Some Problems of Statistical Estimation

Samvel B. GASPARYAN

Yerevan State University

GasparyanSB@gmail.com

AUA, 09 September 2016

Where we use Statistical Estimation?

- Suppose that we have electrons moving in a wire and we want to measure (in amperes) the flow of the electric charge (the electric current, հոսանքի ուժը). Denote it by θ.

Where we use Statistical Estimation?

- Suppose that we have electrons moving in a wire and we want to measure (in amperes) the flow of the electric charge (the electric current, հոսանքի ուժը). Denote it by θ.
- The first measurement gave as the value 5.4 amperes.

Where we use Statistical Estimation?

- Suppose that we have electrons moving in a wire and we want to measure (in amperes) the flow of the electric charge (the electric current, հոսանքի ուժը). Denote it by θ.
- The first measurement gave as the value 5.4 amperes.
- We know that measurements contain errors with 3 characteristics
  1. Errors are small.
  2. Errors are random - we cannot in advance characterize them.
  3. There are no systematic errors - our measurements can be greater as well as smaller than the true value θ.

Where we use Statistical Estimation? (cont.)

- The measurement gave as an approximate value of the electric current $\theta \approx 5.4$

Where we use Statistical Estimation? (cont.)

- The measurement gave as an approximate value of the electric current $\theta \approx 5.4$
- What can be done to have a more precise value of θ?

Where we use Statistical Estimation? (cont.)

- The measurement gave as an approximate value of the electric current $\theta \approx 5.4$
- What can be done to have a more precise value of θ?
- Of course, we have to do more measurements.

What is Statistical Estimation?

- Suppose that after 5 measurements values are
  5.4 5.32 5.68 5.26 5.1

What is Statistical Estimation?

- Suppose that after 5 measurements values are
  5.4 5.32 5.68 5.26 5.1
- The central question of the statistical estimation is:

What is Statistical Estimation?

- Suppose that after 5 measurements values are
  5.4 5.32 5.68 5.26 5.1
- The central question of the statistical estimation is:
- What we have to do with the data to obtain a more precise estimate for θ?

A first look at data

```r
x=c(5.4,5.32,5.68,5.26,5.1)
y=c(x[1],min(x),max(x),median(x),mean(x))
names(y)=c("First","Min","Max","Median","Mean")
sort(x)
```

```
## [1] 5.10 5.26 5.32 5.40 5.68
```

```r
y
```

```
##   First     Min     Max  Median    Mean
##   5.400   5.100   5.680   5.320   5.352
```

Simualtions

- To do Statistical Estimation we need data, which we obtain by simulations (Կեղծակերպություն).

Simualtions

- To do Statistical Estimation we need data, which we obtain by simulations (Կեղծակերպություն).
- We use simulations to illustrate theoretical results.

Simualtions

- To do Statistical Estimation we need data, which we obtain by simulations (Կեղծակերպություն).
- We use simulations to illustrate theoretical results.
- We simulate that we do not know the true value (in fact we do know) and check whether our estimates are really close to the true value or not.

Model 1 - Normal distribution

- Suppose we have n observations from a normal distribution $\mathcal{N}(\theta, 1)$

$$X_n = (X_1, \cdots, X_n)$$

Model 1 - Normal distribution

- Suppose we have n observations from a normal distribution $\mathcal{N}(\theta, 1)$

$$X_n = (X_1, \cdots, X_n)$$

- Using the date we have to construct a good estimator (one which approaches to the true value as the number of observations become bigger).

Model 1 - Normal distribution

- Suppose we have n observations from a normal distribution $\mathcal{N}(\theta, 1)$

$$X_n = (X_1, \cdots, X_n)$$

- Using the date we have to construct a good estimator (one which approaches to the true value as the number of observations become bigger).
- We will be interested also in construction of the best possible estimator.

- The obvious choice (because of the law of large numbers) is the average of the data

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Model 1 - Normal distribution (part 2)

- The obvious choice (because of the law of large numbers) is the average of the data

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- But we can also construct another estimator

$$\bar{\theta}_n = \ln\left(\frac{1}{n}\sum_{i=1} e^{X_i}\right) - \frac{1}{2}.$$

- The obvious choice (because of the law of large numbers) is the average of the data

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

- But we can also construct another estimator

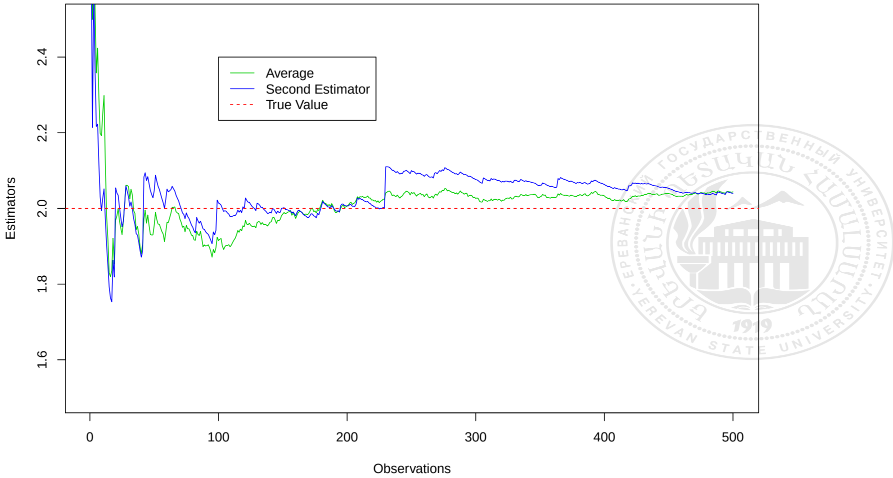
$$\bar{\theta}_n = \ln\left(\frac{1}{n}\sum_{i=1} e^{X_i}\right) - \frac{1}{2}.$$

- We will see that both of them are good estimators. We'll figure out which one is better and how to find the best one.

Model 1 - The First Properties of Estimators

```
theta=2
n=500
X=rnorm(n,mean=theta,sd=1)
th1=cumsum(X)/(1:n)
th=cumsum(exp(X))/(1:n)
th2=log(th)-0.5
plot(1:n,th1,'l',col=3,ylim=c(1.5,2.5),
     xlab="Observations",ylab=c("Estimators"))
lines(1:n,th2,col=4)
abline(h=theta,lty=2,col=2)
legend(100,2.4,c("Average","Second Estimator",
      "True Value"),lty=c(1,1,2),col=c(3,4,2))
```

Model 1 - The First Properties of Estimators (part 2)



Model 1 - Comparison of Estimators

- To compare estimators first of all we look at their convergence rate.

Model 1 - Comparison of Estimators

- To compare estimators first of all we look at their convergence rate.
- For both estimators the convergence rate is $\sqrt{n}$

$$\sqrt{n}(\hat{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, \sigma_1^2), \quad \sqrt{n}(\bar{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, \sigma_2^2).$$

- To compare estimators first of all we look at their convergence rate.
- For both estimators the convergence rate is $\sqrt{n}$

$$\sqrt{n}(\hat{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, \sigma_1^2), \quad \sqrt{n}(\bar{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, \sigma_2^2).$$

- Next, we have to look at the asymptotic variances σ_1^2, σ_2^2 (smaller better).

Model 1 - Comparison of Estimators

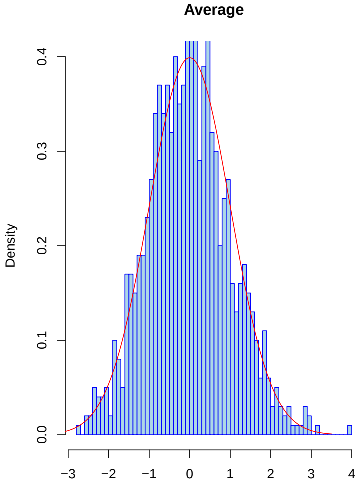
- To compare estimators first of all we look at their convergence rate.
- For both estimators the convergence rate is $\sqrt{n}$

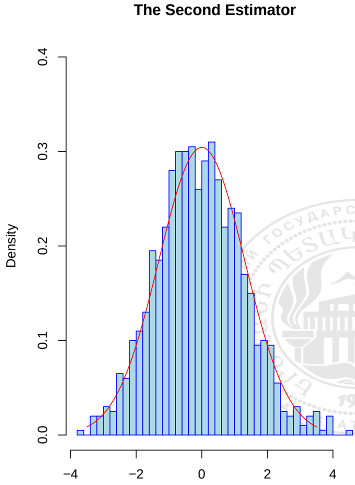
$$\sqrt{n}(\hat{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, \sigma_1^2), \quad \sqrt{n}(\bar{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, \sigma_2^2).$$

- Next, we have to look at the asymptotic variances σ_1^2, σ_2^2 (smaller better).
- Here $\sigma_1^2 = 1$, $\sigma_2^2 = e - 1$, that is

$$\sigma_1^2 < \sigma_2^2.$$

```r
theta=2;m=1000;n=10000
th1=numeric();th2=numeric()
for(i in 1:m){
X=rnorm(n,mean=theta,sd=1)
th1[i]=mean(X)
th=mean(exp(X))
th2[i]=log(th)-0.5
}
int=seq(-3.5,3.5,0.001);s=exp(1)-1
par(mfrow=c(1,2))
hist(sqrt(n)*(th1-theta),nclass=50,freq=FALSE,
     main="Average",border="blue",col="lightblue",
     xlab="",ylim=c(0,0.4))
lines(int,dnorm(int,mean=0,sd=1),col=2)
hist(sqrt(n)*(th2-theta),nclass=50,freq=FALSE,
     main="The Second Estimator",border="blue",
     col="lightblue",xlab="",ylim=c(0,0.4))
lines(int,dnorm(int,mean=0,sd=sqrt(s)),col=2)
```


Average
Density
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3
4


The Second Estimator
Density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Asymptotically the best estimator

- To find asymptotically the best estimator we have to find an estimator with the highest rate of convergence and the smallest asymptotic variance.

- To find asymptotically the best estimator we have to find an estimator with the highest rate of convergence and the smallest asymptotic variance.
- In the Model 1 the highest rate of convergence is $\sqrt{n}$ and the smallest asymptotic variance is 1.

Asymptotically the best estimator

- To find asymptotically the best estimator we have to find an estimator with the highest rate of convergence and the smallest asymptotic variance.
- In the Model 1 the highest rate of convergence is $\sqrt{n}$ and the smallest asymptotic variance is 1.
- Therefore, in the Model 1 the asymptotically best estimator is the average.

Model 2 - Uniform distribution

- Suppose we have n observations from a uniform distribution $\mathbb{U}(0, \theta)$

$$X_n = (X_1, \cdots, X_n).$$

- Suppose we have n observations from a uniform distribution $\mathbb{U}(0,\theta)$

$$X_n = (X_1, \cdots, X_n).$$

- Here too the average (corrected by the factor 2) will be a good estimator for θ

$$\hat{\theta}_n = \frac{2}{n}\sum_{i=1}^{n} X_i.$$

Model 2 - Uniform distribution

- Suppose we have n observations from a uniform distribution $\mathbb{U}(0, \theta)$

$$X_n = (X_1, \cdots, X_n).$$

- Here too the average (corrected by the factor 2) will be a good estimator for θ

$$\hat{\theta}_n = \frac{2}{n} \sum_{i=1}^{n} X_i.$$

- As the second estimator choose the maximal value of observations

$$\bar{\theta}_n = \max_{1 \leq i \leq n} X_i = X_{(n)}.$$

- To compare these two estimators we need to find their rates of convergences.

Model 2 - Uniform distribution (cont.)

- To compare these two estimators we need to find their rates of convergences.
- The following convergences hold

$$\sqrt{n}(\hat{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, \theta^2/3), \quad n(\theta - \bar{\theta}_n) \Longrightarrow \mathbb{E}\left(\frac{1}{\theta}\right),$$

where $\mathbb{E}$ denotes the exponential distribution.

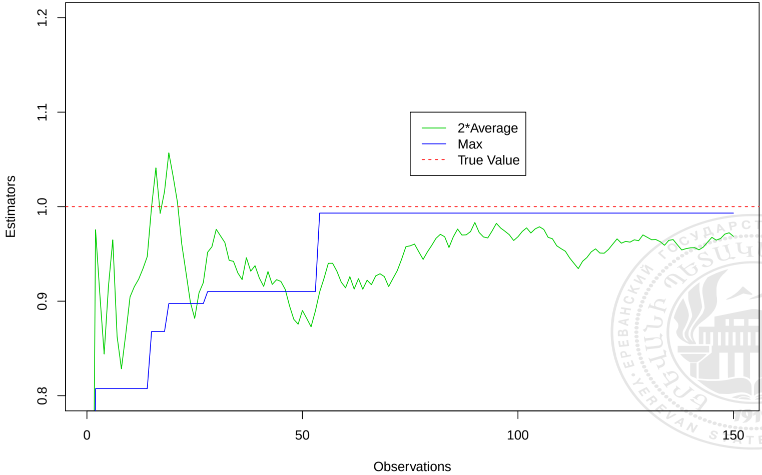
Model 2 - Uniform distribution (cont.)

- To compare these two estimators we need to find their rates of convergences.
- The following convergences hold

$$\sqrt{n}(\hat{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0, \theta^2/3), \quad n(\theta - \bar{\theta}_n) \Longrightarrow \mathbb{E}\left(\frac{1}{\theta}\right),$$

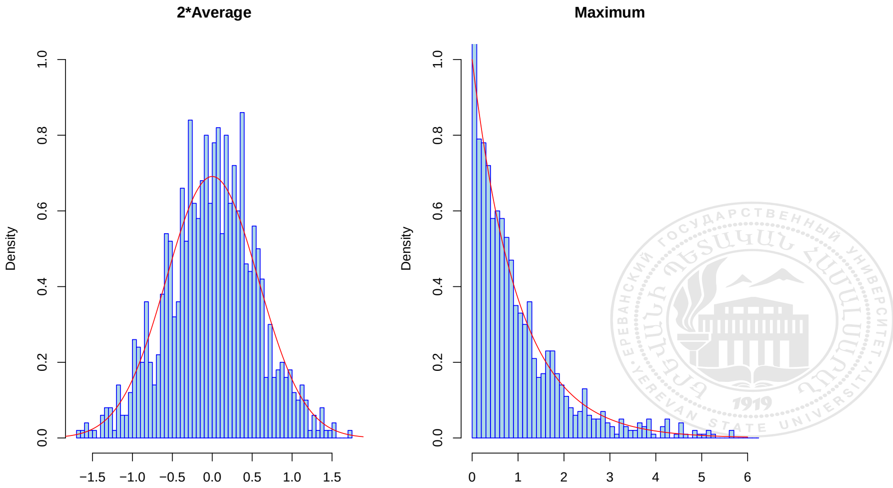
  where $\mathbb{E}$ denotes the exponential distribution.
- This means that the rate of convergence for the average is $\sqrt{n}$, but for the maximum the rate of convergence is n, hence the latter is better.

```r
set.seed(3)
theta=1
n=150
X=runif(n,0,theta)
th1=2*cumsum(X)/(1:n)
th2=numeric()
for(i in 1:n){
th2[i]=max(X[1:i])
}
plot(1:n,th1,'l',col=3,ylim=c(0.8,1.2),
xlab="Observations",ylab=c("Estimators"))
lines(1:n,th2,col=4)
abline(h=theta,lty=2,col=2)
legend(75,1.1,c("2*Average","Max",
"True Value"),lty=c(1,1,2),col=c(3,4,2))
data=c(max(X),2*mean(X),theta)
names(data)=c("Max","2*Aver","True")
data
```




```
##        Max    2*Aver      True
## 0.9932220 0.9685582 1.0000000
```

```
theta=1;m=1000;n=10000;s=theta^2/3
th1=numeric();th2=numeric()
for(i in 1:m){
X=runif(n,0,theta)
th1[i]=2*mean(X)
th2[i]=max(X)
}
int=seq(-3.5,3.5,0.001)
int2=seq(0,6,0.001)
par(mfrow=c(1,2))
hist(sqrt(n)*(th1-theta),nclass=50,freq=FALSE,
     main="2*Average",border="blue",col="lightblue",
     xlab="",ylim=c(0,1))
lines(int,dnorm(int,mean=0,sd=sqrt(s)),col=2)
hist(n*(theta-th2),nclass=50,freq=FALSE,main="Maximum",
     border="blue",col="lightblue",xlab="",
     ylim=c(0,1),xlim=c(0,6))
lines(int2,dexp(int2,1/theta),col=2)
```


2*Average
Density
0.0
0.2
0.4
0.6
0.8
1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Maximum
Density
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6

Model 3 - Truncated Exponential Distribution

- Suppose we have n observations from $X_n = (X_1, \cdots, X_n)$ form the density

$$f(x,\theta) = e^{-(x-\theta)}\mathbb{1}_{[\theta,+\infty)}(x),\ x \in \mathbf{R}, \theta > 0,$$

or, which is the same, as having the distribution function

$$F(x,\theta) = (1 - e^{-(x-\theta)})\mathbb{1}_{[\theta,+\infty)}(x),\ x \in \mathbf{R}, \theta > 0.$$

Model 3 - Truncated Exponential Distribution

- Suppose we have n observations from $X_n = (X_1, \cdots, X_n)$ form the density

$$f(x,\theta) = e^{-(x-\theta)} \mathbb{1}_{[\theta,+\infty)}(x),\ x \in \mathbf{R}, \theta > 0,$$

or, which is the same, as having the distribution function

$$F(x,\theta) = (1 - e^{-(x-\theta)}) \mathbb{1}_{[\theta,+\infty)}(x),\ x \in \mathbf{R}, \theta > 0.$$

- Choose two estimators as follows

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i - 1, \quad \bar{\theta}_n = \min_{1 \le i \le n} X_i = X_{(1)}.$$

Model 3 - Truncated Exponential Distribution

- Suppose we have n observations from $X_n = (X_1, \cdots, X_n)$ form the density

$$f(x,\theta) = e^{-(x-\theta)} \mathbb{1}_{[\theta,+\infty)}(x),\ x \in \mathbf{R}, \theta > 0,$$

or, which is the same, as having the distribution function

$$F(x,\theta) = (1 - e^{-(x-\theta)}) \mathbb{1}_{[\theta,+\infty)}(x),\ x \in \mathbf{R}, \theta > 0.$$

- Choose two estimators as follows

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i - 1, \quad \bar{\theta}_n = \min_{1 \le i \le n} X_i = X_{(1)}.$$

- The first estimator is asymptotically normal

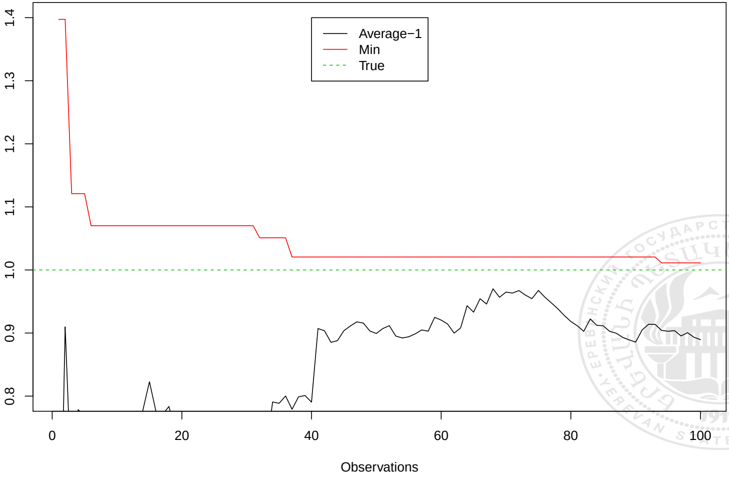
$$\sqrt{n}(\hat{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0,1),$$

Model 3 - Truncated Exponential Distribution

- Suppose we have n observations from $X_n = (X_1, \cdots, X_n)$ form the density
$$f(x,\theta) = e^{-(x-\theta)}\mathbb{1}_{[\theta,+\infty)}(x),\ x \in \mathbf{R}, \theta > 0,$$
or, which is the same, as having the distribution function
$$F(x,\theta) = (1 - e^{-(x-\theta)})\mathbb{1}_{[\theta,+\infty)}(x),\ x \in \mathbf{R}, \theta > 0.$$
- Choose two estimators as follows
$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i - 1, \quad \bar{\theta}_n = \min_{1\le i\le n} X_i = X_{(1)}.$$
- The first estimator is asymptotically normal
$$\sqrt{n}(\hat{\theta}_n - \theta) \Longrightarrow \mathcal{N}(0,1),$$
- for the second estimator we can calculate the non-asymptotic variance of the difference of the estimator and the parameter θ
$$n(\bar{\theta}_n - \theta) \text{ is from } \mathbb{E}(1),\ \forall n \in \mathcal{N}.$$

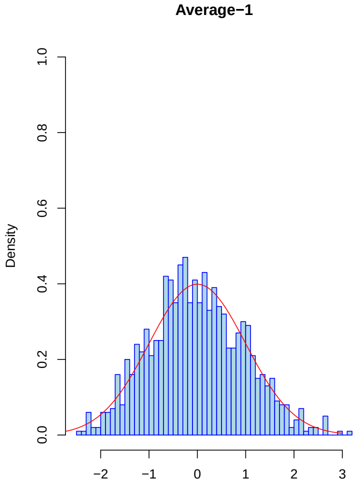
For simulations of a random variable distributed as $F(x,\theta)$ we have to calculate the inverse of the distribution function $F^{-1}(y) = \theta - \log(1-y)$, $y \in [0,1]$, then simulate a r.v. ξ from standard unidorm distribution, then $F^{-1}(\xi)$ will be distributed as $F(x,\theta)$.

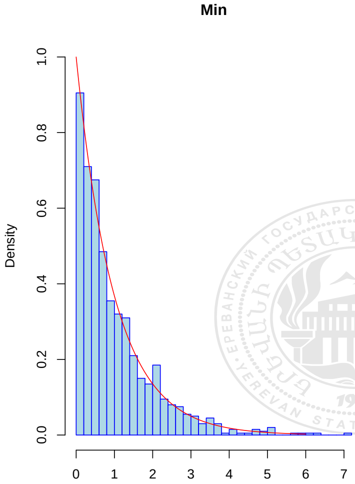
```
set.seed(1000);n=100;th=1;th2=numeric()
rF=function(n)  th-log(1-runif(n,0,1))
X=rF(n)
th1=cumsum(X)/(1:n)-1
for(i in 1:n) th2[i]=min(X[1:i])
plot(1:n,th1,'l',ylim=c(0.8,1.4),
     xlab="Observations",ylab="")
lines(1:n,th2,col=2)
abline(h=th,col=3,lty=2)
legend(40,1.4,c("Average-1","Min","True"),
       col=c(1,2,3),lty=c(1,1,2))
data=c(th1[n],th2[n],th)
names(data)=c("Average-1","Min","True");data
```




```
## Average-1       Min      True
## 0.8895571 1.0112123 1.0000000
```

```
n=10000;m=1000;th=1;th1=numeric();th2=numeric()
rF=function(n)  th-log(1-runif(n,0,1))
for(i in 1:m){
X=rF(n)
th1[i]=mean(X)-1
th2[i]=min(X)
}
int=seq(-3,3,0.001);int2=seq(0,6,0.001)
y=sqrt(n)*(th1-th);z=n*(th2-th)
par(mfrow=c(1,2))
hist(y,freq=FALSE,nclass=50,col="lightblue",
     border="blue",main="Average-1",ylab="",ylim=c(0,1))
lines(int,dnorm(int),col=2)
hist(z,freq=FALSE,nclass=50,col="lightblue",
     border="blue",main="Min",ylab="",ylim=c(0,1))
lines(int2,dexp(int2,1),col=2)
```


Average−1
Density
0.0
0.2
0.4
0.6
0.8
1.0
−2
−1
0
1
2
3


Min
Density
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
7

How theory works for the real Data?

- *Stephen M. Stigler (1977), "Do robust estimators work with real data?", Annals of Statistics, 5, 1055-1098* in his paper took historical data of measurments of the light in the *air*, and compared the performences of 11 estimators.

How theory works for the real Data?

- *Stephen M. Stigler (1977), "Do robust estimators work with real data?", Annals of Statistics, 5, 1055-1098* in his paper took historical data of measurments of the light in the *air*, and compared the performences of 11 estimators.
- As we know, if we have independent measurements, then the best estimator is the mean (if the error is normally distributed).

- *Stephen M. Stigler (1977), "Do robust estimators work with real data?", Annals of Statistics, 5, 1055-1098* in his paper took historical data of measurments of the light in the *air*, and compared the performences of 11 estimators.
- As we know, if we have independent measurements, then the best estimator is the mean (if the error is normally distributed).
- Here we do not have the distribution of the error and have only 100 observations. So, his answer is that the 15%-trimmed mean has the best performence.

```r
#install.packages("HistData")
library(HistData)
```

```
## Warning: package 'HistData' was built under R version 3.
```

```r
data(Michelson)
head(Michelson)
```

```
##   velocity
## 1      850
## 2      740
## 3      900
## 4     1070
## 5      930
## 6      850
```

```r
length(Michelson$velocity)
```

```
## [1] 100
```

Չորս գնահատականներ

1. 10%-կտրված միջինը **(trimmed mean)** սահմանվում է

$$\bar{X}_{.1} = \frac{X_{11} + \cdots + X_{90}}{80}.$$

2. 15%-կտրված միջինն է

$$\bar{X}_{.15} = \frac{X_{16} + \cdots + X_{85}}{70}.$$

3. 25%-կտրված միջինն է

$$\bar{X}_{.25} = \frac{X_{26} + \cdots + X_{75}}{50}.$$

4. Էդջուորթը **(Edgeworth)** առաջին քառորդիչի, կիսորդիչի և երրորդ քառորդիչի կշիռներով միջինն է՝ 5:6:5 հարաբերակցությամբ

$$X_{\mathbb{E}} = \frac{5 * Q1 + 6 * Q2 + 5 * Q3}{16}.$$

Գնահատականների համեմատումը

Ավելացնենք նաև միջինը և կիսորդիչը: Իրական արժեքը 734.5 է:

```
x=Michelson$velocity;x=sort(x)
E=(quantile(x)[2]*5+quantile(x)[3]*6+quantile(x)[3]*5)/16
data=c(sum(x[11:90])/80,sum(x[16:85])/70,sum(x[26:75])/50,
       E,mean(x),median(x))
names(data)=c("10%","15%","25%",
              "E","Mean","Median")
data
```

```
##      10%      15%      25%        E     Mean   Median
## 852.2500 851.4286 849.0000 836.7188 852.4000 850.0000
```