Math464/564 O. Knill, **March 1997**

Homework 7

Due: Tuesday, March 11, 1997 in class

Topic: Expectation and Variance, Correlation Probability generating function, Correlation coefficient.

Remark. This HW contains 5 questions. Each question can give 10 points, so that you can get a maximal credit of 50 points for this HW.

1) (10 points) Topic: **Probability generating function**. a) (5 points) Compute the expectation and variance of a (n, p)-Bernoulli distributed random variable using the probability generating function.
b) (5 points) Compute the expectation and variance of a Poisson distributed random variable using the probability generating function.

2) (10 points) Topic: **Covariance and independence**. Given two independent random variables X, Y satisfying $\mathrm{Var}[X] = \mathrm{Var}[Y] = 1$.

a) Compute $\mathrm{Var}[3X + 5Y]$.

b) Compute the correlation coefficient $\rho(2X + Y, X)$.

3) (10 points) Topic: **Expectation and Variance**. The probability space in roulette is defined by $(\Omega = \{0, 1, 2, \dots, 36\}, \mathcal{A} = \{A \subset \Omega\}, \mathrm{P}[\{i\}] = 1/37)$. A new house in Las Vegas offers to play the following option: as usually you bet 1 dollar. If the ball hits a number different from zero which is divisible by 5 then you get back 5 dollars, (you win then $4 = 5 - 1$ dollars). In any other case, you loose your dollar.

a) (5 points) Determine $\mathrm{E}[X]$, which is the expected win or loss.

b) (5 points) Compute $\mathrm{Var}[X]$ which is a measure for the risk of this option.

4) (10 points) Topic: **Correlation**. Given two discrete random variables X, Y which have both positive variance.
a) For each $\theta \in [0, 2\pi]$, define

$$\begin{aligned} X_\theta &= X\cos(\theta) - Y\sin(\theta) , \\ Y_\theta &= X\sin(\theta) + Y\cos(\theta) . \end{aligned}$$

Verify that there exists a value of θ for which X_θ, Y_θ are uncorrelated.

5) (10 points) Topic: **Chebychev inequality**. We throw a dime 100 times. Let X be the number of heads.

Give with Chebychev's inequality an upper bound for the probability that $|X - 50| \geq 10$.

REMINDER OF SOME DEFINITIONS

A **random variable** on a probability space $(\Omega, \mathcal{A}, P)$ is a function $X : \Omega \to \mathbb{R}$ such that for all $x \in \mathbb{R}$, the event $\{X = x\}$ is in $\mathcal{A}$. The random variable X is **discrete** if $X(\Omega)$ is a discrete set.

The **expectation** of a random variable X is defined as

$$\mathrm{E}[X] = \sum_{x \in X(\Omega)} x \cdot \mathrm{P}[X = x] \; .$$

The **variance** of a random variable X is defined as

$$\mathrm{Var}[X] = \mathrm{E}[\,(X - \mathrm{E}[X])^2\,] \; .$$

Convenient is the formula $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$.
The **standard deviation** of X is

$$\sigma[X] = \sqrt{\mathrm{Var}[X]} \; .$$

The **covariance** of two random variables X, Y is defined as

$$\mathrm{Cov}[X, Y] = \mathrm{E}[(X - \mathrm{E}[X]) \cdot (Y - \mathrm{E}[Y])] \; .$$

We know $\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$.

Given two random variables X, Y with $\mathrm{Var}[X] > 0, \mathrm{Var}[Y] > 0$, the **correlation** between X and Y is defined as

$$\mathrm{Corr}[X, Y] = \frac{\mathrm{Cov}[X, Y]}{\sigma[X]\sigma[Y]} \; .$$

If $\mathrm{Corr}[X, Y] = 0$, then X, Y are called **uncorrelated**.

The **correlation coefficient** $\rho(X, Y)$ of two random variables with $\mathrm{Var}[X] > 0, \mathrm{Var}[Y] > 0$ is defined as

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma[X]\sigma[Y]} \; .$$

Two random variables X, Y are **independent** if for all $a, b \in \mathbb{R}$

$$\mathrm{P}[X = x; Y = y] = \mathrm{P}[X = x] \cdot \mathrm{P}[Y = y] \; .$$

The **regression line** of two random variables X, Y is the line $y = ax + b$, where

$$a = \frac{\mathrm{Cov}[X, Y]}{\mathrm{Var}[X]}, b = \mathrm{E}[Y] - a\mathrm{E}[X] \; .$$

REMINDER OF SOME FACTS

$\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$ if X and Y are independent.

If X and Y are independent, then $\mathrm{Cov}[X, Y] = 0$.

Schwartz inequality $\mathrm{E}[XY]^2 \leq E[X^2]\mathrm{E}[Y^2]$. It implies that $|\rho(X, Y)| \leq 1$.

Chebychev inequality. $\mathrm{P}[|X - E[X]| \geq \epsilon] \leq \mathrm{Var}[X]/\epsilon^2$

If $y = ax + b$ is the regression line of two random variables X and Y, then the random variable $\tilde{Y} = aX + b$ minimizes $\mathrm{Var}[Y - \tilde{Y}]$ under the constraint $\mathrm{E}[Y] = \mathrm{E}[\tilde{Y}]$. It is the best guess for Y, when knowing only $\mathrm{E}[Y]$ and $\mathrm{Cov}[X, Y]$. We check $\mathrm{Cov}[X, Y] = \mathrm{Cov}[X, \tilde{Y}]$.