Problem set 5, Probability and statistics, 10 March, 2025

Markov-inequality Let $g : \mathbb{R} \to \mathbb{R}$ be a monotonically increasing, positive function, $X \geq 0$ a random variable, for which $EX < \infty$ and $\varepsilon > 0$ Then

$$P(X \geq \varepsilon) \leq \frac{E\left(g(X)\right)}{g(\varepsilon)}.$$

Especially, if $g(x) = x$, then

$$P(X \geq \varepsilon) \leq \frac{EX}{\varepsilon}.$$

Chebyshev-inequality Let X be an arbitrary random variable, for which $Var(X) < \infty$ and $\varepsilon > 0$. Then

$$P(|X - EX| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}.$$

Covariance of two random variables:

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

The joint density function of (X, Y) is $f(x, y)$ if for every appropriate subset $A \subseteq \mathbb{R}^2$ we have

$$\mathbb{P}(X \in A) = \int_A f(x, y)\, dx\, dy.$$

Calculation of covariance if the joint density function is $f(x, y)$:

$$\operatorname{cov}(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f(x, y)\, dx\, dy - \left(\int_{-\infty}^{\infty} x \cdot f_X(x)\, dx\right) \cdot \left(\int_{-\infty}^{\infty} y \cdot f_Y(y)\, dy\right),$$

where $f_X(x)$ and $f_Y(y)$ are the marginal densities that can be computed as $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$.

If random variables X and Y are independent, then $\operatorname{cov}(X, Y) = 0$. On the other hand, it may happen that $\operatorname{cov}(X, Y) = 0$, but X and Y are not independent.

The (Pearson) **correlation coefficient** of X and Y (this is always at least -1 and at most 1):

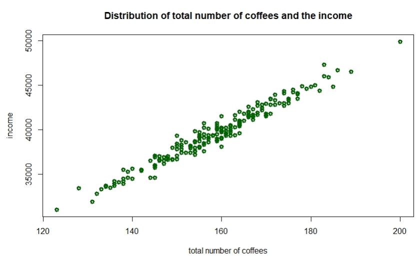
$$R(X, Y) = \frac{\operatorname{cov}(X, Y)}{sd(X)sd(Y)}.$$

$R(X, Y) = 1$ means that $Y = aX + b$ for some a, b, with $a > 0$:

---

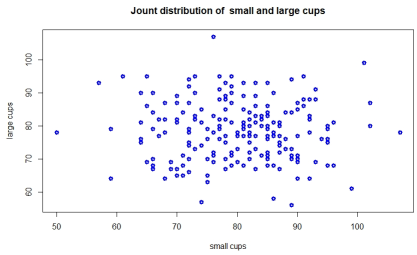
1. A shop is selling smaller and larger cups of coffee. The price of the small version is 200, the price of the large version is 300. Let X be the number of small coffees sold in a day, and Y the number of large coffees sold in a day. We assume that X and Y are independent, have Poisson distribution, and both have expectation 80. Determine the correlation coefficient of the total number of coffees sold in a day and the income (the total price of coffees sent in a day).

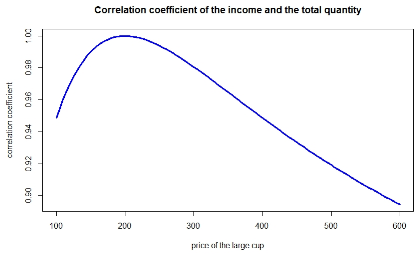
   What will be different if the price of the large version is not 300, but $350, 400, 450$ etc.?




```
> curve((200+x)/sqrt(80000+2*x^ 2), from=100, to=600, lwd="3",
main="Correlation coefficient of the income and the total quantity",
xlab="price of the large cup", ylab="correlation coefficient", col="blue")
```

2. There are 15 boys and 18 girls in a school class. Suppose that students are absent from school independently of each other, everyone with probability $1/10$ on a given day. Determine the covariance and correlation coefficient of the number of girls who are absent and the total number of absent students.





3. We measured temperature at a given place with two different tools. Suppose that the results are independent, normal variables with expectation 6 and standard deviation 2. Let X be the first result, Y the second (with the second tool). Determine the following quantities:

   $\mathrm{cov}(X, X+Y), \mathrm{cov}(X, \frac{X+Y}{2})$, $\mathrm{cov}(X-Y, X+Y)$, $R(X, \frac{X+Y}{2})$.

4. Suppose that the joint density function of (X, Y) is given as follows.

$$f(x,y) = \begin{cases} x+y, & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1; \\ 0 & \text{otherwise.} \end{cases}$$

   Determine the correlation coefficient of X and Y.

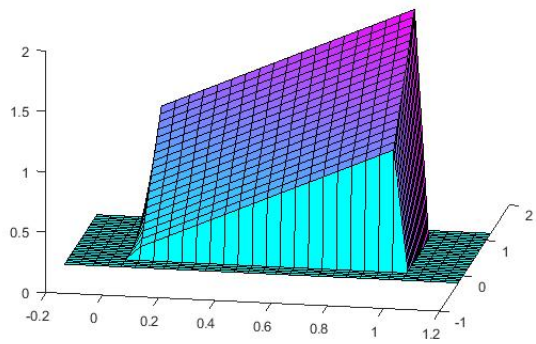


Figure 1: Joint density function on $[0,1] \times [0,1]$

Homework until 16 March, 23:59 Suppose that a server breaks down on each day independently with probability 0.1. Let X be the number of days when it breaks down in September, Y the same in October, and Z the same in November. Calculate the correlation coefficient of X and $X+Y+Z$.