Assignment 8 for Statistical Computing and Empirical Methods: Continuous random variables, independence and laws of large numbers

Henry Reeve

Introduction

This document describes your eighth assignment for Statistical Computing and Empirical Methods (Unit EMATM0061) on the MSc in Data Science. Before starting the assignment it is recommended that you first watch video lecture 8 entitled “Continuous random variables, independence and laws of large numbers”.

1 The Gaussian distribution

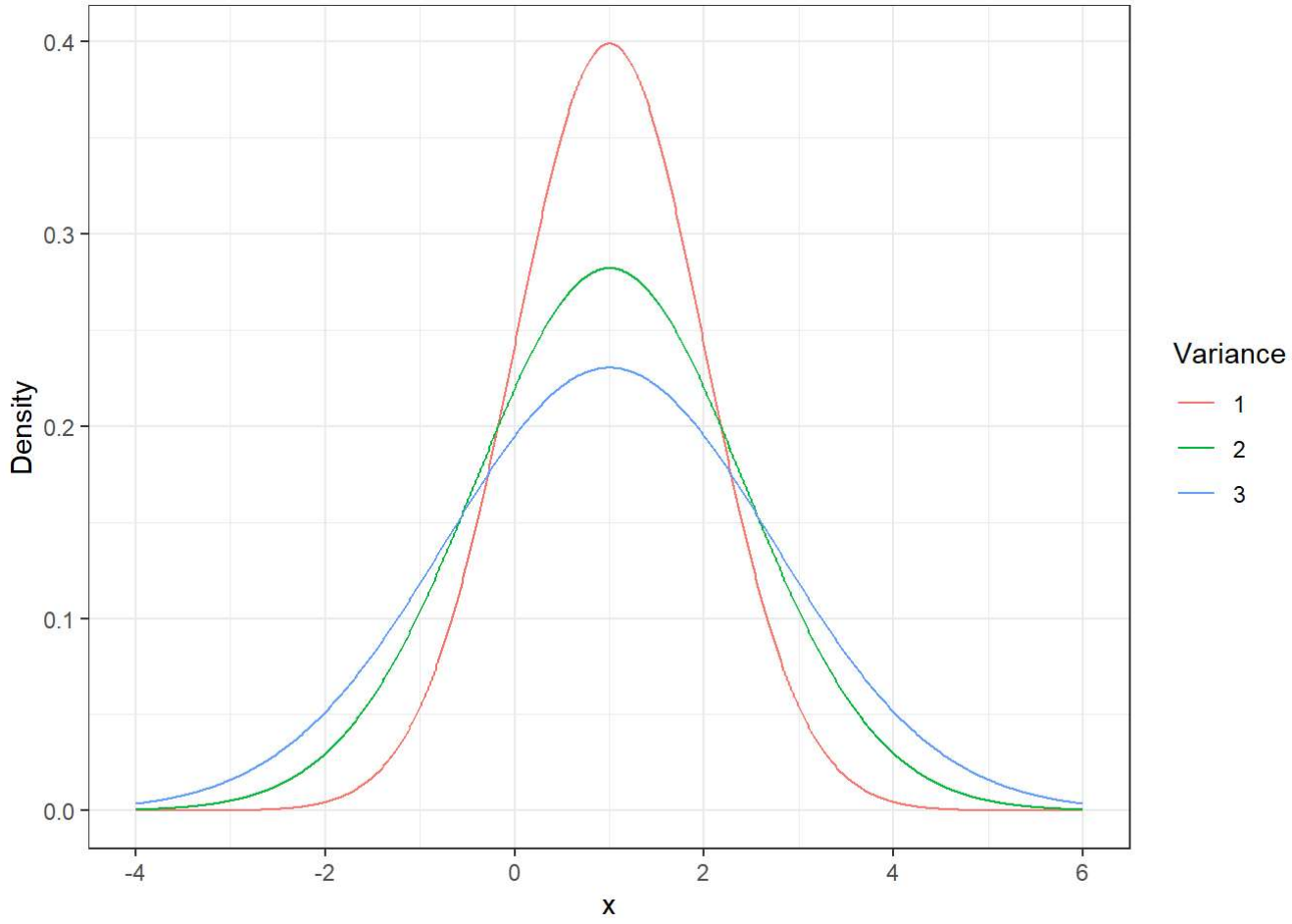
Write out the probability density function of a Gaussian random variable with mean μ and standard deviation $\sigma > 0$.

Use the help function to look up the following four functions: **dnorm()**, **pnorm()**, **qnorm()** and **rnorm()**.

Generate a plot which displays the probability density function for three Gaussian distributions $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_3^2)$ with $\mu_1 = \mu_2 = \mu_3 = 1$ and variances $\sigma_1^2 = 1$, $\sigma_2^2 = 2$ and $\sigma_3^2 = 3$. Your plot should look something like this:

A

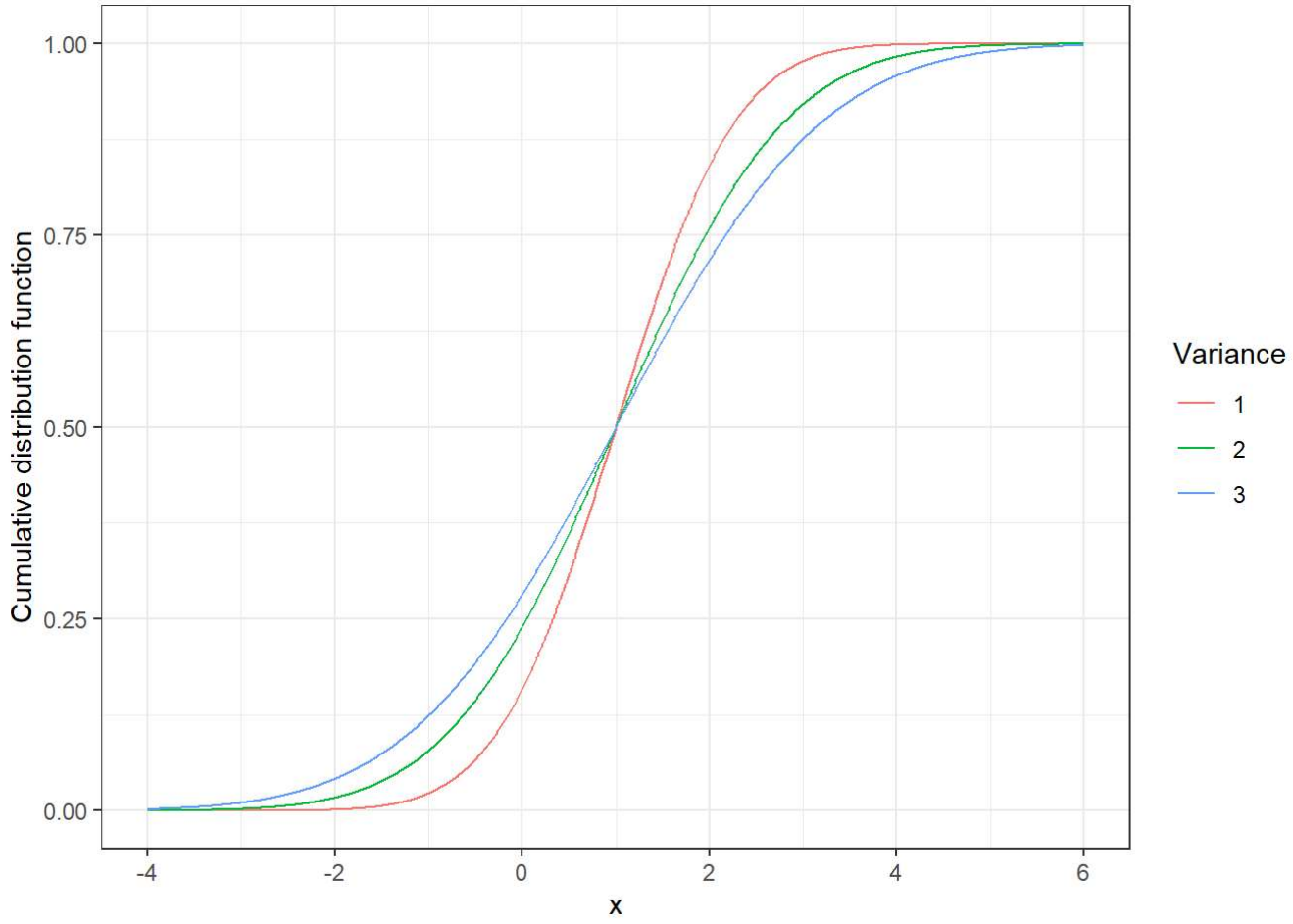
```
x<-seq(-4,6,0.01)

normal_densities_by_x<-data.frame(x=x,density=dnorm(x,mean=1,sd=sqrt(1)),var=1)%>%
  rbind(data.frame(x=x,density=dnorm(x,mean=1,sd=sqrt(2)),var=2))%>%
  rbind(data.frame(x=x,density=dnorm(x,mean=1,sd=sqrt(3)),var=3))


ggplot(normal_densities_by_x,aes(x,y=density,color=as.character(var)))+geom_line()+
  theme_bw()+labs(color="Variance",x="x",y="Density")
```



Generate a similar plot for the cumulative distribution function for three Gaussian distributions $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_3^2)$ with $\mu_1 = \mu_2 = \mu_3 = 1$ and variances $\sigma_1^2 = 1$, $\sigma_2^2 = 2$ and $\sigma_3^2 = 3$.

A

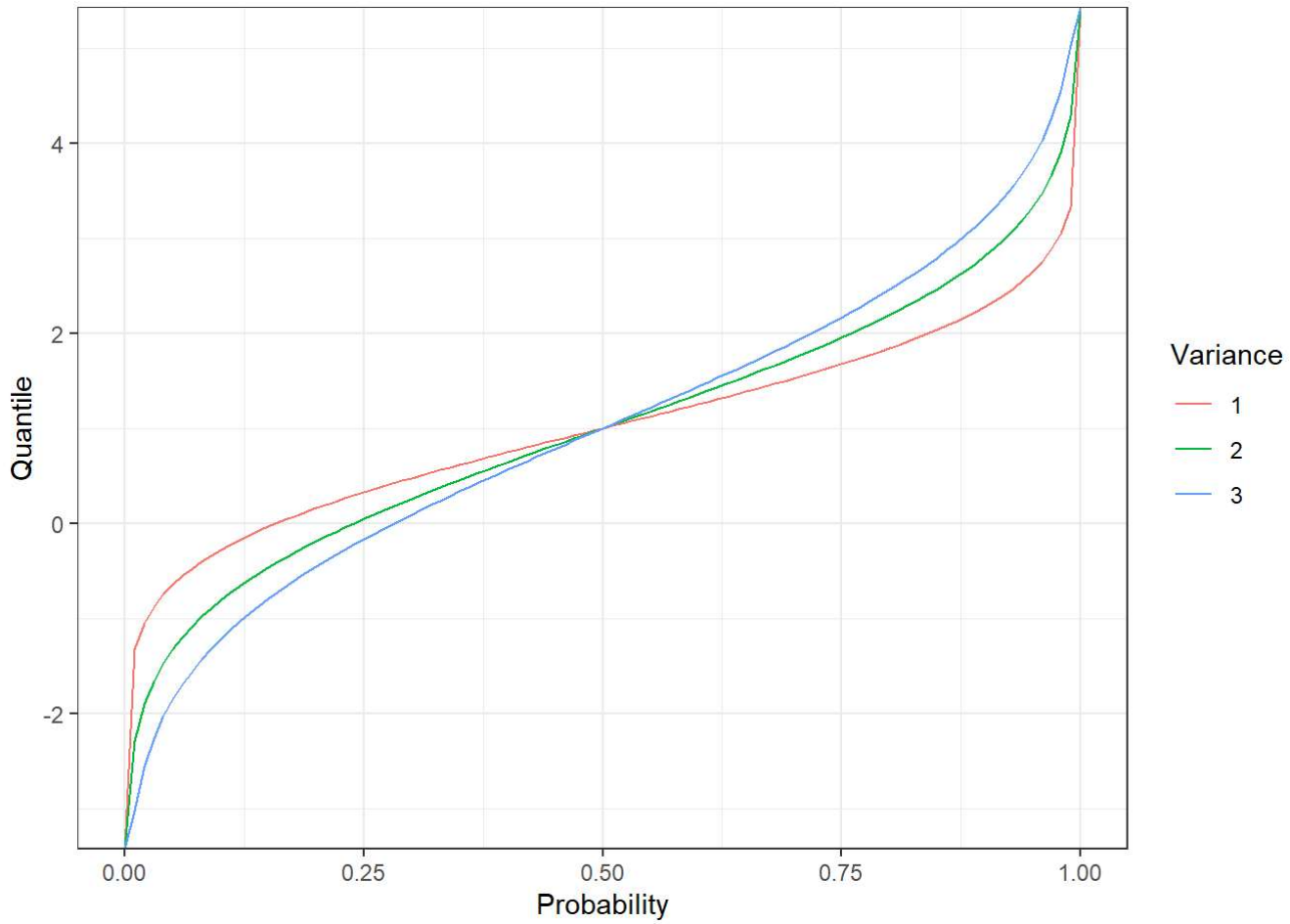
```
normal_cdf_by_x<-data.frame(x=x,cdf=pnorm(x,mean=1,sd=sqrt(1)),var=1)%>%
  rbind(data.frame(x=x,cdf=pnorm(x,mean=1,sd=sqrt(2)),var=2))%>%
  rbind(data.frame(x=x,cdf=pnorm(x,mean=1,sd=sqrt(3)),var=3))


ggplot(normal_cdf_by_x,aes(x,y=cdf,color=as.character(var)))+geom_line()+
  theme_bw()+labs(color="Variance",x="x",y="Cumulative distribution function")
```



Next generate a plot for the quantile function for the same three Gaussian distributions. Describe the relationship between the quantile function and the cumulative distribution function.

A

```
probs=seq(0,1,0.01)

normal_cdf_by_x<-data.frame(p=probs,q=qnorm(probs,mean=1,sd=sqrt(1)),var=1)%>%
  rbind(data.frame(p=probs,q=qnorm(probs,mean=1,sd=sqrt(2)),var=2))%>%
  rbind(data.frame(p=probs,q=qnorm(probs,mean=1,sd=sqrt(3)),var=3))


ggplot(normal_cdf_by_x,aes(x=p,y=q,color=as.character(var)))+geom_line()+
  theme_bw()+labs(y="Quantile",x="Probability",color="Variance")
```



Now use **rnorm()** generate a random independent and identically distributed sequence $Z_1, \cdots, Z_n \sim \mathcal{N}(0, 1)$ so that each $Z_i \sim \mathcal{N}(0, 1)$ has standard Gaussian distribution with $n = 100$. Make sure your code is reproducible by using the **set.seed()** function. Store your random sample in a vector called “standardGaussianSample”.

```
set.seed(123)
standardGaussianSample<-rnorm(100)
```

Without calling the rnorm() function again, use your existing sample stored in “standardGaussianSample” to generate a sample of size n of the form $Y_1, \cdots, Y_n \sim \mathcal{N}(1, 3)$ with mean $\mu = 1$ and variance $\sigma^2 = 3$. Store your second sample in a vector called “mean1Var3GaussianSampleA”.

```
mean1Var3GaussianSampleA<-1+sqrt(3)*standardGaussianSample
```

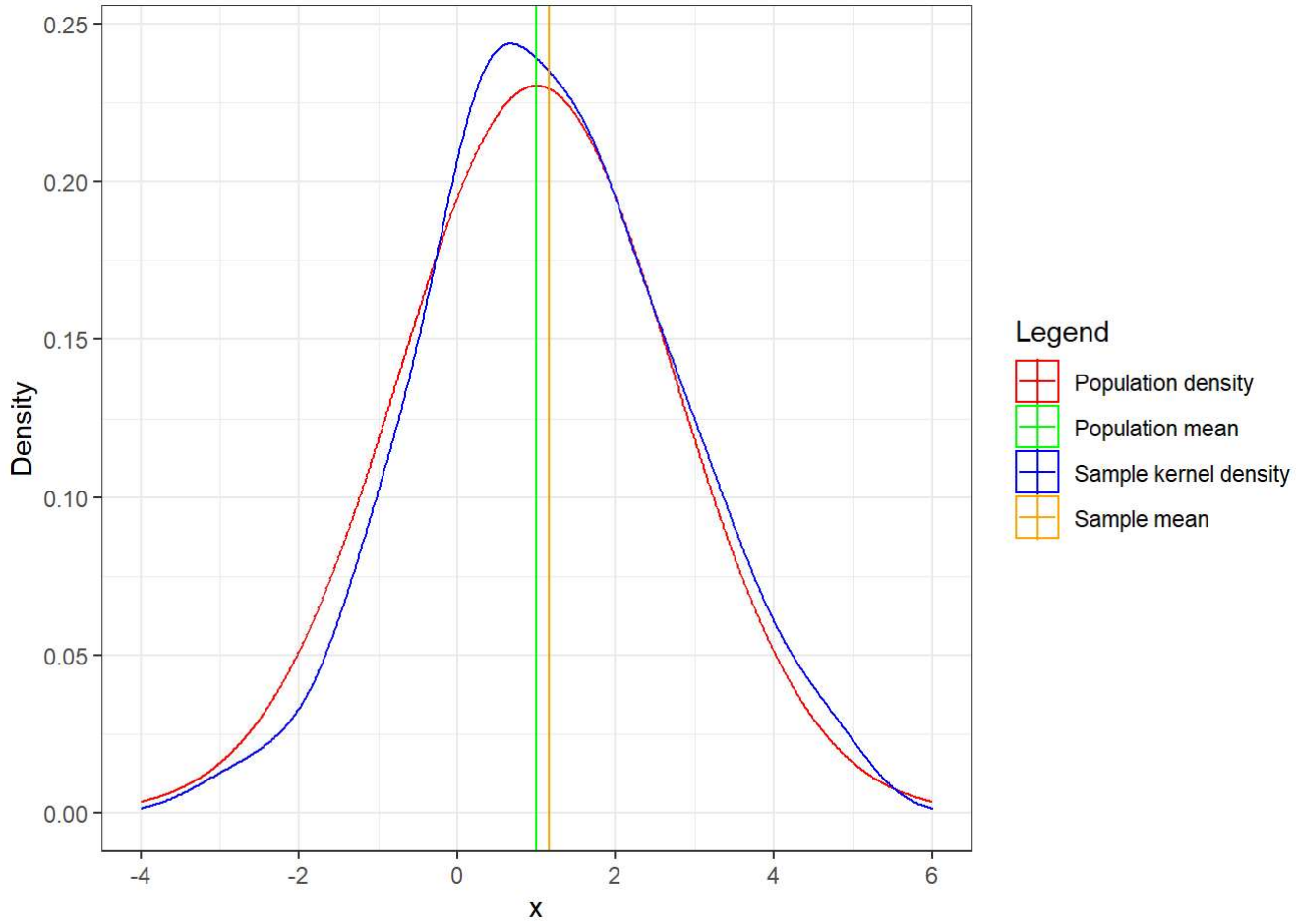
Reset the random seed to the same value as before using the **set.seed()** function and generate an i.i.d. sample of the form $Y_1, \cdots, Y_n \sim \mathcal{N}(1, 3)$ using the **rnorm()** function by setting the mean and standard deviation. Store this sample in a vector called “mean1Var3GaussianSampleB”. Compare the vectors mean1Var3GaussianSampleA and mean1Var3GaussianSampleB.

```
set.seed(123)
mean1Var3GaussianSampleB<-rnorm(100,1,sqrt(3))

all.equal(mean1Var3GaussianSampleA,mean1Var3GaussianSampleB)
```

```
## [1] TRUE
```

Now generate a graph which includes both a kernel density plot for your sample mean1Var3GaussianSampleA and the population density (the probability density function) generated using **dnorm()**. You can also include two vertical lines which display both the population mean and the sample mean. Your plot should something like the following:

```
colors<-c("Population density"="red", "Sample kernel density"="blue", "Population mean"=
"green", "Sample mean"="orange")

g<-ggplot()+labs(x="x",y="Density",color="Legend")+theme_bw()+
  geom_line(data=(normal_densities_by_x%>%
  filter(var==3)),
  aes(x,y=density,color="Population density")) # create plot of theoretical density

g<-g+geom_density(data=data.frame(x=mean1Var3GaussianSampleA),aes(x=x,color="Sample kerne
l density")) # add in kernel density plot from real sample

g<-g+geom_vline(aes(xintercept=1,color="Population mean"))
g<-g+geom_vline(aes(xintercept=mean(mean1Var3GaussianSampleA),color="Sample mean"))
g<-g+scale_color_manual(values=colors)

g
```



2 Bayes theorem

Suppose that there is a rare medical condition and an associated test for that condition. Let X and T be a pair of binary random variables corresponding to a randomly selected member of the population. More precisesly both X and T have outcome space $\{0, 1\}$, with X defined by

$$X = \begin{cases} 1 & \text{if the person has the condition} \\ 0 & \text{if the person doesn't have the condition.} \end{cases}$$

Similarly, T is defined by

$$T = \begin{cases} 1 & \text{if the test is positive} \\ 0 & \text{if the test is negative.} \end{cases}$$

Suppose that the probability that a random person within the population has the condition is $\mathrm{P}(X = 1) = 0.001$. Suppose further that a person who has the condition will have a positive test result with probability $\mathrm{P}(T = 1 \,|\, X = 1) = 0.95$. Similarly, a person who doesn't have the condition will have a negative test result with probability $\mathrm{P}(T = 0 \,|\, X = 0) = 0.95$.

Compute the probability that a person who has a positive test result, actually has the condition. That is, compute $\mathrm{P}(X = 1 \mid T = 1)$.

A

Note that $\mathrm{P}(X = 0) = 1 - \mathrm{P}(X = 1) = 1 - 0.001$ and
$\mathrm{P}(T = 1 \mid X = 0) = 1 - \mathrm{P}(T = 0 \mid X = 0) = 1 - 0.95$. Thus, by Bayes' theorem we have

$$\begin{aligned}\mathrm{P}(X = 1 \mid T = 1) &= \frac{\mathrm{P}(T = 1 \mid X = 1) \cdot \mathrm{P}(X = 1)}{\mathrm{P}(T = 1 \mid X = 1) \cdot \mathrm{P}(X = 1) + \mathrm{P}(T = 1 \mid X = 0) \cdot \mathrm{P}(X = 0)} \\ &= \frac{0.95 \cdot 0.001}{0.95 \cdot 0.001 + (1 - 0.95) \cdot (1 - 0.001)} = 0.01866405.\end{aligned}$$

3 The exponential distribution

Let $\lambda > 0$ be a positive real number. An exponential random variable X with parameter λ is a continuous random variable with density $p_\lambda : \mathrm{R} \to (0, \infty)$ defined by

$$p_\lambda(x) := \begin{cases} 0 & \text{if } x < 0 \\ \lambda e^{-\lambda x} & \text{if } x \geq 0. \end{cases}$$

First prove that p_λ is a well-defined probability density function.

A

First note that $p_\lambda(x) \geq 0$ for all $x \in \mathrm{R}$ and

$$\int_{-\infty}^{\infty} p_\lambda(x) = \int_0^\infty \lambda e^{-\lambda x} = \lambda \cdot \left[-\lambda^{-1} \cdot e^{-\lambda x}\right]_0^\infty = 1.$$

Compute the population mean and variance of an exponential random variable X with parameter λ.

A

Using integration by parts we have,

$$\begin{aligned}\mathrm{E}[X] &= \int_{-\infty}^{\infty} x p_\lambda(x) dx = \int_0^\infty \lambda x e^{-\lambda x} dx \\ &= \left[-x e^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x} dx \\ &= \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^\infty = \frac{1}{\lambda}.\end{aligned}$$

Using integration by parts again we have,

$$\begin{aligned}\mathrm{E}[X^2] &= \int_{-\infty}^{\infty} x^2 p_\lambda(x)dx = \int_0^\infty \lambda x^2 e^{-\lambda x}dx \\ &= \left[-x^2 e^{-\lambda x}\right]_0^\infty + 2\int_0^\infty x e^{-\lambda x}dx \\ &= \frac{2}{\lambda}\cdot\int_0^\infty \lambda x e^{-\lambda x}dx = \frac{2}{\lambda}\cdot\mathrm{E}[X] = \frac{2}{\lambda^2}.\end{aligned}$$

Hence, $\mathrm{Var}(X) = \mathrm{E}[X^2] - E[X]^2 = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2$.

Compute the cumulative distribution function and the quantile function for exponential random variables with parameter λ.

A

The cumulative distribution function is given by

$$F_\lambda(x) = \int_{-\infty}^x p_\lambda(t)dt = \begin{cases} 0 & \text{if } x \le 0 \\ \int_0^x \lambda e^{-\lambda t}dt & \text{if } x > 0. \end{cases}$$

Moreover, we have

$$\int_0^x \lambda e^{-\lambda t}dt = \left[-e^{-\lambda t}\right]_0^x = 1 - e^{-\lambda x}.$$

Thus, the cumulative distribution function is given by

$$F_\lambda(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1 - e^{-\lambda x} & \text{if } x > 0. \end{cases}$$

The quantile function is given by

$$\begin{aligned}F_\lambda^{-1}(p) &:= \inf\left\{x \in \mathrm{R} : F_\lambda(x) \le p\right\} \\ &= \begin{cases} -\infty & \text{if } p = 0 \\ -\frac{1}{\lambda}\ln(1-p) & \text{if } p \in (0, 1]. \end{cases}\end{aligned}$$

4 Transformations of continuous random variables

Suppose that X is a continuous real-valued random variable with density $p_X : \mathrm{R} \to (0, \infty)$. Suppose that $\varphi : \mathrm{R} \to \mathrm{R}$ is a function and define $Y = \varphi(X)$ by

$$\mathrm{P}(\varphi(X) \in A) = \mathrm{P}\Big(X \in \varphi^{-1}(A)\Big),$$

for $A \subseteq \mathrm{R}$. Is $Y = \varphi(X)$ a discrete or continuous random variable?

A

Y can be either discrete, for example if $\varphi \equiv 0$ or continuous, for example if $\varphi(x) = x$ for all $x \in \mathrm{R}$.

Suppose that φ is a strictly increasing differentiable function with derivative $\partial\varphi/\partial x > 0$. Show that $\varphi(X)$ is a continuous random variable and compute its density.

A

Since $\partial\varphi/\partial x > 0$, φ is strictly increasing. Hence, given $a < b$ we have

$$\begin{aligned}\mathrm{P}[a < Y < b] &= \mathrm{P}[a < \varphi(X) < b] \\ &= \mathrm{P}[\varphi^{-1}(a) < X < \varphi^{-1}(b)] \\ &= \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} p_X(x)dx \qquad = \int_a^b p_X(\varphi^{-1}(y))\left(\frac{\partial\varphi}{\partial x}\Big|_y\right)^{-1} dy.\end{aligned}$$

Thus, $Y = \varphi(X)$ has density $y \mapsto p_X(\varphi^{-1}(y))\left(\frac{\partial\varphi}{\partial x}\Big|_y\right)^{-1}$.

What is the probability density function of the random variable $\varphi(X)$ when φ is a strictly decreasing differentiable function with $\partial\varphi/\partial x < 0$?

A

Since $\partial\varphi/\partial x < 0$, φ is strictly decreasing. Hence, given $a < b$ we have

$$\begin{aligned}\mathrm{P}[a < Y < b] &= \mathrm{P}[a < \varphi(X) < b] \\ &= \mathrm{P}[\varphi^{-1}(b) < X < \varphi^{-1}(a)] \\ &= \int_{\varphi^{-1}(b)}^{\varphi^{-1}(a)} p_X(x)dx \qquad = -\int_a^b p_X(\varphi^{-1}(y))\left(\frac{\partial\varphi}{\partial x}\Big|_y\right)^{-1} dy.\end{aligned}$$

Thus, $Y = \varphi(X)$ has density $y \mapsto -p_X(\varphi^{-1}(y))\left(\frac{\partial\varphi}{\partial x}\Big|_y\right)^{-1}$.

Thus, in general if φ is monotone, Y is continuous with density $y \mapsto p_X(\varphi^{-1}(y))\left|\left(\frac{\partial \varphi}{\partial x}\,|_{\,y}\right)^{-1}\right|$

.

Suppose now that $\alpha \in \mathrm{R}\backslash\{0\}$, $\beta \in \mathrm{R}$. Define $\varphi_{\alpha,\beta}\colon \mathrm{R} \to \mathrm{R}$ by $\varphi_{\alpha,\beta}(z) = \alpha \cdot z + \beta$. Compute the probability density function for the random variable $Y_{\alpha,\beta} = \varphi_{\alpha,\beta}(X)$.

A

By the above Y has density $y \mapsto |\alpha|^{-1} p_X((y-\beta)/\alpha)$.

When $\varphi = \varphi_{\alpha,\beta}$ we have $\mathrm{E}\left[\varphi_{\alpha,\beta}(X)\right] = \varphi_{\alpha,\beta}(\mathrm{E}[X])$. Is this case for other functions $\varphi\colon \mathrm{R} \to \mathrm{R}$ in place of $\varphi_{\alpha,\beta}$?

A

Not necessarily. For example given an exponential random variable X and $\varphi(x) = x^2$ we have $\mathrm{E}[\varphi(X)] = 2\varphi(\mathrm{E}[X])$.

As an optional extra look up Jensen's inequality.

A

Jensen's inequality says that when $\varphi\colon \mathrm{R} \to \mathrm{R}$ is a convex function we have $\varphi(\mathrm{E}[X]) \leq \mathrm{E}[\varphi(X)]$.

5 The Binomial distribution and the central limit theorem

Two important discrete distributions are the Bernoulli distribution and the Binomial distribution. We say that a random variable X has Bernoulli distribution with parameter $q \in [0, 1]$ if X has outcome space $\{0, 1\}$ with $\mathrm{P}(X = 1) = q$. This is often abbreviated as $X \sim \mathcal{B}(q)$. For more details on the Bernoulli distribution we refer to the lecture.

Given $k \in \mathrm{N}$ and $q \in [0, 1]$, we say that Z is Binomially distributed random variable with parameters k and q if $Z = X_1 + \cdots + X_k$ where $X_i \sim \mathcal{B}(q)$ and $X_1, \cdots, X_k$ are independent and identically distributed. This is often abbreviated as $Z \sim \mathrm{BINOM}(k, q)$.

Compute the expectation and variance of $Z \sim \mathrm{BINOM}(k, q)$. You can use the following two useful facts:

1. Given any sequence of random variables $W_1, \cdots, W_k$ we have

   $\mathrm{E}\left[\sum_{i=1}^{k} W_i\right] = \sum_{i=1}^{k} \mathrm{E}\left[W_i\right]$.

2. Given a sequence of **independent** random variables $W_1, \cdots, W_k$ we have

$$\text{Var}\left(\sum_{i=1}^{k} W_i\right) = \sum_{i=1}^{k} \text{Var}\left(W_i\right).$$

A

From 1. we have

$$\text{E}[Z] = \text{E}\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} \text{E}[X_i] = \sum_{i=1}^{k} q = kq.$$

Form 2. we have

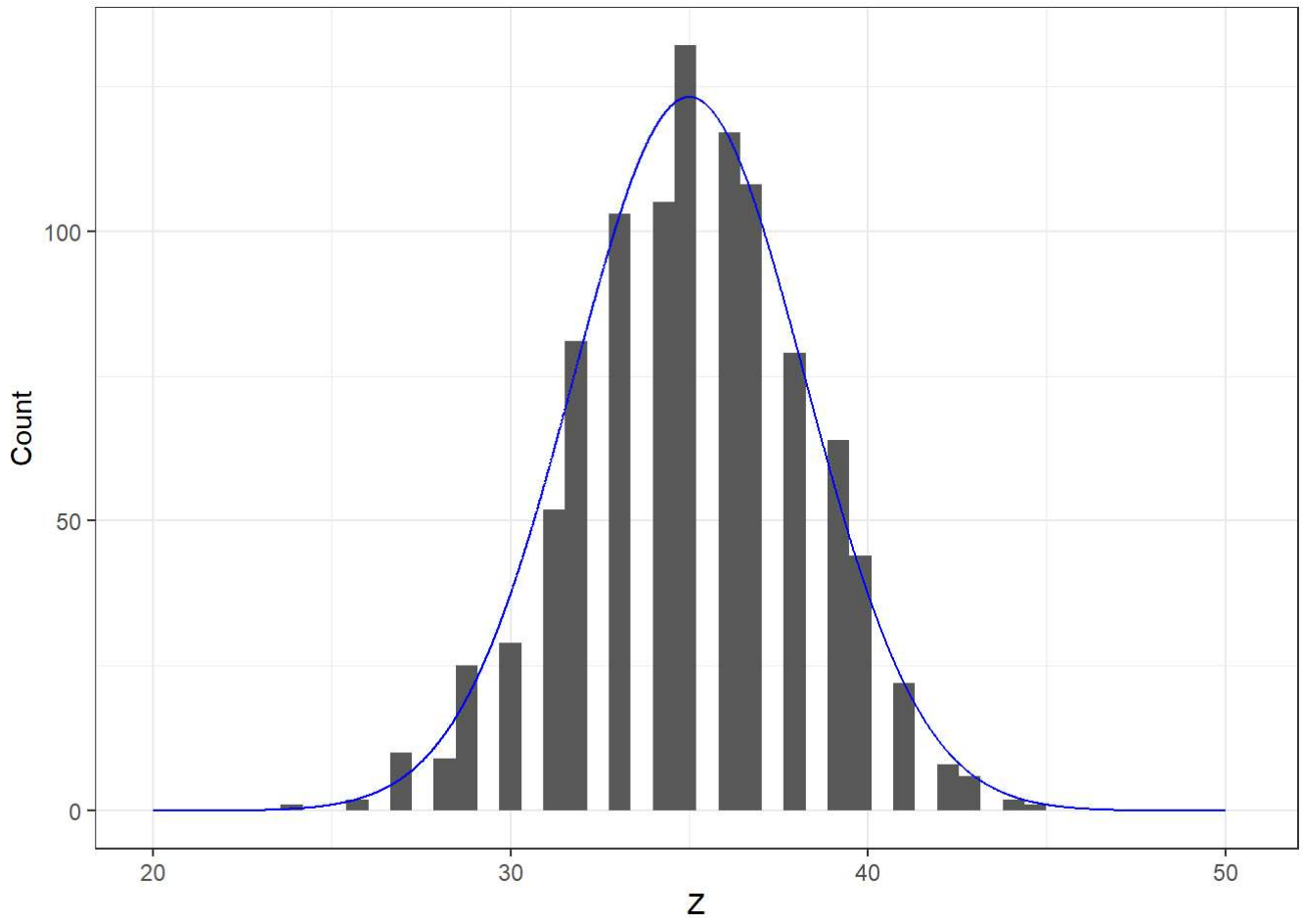
$$\text{Var}(Z) = \text{Var}\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} \text{Var}(X_i) = \sum_{i=1}^{k} q(1-q) = kq(1-q).$$

Is it always true that $\text{Var}\left(\sum_{i=1}^{k} W_i\right) = \sum_{i=1}^{k} \text{Var}\left(W_i\right)$, even if $W_1, \cdots, W_k$ are not independent?

A No - Consider the case where $k = 2$ and $W_2 = -W_1$.

Use the **rbinom()** function to generate a sample of size n consisting of independent Binomial random variables $Z_1, \cdots, Z_n \sim \text{BINOM}(k, q)$, with $n = 1000$, $k = 50$ and $q = 0.7$.

Generate a histogram for your sample. In addition, plot a rescaled density for the normal distribution with mean $\mu = k \cdot q = 700$ and variance $\sigma^2 = k \cdot q \cdot (1-q) = 210$. You should rescale by $n = 1000$. Your plot should something like this:



As a challenging optional extra, try to explain this behavior using the central limit theorem?

A

Each $Z_i = \sum_{j=1}^{k} X_j$ where $X_j \sim \mathcal{B}(q)$, with $\mathrm{E}[X_j] = q$ and $\mathrm{Var}(X_j) = q(1-q)$. By the central limit theorem for large values of k if we take

$$W_i := \sqrt{\frac{k}{q(1-q)}}\left(\frac{1}{k}\sum_{j=1}^{k} X_j - q\right) = \sqrt{\frac{k}{\sigma^2}}\left(\frac{1}{k}\sum_{j=1}^{k} X_j - \mu\right).$$

Then W_i is approximately distributed as a standard Gaussian $\mathcal{N}(0, 1)$. This means that Z_i can be approximated by a Gaussian random variable with mean $k \cdot q$ and variance $k \cdot q(1-q)$.

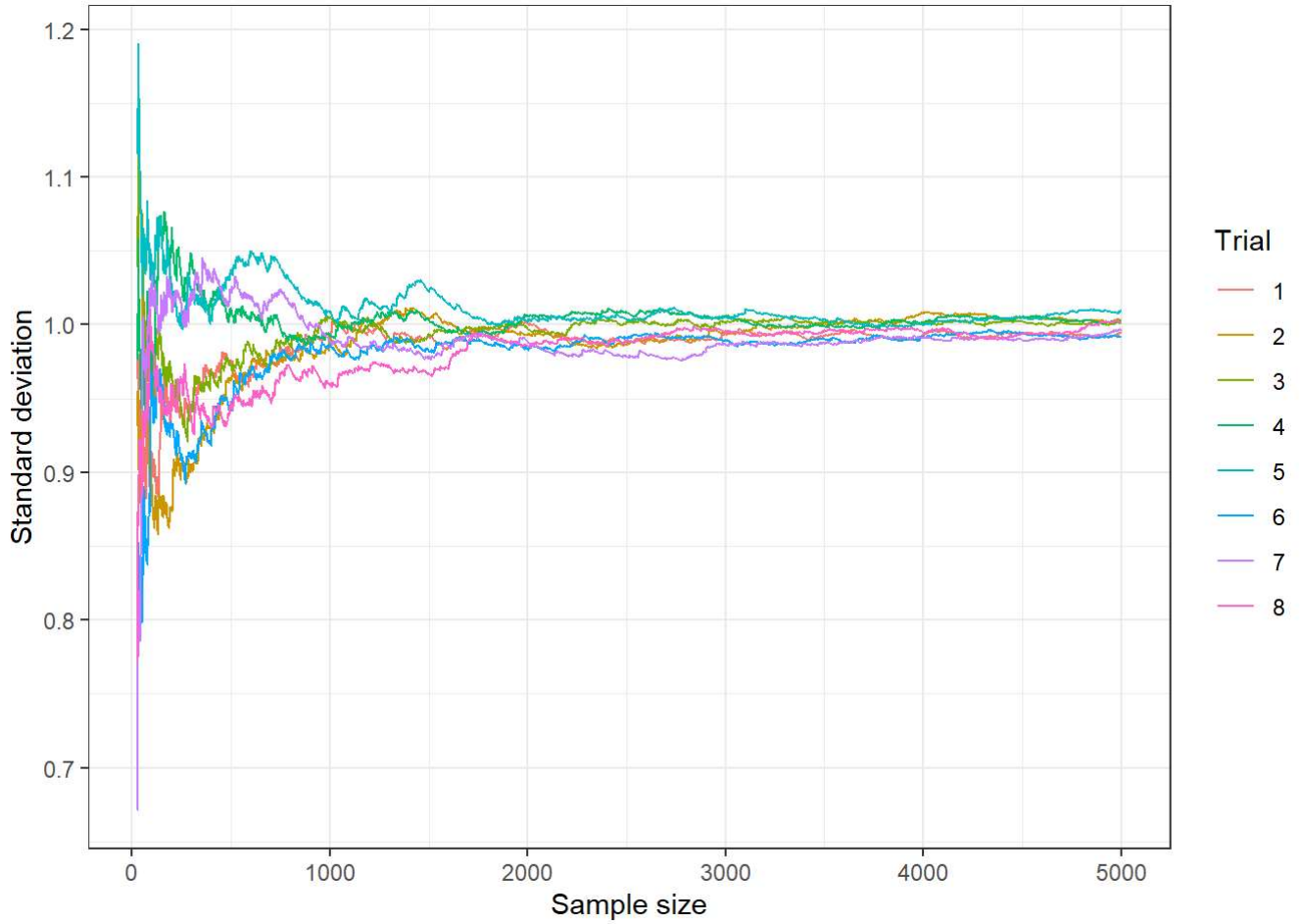
6 Why do we rescale the median absolute deviation?

In our lecture on exploratory data analysis we introduced the Median Absolute Deviation. This can be computed within R using the **mad()** function. Use the **help()** function to investigate the arguments for this function. Note that there is an optional “constant”

argument that defaults to 1.4826. In this question we will look at why this is the case.

The following code generates the sample standard deviation as a function of the sample size for some randomly generated Gaussian data.

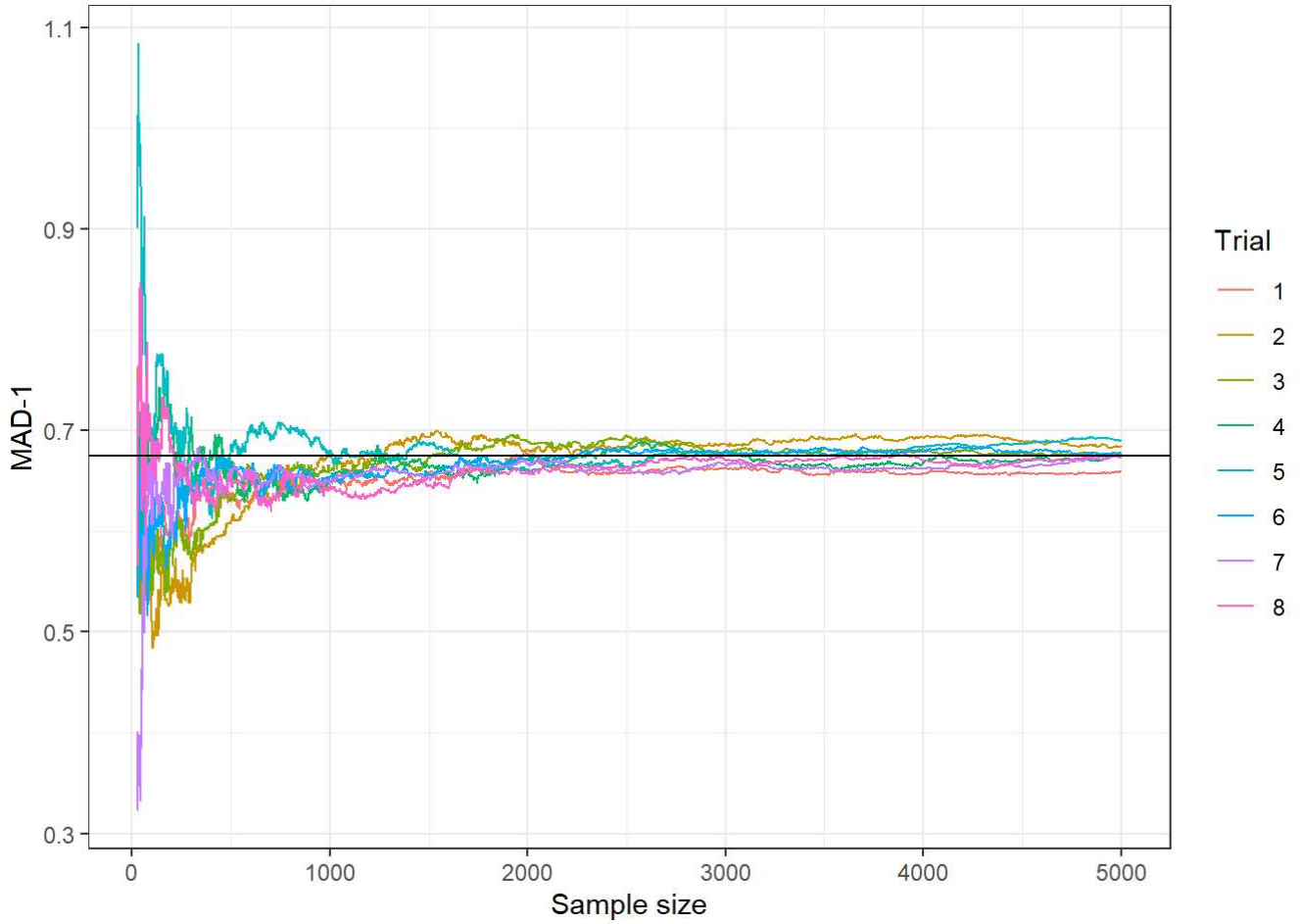
```
total_sample_size<-5000 # set the total sample size

num_trials<-8 # set the number of trials

set.seed(123) # set the random seed

# generate a sample standard deviations as function of the sample size
# for some randomly generated Gaussian data
gaussian_sample_sd_by_sample_size<-function(){

  return(
    data.frame(sample_size=seq(total_sample_size),X=rnorm(total_sample_size))%>% #generat
e normal data
           mutate(sd=map_dbl(row_number(),~sd(X[1:.x]))) # sd of the initial segment
  )
}

df<-data.frame(trial=seq(num_trials))%>%

         mutate(data=map(trial, ~gaussian_sample_sd_by_sample_size()))%>% # apply simula
tion for each trial
  unnest(cols=data) # unnest over the different trials



ggplot(df%>%
        filter(sample_size>25),aes(x=sample_size,y=sd,color=as.character(trial)))+
  geom_line()+theme_bw()+labs(color="Trial",x="Sample size", y="Standard deviation") # Pl
ot results
```



By default **mad()** includes a scale factor of 1.4826. However, we can set the scale factor to 1 by applying constant=1. We shall denote this form of the median absolute deviation by “MAD-1”. Copy and modify the above code to generate a data frame which contains a column for the median absoluted deviation with a scale factor of one (MAD-1). Plot your data and include a horizontal line at $0.6744908 = 1/1.4826$. Your plot should look something like this:

A

```
set.seed(123) # set the random seed

# generate a sample MAD with scale factor of one as a function of the sample size
# for some randomly generated Gaussian data

gaussian_mad_1_by_sample_size<-function(){

  return(
    data.frame(sample_size=seq(total_sample_size),X=rnorm(total_sample_size))%>% #generat
e normal data
           mutate(mad_1=map_dbl(row_number(),~mad(X[1:.x],constant=1)))# sd of the initi
al segment
             )
}

df<-data.frame(trial=seq(num_trials))%>%

          mutate(data=map(trial, ~gaussian_mad_1_by_sample_size()))%>% # apply simulation
for each trial
  unnest(cols=data) # unnest over the different trials



ggplot(df%>%
         filter(sample_size>25),aes(x=sample_size,y=mad_1,color=as.character(trial)))+
  geom_line()+theme_bw()+labs(color="Trial",x="Sample size", y="MAD-1")+# Plot results
  geom_hline(yintercept=0.6744908)# Include a horizontal line at 1/1.4826
```



7 Chebyshev's law of large numbers

This is an entirely optional and much more advanced exercise. It is intended only for those who already have significant experience of probability. You can safely leave this out if you have insufficient time.

Firstly, give a proof of Chebyshev's law of large numbers.

A

First observe that for independent X_i, X_j we have

$$\mathrm{E}\Big[\big(X_i-\mu\big)\big(X_j-\mu\big)\Big] = 0.$$

Thus, we have

$$\mathrm{E}\left[\left\{\sum_{i=1}^{n}\big(X_i-\mu\big)\right\}^2\right] = \sum_{i=1}^{n}\mathrm{E}\Big[\big(X_i-\mu\big)^2\Big] = \sum_{i=1}^{n}\mathrm{Var}\big(X_i\big) = n\sigma^2.$$

Thus, for any $\epsilon > 0$ we have

$$\mathrm{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \epsilon\right] = \mathrm{P}\left[\left|\sum_{i=1}^{n}\left(X_i - \mu\right)\right| > n\epsilon\right]$$

$$= \mathrm{P}\left[\left\{\sum_{i=1}^{n}\left(X_i - \mu\right)\right\}^2 > (n\epsilon)^2\right]$$

$$\leq \frac{1}{(n\cdot\epsilon)^2}\cdot \mathrm{E}\left[\left\{\sum_{i=1}^{n}\left(X_i - \mu\right)\right\}^2\right] = \frac{\sigma^2}{n\cdot\epsilon^2}.$$

Moreover, as $n \to \infty$ the above converges to zero for any $\epsilon > 0$.

Secondly, look up Hoeffding's inequality and explain the advantage of Hoeffding's inequality over the law of large numbers?

Hoeffding's inequality gives results for all sample sizes $n \in \mathrm{N}$ and not just the limit. However, Hoeffding's inequality does require stronger conditions on the random variable (either bounded or at least "sub-Gaussian").