Overview

- Distribution of Sample Mean
- Weak Law of Large Numbers
- Confidence Intervals
- Confidence Intervals: Boostrapping
- Confidence Intervals: Unicorns Example Revisited

Distribution of Sample Mean

Consider N random variables $X_1, \cdots X_N$.

- Let's consider $\bar{X} = \frac{1}{N} \sum_{k=1}^{N} X_k$.
- $\bar{X}$ is called the "sample mean" or the "empirical mean".
- $\bar{X}$ is a random variable.

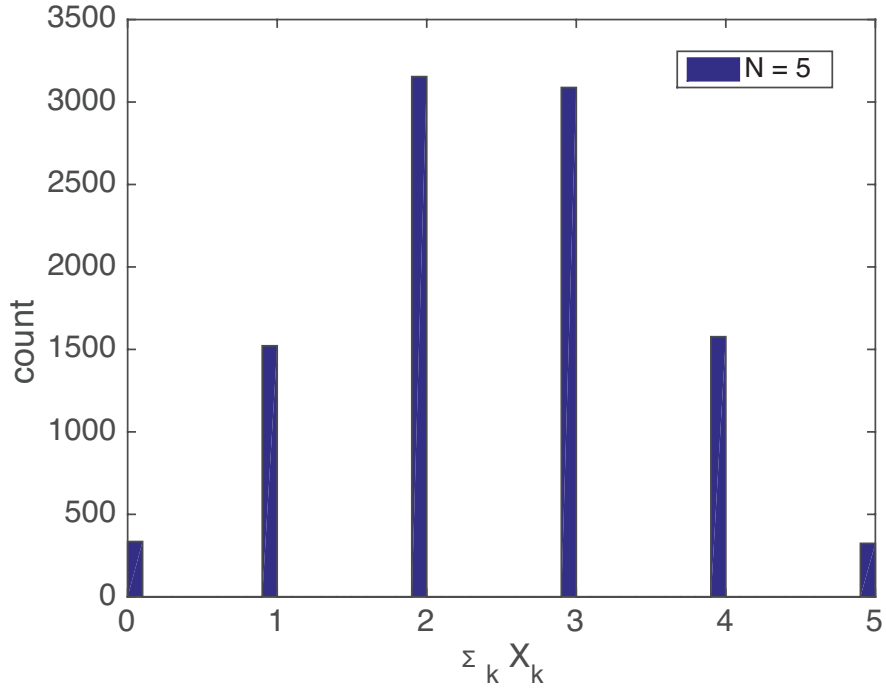
Suppose we observe values for $X_1, \cdots X_N$ and calculate the empirical mean of the observed values. That gives us one value for $\bar{X}$. But the value of $\bar{X}$ changes depending on the observed values.

- Suppose we toss a fair coin $N = 5$ times and get H, H, H, T, T. Let $X_k = 1$ when come up heads. Then $\frac{1}{N} \sum_{k=1}^{N} X_k = \frac{3}{5}$
- Suppose we toss the coin another $N = 5$ times and get T, T, H, T, H. Now $\frac{1}{N} \sum_{k=1}^{N} X_k = \frac{2}{5}$

Distribution of Sample Mean

Toss fair coin $N = 5$ times and calculate $\sum_{k=1}^{N} X_k$. Repeat, and plot histogram of values. Its binomial $Bin(N, \frac{1}{2})$.

- X=[];for i=1:10000,X=[X,sum((rand(1,5)<0.5))]; end; hist(X,50)



Distribution of Sample Mean

Random variable $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$.

- Suppose the X_k are **independent and identically distributed** (i.i.d)
- Each X_k has mean $E(X_k) = \mu$ and variance $Var(X_k) = \sigma^2$.

Then we can calculate the mean of $\bar{X}$ as:

$$E(\bar{X}) = E(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{1}{N}\sum_{k=1}^{N} E(X_k) = \mu$$

NB: recall linearity of expectation: $E(X + Y) = E(X) + E(Y)$ and $E(aX) = aE[X]$

- We say $\bar{X}$ is an **unbiased estimator** of μ since $E[\bar{X}] = x$

Distribution of Sample Mean

We can calculate the variance of $\bar{X}$ as:

$$var(\bar{X}) = var(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{1}{N^2} var(\sum_{k=1}^{N} X_k)$$

$$= \frac{1}{N^2}\sum_{k=1}^{N} var(X_k) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$$

NB: recall $Var(aX) = a^2 Var(X)$ and $Var(X + Y) = Var(X) + Var(Y)$ when X, Y independent.

- As N increases, the variance of $\bar{X}$ falls.
- $Var(NX) = N^2 Var(X)$ for random variable X.
- But when add together **independent** random variables $X_1 + X_2 + \cdots$ the variance is only $NVar(X)$ rather than $N^2 Var(X)$
- This is due to **statistical multiplexing**. Small and large values of X_i tend to cancel out for large N.

Weak Law of Large Numbers[1]

Consider N independent identically distributed (i.i.d) random variables $X_1, \cdots X_N$ each with mean μ and variance σ^2. Let $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$. For any $\epsilon > 0$:

$$P(|\bar{X} - \mu| \geq \epsilon) \to 0 \text{ as } N \to \infty$$

That is, $\bar{X}$ **concentrates** around the mean μ as N increases.
Proof:

- $E(\bar{X}) = E(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{1}{N}\sum_{k=1}^{N} E(X_k) = \mu$
- $var(\bar{X}) = var(\frac{1}{N}\sum_{k=1}^{N} X_k) = \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$
- By Chebyshev's inequality: $P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}$

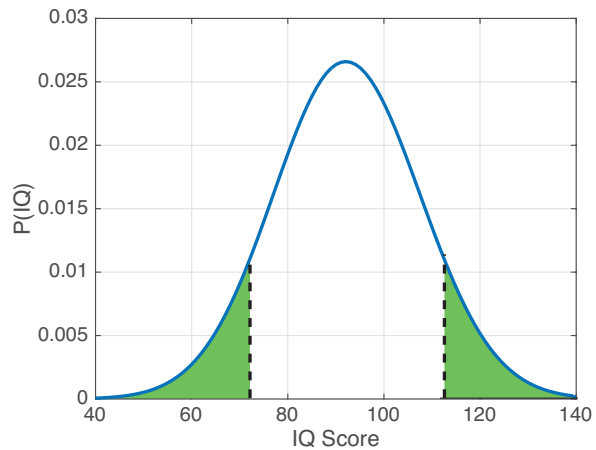
[1] There is also a **strong law of large numbers**, but we won't deal with that here.

Who cares ?

- Suppose we have an event E
- Define indicator random variable X_i equal to 1 when event E is observed in trial i and 0 otherwise
- Recall $E[X_i] = P(E)$ is the probability that event E occurs.
- $\bar{X} = \frac{1}{N} \sum_{k=1}^{N} X_k$ is then the relative frequency with which event E is observed over N experiments.
- And ...
  $P(|\bar{X} - \mu| \geq \epsilon) \to 0$ as $N \to \infty$
  tells us that this observed relative frequency $\bar{X}$ converges to the probability $P(E)$ of event E as N grows large.
- So the law of large numbers formalises the intuition of probability as frequency when an experiment can be repeated many times. But probability still makes sense even if cannot repeat an experiment many times – all our analysis still holds.

Confidence Intervals

- Recall that when a random variable lies in an interval $a \leq X \leq b$ with a specified probability we call this a confidence interval e.g. $p - 0.05 \leq Y \leq p + 0.05$ with probability at least 0.95.



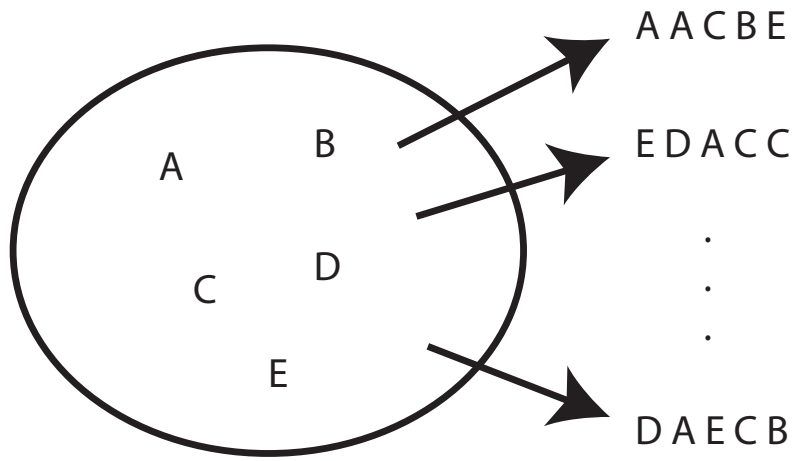
- Chebyshev inequality allows us to calulate confidence intervals given the mean and variance of a random variable.
- For sample mean $\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k$, Chebyshev inequality tells us $P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}$ where μ is mean of X_k and σ^2 is its variance.
- E.g. When $\epsilon = \frac{\sigma}{\sqrt{0.05N}}$ then $\frac{\sigma^2}{N\epsilon^2} = 0.05$ and Chebyshev tells us that $\mu - \frac{\sigma}{\sqrt{0.05N}} \leq \bar{X} \leq \mu + \frac{\sigma}{\sqrt{0.05N}}$ with probability at least 0.95.

Confidence Intervals: Bootstrapping

Mean μ, variance σ^2 and number of points N summarises our N data points using three numbers. But we have $N \gg 3$ data points. Can we use these to also empirically estimate the distribution of the sample mean ? Yes !

- Makes use of the fact that computing power is cheap.
- The N data points are drawn independently from the same probability distribution F
- So the idea is to use these N data points as a surrogate for F. To generate new samples from F we draw uniformly at random from our N data points. This is **sampling with replacement**.
- Suppose our data is $\{A, B, C, D, E, F\}$. Select one point uniformly at random e.g. B. Select a second point uniformly at random, might be B again or might be something else. And so on until we get desired number of samples.
- Bootstrapping is an example of a **resampling method**

Confidence Intervals: Bootstrapping



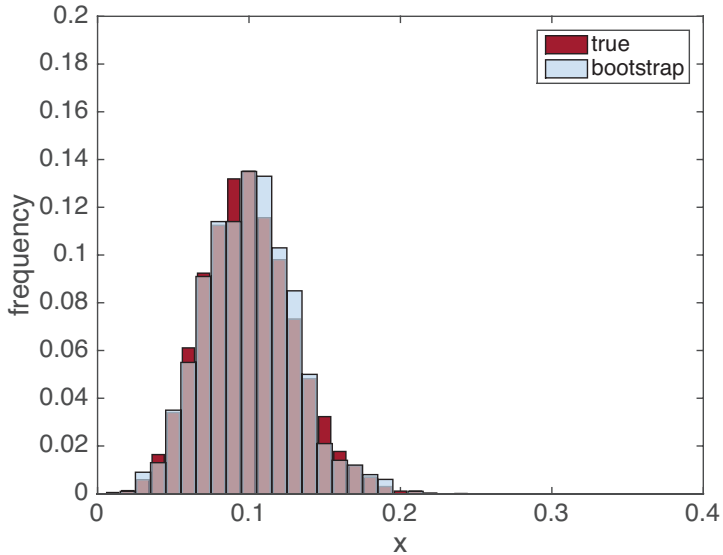
Bootstrapping:

- Draw a sample of N data points uniformly at random from data, with replacement.
- Using this sample estimate the mean $X_1 = \frac{1}{N}\sum_{i=1}^{N} X_{1,i}$
- Repeat, to generate a set of estimates X_1, $X_2, \cdots$.
- The distribution of these estimates approximates the distribution of the sample mean (its not exact)

Confidence Intervals: Bootstrapping

Example: coin tosses yet again.

- Toss $N = 100$ biased coins, lands heads with prob $p = 0.1$. $X_i = 1$ if i'th toss is heads, 0 otherwise
- Sample with replacement from the $N = 100$ data points.
- Calculate sample mean $X = \frac{1}{N} \sum_{i=1}^{N} X_i$
- Repeat 1000 times and plot observed distribution of X.



Confidence Intervals: Bootstrapping

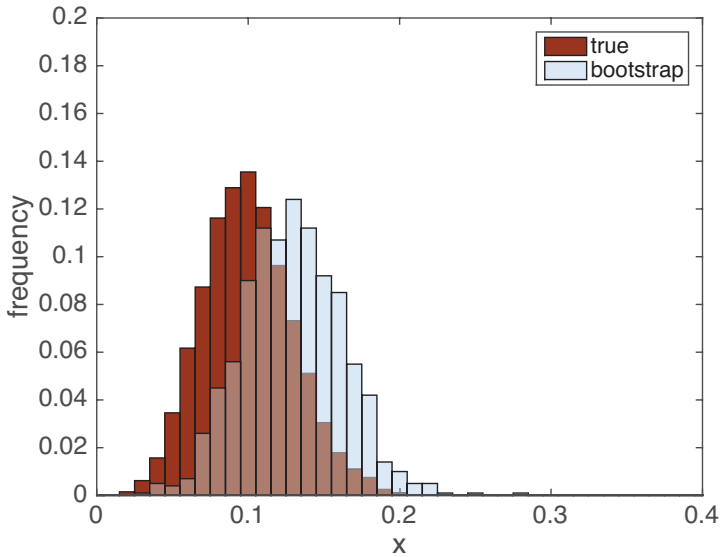
Matlab code used to generate above plot:

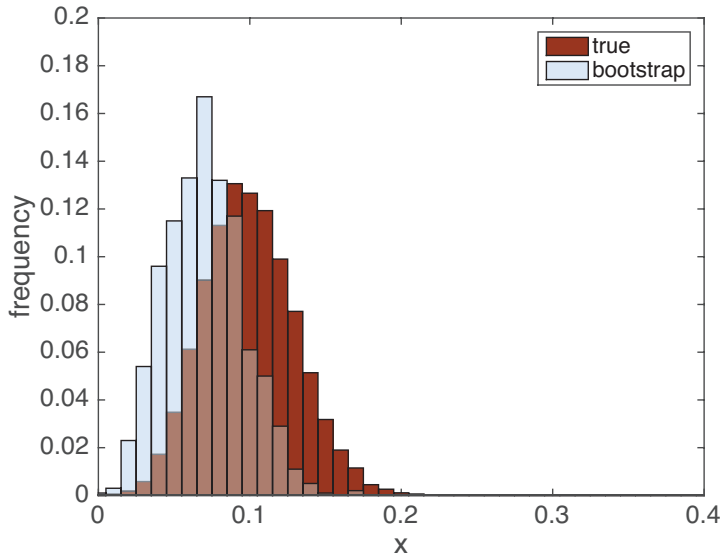
```
p=0.1; n=100;
y=[];
for i=1:10000,
    r=rand(1,n); x=(r<=p); y=[y;sum(x)/n];
end;
xx=[0:1/n-eps:1]; nn=histc(y,xx,1);
hold off, bar(xx,nn/sum(nn),1)

r=rand(1,n); x=(r<=p);
y1=[];
for i=1:1000,
    y1= [y1; sum(x(randi(n,1,n)))/n];
end
nn1=histc(y1,xx);
hold on, bar(xx,nn1/sum(nn1),1,'g')
axis([0 0.4 0 0.2])
```

Confidence Intervals: Bootstrapping

Note: bootstrap estimate of the distribution is only approximate.

- Different data leads to different estimates of the the distribution, e.g. here are two more runs of the coin toss example.
- But v handy all the same.
- Using our empirical estimate of the distribution of the sample mean X we can estimate confidence intervals etc.

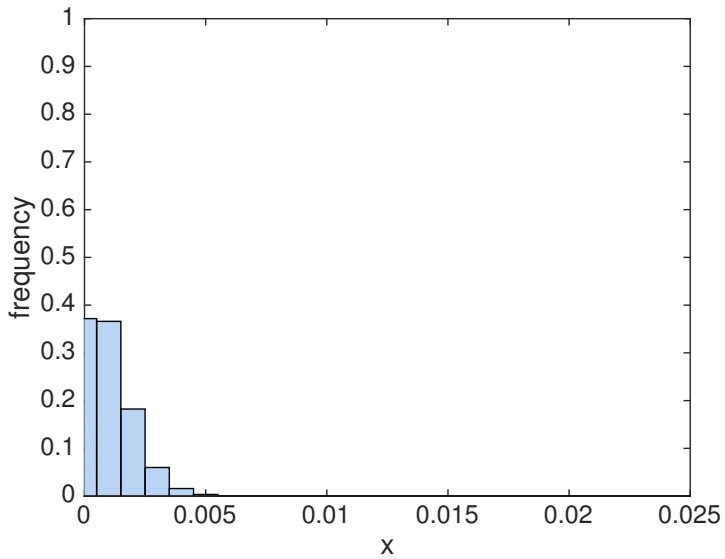




Confidence Intervals: Unicorns Example Revisited

Bootstrapping doesn't require data to be normally distributed.

- How about those pet unicorns ?
- Data consists $N = 1000$ survey results. $X_i = 1$ if answer "yes" and 0 otherwise. We have 999 of the X_i's equal to 0 and one equal to 1. Bootstrap estimate of distribution of $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$:



- Makes clear that number of unicorns is likely to be 0.

Confidence Intervals: Unicorns Example Revisited

- Use bootstrap distribution to estimate confidence intervals
- Estimate $P(\bar{X} < 1/4.5 \times 10^3) \approx 0.4$ and $P(\bar{X} < 10/4.5 \times 10^3) \approx 0.9$ i.e. if scale up up $4.5 \times 10^6/N = 4.5 \times 10^3$ then prob no pet unicorns is estimated at about 0.4 and of less than 10 unicorns at about 0.9.
- Compare with original estimate of 4500 pet unicorns based on mean value alone.

```
x=[1,zeros(1,999)];
y1=[]; n=length(x);
for i=1:20000, y1=[y1; sum(x(randi(n,1,n)))/n]; end
xx=[0:1/n-eps:0.025];
nn1=histc(y1,xx);
bar(xx*4.5e6,nn1/sum(nn1),1)
axis([0 0.01*4.5e6 0 1])
sum(y1<10*0.22e-03)/sum(nn1)
```