Seminar Nr. 6, Numerical Characteristics of Random Variables

Theory Review

Expectation:
- if $X\begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$ is discrete, then $E(X)=\sum\limits_{i\in I} x_i p_i$.
- if X is continuous with pdf f, then $E(X)=\int\limits_{\mathbb{R}} x f(x)dx$.

Variance: $V(X)=E\left((X-E(X))^2\right)=E\left(X^2\right)-(E(X))^2$.

Standard Deviation: $\sigma(X)=\sqrt{V(X)}$.

Moments:
- **moment of order k**: $\nu_k=E\left(X^k\right)$.
- **absolute moment of order k**: $\overline{\nu_k}=E\left(|X|^k\right)$.
- **central moment of order k**: $\mu_k=E\left((X-E(X))^k\right)$.

Properties:
1. $E(aX+b)=aE(X)+b$, $V(aX+b)=a^2V(X)$
2. $E(X+Y)=E(X)+E(Y)$
3. if X and Y are independent, then $E(XY)=E(X)E(Y)$ and $V(X+Y)=V(X)+V(Y)$
4. if $h:\mathbb{R}\to\mathbb{R}$ is a measurable function, X a random variable;
- if X is discrete, then $E\left(h(X)\right)=\sum\limits_{i\in I} h(x_i)p_i$
- if X is continuous, then $E\left(h(X)\right)=\int\limits_{\mathbb{R}} h(x)f(x)dx$

Covariance: $\text{cov}(X,Y)=E\left((X-E(X))(Y-E(Y))\right)$

Correlation Coefficient: $\rho(X,Y)=\dfrac{\text{cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$

Properties:
1. $\text{cov}(X,Y)=E(XY)-E(X)E(Y)$
2. $V\left(\sum\limits_{i=1}^{n} a_iX_i\right)=\sum\limits_{i=1}^{n} a_i^2V(X_i)+2\sum\limits_{1\le i<j\le n} a_ia_j\text{cov}(X_i,X_j)$
3. X,Y independent => $\text{cov}(X,Y)=\rho(X,Y)=0$ (X and Y are $uncorrelated$)
4. $-1\le\rho(X,Y)\le 1$; $\rho(X,Y)=\pm1$ <=> $\exists\, a,b\in\mathbb{R},\ a\neq 0$ s.t. $Y=aX+b$

Let (X,Y) be a continuous random vector with pdf $f(x,y)$, let $h:\mathbb{R}^2\to\mathbb{R}^2$ a measurable function, then

$$E\left(h(X,Y)\right)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} h(x,y)f(x,y)dxdy$$

---

1. Every day, the number of network blackouts has the following pdf

$$X\begin{pmatrix} 0 & 1 & 2 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}.$$

A small internet trading company estimates that each network blackout costs them $\$500$.

a) How much money can the company expect to lose each day because of network blackouts?
b) What is the standard deviation of the company's daily loss due to blackouts?

2. About ten percent of computer users in a public library do not close Windows properly. On the average, how many users *do* close Windows properly before someone *does not*?

3. (Refer to Problem 1 in Sem. 5) The lifetime, in years, of some electronic component is a random variable with density

$$f(x) = \begin{cases} \dfrac{3}{x^4}, & \text{for } \quad x \geq 1 \\ 0, & \text{for } \quad x < 1. \end{cases}$$

How many years, on the average, can we expect that electronic equipment to last?

4. (Optimal portfolio) A businessman wants to invest $\$600$ and has two companies to choose from, company A, where shares cost $\$20$ each and company B, where shares cost $\$30$ per share. The market analysis shows that for company A the return per share is distributed as follows: lose $\$1$ with probability 0.2, win $\$2$ with probability 0.6, or win/lose nothing. For company B: lose $\$1$ with probability 0.3, win $\$3$ with probability 0.6, or win/lose nothing. The returns from the two companies are independent. In order to maximize the expected return and minimize the risk, which way is better to invest:
a) all money in company A;
b) all money in company B;
c) half the amount in each?

5. (Reduced Variables). Let X be a random variable with mean $E(X)$ and standard deviation $\sigma(X) = \sqrt{V(X)}$. Find the mean and variance of $Y = \dfrac{X - E(X)}{\sigma(X)}$.

6. The joint density function of the vector (X, Y) is $f(x, y) = x + y,\ (x, y) \in [0, 1] \times [0, 1]$. Find
a) the means and variances of X and Y;
b) the correlation coefficient $\rho(X, Y)$.

7. Let X be a discrete random variable with pdf $X \begin{pmatrix} -1 & 0 & 1 \\ \sin^2 a & \cos 2a & \sin^2 a \end{pmatrix}$, $a \in \left(0, \frac{\pi}{4}\right)$. For any $k \in \mathbb{N}^*$, let $Y_k = X^{2k-1}$ and $Z_k = X^{2k}$. Find $\rho(Y_k, Z_k)$. (In particular, X and X^2 are uncorrelated, but *not* independent).

Bonus Problems

8. Two independent customers are scheduled to arrive in the afternoon.Their arrival times are uniformly distributed between 2 pm and 8 pm. What is the expected time of
a) the first (earlier) arrival;
b) the last (later) arrival?

9. In an office n different letters are placed randomly into n addressed envelopes. Let Z_n denote the number of correct mailings. For each $k \in \{1, \ldots, n\}$, let X_k be the random variable defined by

$$X_k = \begin{cases} 1, & \text{if the } k\text{-th letter is placed correctly} \\ 0, & \text{otherwise .} \end{cases}$$

a) Find $E(X_k)$ and $V(X_k)$ for each $k \in \{1, \ldots, n\}$.
b) Find $E(Z_n)$ and $V(Z_n)$.
c) How many correct mailings are to be expected?