HW3 has been posted
It is due next Tuesday (4/13) before the lecture

Gradescope questions to Ananthan
Mean & Variance of a Linear Function

\[ Y = c_1X_1 + c_2X_2 + \ldots + c_pX_p \]

\[ E(Y) = c_1E(X_1) + c_2E(X_2) + \ldots + c_pE(X_p) \] \hspace{1cm} (5-25)

\[ V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \ldots + c_p^2V(X_p) + 2\sum_{i<j} c_ic_j\text{cov}(X_iX_j) \] \hspace{1cm} (5-26)

If \( X_1, X_2, \ldots, X_p \) are independent, then \( \text{cov}(X_iX_j) = 0 \),

\[ V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \ldots + c_p^2V(X_p) \] \hspace{1cm} (5-27)
Descriptive statistics:
Sample mean and its variance

Standard error vs
Standard deviation
Some Definitions

• The random variables $X_1, X_2, ..., X_n$ are a random sample of size $n$ if:
  a) The $X_i$ are independent random variables.
  b) Every $X_i$ has the same probability distribution.

Such $X_1, X_2, ..., X_n$ are also called independent and identically distributed (or i.i.d.) random variables.

• A **statistic** is any function of the observations in a random sample.

• The probability distribution of a statistic is called a **sampling distribution**.
Statistic #1: Sample Mean

If the $n$ observations in a random sample are denoted by $x_1, x_2, \ldots, x_n$, the sample mean is

$$
\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (6-1)
$$
**IMPORTANT:**

Sample mean $\overline{X}$ is drawn from a random variable

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$$E(\overline{X}) = \frac{n \cdot E(X_i)}{n} = \frac{n \cdot \mu}{n} = \mu$$

$$V(\overline{X}) = \frac{n \cdot V(X_i)}{n^2} = \frac{n \cdot \sigma^2}{n} = \frac{\sigma^2}{n}$$

**Stand. dev. $(\overline{X}) = \frac{\sigma}{\sqrt{n}}$**
Matlab exercise

• Do a numerical experiment: generate a sample of size n by rolling n fair dice

• Calculate the sample mean \( \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} \)

• Repeat Stats=100,000 times

• Generate PMFs of sample means for different samples sizes: n=1, n=2, n=3, n=5, and n=10

• Plot them in the same (semi-logarithmic) figure

• What do you see?

• Template is at the website: central_limit_theorem_template.m
Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size $n$ is taken from a population (either finite or infinite) with mean $\mu$ and finite variance $\sigma^2$, and if $\bar{X}$ is the sample mean, then the limiting form of the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

(7-1)

for large $n$, is the standard normal distribution. If $X_1, X_2, \ldots, X_n$ are themselves normally distributed - for any $n$
Sampling Distributions of Sample Means

Figure 7-1  Distributions of average scores from throwing dice.
Mean = (6+1)/2=3.5
Sigma^2 = [(6-1+1)^2-1]/12=2.92
Sigma=1.71

Formulas
\[ \mu = \frac{b + a}{2} \]
\[ \sigma_X^2 = \frac{(b - a + 1)^2 - 1}{12} \]
\[ \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n} \]
Example 7-1: Resistors

An electronics company manufactures resistors having a mean resistance of 100 ohms and a standard deviation of 10 ohms. What is the approximate probability that a random sample of $n = 25$ resistors will have an average resistance of less than 95 ohms?
Example 7-1: Resistors

An electronics company manufactures resistors having a mean resistance of 100 ohms and a standard deviation of 10 ohms. What is the approximate probability that a random sample of \( n = 25 \) resistors will have an average resistance of less than 95 ohms?

\[
\mu = 100 \text{ ohms }, \quad \sigma = 10 \text{ ohms }, \quad n = 25
\]

\[
\mu_{\bar{x}} = \mu \cdot \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = \frac{10}{5} = 2 \text{ ohms}
\]

\[
Z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{95 - \mu_{\bar{x}}}{2} = \frac{95 - 100}{2} = -2.5
\]

\[
\Pr(\bar{x} < 95) = \Phi(Z_{\bar{x}}) = \Phi(-2.5) = 0.0062
\]
Example 7-1: Resistors

An electronics company manufactures resistors having a mean resistance of 100 ohms and a standard deviation of 10 ohms. What is the approximate probability that a random sample of $n = 25$ resistors will have an average resistance of less than 95 ohms?

Answer:

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2.0$$

$$\Phi \left( \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \right) = \Phi \left( \frac{95 - 100}{2} \right)$$

$$= \Phi (-2.5) = 0.0062$$
Two Populations

We have two independent populations. What is the distribution of the difference of their sample means?

The sampling distribution of $\bar{X}_1 - \bar{X}_2$ has the following mean and variance:

$$
\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2
$$

$$
\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}
$$
Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means $\mu_1$ and $\mu_2$, and variances $\sigma_1^2$ and $\sigma_2^2$,

And if $X_{\bar{1}}$ and $X_{\bar{2}}$ are the sample means of two independent random samples of sizes $n_1$ and $n_2$ from these populations:

Then the sampling distribution of:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$  \hspace{1cm} (7-4)$$

is approximately standard normal, if the conditions of the central limit theorem apply.

If the two populations are normal, then the sampling distribution is exactly standard normal.
Example 7-3: Aircraft Engine Life

The effective life of a component used in jet-turbine aircraft engines is a random variable with $\mu_{\text{old}}=5000$ hours and $\sigma_{\text{old}}=40$ hours (old). The engine manufacturer introduces an improvement into the manufacturing process for this component that changes the parameters to $\mu_{\text{new}}=5050$ hours and $\sigma_{\text{new}}=30$ hours (new).

Random samples of 16 components manufactured using “old” process and 25 components using “new” process are chosen.

What is the probability new sample mean is at least 25 hours longer than old?
Example 7-3: Aircraft Engine Life

The effective life of a component used in jet-turbine aircraft engines is a random variable with \( \mu_{\text{old}} = 5000 \) hours and \( \sigma_{\text{old}} = 40 \) hours (old). The engine manufacturer introduces an improvement into the manufacturing process for this component that changes the parameters to \( \mu_{\text{new}} = 5050 \) hours and \( \sigma_{\text{new}} = 30 \) hours (new).

Random samples of 16 components manufactured using “old” process and 25 components using “new” process are chosen.

What is the probability new sample mean is at least 25 hours longer than old?

\[
\bar{x}_{\text{old}} = \frac{6000}{\sqrt{16}} = 10 \text{ hrs}
\]
\[
\bar{x}_{\text{new}} = \frac{71}{\sqrt{25}} = 7 \text{ hrs}
\]
\[
\bar{x}_{\text{tot}} = \sqrt{6 \bar{x}^2 + 7} = \sqrt{-100 + 36} \approx 11.7 \text{ hrs}
\]
\[
\mu_{\text{new}} - \mu_{\text{old}} = 50 \text{ hrs}
\]
\[
Z = \frac{25 - (50)}{11.7} = -2.14
\]
\[
\text{Prob}(Z > -2.14) = 0.9840
\]
Example 7-3: Aircraft Engine Life

The effective life of a component used in jet-turbine aircraft engines is a normal-distributed random variable with parameters shown (old). The engine manufacturer introduces an improvement into the manufacturing process for this component that changes the parameters $\mu$ and $\sigma$ as shown (new). Random samples are selected from the “old” process and “new” process as shown.

What is the probability new sample mean is at least 25 hours longer than old?

<table>
<thead>
<tr>
<th>Process</th>
<th>Old (1)</th>
<th>New (2)</th>
<th>Diff (2-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>5,000</td>
<td>5,050</td>
<td>50</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>40</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td>$n$</td>
<td>16</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

| Calculations | $s/\sqrt{n}$ | 10 | 6 | 11.7 |
|              | $z$         |    |   | -2.14 |
|              | $P(\bar{x}_{2}-\bar{x}_{1} > 25) = P(Z > z ) = 0.9840$ |    |   |      |

Figure 7-4 Sampling distribution of the sample mean difference.
Questions
Found in Google Autocomplete

- Why are whales jumping?
- Why are witches green?
- Why are there mirrors above beds?
- Why do I say uh when I mean yes?
- Why is sea salt better?
- Why are there trees in the middle of fields?
- Why is there not a Pokemon MMO?
- Why is there laughing in TV shows?
- Why are there doors on the freeway?
- Why are there so many schochiko running why are there any countries in Antarctica?
- Why are there scary sounds in Minecraft?
- Why is there sticking in my stomach?
- Why are there two slashes after HTTP?
- Why are there celebrities?
- Why do snakes exist?
- Why do oysters have pearls?
- Why are ducks called ducks?
- Why do they call it the clap?
- Why are Kyle and Carman friends?
- Why is there an arrow on Mmmm's head?
- Why are these messages blue?
- Why are there mustaches on clothes?
- Why are there mustaches on cars?
- Why are there mustaches everywhere?
- Why are there so many birds in Ohio?
- Why is there so much rain in Ohio?
- Why is Ohio weather so weird?
- Why are there male and female bikes?
- Why are there tiny spiders in my house?
- Why do spiders come inside?
- Why are there huge spiders in my house?
- Why are there lots of spiders in my house?
- Why are there spiders in my room?
- Why are there so many spiders in my room?
- Why do spider bites itch?
- Why is dying so scary?
- Why is Mt. Vesuvius there?
- Why do they say t minus?
- Why are there ocelots?
- Why are wrestlers always wet?
- Why are oceans becoming more acidic?
- Why is Arwen dying?
- Why aren't my quail laying eggs?
- Why aren't my quail eggs hatching?
- Why aren't there any foreign military bases in America?

Credit: XKCD
comics

- Why aren't my arms growing?
- Why aren't there so many crows in Rochester, Minnesota?
- Why is there no light in my laptop?
- Why is there an owl in my backyard?
- Why is there an owl outside my window?
- Why is there an owl on the dollar bill?
- Why do owls attack people?
- Why are AK-47s so expensive?
- Why are there helicopters circling my house?
- Why are there gods?
- Why are there two spoons?
- Why is MT Value there?
- Why is isolation bad?
- Why do knees click?
- Why aren't there e grades?
- Why is sex so important?
- Why do my parents look at me?
- Why do people cry when I die?
- Why is life so boring?
- Why aren't there gins in fireworks?
- Why aren't there guns in Harry Potter?
- Why are ultrasounds important?
- Why do I feel dizzy when I'm hungry?
- Why is there a red line through HTTPS?
Descriptive statistics:
Point estimation:
Some Definitions

• The random variables $X_1, X_2, \ldots, X_n$ are a random sample of size $n$ if:
  a) The $X_i$ are independent random variables.
  b) Every $X_i$ has the same probability distribution.

Such $X_1, X_2, \ldots, X_n$ are also called independent and identically distributed (or i. i. d.) random variables.

• A statistic is any function of the observations in a random sample.

• The probability distribution of a statistic is called a sampling distribution.
Point Estimation

• A sample was collected: $X_1, X_2, \ldots, X_n$
• We suspect that sample was drawn from a random variable distribution $f(x)$
• $f(x)$ has $k$ parameters that we do not know
• Point estimates are estimates of the parameters of the $f(x)$ describing the population based on the sample
  — For exponential PDF: $f(x) = \lambda \exp(-\lambda x)$ one wants to estimate $\lambda$
  — For Bernoulli PDF: $p^x (1-p)^{1-x}$ one wants to estimate $p$
  — For normal PDF one wants to estimates both $\mu$ and $\sigma$
• Point estimates are uncertain: therefore we can talk of averages and standard deviations of point estimates
A point estimate of some parameter $\theta$ describing population random variable is a single numerical value $\hat{\theta}$ depending on all values $x_1, x_2, \ldots x_n$ in the sample. The sample statistic (whis a random variable $\Theta$ defined by a function $\hat{\Theta}(X_1, X_2, \ldots X_n)$) is called the point estimator.

- There could be multiple choices for the point estimator of a parameter.
- To estimate the mean of a population, we could choose the:
  - Sample mean
  - Sample median
  - Peak of the histogram
  - $\frac{1}{2}$ of (largest + smallest) observations of the sample.
- We need to develop criteria to compare estimates using statistical properties.
Unbiased Estimators Defined

The point estimator $\hat{\Theta}$ is an unbiased estimator for the parameter $\theta$ if:

$$E(\hat{\Theta}) = \theta \quad (7-5)$$

If the estimator is not unbiased, then the difference:

$$E(\hat{\Theta}) - \theta \quad (7-6)$$

is called the bias of the estimator $\hat{\Theta}$. 
Mean Squared Error

The **mean squared error** of an estimator $\hat{\theta}$ of the parameter $\theta$ is defined as:

$$\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2 \quad (7-7)$$

Can be rewritten as

$$= E[\hat{\theta} - E(\hat{\theta})]^2 + [\theta - E(\hat{\theta})]^2$$

$$= V(\hat{\theta}) + (\text{bias})^2$$
Methods of Point Estimation

• We will cover two popular methodologies to create point estimates of a population parameter.
  – Method of moments
  – Method of maximum likelihood
• Each approach can be used to create estimators with varying degrees of biasedness and relative MSE efficiencies.
Method of moments for point estimation
What are moments?

• A k-th moment of a random variable is the expected value $E(X^k)$
  
  – First moment: $\mu = \int_{\infty}^{+\infty} xf(x) \, dx$
  
  – Second moment: $\mu^2 + \sigma^2 = \int_{\infty}^{+\infty} x^2 f(x) \, dx$

• A population moment relates to the entire population

• A sample moment is calculated like its population moments but for a finite sample
  
  – Sample first moment = sample mean = $\frac{1}{n} \sum_{i=1}^{n} x_i$
  
  – Sample k-th moment $\frac{1}{n} \sum_{i=1}^{n} x_i^k$
Moment Estimators

Let $X_1, X_2, ..., X_n$ be a random sample from either a probability mass function or a probability density function with $m$ unknown parameters $\theta_1, \theta_2, ..., \theta_m$.

The moment estimators $\hat{\Theta}_1, \hat{\Theta}_2, ..., \hat{\Theta}_m$ are found by equating the first $m$ population moments to the first $m$ sample moments and solving the resulting simultaneous equations for the unknown parameters.
Exponential Distribution: Moment Estimator-1\textsuperscript{st} moment

• Suppose that $x_1, x_2, \ldots, x_n$ is a random sample from an exponential distribution $f(x) = \lambda \exp(-\lambda x)$ with parameter $\lambda$.

• There is only one parameter to estimate, so equating population and sample first moments, we have one equation: $E(X) = \bar{x}$.

• $E(X) = 1/\lambda$ thus $\lambda = 1/\bar{x}$ is the 1\textsuperscript{st} moment estimator.
Matlab exercise

• Generate 100,000 exponentially distributed random numbers with $\lambda=3$: $f(x) = \lambda \exp(-\lambda x)$
  – Use `random('Exponential',...)` but read the manual to know how to introduce parameters.

• Get a moment estimate of lambda based on the 1$^{\text{st}}$ moment

• Get a moment estimate of lambda based on the 2$^{\text{nd}}$ moment
  – Second moment of the exponential distribution is $E(X^2) = E(X)^2 + \text{Var}(X) = 1/\lambda^2 + 1/\lambda^2 = 2/\lambda^2$

• Get a moment estimate of lambda based on the 3$^{\text{rd}}$ moment
  – Second moment of the exponential distribution is $E(X^3) = 3!/\lambda^3 = 6/\lambda^3$. Most generally, $E(X^p) = p!/\lambda^p$
Method of Maximum Likelihood for point estimation
Maximum Likelihood Estimators

• Suppose that $X$ is a random variable with probability distribution $f(x, \theta)$, where $\theta$ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size $n$. Then the likelihood function of the sample is the probability to get it in a random variable with PDF $f(x, \theta)$:

$$L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta) \quad (7-9)$$

• Note that the likelihood function is now a function of only the unknown parameter $\theta$. The maximum likelihood estimator (MLE) of $\theta$ is the value of $\theta$ that maximizes the likelihood function $L(\theta)$.

• Usually it is easier to work with logarithms: $l(\theta) = \ln L(\theta)$
Exponential MLE:

\[ f(x_i) = \lambda e^{-\lambda x_i} \]

\[ \mathcal{L}(\lambda) = P(x_1, x_2, \ldots, x_n | \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \]

\[ = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \]

\[ \ln \mathcal{L}(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i \]

\[ \frac{d \ln \mathcal{L}(\lambda)}{d \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \]

\[ \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{X}} \]

Same as 1st moment estimator
Example 7-11: Exponential MLE

Let $X$ be a exponential random variable with parameter $\lambda$. The likelihood function of a random sample of size $n$ is:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\ln L(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{d \ln L(\lambda)}{d \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$$\hat{\lambda} = n / \sum_{i=1}^{n} x_i = 1 / \bar{X} \quad \text{(same as moment estimator)}$$
Bernoulli MLE

\[
    f(x_i, p) = p^{x_i} (1-p)^{1-x_i}
\]

\[
    L(p) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} = \prod_{i=1}^{n} p^{x_i} (1-p)^{h-x_i}
\]

\[
    = p^{\sum x_i} (1-p)^{h-\sum x_i}
\]

\[
    \ln L(p) = \left( \sum x_i \right) \ln p + \left( h - \sum x_i \right) \ln (1-p)
\]

\[
    \frac{d \ln L(p)}{dp} = \frac{\sum x_i}{p} - \frac{h - \sum x_i}{1-p} = 0 \Rightarrow \hat{p} = \frac{\sum x_i}{n}
\]

\[
    0 = \frac{(1-\hat{p})\sum x_i - \hat{p}(h-\sum x_i)}{\hat{p}(1-\hat{p})}
\]

\[
    \hat{p} = \frac{\sum x_i}{n}
\]
Example 7-9: Bernoulli MLE

Let $X$ be a Bernoulli random variable. The probability mass function is $f(x;p) = p^x(1-p)^{1-x}$, $x = 0, 1$ where $P$ is the parameter to be estimated. The likelihood function of a random sample of size $n$ is:

$$L(p) = p^{x_1}(1-p)^{1-x_1} \cdot p^{x_2}(1-p)^{1-x_2} \cdots \cdot p^{x_n}(1-p)^{1-x_n}$$

$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^\sum_{i=1}^{n} x_i(1-p)^{n-\sum_{i=1}^{n} x_i}$$

$$\ln L(p) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln (1-p)$$

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{(1-p)} = 0$$

$$\boxed{p = \frac{\sum_{i=1}^{n} x_i}{n}}$$
Normal MLE for $\mu$

\[
f(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)
\]

\[
L(\mu, \sigma) = \left(\frac{1}{\sqrt{2\pi} \sigma}\right)^n \exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right)
\]

\[
\ln L(\mu, \sigma) = -n \ln(\sqrt{2\pi}) - \frac{1}{2\sigma^2} \sum(x_i - \mu)^2
\]

\[
\frac{d \ln L(\mu, \sigma)}{d\mu} = \frac{1}{\sigma^2} \sum(x_i - \mu) = 0 \quad \text{at} \quad \hat{\mu} = \frac{\sum x_i}{n}
\]
Example 7-10: Normal MLE for $\mu$

Let $X$ be a normal random variable with unknown mean $\mu$ and known variance $\sigma^2$. The likelihood function of a random sample of size $n$ is:

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2}$$

$$\ln L(\mu) = \frac{-n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\frac{d \ln L(\mu)}{d \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\sum_{i=1}^{n} x_i\quad \mu = \frac{i=1}{n} = \bar{X} \quad \text{(same as moment estimator)}$$
MLE for Poisson distribution

\[ f(x_1, \ldots, x_n | \lambda) = \frac{e^{-\lambda} \lambda^{x_1}}{x_1!} \cdots \frac{e^{-\lambda} \lambda^{x_n}}{x_n!} \]

\[ = \frac{e^{-n\lambda} \lambda^{\Sigma_1^n x_i}}{x_1! \cdots x_n!} \]

\[ \log f(x_1, \ldots, x_n | \lambda) = -n\lambda + \sum_{i=1}^{n} x_i \log \lambda - \log c \]

where \( c = \prod_{i=1}^{n} x_i! \) does not depend on \( \lambda \), and

\[ \frac{d}{d\lambda} \log f(x_1, \ldots, x_n | \lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} \]

By equating to zero, we obtain that the maximum likelihood estimate \( \hat{\lambda} \) equals

\[ \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} \]
Matlab exercise

• Generate 100,000 exponentially distributed random numbers with $\lambda=3$: $f(x) = \lambda \exp(-\lambda x)$
  – Use `random('Exponential',...)` but read the manual to know how to introduce parameters.

• Get a moment estimate of lambda based on the 1$^{st}$ moment

• Get a moment estimate of lambda based on the 2$^{nd}$ moment
  – Second moment of the exponential distribution is $E(X^2) = E(X)^2 + \text{Var}(X) = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2}$
Prevalence (with 95% CI bars) of obesity among New York City public elementary schoolchildren, by sex and race/ethnicity, 2003.

(source: CDC.GOV)

What do those bars actually mean?
Patterns of somatic mutation in human cancer genomes

The numbers of passenger and driver mutations present can be estimated from these results (see Supplementary Methods). Of the 921 base substitutions in the primary screen, 763 (95% confidence interval, 675–858) are estimated to be passenger mutations. Therefore, the large majority of mutations found through sequencing cancer genomes are not implicated in cancer development, even when the search has been targeted to the coding regions of a gene family of high candidature. However, there are an estimated 158 driver mutations (95% confidence interval, 63–246), accounting for the observed positive selection pressure. These are estimated to be distributed in 119 genes (95% confidence interval, 52–149). The number of samples containing a driver mutation is estimated to be 66 (95% confidence interval, 36–77). The results, therefore, provide statistical evidence for a large set of mutated protein kinase genes implicated in the development of about one-third of the cancers studied.
- We have talked about how a parameter can be estimated from sample data. However, it is important to understand how good is the estimate obtained.

- Bounds that represent an interval of plausible values for a parameter are an example of an interval estimate.