Solutions to problem set 4, 3 March, 2024

Definition 1 *A random variable has normal distribution with mean m and variance σ^2 if its density function is defined by*

$$f(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

This is denoted by $X \sim N(m, \sigma^2)$

A random variable Z has standard normal distribution if it has normal distribution with mean $m = 0$ and variance $\sigma = 1$, that is, $Z \sim N(0,1)$. Its density function is given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\cdot\exp\left(-\frac{x^2}{2}\right).$$

That is, for every $a < b$ we have

$$\mathbb{P}(a \le X \le b) = \int_a^b f(x)\,dx = \int_a^b \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx.$$

The cumulative distribution function of the standard normal distribution is denoted by Φ:

$$\Phi(t) = \mathbb{P}(Z \le t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}}\cdot\exp\left(-\frac{x^2}{2}\right)dx.$$

We can use that $\Phi(-x) = 1 - \Phi(x)$, and $\Phi(0.5) = 0.6915$, $\Phi(1) = 0.8413$, $\Phi(2) = 0.9772$, $\Phi(3) = 0.9987$.

Definition 2 *Random variables $X_1, \ldots, X_n$ are independent if*

$$\mathbb{P}(X_1 \le t_1, X_2 \le t_2, \ldots, X_n \le t_n) = \mathbb{P}(X_1 \le t_1)\cdot\mathbb{P}(X_2 \le t_2)\cdot\ldots\cdot\mathbb{P}(X_n \le t_n)$$

holds for all real numbers $t_1, t_2, \ldots, t_n$.

Proposition 3 *Let X and Y be independent random variables with normal distribution. Then*

a) for real numbers $c \ne 0$ and d, the random variable $cX + d$ has normal distribution with mean $c \cdot \mathbb{E}(X) + d$ and variance $c^2\mathrm{Var}(X)$;

b) for every real number t we have $(X - m)/\sigma$ has standard normal distribution, and

$$\mathbb{P}(X \le t) = \mathbb{P}\left(\frac{X-m}{\sigma} \le \frac{t-m}{\sigma}\right) = \Phi\left(\frac{t-m}{\sigma}\right).$$

c) $X + Y$ has normal distribution, and $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$, and $\mathrm{Var}(X+Y) = \mathrm{Var(X)} + \mathrm{Var}(Y)$.

d) If $X_1, X_2, \ldots, X_n$ are independent normal variables with mean m and variance σ^2, then their mean also has a normal distribution:

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n} \sim N\left(m, \frac{\sigma^2}{n}\right).$$

---

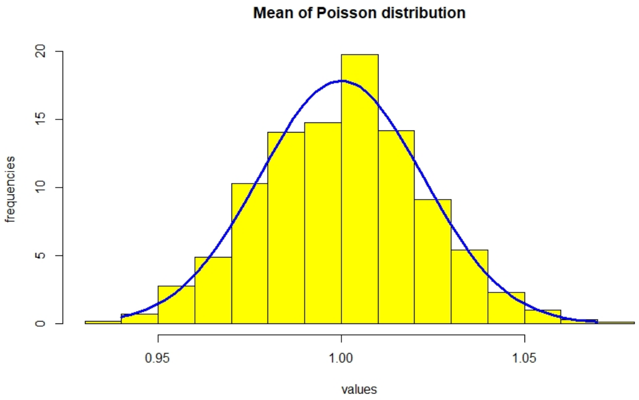
1. Generate $m = 1000$ independent, different samples, each consisting of $n = 2000$ independent random variables with
   (a) Poisson distribution with parameter $\lambda = 1$;
   (b) exponential distribution with parameter 1;
   (c) uniform distribution on the interval $[0, 2]$.
   (d) uniform distribution on the interval $[-3, 5]$

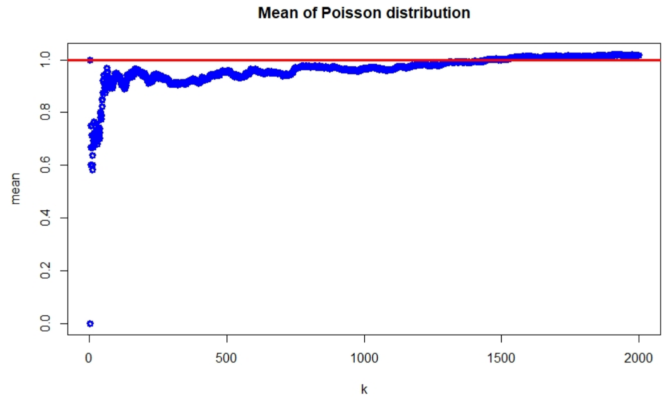
   Calculate the mean of each sample, and make a histogram (on four different figures) of these means.

   Then, let us take just one sample consisting of $n = 2000$ elements (with the distributions given above), and plot the mean of the first k elements as a function of k (here $k = 1, 2, \ldots, 2000$). Are the four figures similar? What are the differences?

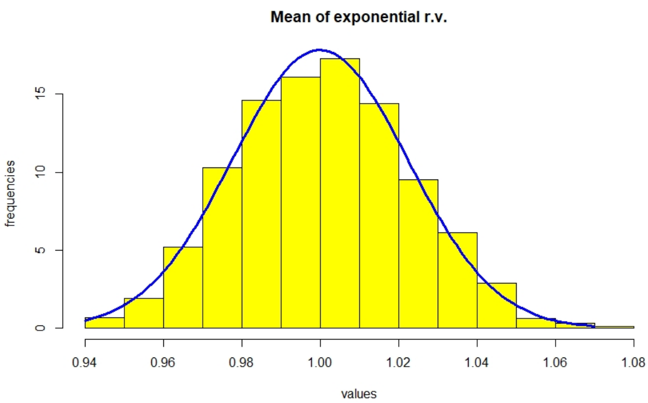
   Poisson distribution:

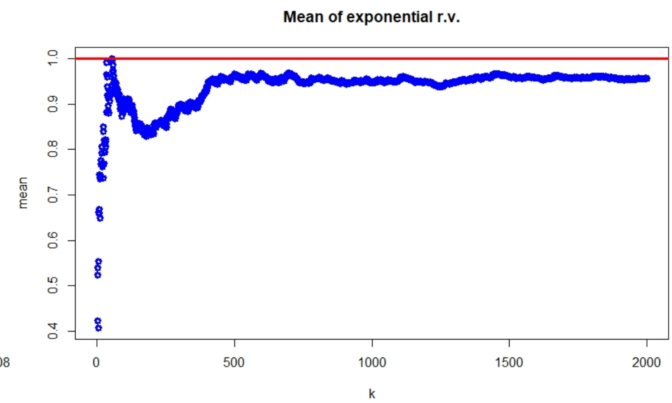
```
> samplea=matrix(rpois(1000*2000, lambda=1), nrow=1000)
> meana=rowMeans(samplea)
```






```
> hist(meana, col="yellow", main="Mean of Poisson distribution",
xlab="values", ylab="frequencies", freq=F)
> curve(dnorm(x, mean=1, sd=1/sqrt(2000)), from=0.94, to=1.07, lwd="3", col="blue", add=T)
> plot(cumsum(samplea[1,])/(1:2000), col="blue", lwd="3", xlab="k", ylab="mean",
main="Mean of Poisson distribution")
> lines(abline(a=1, b=0, col="red", lwd="3"))
```





Exponential distribution:

```
sampleb=matrix(rexp(1000*2000, rate=1), nrow=1000)
meana=rowMeans(sampleb)
hist(meana, col="yellow", main="Mean of exponential r.v.", xlab="values",
ylab="frequencies", freq=F)
curve(dnorm(x, mean=1, sd=1/sqrt(2000)), from=0.94, to=1.07, lwd="3", col="blue", add=T)
plot(cumsum(sampleb[1,])/(1:2000), col="blue", lwd="3", xlab="k", ylab="mean",
main="Mean of exponential r.v.")
lines(abline(a=1, b=0, col="red", lwd="3"))
```

2. Suppose that the temperature in Budapest, on October 31 at midnight has normal distribution X with mean 1 and variance 4 (that is, $X \sim N(1,4)$), in Celsius degree).
a) What is the probability that the temperature at midnight on the given day is below 0° C?

$\mathbb{P}(X < 0) = \mathbb{P}\big(\frac{X-1}{2} < -\frac{1}{2}\big) = \Phi(-0.5) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085.$

```
> pnorm(0, mean=1, sd=2)
[1] 0.3085375
```

b) What is the probability that the temperature is between $-1°C$ and $3°C$?

$\mathbb{P}(Z > 1) = 1 - \Phi(1)$ if $Z \sim N(0,1)$

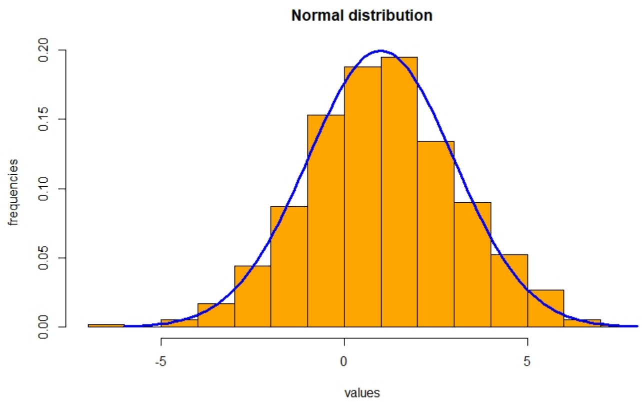
$\mathbb{P}(-1 < X < 3) = \mathbb{P}\big(-1 < \frac{X-1}{2} < 1\big) = \mathbb{P}(X < 3) - \mathbb{P}(X < -1) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 0.6826$, where we used that $\Phi(-x) = 1 - \Phi(x)$ holds for every x.

```
> pnorm(3, mean=1, sd=2)-pnorm(-1, mean=1, sd=2)
```

```
[1] 0.6826895
```

c) Generate a random independent sample of size 1000 from the normal distribution with mean 1 and variance 4. Make a histogram, calculate the mean and the standard deviation, and determine the proportion of the elements between -1 and 3.

```
> temp=rnorm(1000, mean=1, sd=2)
> hist(temp, col="orange", main="Normal distribution", xlab="values",
ylab="frequencies", freq=F)
> curve(dnorm(x, mean=1, sd=2), from=-6, to=8, lwd="3", col="blue", add=T)
> length(temp[(-1<temp)&(temp<3)])/1000
[1] 0.704
```



3. Suppose that the volume of mineral water in a bottle has normal distribution, with standard deviation $0,01$ (in litres). What is the mean of the distribution if the probability that the volume is less than $0,5$ litres is 2%?

Let X be the volume of the water, its mean denoted by m (in litres). Then the condition says that

$$\mathbb{P}(X < 0.5) = 0.02.$$

Since X has normal distribution with mean m and standard deviation $\sigma = 0.01$, we obtain that

$$\mathbb{P}(X < 0.5) = \Phi\left(\frac{0.5 - m}{0.01}\right) = 0.02 \quad \Rightarrow \quad \frac{0.5 - m}{0.01} = \Phi^{-1}(0.02) = -\Phi^{-1}(0.98) = -2.05$$

$$0.5 - m = 0.01 \cdot -2.05 \quad \Rightarrow \quad m = 0.5 + 0.01 \cdot 2.05 = 0.52.$$

```
> qnorm(0.02)
[1] -2.053749
> qnorm(0.98)
[1] 2.053749
```

4. We install a programme which consists of $n = 68$ components. The time of the installation of each file has mean $m = 10$ and $\sigma = 2$ (in seconds), and these are independent of each other.

a) Calculate the probability that the sum of the installation time is at most 12 minutes, supposing that the installation times have normal distribution.

b) Approximate the same probability without assuming normal distribution.

c) The next version consists of k files, with the same rules as above. We also know that the probability that the total time is less than 10 minutes is 95% percent. Determine the value of k.

Let us denote by S_n the sum of the installation times of n components ($n = 68$), in seconds. Since the sum of independent normal distribution has normal distribution, S_n has normal distribution, with mean $n\mu$ and standard deviation $\sqrt{n}\sigma$. Hence

$$\mathbb{P}(S_{68} \leq 720) = \mathbb{P}(S_n < 720) = \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} < \frac{720 - 680}{2\sqrt{68}}\right) = \Phi(2.42) = 99.2\%$$

b) Postponed till next week.

c)

$$0.95 = \mathbb{P}(S_n < 600) = \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} < \frac{600 - 10n}{2\sqrt{n}}\right) = \Phi\left(\frac{600 - 10n}{2\sqrt{n}}\right)$$

Since $\Phi(1.645) = 0.95$, we have $1.645 = \dfrac{600 - 10n}{2\sqrt{n}}$. By solving this, we have that $n \leq 57.5$, hence the number of components has to be at most 58.

5. Suppose that $X_1, X_2, \ldots,$ are independent random variables with Poisson distribution, and the variance of each X_j is equal to 3.

   a) Find the almost sure limit of $\frac{X_1+X_2+\ldots+X_n}{n}$.

   b) Find the limit of $\mathbb{P}\left(\frac{X_1+\ldots+X_n-3\cdot n}{\sqrt{n}} \leq 1\right)$ as n tends to infinity.

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = 3\right) = 1.$$

6. Suppose that $X_1, X_2, \ldots,$ are independent random variables with exponential distribution, and we also know that $\mathbb{P}(X_j \geq 1) = e^{-3}$.

   a) Find the almost sure limit of $\frac{X_1+X_2+\ldots+X_n}{n}$.

   b) Find the limit of $\mathbb{P}\left(\frac{X_1+\ldots+X_n-\frac{1}{3}\cdot n}{\sqrt{n}} \leq 1\right)$ as n tends to infinity.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}; & x > 0; \\ 0; & x < 0. \end{cases}$$

Random variables X, Y are identically distributed if their cumulative distribution function is the same: $\mathbb{P}(X \leq t) = \mathbb{P}(Y \leq t)$ holds for every t.

Theorem 4 (Law of large numbers) *Let $X_1, X_2, \ldots$ be independent identically distributed random variables. Suppose furthermore that $m = \mathbb{E}(X_1) < \infty$. Then*

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathbb{E}(X_1) = m$$

holds with probability 1 as $n \to \infty$.

Theorem 5 (Central limit theorem) *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables, for which $\mathbb{E}(X_1) = m$ and s.d.$(X_1) = \sigma < \infty$, that is, their variance is finite. Then for every real number t we have*

$$\mathbb{P}\left(\frac{X_1 + X_2 + \ldots + X_n - n \cdot m}{\sigma\sqrt{n}} \leq t\right) \to \mathbb{P}(Z \leq t) \qquad (n \to \infty),$$

where Z has standard normal distribution, that is,

$$\mathbb{P}(Z \leq t) = \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$