Chapter 8

• Sample $X_1, X_2, ..., X_n$ – independent random variables with the same distribution $f(x)$. The joint distribution of the sample is

$$f(x_1, x_2, ..., x_n) = f(x_1)f(x_2) \cdot ... \cdot f(x_n).$$

Any function of $X_1, X_2, ..., X_n$ is called a statistic.

• Some important statistics

The sample mean (describes location)

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The sample variance (describes variability)

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

The order statistics (describe both location and variability)

$$X_{(1)} \le X_{(2)} \le ... \le X_{(n)},$$

where $X_{(1)} = \min_i X_i$, $X_{(2)}$ is the second minimal, ...,$X_{(n)} = \max_i X_i$.

The sample median (location)

$$\tilde{X} = X_{((n+1)/2)}$$

if n is odd and

$$\tilde{X} = \frac{1}{2}(X_{(n/2)} + X_{(n/2+1)})$$

if n is even.

The sample standard deviation S (variability).

The range (variability)

$$R = X_{(n)} - X_{(1)}.$$

• Distribution of the sample mean.

If $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$, then

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

exactly.

In the general case (X-s are not normally distributed)

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

approximately (asymptotically) if n is not too small. This is due to the

Central Limit Theorem: If $\sigma^2 < \infty$, then

$$P\left(a \le \sqrt{n}\frac{\bar{X} - \mu}{\sigma} \le b\right) \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx$$

for any $a < b$ as $n \to \infty$.