Задача 2.4

In [2]:

```
%matplotlib inline
import scipy.stats as st
import numpy as np
import matplotlib.pyplot as plt
import math
import statsmodels.api as sm
```

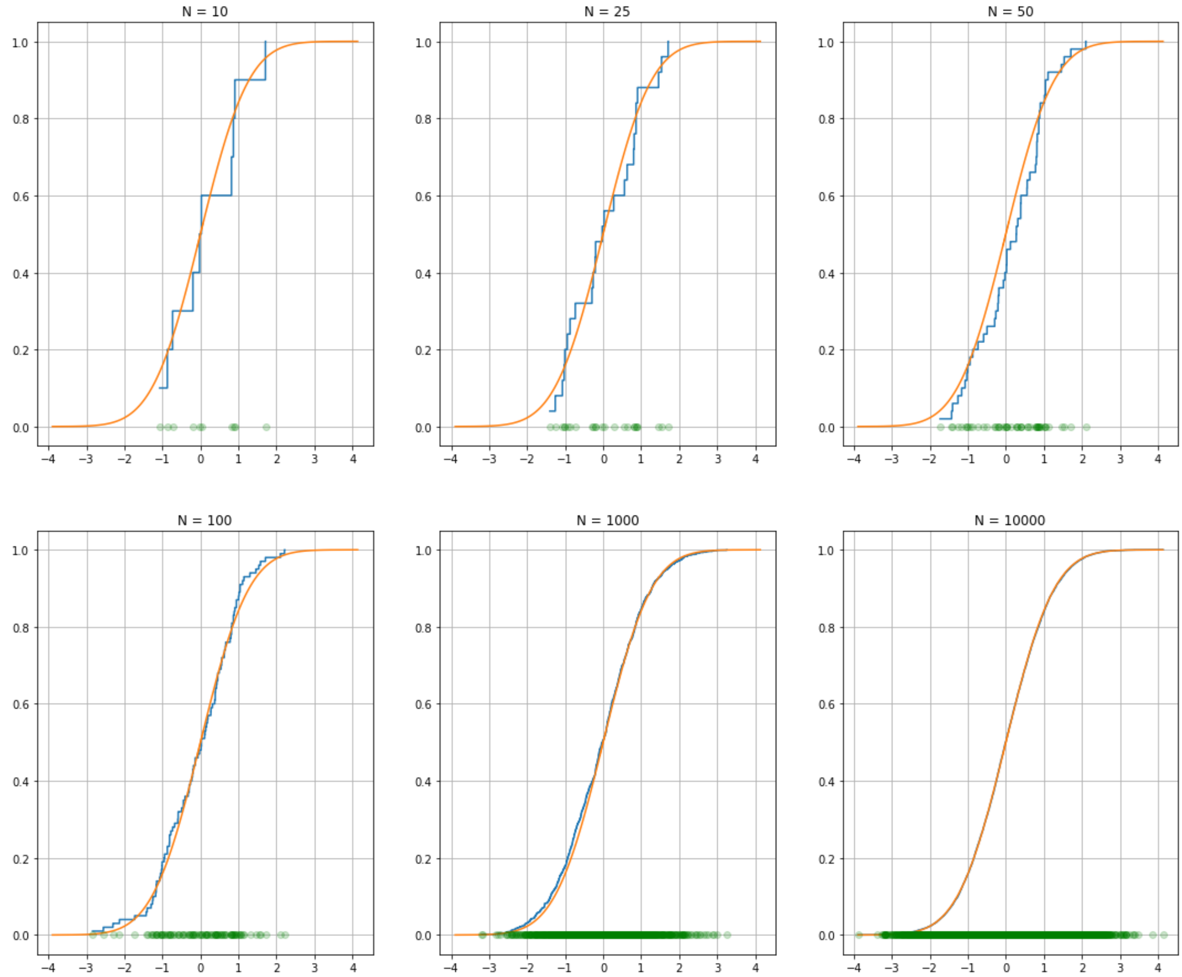
In [3]:

```
teta = 1.0 #параметр распредения
norm = st.norm(loc=0, scale=1.0)
sample = norm.rvs(size=10000)
```

Строим эмпирическую функцию распределения $\hat{F}(x)$

In [10]:

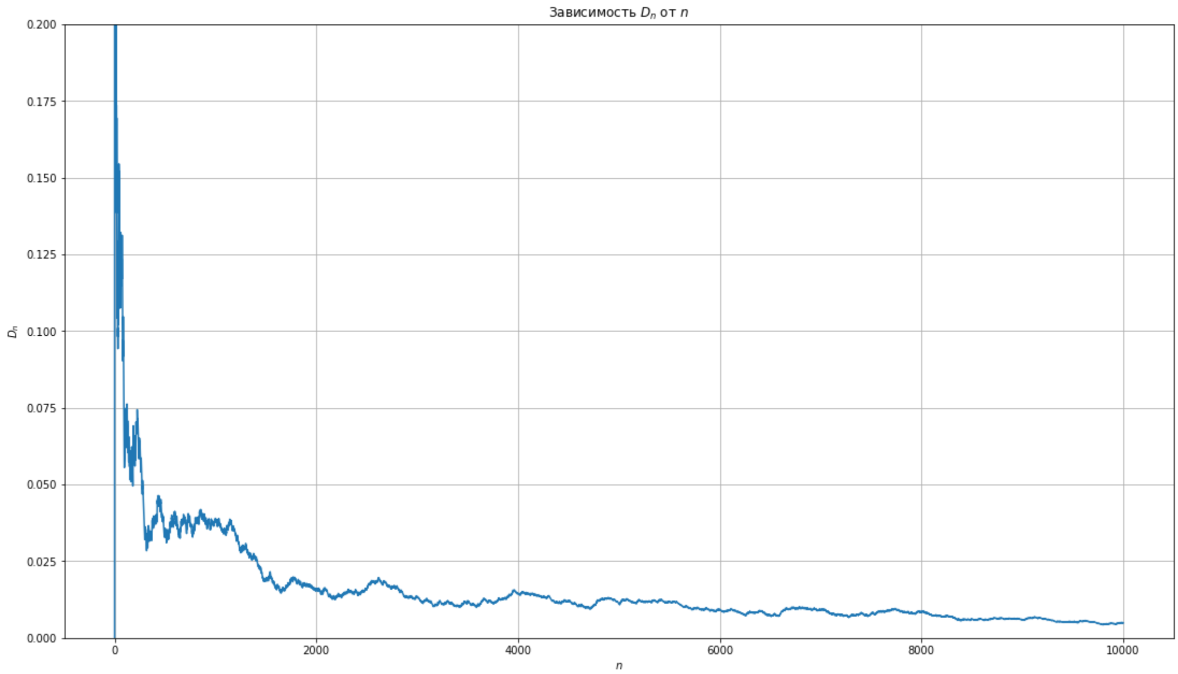
```
sp_n = {10, 25, 50, 100, 1000, 10000}
stats = np.zeros((6,10000))
D = np.zeros(10001)
D_sn = np.zeros(10001)
i = 0
for n in range(1,10001):
    ecdf = sm.distributions.ECDF(sample[:n])
    z = np.linspace(min(sample), max(sample),10000)
    x = np.linspace(min(sample[:n]), max(sample[:n]), 10000)
    y = ecdf(x)
    D[n] = np.amax(abs(ecdf(z) - norm.cdf(z)))
    D_sn[n] = D[n]*math.sqrt(n)
    if n in sp_n:
        stats[i] = y
        i += 1
        plt.subplot (2, 3, i)
        plt.step(x, y)
        plt.step(z, norm.cdf(z))
        plt.plot(sample[:n], np.zeros(n), 'go', alpha = 0.2)
        plt.title ("N = " + str(n))
        fig = plt.gcf()
        plt.grid(True)
        fig.set_size_inches(18.5, 15.5)
```



Построим график зависимости статистики $D_n = \sup_{x \in \mathbb{R}} \left|\hat{F}(x) - F(x)\right|$

In [11]:

```
plt.plot([n for n in range(0,10001)], D)
plt.title (r'Зависимость $D_{n}$ от $n$')
plt.xlabel(r'$n$')
plt.ylabel(r'$D_{n}$')
fig = plt.gcf()
plt.grid(True)
fig.set_size_inches(18.5, 10.5)
plt.ylim (0.0, 0.2)
plt.show()
```



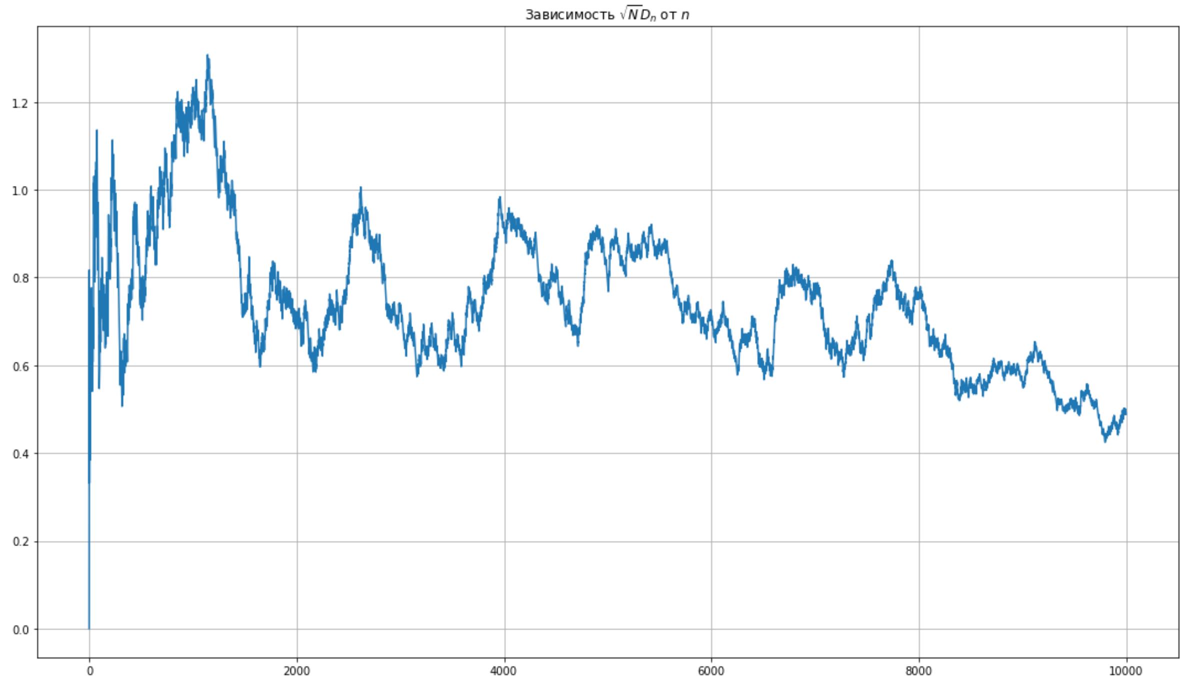
Согласно теореме Гливенко — Кантелли: Пусть $X_1, \ldots, X_n, \ldots$ - бесконечная выборка из распределения, задаваемого функцией распределения F. Пусть $\hat{F}$ - выборочная функция распределения, построенная на первых n элементах выборки. Тогда : $\lim\limits_{n\to\infty} \sup\limits_{x\in\mathbb{R}} \left|\hat{F}(x) - F(x)\right| = 0$

График иллюстрирует формулировку теоремы

Построим график зависимости статистики $\sqrt{n}D_n$

In [7]:

```
plt.plot([n for n in range(0,10001)], D_sn)
plt.title (r'Зависимость $\sqrt{n}D_{n}$ от $n$')
plt.xlabel(r'$n$')
plt.ylabel(r'$\sqrt{n}D_{n}$')
fig = plt.gcf()
plt.grid(True)
fig.set_size_inches(18.5, 10.5)
plt.show()
```



Теорема Колмогорова уточняет скорость сходимости выборочной функции распределения к её теоретическому аналогу. Пусть $X_1, \ldots, X_n, \ldots$ — бесконечная выборка из распределения, задаваемого непрерывной функцией распределения $F(x)$. Пусть $F_n(x)$ — выборочная функция распределения, построенная на первых n элементах выборки. Тогда : $\sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to K$ по распределению при $n \to \infty$, где K — случайная величина, имеющая распределение Колмогорова.

Вывод

Таким образом, мы наблюдаем, что скорость сходимости выборочной функции распределения к её теоретическому аналогу имеет порядок $1/\sqrt{n}$