# Law of large numbers

Tail bound

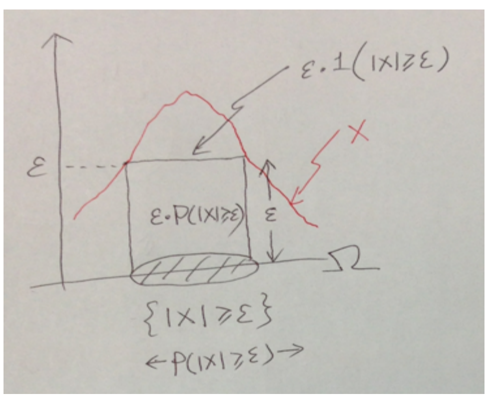
[Markov's inequality] $\qquad \mathbb{P}(|X| \geq \varepsilon) \leq \dfrac{\mathbb{E}|X|}{\varepsilon}$

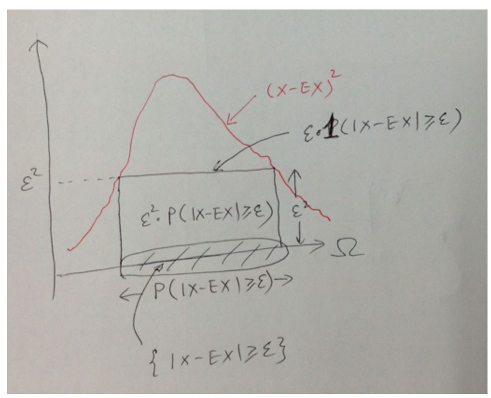
[Chebyshev's inequality] $\qquad \mathbb{P}(|X - EX| \geq \varepsilon) \leq \dfrac{Var(X)}{\varepsilon^2}$

[One-sided Chebyshev's inequality] $\qquad \mathbb{P}(X - \mathbb{E}X \geq \quad \varepsilon) \leq \dfrac{Var(X)}{\varepsilon^2 + Var(X)}$

$$\mathbb{P}(X - \mathbb{E}X \leq -\varepsilon) \leq \frac{Var(X)}{\varepsilon^2 + Var(X)}$$

[Chernoff's bound] $\qquad \mathbb{P}(X \geq \varepsilon) \leq \mathbf{min}_{t>0} \dfrac{\mathbb{E}e^{tX}}{e^{t\varepsilon}}$




For any $b > 0$

$$\begin{aligned} X - \mathbb{E}X \geq \varepsilon \quad &\Leftrightarrow \quad X - \mathbb{E}X + b \geq \varepsilon + b \\ &\Rightarrow \quad (X - \mathbb{E}X + b)^2 \geq (\varepsilon + b)^2 \end{aligned}$$

Hence, by Markov's inequality with $b = \frac{\sigma^2}{\varepsilon}$

$$\mathbb{P}(X - \mathbb{E}X \geq \varepsilon) \leq \frac{\mathbb{E}(X - \mathbb{E}X + b)^2}{(\varepsilon + b)^2} = \frac{\sigma^2 + b^2}{(\varepsilon + b)^2} = \frac{\sigma^2}{\varepsilon^2 + \sigma^2}$$

Example - Tail bound - Part 1

| $B(1000, 0.01)$ | Markov | Chebyshev | one-sided | Chernoff | CLT approx |
|---|---|---|---|---|---|
| $\mathbb{P}(X \geq 20)$ | 0.5 | 0.0990 | 0.0901 | 0.0210 | $7.4094 \times 10^{-4}$ |
| $\mathbb{P}(X \geq 100)$ | 0.1 | 0.0012 | 0.0012 | $1.2204 \times 10^{-61}$ | 0 |

$$\mathbb{E}X = np = 10, \quad Var(X) = npq = 9.9$$

[Markov] $\quad \mathbb{P}(X \geq 20) \leq \frac{\mathbb{E}X}{20} = \frac{10}{20} = 0.5$

[Chebyshev] $\quad \mathbb{P}(X \geq 20) \leq \mathbb{P}(|X - \mathbb{E}X| \geq 10) \leq \frac{Var(X)}{10^2} = \frac{9.9}{10^2} = 0.0990$

[One-sided Chebyshev] $\quad \mathbb{P}(X \geq 20) \leq \frac{Var(X)}{10^2 + Var(X)} = \frac{9.9}{10^2 + 9.9} = 0.0901$

To calculate Chernoff's bound we need the MGF of $X$. Let $X$ be $X = \sum_{i=1}^n X_i$, where $X_i$ be iid $B(p)$. Then, we have the following upper bound of the MGF of $X$;

$$\mathbb{E}e^{tX} = \left(\mathbb{E}e^{tX_1}\right)^n = (e^t \cdot p + 1 \cdot (1-p))^n = (1 + p(e^t - 1))^n \leq (e^{p(e^t-1)})^n = e^{np(e^t-1)}$$

So, with $t$ such that $e^t = 2$

$$\text{[Chernoff]} \quad \mathbb{P}(X \geq 20) \leq \mathbf{min}_{t>0} \frac{\mathbb{E}e^{tX}}{e^{20t}} \leq \mathbf{min}_{t>0} \frac{e^{10(e^t-1)}}{e^{20t}} \leq \frac{e^{10(2-1)}}{e^{20\log 2}} = 0.0210$$

We can also use the CLT approximation.

$$\mathbb{P}(X \geq 20) = \mathbb{P}\left(\frac{X-10}{\sqrt{9.9}} \geq \frac{20-10}{\sqrt{9.9}}\right) \approx N\left(-\frac{20-10}{\sqrt{9.9}}\right) = 7.4094 \times 10^{-4}$$

Example - Tail bound - Part 2

| $Po(100)$ | Markov | Chebyshev | one-sided | Chernoff | CLT approx |
|---|---|---|---|---|---|
| $\mathbb{P}(X \geq 200)$ | 0.5 | 0.0100 | 0.0099 | $1.6728 \times 10^{-17}$ | $7.6199 \times 10^{-24}$ |
| $\mathbb{P}(X \geq 110)$ | 0.9091 | 1 | 0.5 | 0.6162 | 0.1587 |

$$\mathbb{E}X = \lambda = 100, \quad Var(X) = \lambda = 100$$

[Markov] $$\mathbb{P}(X \geq 200) \leq \frac{\mathbb{E}X}{200} = \frac{100}{200} = 0.5$$

[Chebyshev] $$\mathbb{P}(X \geq 200) \leq \mathbb{P}(|X - \mathbb{E}X| \geq 100) \leq \frac{Var(X)}{100^2} = 0.0100$$

[One-sided Chebyshev] $$\mathbb{P}(X \geq 200) \leq \frac{Var(X)}{100^2 + Var(X)} = \frac{9.9}{10^2 + 9.9} = 0.0099$$

To calculate Chernoff's bound we need the MGF of $X$. With $\lambda = 100$

$$\mathbb{E}e^{tX} = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

So, with $t$ such that $e^t = 2$

[Chernoff] $$\mathbb{P}(X \geq 200) \leq \mathbf{min}_{t>0} \frac{\mathbb{E}e^{tX}}{e^{200t}} \leq \mathbf{min}_{t>0} \frac{e^{\lambda(e^t-1)}}{e^{200t}} \leq \frac{e^{100}}{e^{200 \log 2}} = 1.6728 \times 10^{-17}$$

We can also use the CLT approximation.

$$\mathbb{P}(X \geq 200) = \mathbb{P}\left(\frac{X - 100}{\sqrt{100}} \geq 10\right) \approx N(-10) = 7.6199 \times 10^{-24}$$

Weak and strong convergence

$$X_n \approx X$$

Weak convergence

$$X_n \to X \quad \text{in probability} \quad \text{if} \quad \mathbb{P}\left(|X_n - X| > \varepsilon\right) \to 0 \quad \text{for any } \varepsilon > 0$$

Strong convergence

$$X_n \to X \quad \text{a.s. (almost surely)} \quad \text{if} \quad \mathbb{P}\left(X_n \to X \text{ as } n \to \infty\right) = 1$$

Law of large numbers (LLN)

$$N \text{ iid samples from PDF/PMF } f(x) \qquad X_i$$

Weak law of large numbers

$$\text{If } \mathbb{E}|X_i| < \infty, \text{ then} \qquad \int x f(x) dx \approx \frac{1}{N}\sum_{i=1}^{N} X_i \quad \text{in prob}$$

$$\text{If } \mathbb{E}|g(X_i)| < \infty, \text{ then} \qquad \int g(x) f(x) dx \approx \frac{1}{N}\sum_{i=1}^{N} g(X_i) \quad \text{in prob}$$

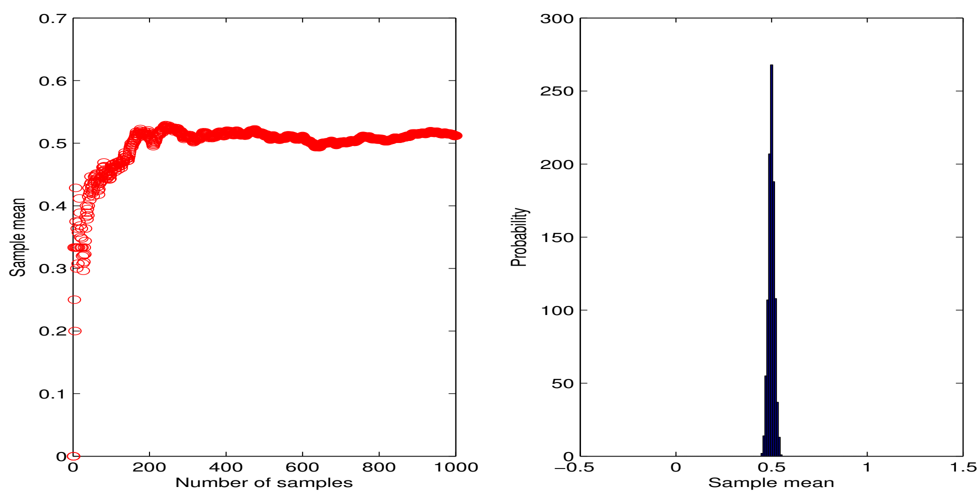
Strong law of large numbers

$$\text{If } \mathbb{E}|X_i| < \infty, \text{ then} \qquad \int x f(x) dx \approx \frac{1}{N}\sum_{i=1}^{N} X_i \quad \text{a.s.}$$

$$\text{If } \mathbb{E}|g(X_i)| < \infty, \text{ then} \qquad \int g(x) f(x) dx \approx \frac{1}{N}\sum_{i=1}^{N} g(X_i) \quad \text{a.s.}$$

Which one is the weak law and which one is the strong law?

In the left we flip a fair coin many times and record the sample mean as a function of the number of flips. In the right we flip a fair coin 1000 times and record the sample mean. We do this 1000 times and draw the histogram of the sample mean of the 1000 flips. Which one is the weak law and which one is the strong law?




```
clear all; close all; clc; rng('default');

% Binomial parameter
n=1; p=0.5;

% Number of simulation
N=1000; M=1000;

subplot(121) % Strong law of large numbers
x=random('Binomial',n*ones(1,N),p*ones(1,N));
Sample_Mean=cumsum(x)./(1:N);
plot(1:N,Sample_Mean,'or');
xlabel('Number of samples'); ylabel('Sample mean');

subplot(122) % Weak law of large numbers
s=random('Binomial',N*ones(1,M),p*ones(1,M));
Sample=s/N;
hist(Sample,0:0.01:1)
xlabel('Sample mean'); ylabel('Probability');
```

## Proof of LLN

### Proof of weak law of large numbers (with finite 2nd moment)

$$
\begin{array}{ll}
\text{Mean} & \mathbb{E}\left(\frac{S_n}{n}\right) = \mu \\
\text{Variance} & Var\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n} \\
\text{Chebyshev} & \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{Var(\frac{S_n}{n})}{\varepsilon^2} = \frac{\sigma^2/n}{\varepsilon^2} = Cn^{-1} \to 0
\end{array}
$$

### Proof of strong law of large numbers (with finite 4th moment)

$$
\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{E}\left(\frac{S_n - n\mu}{n}\right)^4 \leq C \sum_{n=1}^{\infty} n^{-2} < \infty \quad & \Rightarrow \quad \sum_{n=1}^{\infty} \left(\frac{S_n - n\mu}{n}\right)^4 < \infty \quad \text{a.s.} \\
& \Rightarrow \quad \left(\frac{S_n - n\mu}{n}\right)^4 \to 0 \quad \text{a.s.} \\
& \Rightarrow \quad \frac{S_n}{n} \to \mu \quad \text{a.s.}
\end{aligned}
$$

**Example - Coupon collector problem**

Let $T_n$ be time to collect all $n$ different coupons. Then

$$\frac{T_n}{n\log n} \to 1 \quad \text{in probability}$$

Let $\tau_i$ be the minimum number of the happy meals that I have to eat to collect the $i$-th new toy after I get the $(i-1)$-th new toy. Then,

$$\begin{aligned} &(1) \qquad T_n = \sum_{i=1}^{n} \tau_i \\ &(2) \qquad \tau_i \text{ is } Geo(\tfrac{N-(i-1)}{N}) \\ &(3) \qquad \tau_i \text{ are independent} \end{aligned}$$

**[Step 1]** Compute the mean.

$$\mathbb{E}T_n = \sum_{i=1}^{n} \mathbb{E}\tau_i = n\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) \sim n\log n$$

**[Step 2]** Compute the variance.

$$Var(T_n) = \sum_{i=1}^{n} Var(\tau_i) \le Cn^2$$

**[Step 3]** Apply Chebyshev's inequality.

$$\mathbb{P}\left(\left|\frac{T_n - ET_n}{a_n}\right| > \varepsilon\right) \le \frac{Var(T_n)}{\varepsilon^2 a_n^2} \to 0$$

where we choose $a_n$ as $a_n = n\log n$ so that

$$\begin{aligned} &(A) \qquad \frac{Var(T_n)}{a_n^2} \to 0 \\ &(B) \qquad \frac{ET_n}{a_n} \to 1 \end{aligned}$$

CLT and LLN

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \approx N(0,1)$$

If we know $\sigma$

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \begin{cases} \stackrel{d}{=} & N(0,1) \quad \text{if } X_i \text{ are iid } N(\mu,\sigma^2) \quad \text{by property of normal} \\ \approx & N(0,1) \quad \text{if } X_i \text{ are iid with } \mathbb{E}X_i^2 < \infty \quad \text{by CLT} \end{cases}$$

If we don't know $\sigma$

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \begin{cases} \stackrel{d}{=} & t_{n-1} \quad \text{if } X_i \text{ are iid } N(\mu,\sigma^2) \quad \text{by property of normal} \\ \approx & N(0,1) \quad \text{if } X_i \text{ are iid with } \mathbb{E}X_i^2 < \infty \quad \text{by CLT and LLN} \end{cases}$$

**Monte Carlo to estimate $\pi$**

Draw $n$ random points $X_i$ from $[-1,1]^2$ and record $R_i$ whether the point is inside of the unit circle.

$$R_i = \begin{cases} 1 & \text{if } X_i \text{ is inside of the unit circle} \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$R_i \text{ is iid } B(p),\ p = \frac{\pi}{4}.$$

Therefore, by the weak or strong law of large numbers we have for large $n$

$$\frac{\sum_{i=1}^n R_i}{n} \approx \frac{\pi}{4} \qquad \Rightarrow \qquad \pi \approx \frac{4\sum_{i=1}^n R_i}{n}$$

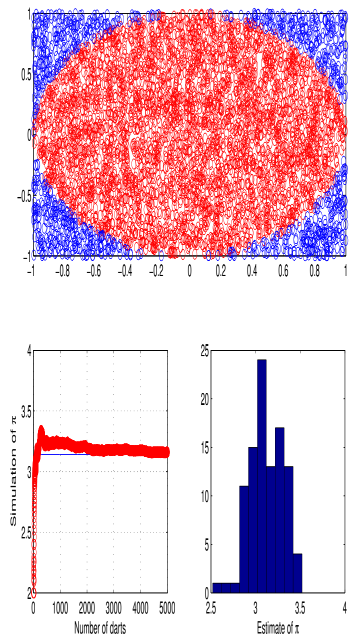



Figure 1: Monte Carlo simulation to estimate $\pi$. On the top we draw 5000 random darts to $[-1,1]^2$. We color red on the darts inside the unit circle and blue on the darts out side the unit circle. On the bottom left for each $i$ we estimate $\pi$ using the first $i$ random draw. As we note, the estimate are getting better as we have more samples. On the bottom right we draw 100 random darts and we estimate $\pi$. We do this 100 times and make a histogram.

```
clear all; close all; clc; rng('default');

n=5000; % Number of darts for each estimate
x=2*rand(2,n)-1; % Uniform random samples from [-1,1]^2
r2=sum(x.^2); % Squre distance from the origin
N_Circle=sum(r2<=1); % Number of random samples inside unit circle
Estimated_pi=4*N_Circle/n % Estimate pi

indicator=zeros(1,n);
indicator(r2<=1)=1;

subplot(2,2,1:2)
plot(x(1,indicator==1),x(2,indicator==1),'or'); hold on
plot(x(1,indicator==0),x(2,indicator==0),'o');

subplot(2,2,3)
plot(1:n,pi,'-',1:n,4*cumsum(indicator)./(1:n),'or'); grid on;
axis([0 n 2 4])
xlabel('Number of darts'); ylabel('Simulation of \pi')

subplot(2,2,4)
n=100; % Number of darts for each estimate
m=100; % Number of estimates computed using n dart
x=2*rand(2,n,m)-1; % Uniform random samples from [-1,1]^2
r2=sum(x.^2); % Squre distance from the origin
N_Circle=sum(r2<=1); % Number of random samples inside unit circle
Estimated_pi=4*N_Circle/n; % Estimate pi
Estimated_pi=Estimated_pi(:);
hist(Estimated_pi); xlabel('Estimate of \pi');
```

### Buffon's needle

On a paper we draw parallel lines 1 units apart. We drop a needle of length 1 onto the paper $n$ times and record $R_i$ whether the needle intersect the line.

$$R_i = \begin{cases} 1 & \text{if the needle intersect the line at the } i\text{-th drop} \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$R_i \text{ is iid } B(p),\ p = \frac{2}{\pi}.$$

Therefore, by the weak or strong law of large numbers we have for large $n$

$$\frac{\sum_{i=1}^n R_i}{n} \approx \frac{2}{\pi} \qquad \Rightarrow \qquad \pi \approx \frac{2n}{\sum_{i=1}^n R_i}$$

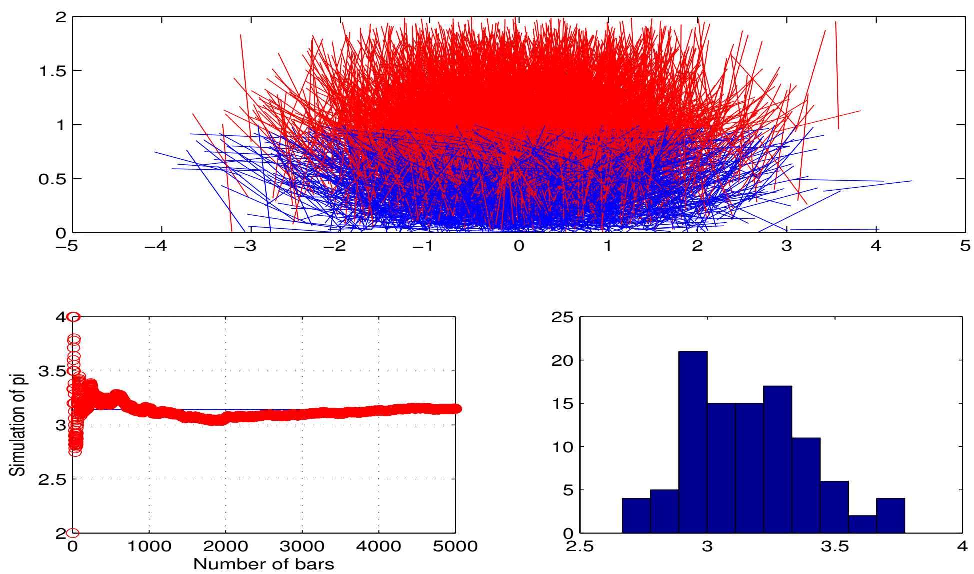



Figure 2: Buffon's needle, a simulation to estimate $\pi$ On the topt we draw 5000 random bars such that the lower end of the bar are between 0 and 1. We color red on the bars that cross $y = 1$ and blue on the bars that don't cross $y = 1$. On the bottom left for each $i$ we estimate $\pi$ using the first $i$ random draw. As we note, the estimate are getting better as we have more samples. On the bottom right we draw 100 random bars and we estimate $\pi$. We do this 100 times and make a histogram.

```
clear all; close all; clc; rng('default');

n=5000; % Number of random samples generated
x=rand(2,n); % First row = Height of lower end; Second row = Angle/pi;
h=x(1,:)+sin(pi*x(2,:)); % Height of higher end
N_Bar=sum(h>=1); % Number of random samples hit the upper bar at y=1
Estimated_pi=2*n/N_Bar % Estimate pi

indicator=zeros(1,n);
indicator(h>=1)=1;

subplot(2,2,1:2)
for i=1:n

    temp=randn(1,1);
    plot_x=[temp temp+cos(pi*x(2,i))];
    plot_y=[x(1,i) x(1,i)+sin(pi*x(2,i))];

    if (indicator(i)==1),
        plot(plot_x,plot_y,'-r'); hold on
    else
        plot(plot_x,plot_y,'-b'); hold on;
    end

end

subplot(2,2,3)
plot(1:n,pi,'-',1:n,2*(1:n)./cumsum(indicator),'or'); grid on;
axis([0 n 2 4])
xlabel('Number of bars'); ylabel('Simulation of pi')

subplot(2,2,4)
n=100; % Number of darts for each estimate
m=100; % Number of estimates computed using n dart
x=rand(2,n,m); % First row = Height of lower end; Second row = Angle/pi;
h=x(1,:,:)+sin(pi*x(2,:,:)); % Height of higher end
N_Bar=sum(h>=1); % Number of random samples hit the upper bar at y=1
Estimated_pi=2*n/N_Bar; % Estimate pi
Estimated_pi=Estimated_pi(:);
hist(Estimated_pi)
```