MATHEMATICAL STATISTICS, Problems to Lessons 1-2.

1. Let $F(x)$ be the theoretical, and $F_n^*(x)$ be the empirical c.d.f. of the sample. Prove that for any $x \in \mathbb{R}$: $\mathbb{E}(F_n^*(x)) = F(x)$, $\mathbb{D}^2(F_n^*(x)) = \frac{F(x)(1-F(x))}{n}$, and $\lim_{n\to\infty} F_n^*(x) = F(x)$, almost surely.

2. *Empirical density histogram*: for sample size n let us divide the real number line into disjoint intervals Δ_j's of length h_n. Denote by ν_j the number of sample entries in Δ_j.

$$f_n^*(x) := \frac{\nu_j}{nh_n}, \qquad x \in \Delta_j.$$

   Prove that $\int_{-\infty}^{\infty} f_n^*(x)\,dx = 1$. Construct $f_n^*(x)$ with different h_n's and compare it to the theoretical p.d.f. $f(x)$.

   Proposition: if x is a point of continuity of f and $n \to \infty$ in such a way that $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} nh_n = \infty$, then $\lim_{n\to\infty} f_n^*(x) = f(x)$, almost surely.

3. Let $X_1, \dots, X_n$ be i.i.d. r.v.'s with c.d.f. $F(x)$, and $X_1^* \le \dots \le X_n^*$ be the ordered sample. Find a formula for $F_{n;k}(x)$, the c.d.f. of X_k^* $(k = 1, \dots, n)$.

4. Let $X_1, \dots, X_n$ be i.i.d. r.v.'s with continuous c.d.f. $F(x)$, and $X_1^* \le \dots \le X_n^*$ be the ordered sample. By the transformation $Y_i := F(X_i)$, $Y_1, \dots, Y_n$ is an i.i.d. $\mathcal{U}(0,1)$-sample. Show that $Y_k^* = F(X_k^*)$ and find $F_{n;k}(x)$ by means of the c.d.f. $G_{n;k}(y)$ of Y_k^*. Further find the p.d.f. $f_{n;k}(x)$ of X_k^* by means of the p.d.f. $g_{n;k}(y)$ of Y_k^*.

5. Prove that $Y_k^* \sim \mathcal{B}(k, n-k+1)$.

6. Find the s^{th} moment $(s = 1, 2, \dots)$, further the expectation and variance of Y_k^*.

7. Find the joint p.d.f. $g_{n;k_1,\dots,k_r}(y_1, \dots, y_r)$ of $Y_{k_1}^*, \dots, Y_{k_r}^*$ for any r-tuple $1 \le k_1 < k_2 < \dots < k_r \le n$.

8. Find the joint p.d.f. of $Y_1^*, \dots, Y_n^*$, and that of $X_1^*, \dots, X_n^*$ by means of the result obtained in 7. Give a combinatorial explanation too.

9. Let $Y_1^* \le \dots \le Y_n^*$ be $\mathcal{U}(0,1)$ order statistics. Find the joint distribution of the differences

$$U_1 := Y_1^*, \quad U_k := Y_k^* - Y_{k-1}^* \quad (k = 2, \dots, n), \quad U_{n+1} := 1 - Y_n^*$$

   (They are identically distributed but not independent.)

10. Let $X_1^* \le \dots \le X_n^*$ be $\mathcal{E}xp(\lambda)$ order statistics. Find the joint distribution of the differences

$$U_1 := X_1^*, \quad U_k := X_k^* - X_{k-1}^* \quad (k = 2, \dots, n).$$

   (They are independent but not identically distributed.)

11. With the above notation (see Problem 4) prove that

$$\sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| = \sup_{0<y<1} |G_n^*(y) - G(y)|,$$

   where $G(y) = y$ $(0 < y < 1)$ is the c.d.f. of the $\mathcal{U}(0,1)$-distribution. (The same is true without absolute values.) Also prove that the above suprema are, in fact, finite maxima over the order statistics. We remark that $\sqrt{n}\sup_{0<y<1}(G_n^*(y) - G(y))$ approaches the so-called *Brownian bridge* process, if $n \to \infty$. Generate this process by means of computer simulation!