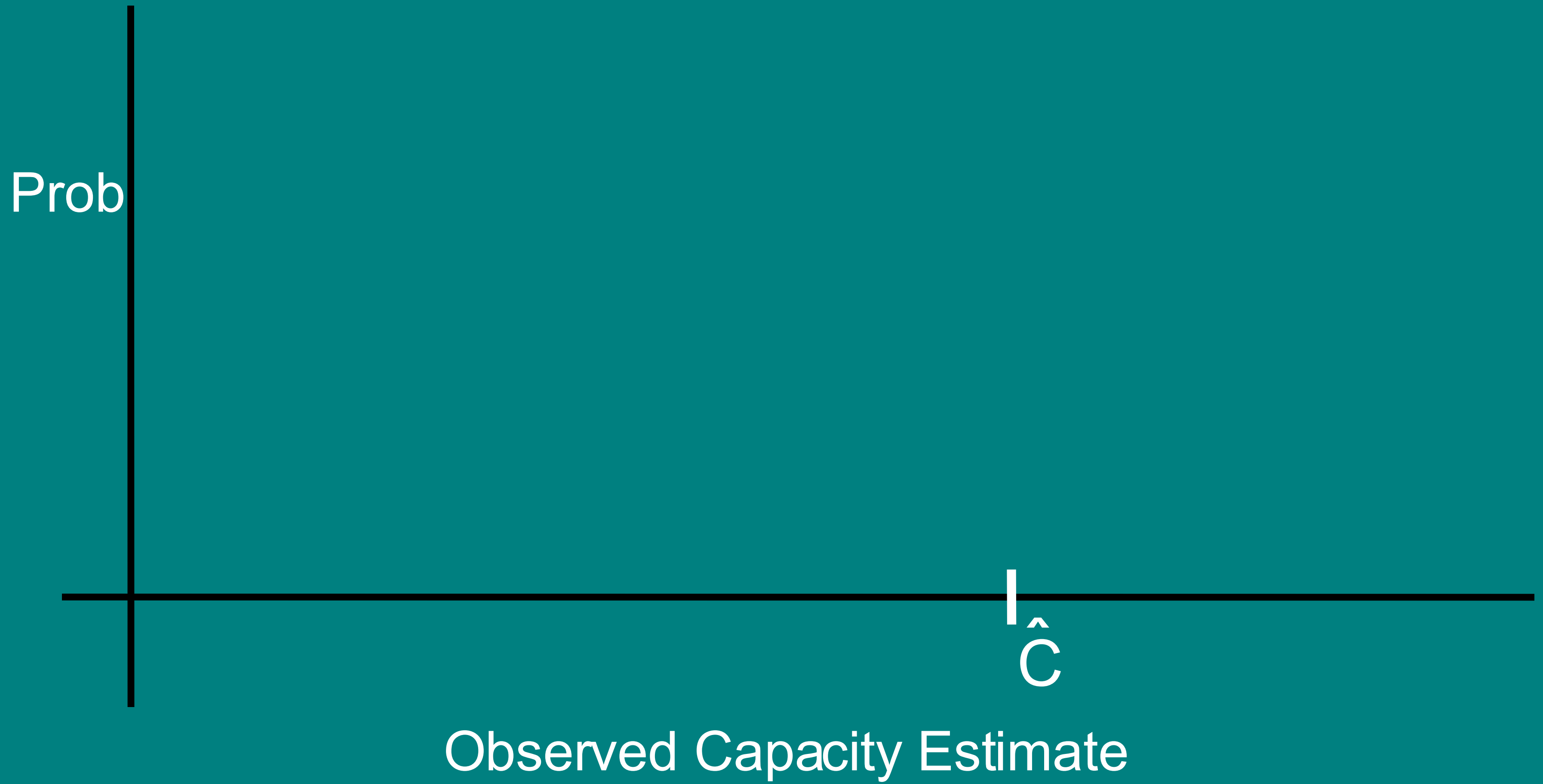
- We define the inputs (I) and outputs (J).
- Run n tests of the system with n >> I.J
- Estimate the matrix and find $\hat{C} = I(\hat{Q}_m(\hat{W}_n),\hat{W}_n)$
- Point Estimate is:

  $$\mathrm{Max}\ (\ 0,\ I(\hat{Q}_m(\hat{W}_n),\hat{W}_n) - (I-1)(J-1)/2n\ )$$

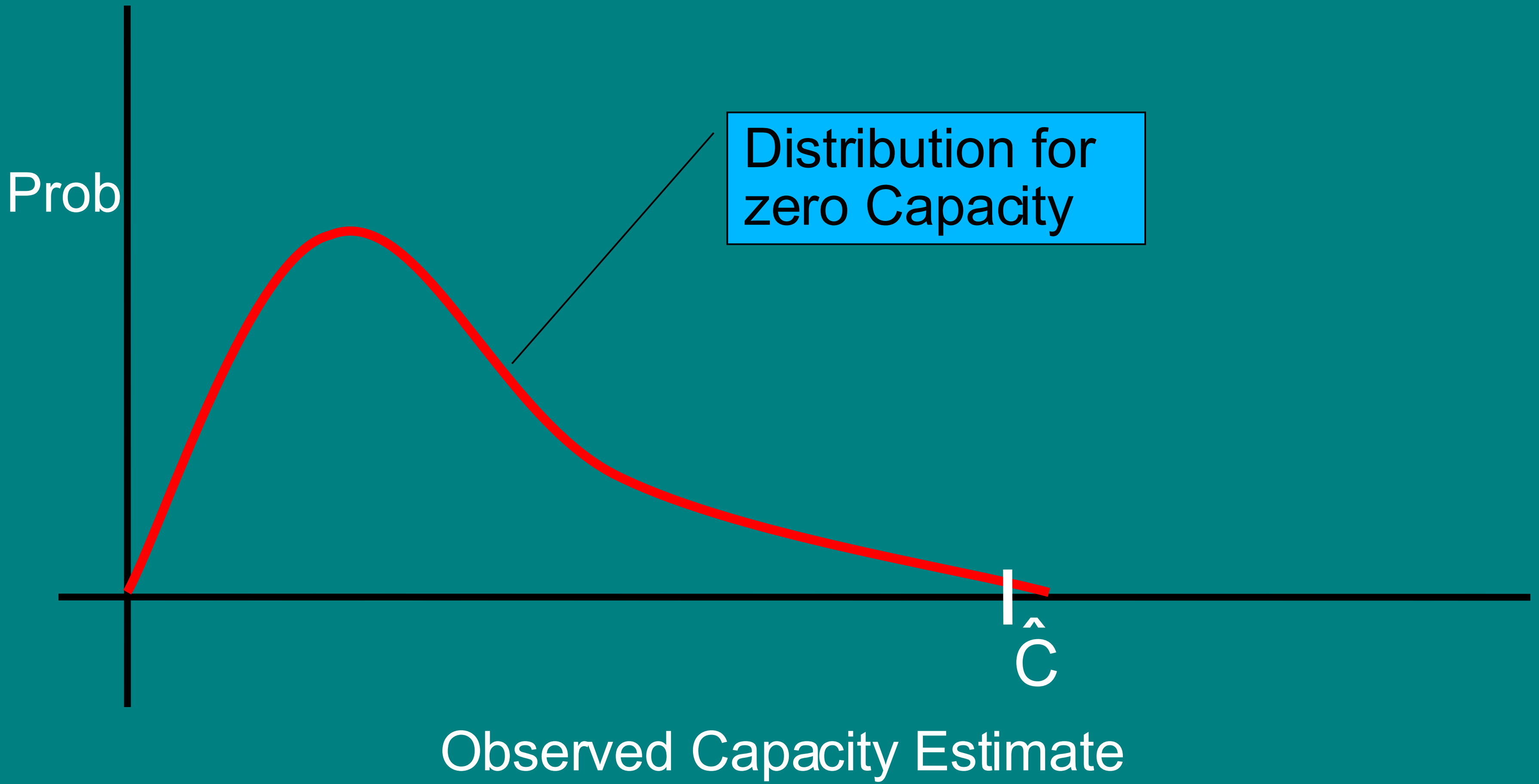
Using the Distributions



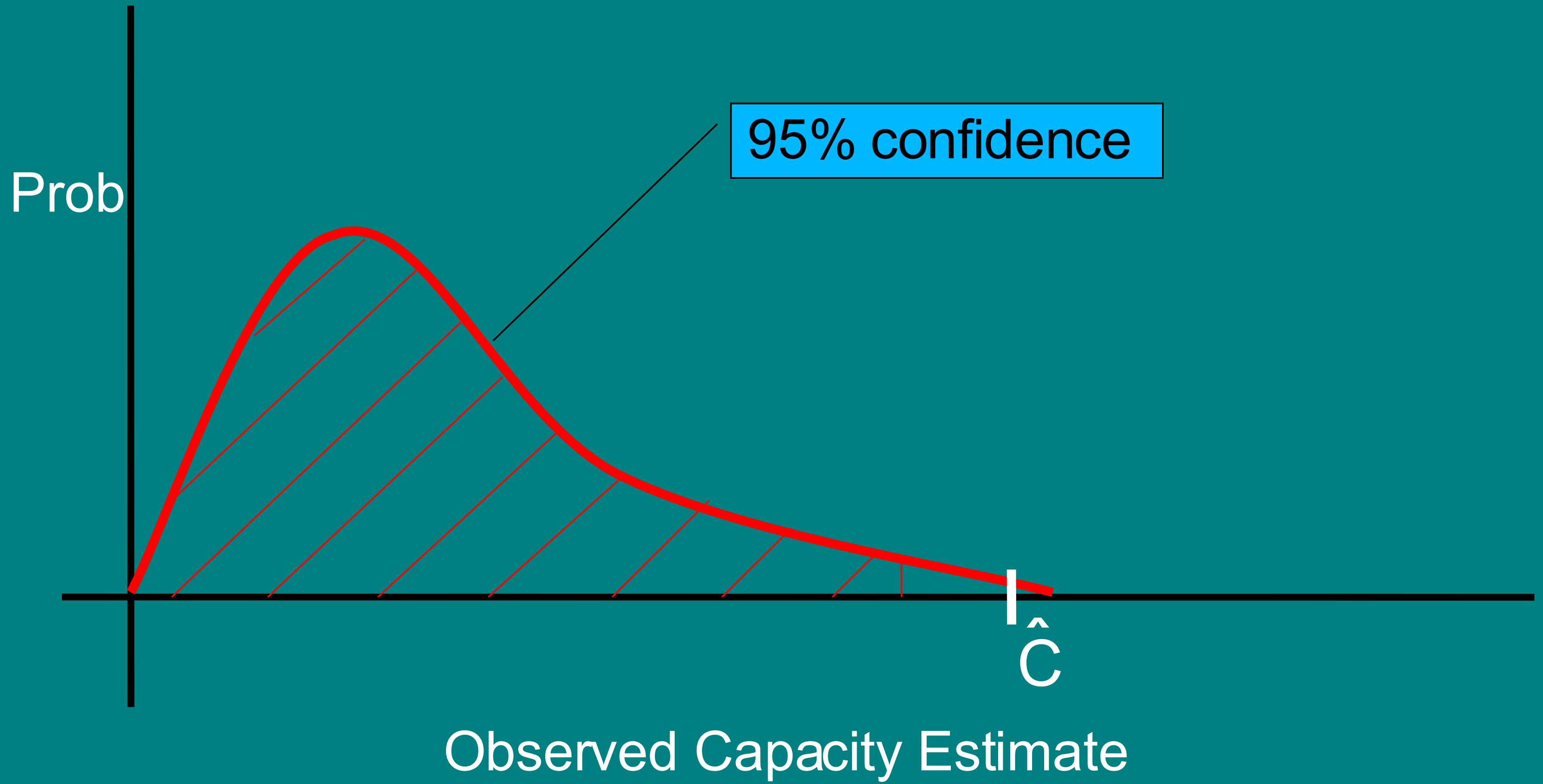
Using the Distributions



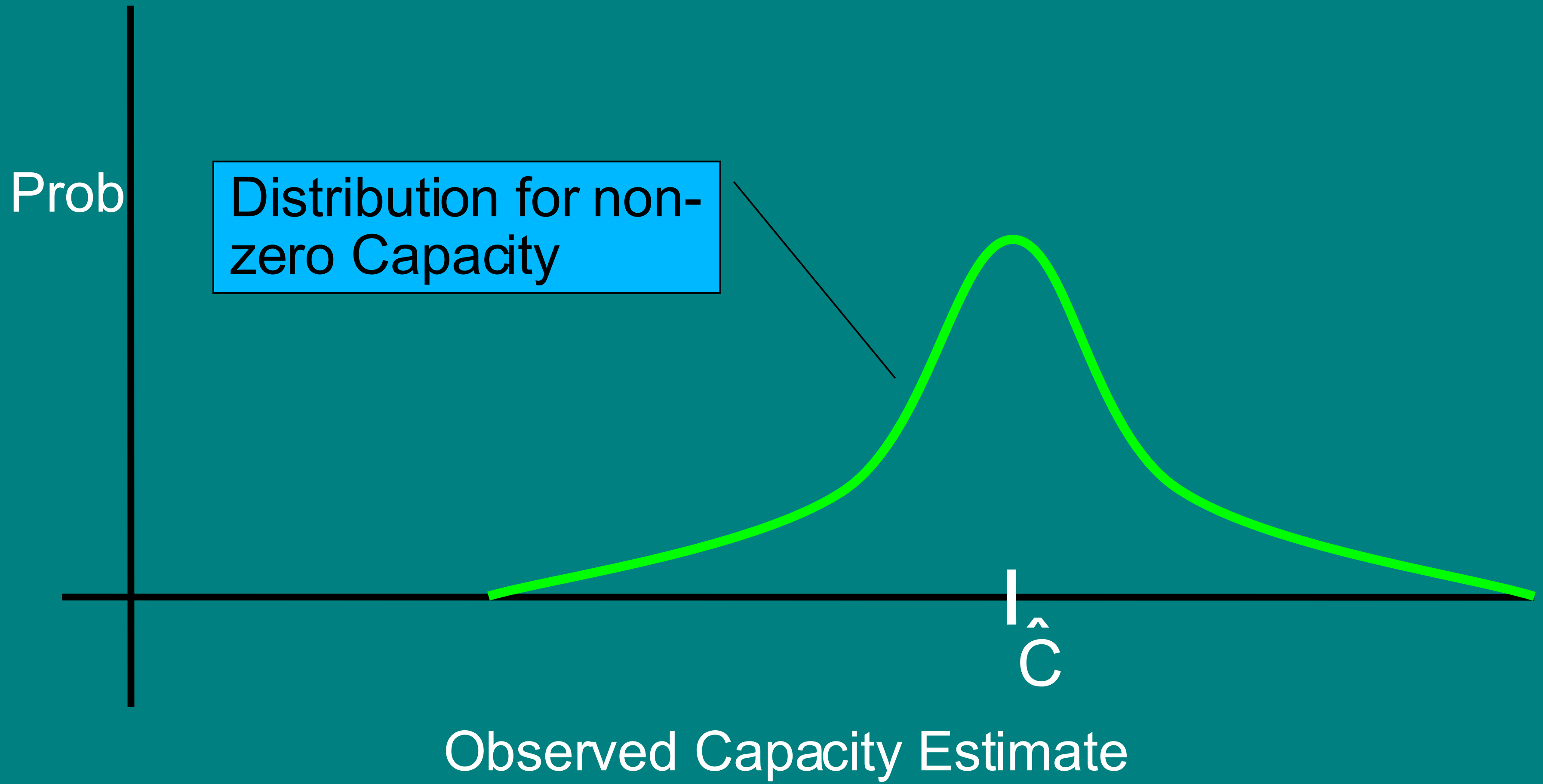
Using the Distributions



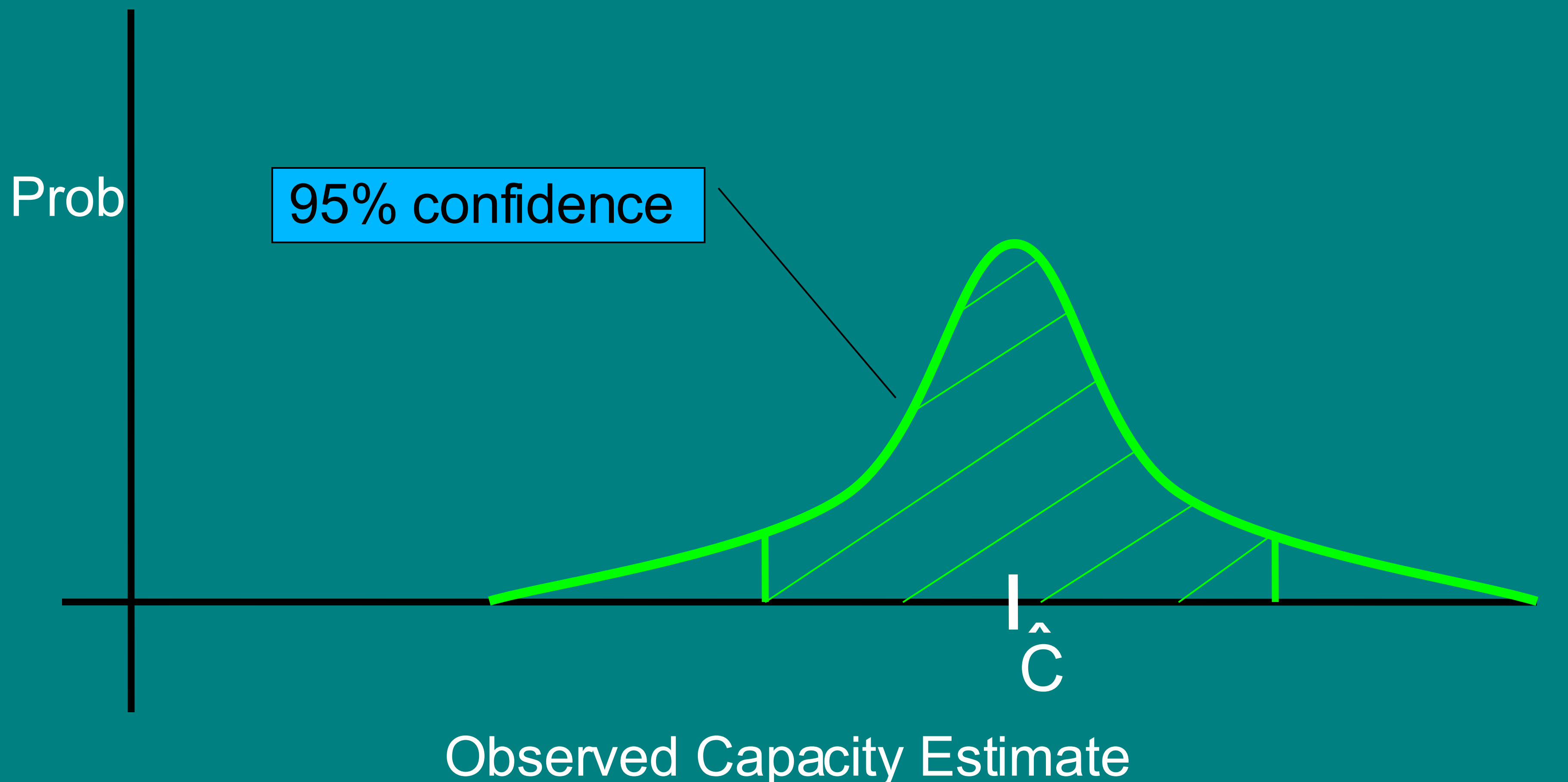
Using the Distributions



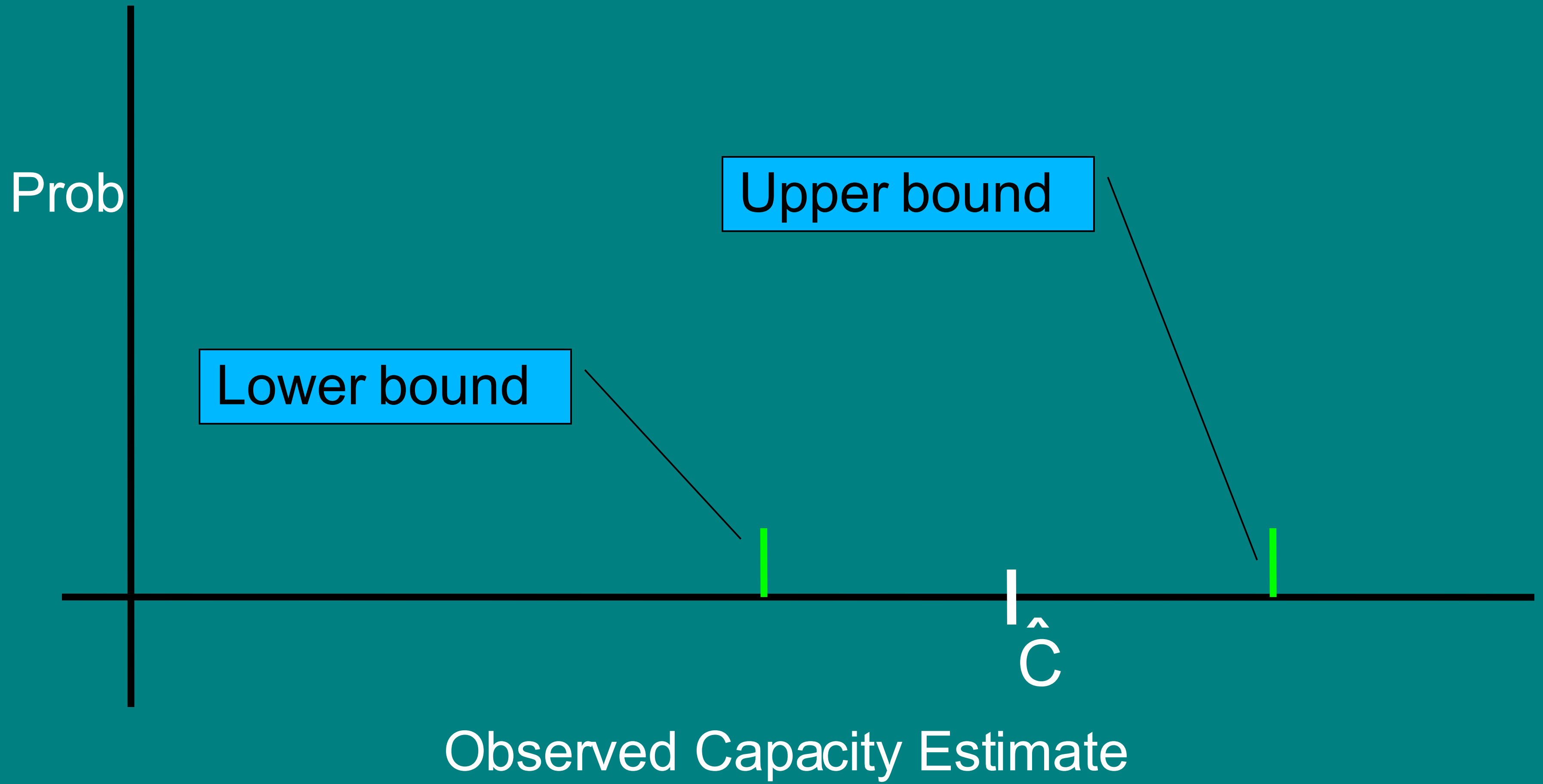
Using the Distributions



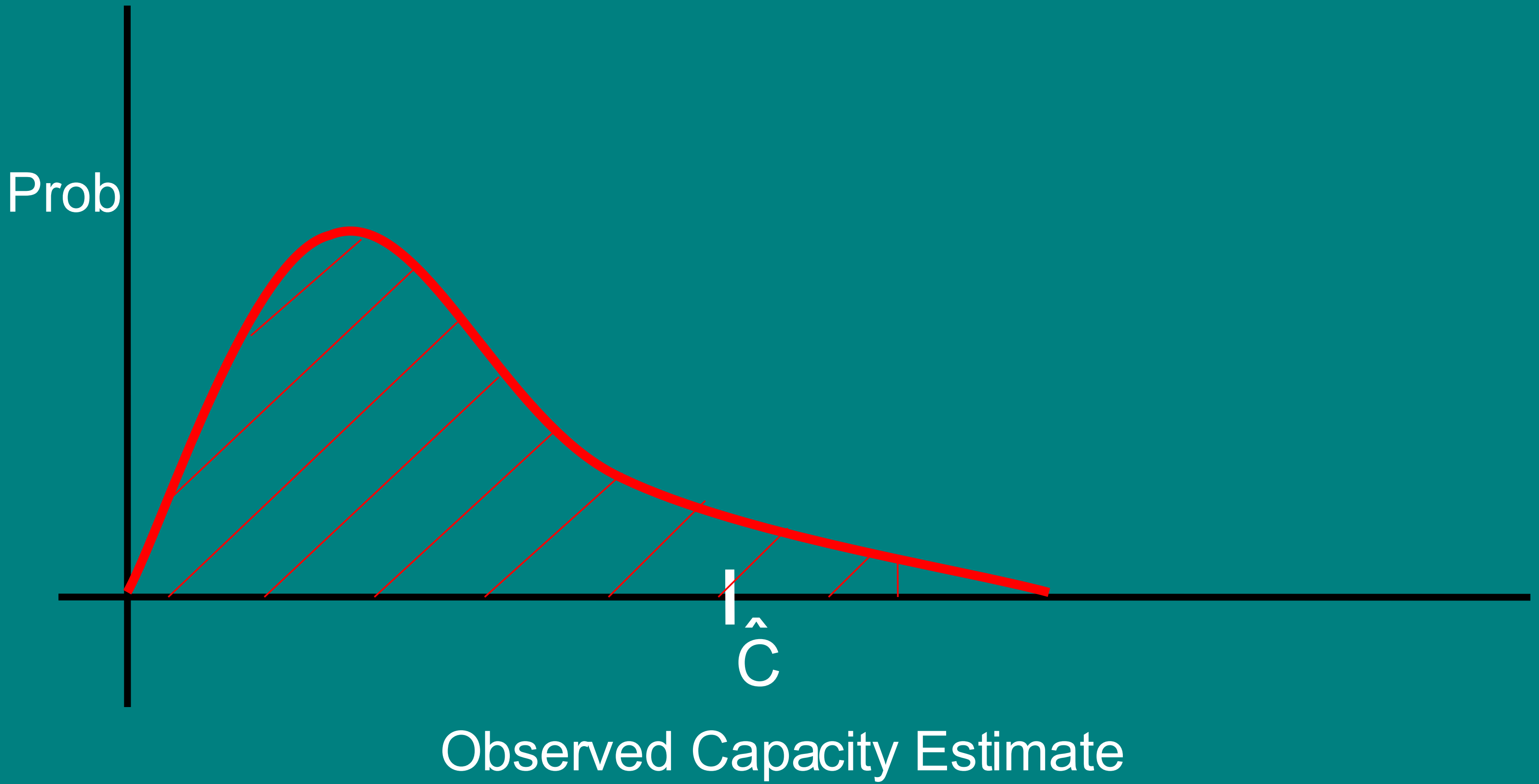
Using the Distributions



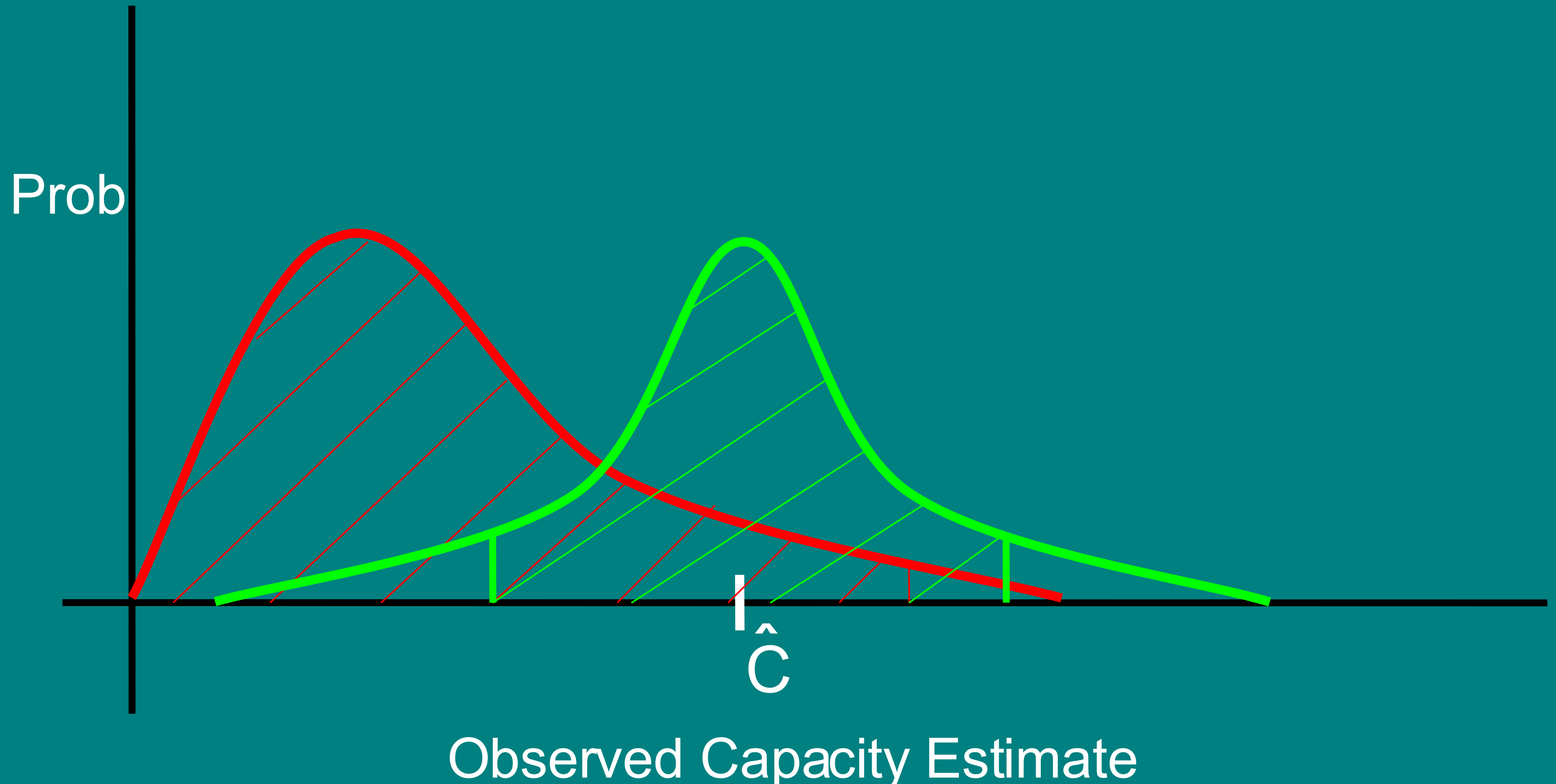
Using the Distributions



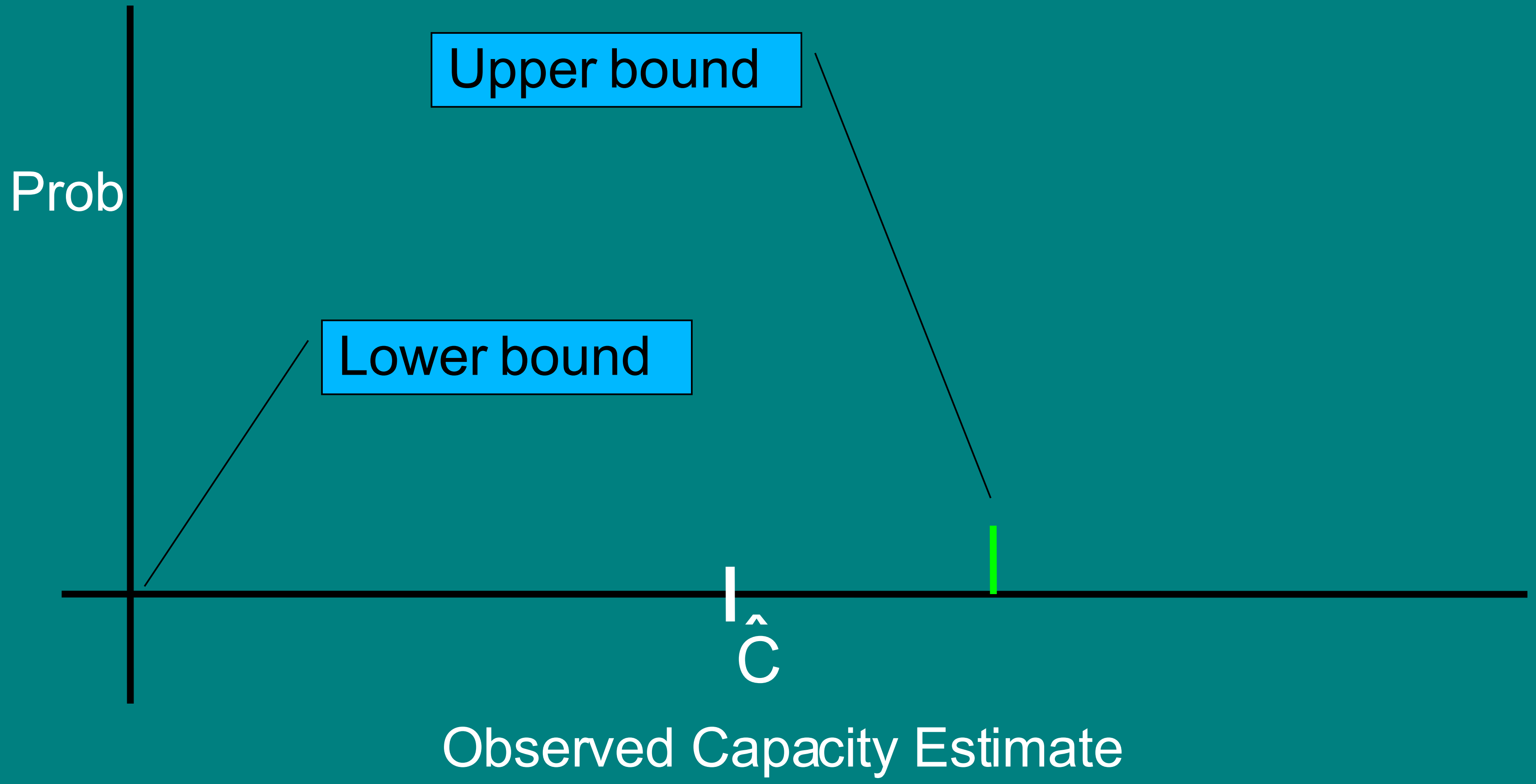
Using the Distributions



Using the Distributions



Using the Distributions



Test for Zero Leakage

- But what if we want to know if the leakage is really zero?
- What distinguishes the zero from the non-zero case is the variance:
  - $O(n^{-1})$ for non zero
  - $O(n^{-2})$ for zero.
- A large enough sample size will always tell these apart.

Test for Zero Leakage

- Run 100 tests of the system and calculate the observed variance “o” in the tests results.
- Test o against the predicated variance for zero and non-zero observations.
- If it matches the zero predication but not the non-zero we can conclude that there is zero leakage.
- If it only matches the non-zero predication then we can find the confidence interval for the results.
- If it matches both then increase the sample size.

Demo:

- Example of using analysing some data sampled form a toy Dining crypos implementation and a real Mixminion mix node.

Comparison with Bayesian Methods

- In our analysis we make no assumptions about how the system is used
  - says nothing about a single run of the system,
  - user actions can be depended on previous user action, Capacity is still the maximum amount of information that can be sent.

- Bayesian analysis makes assumption about the input distribution.
  - Can infer guilt of a user from the actions
  - Not valid if the assumptions aren’t correct or user actions depend on previous actions.

Extensions

- Calculate the distribution of conditional mutual information and the upper bound on capacity.
- Statefull channels, for more complex and interactive systems.
- Continuous mutual information and capacity for continuous data sets.

Overview

- Estimate information leakage statistically from trail runs of a real system.
  - Automatic tool to calculate information leakage.
- We work out bounds on the possible error.
- We present an ***if, and only if***, test for ***zero*** leakage.
- For actuate results you need more samples than the (no. of inputs) x (no. of observations).

Questions?