Uppsala Universitet
Matematiska Institutionen
I. Kaj, R. Müller, Y. Wang

Repetition och viktiga formler för M4

- Understand the definitions of expectation, variance and covariance
- Can perform calculation of expectation, variance and covariance
- Probability Computation, find the distribution of given random variables

1 Expectation, Variance and Covariance

1.1 All about expectation

Expectation of a random variable (väntervärde för slumpvariabler): The *expectation* or *expectated value* of a random variable X is written as $\mathbb{E}[X]$, Roughly, it is the average value of the random variable where each value is weighted according to its probability distribution, and it is used to describe where the center of the mass lies.

Definition. *Let X be a continuous random variable with p.d.f (probability density function) $f_X(x)$, the expectation of X is*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Definition. *Let X be a discrete random variable, the expectation of X is*

$$\mathbb{E}[X] = \sum_x x \mathbb{P}(X = x).$$

Expectation of $g(X)$: Let $g(X)$ be a function of X, then it is another random variable where the randomness arises from X. Imagine observing X for many times, then the long-term average of $g(X)$ would approach its expectation.

Definition. *Let X be a continuous random variable with p.d.f (probability density function) $f_X(x)$, and let g be a function. The expectation of $g(X)$ is*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

Definition. *Let X be a discrete random variable, and let g be a function. The expectation of $g(X)$ is*

$$\mathbb{E}[g(X)] = \sum_x g(x)\mathbb{P}(X = x).$$

Important: do some exercises on how to calculate the expectation of a random variable.

Exercise. *Calculate the expectation of X which follows an exponential distribution with parameter a:*

$$f_X(x) = \frac{1}{a}e^{-\frac{x}{a}}, \;\; x \geq 0.$$

Hint: *Integration by parts. The answer is $\mathbb{E}[X] = a$, did you get it correct? One mistake students usually make is that they often integrate over the whole real line. But keep in mind that you shall only integrate over the support of X.*

Properties of expectation:

- *Let g and h be functions, and let a and b be real constants. For any random variable X (discrete or continuous), we have*

  $$\mathbb{E}[ag(X) + bh(X)] = a\mathbb{E}[g(X)] + b\mathbb{E}[h(X)],$$

  and in particular

  $$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b.$$

- *Let X and Y be any random variable (discrete, continuous, independent, or dependent). Then*

  $$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

  More generally, for random variables $X_1, X_2, \ldots, X_n$ and constants $a_1, a_2, \ldots, a_n$,

  $$\mathbb{E}[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1\mathbb{E}[X_1] + a_2\mathbb{E}[X_2] + \cdots + a_n\mathbb{E}[X_n].$$

*One can see that **linearity** of expectation holds regardless of whether the random variables are independent. It is very useful in algorithms as many problems in computer science can be formulated in terms of a **balls into bins** processes. We will see an example in the quiz. Another example could be calculating the expected time complexity of a randomized quick sort algorithm. But as IT students you probably either already know better than me or you will.*

Remark. *Note that $\mathbb{E}[X] = \mu$ does not imply $\mathbb{E}[X^2] = \mu^2$.*

Expectation for independent random variables *As you probably know by this point, assuming independence brings more properties. Let X and Y be **independent** random variables and g, h be functions, we have*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y],$$
$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)].$$

Remark. *The converse is not true. It is possible for* $\mathbb{E}(XY) = \mathbb{E}[X]\mathbb{E}[Y]$ *to hold even if they are dependent. See Quiz.*

Probability as an expectation *Let A be an event, we can write $\mathbb{P}(A)$ as an expectation as follows. Define the indicator function:*

$$I_A = \begin{cases} 1, & \text{if } A \text{ occurs}, \\ 0, & \text{otherwise}. \end{cases}$$

Then I_A is a random variable and

$$\mathbb{E}[I_A] = 0 \times \mathbb{P}(I_A = 0) + 1 \times \mathbb{P}(I_A = 1) = \mathbb{P}(A).$$

Example. *Colin flips a fair coin for 10 times, and he gets rewarded n kr if he got n heads. What's the expectation of his gain?*
Solution: *Define Colin's gain as X, one can of course write down the distribution of X and calculate the expectation by definition. An alternative is to define the result of the ith flip as I_i, and $I_i = 1$ if it's a head, and 0 if it's a tail. Since we get rewarded $1kr$ for each head we get, one can write*

$$X = I_1 + I_2 + \cdots + I_{10}.$$

By the linearity of expectation, we have

$$\mathbb{E}[X] = \mathbb{E}[I_1 + I_2 + \cdots + I_{10}] = 10\mathbb{P}(\text{getting a head with one flip}) = 5.$$

One can see that sometimes it becomes easier to use indicators in such "count the cases" scenarios.

1.2 Variance (Varians) and Covaraince (Kovarians)

Definition of variance *The variance of a random variable X is a measure of how spread out it is, or, how far the values of X are from their mean, on average.*

Definition. *Let X be a random variable and $\mathbb{E}[X] = \mu$. The variance of X is*

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

The variance is the mean squared deviation of a random variable from its own mean. High variance means we commonly observe values of X a long way from the mean value. If X has low variance means the values of X clustered tightly around their mean.

Theorem. *Significantly more convenient definition*

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Can you show this for continuous random variables? This definition is in particular far easier to program a computer to calculate this (no need to save a record of all divergences). Hence it is the most usually encountered definition.

Standard Deviation (Standardavvikelse): *Standard deviation* σ_X *of a random variable* X *is defined to be the square root of* $\mathrm{Var}(X)$*.*

Covariance *Covariance is a measure of the association or dependence between two random variables* X *and* Y *. Covariance can be either positive or negative. (Variance is always positive.)*

Definition. *Let* X *and* Y *be any random variables, the* <u>*covariance*</u> *between* X *and* Y *is given by*

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

where $\mu_X = \mathbb{E}[X]$, $\mu_Y = \mathbb{E}[Y]$.

- $\mathrm{Cov}(X, Y)$ *will be positive if values of* X *and values of* Y *move towards the same direction, for example, height and weight of a person.*
- $\mathrm{Cov}(X, Y)$ *will be negative if values of* X *and values of* Y *move towards the opposite direction, for example, the time you spend playing video games and your grade.*
- *If* X *and* Y *are independent, we have* $\mathrm{Cov}(X, Y) = 0$*. However, the converse is not true.* $\mathrm{Cov}(X, Y) = 0$ *does NOT imply* X *and* Y *are independent. (How so? Check the definition and think.).*

Properties of variance

- *Let* g *be a function, and let* a *and* b *be constants, for a random variable* X *we have*

  $$\mathrm{Var}(ag(X) + b) = a^2\mathrm{Var}(g(X)).$$

  In particular,

  $$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X).$$

- *Let* X *and* Y *be* <u>*independent*</u> *random variables, then*

  $$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y).$$

For arbitrary random variables X and Y,

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).$$

Remark. *Common mistakes among students are, for example, either*

$$\text{Var}(2X-Y) = 2\text{Var}(X) + \text{Var}(Y),$$

or

$$\text{Var}(2X-Y) = 4\text{Var}(X) - \text{Var}(Y).$$

2 Probability calculation and some applications

Remark. *Do the exercises 420-423 @JR by yourself using what you learned from Module 3. Note that there are mainly two types of problems in probability calculation:*

- *Find certain probability for given events,*
- *Find certain threshold for given probability.*

See the first two questions in Quiz.

At this point I'm sure you know how to find a distribution of a given random variable, let us now practice how to do it, by looking at how to prove the following (important) theorems.

Theorem. *Let X_1, X_2 be two independent binomial random variables with $X_1 \sim Bin(n_1, p)$ and $X_2 \sim Bin(n_2, p)$, define random variable*

$$Y = X_1 + X_2,$$

then

$$Y \sim Bin(n_1 + n_2, p).$$

Note that two important assumptions in this theorem are independence and same p parameter.

Theorem. *Let X_1, X_2 be two independent Poisson random variables with $X_1 \sim Po(m_1)$ and $X_2 \sim Po(m_2)$, define random variable*

$$Y = X_1 + X_2,$$

then

$$Y \sim Po(m_1 + m_2).$$

Note that the important assumption in this theorem is independence.
Now let us try to prove these two theorems. First the binomial distribution:

Proof. Let $0 \leq k \leq n_1 + n_2$, we have

$$\begin{aligned}
\mathbb{P}(Y = k) = \mathbb{P}(X_1 + X_2 = k) &= \sum_{i=0}^{k} \mathbb{P}(X_1 = i, X_2 = k - i) \\
&= \sum_{i=0}^{k} \mathbb{P}(X_1 = i)\mathbb{P}(X_2 = k - i) \quad \text{(by independence)} \\
&= \sum_{i=0}^{k} \binom{n_1}{i} p^i (1-p)^{n_1 - i} \binom{n_2}{k-i} p^{k-i} (1-p)^{n_2 - k + i} \\
&= p^k (1-p)^{n_1 + n_2 - k} \sum_{i=0}^{k} \binom{n_1}{i} \binom{n_2}{k-i} \\
&= \binom{n_1 + n_2}{k} p^k (1-p)^{n_1 + n_2 - k},
\end{aligned}$$

which shows that

$$Y \sim Bin(n_1 + n_2, p).$$

□

Similarly, let us try the second theorem for the Poisson distribution:

Proof. Let $0 \leq k \leq m_1 + m_2$, we have

$$\begin{aligned}
\mathbb{P}(Y = k) = \mathbb{P}(X_1 + X_2 = k) &= \sum_{i=0}^{k} \mathbb{P}(X_1 = i, X_2 = k - i) \\
&= \sum_{i=0}^{k} \mathbb{P}(X_1 = i)\mathbb{P}(X_2 = k - i) \quad \text{(by independence)} \\
&= \sum_{i=0}^{k} \frac{e^{-m_1} m_1^i}{i!} \frac{e^{-m_2} m_2^{k-i}}{(k-i)!} \\
&= e^{-m_1} e^{-m_2} \sum_{i=0}^{k} \frac{m_1^i}{i!} \frac{m_2^{k-i}}{(k-i)!} \frac{k!}{k!} \\
&= \frac{e^{-m_1} e^{-m_2}}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} m_1^i m_2^{k-i} \\
&= \frac{e^{-(m_1+m_2)}}{k!} \sum_{i=0}^{k} \binom{k}{i} m_1^i m_2^{k-i} \\
&= \frac{e^{-(m_1+m_2)}}{k!} (m_1 + m_2)^k,
\end{aligned}$$

which is the pmf (probability mass function) of a Poisson distribution with parameter $m_1 + m_2$. $\square$

2.1 An Important Remark

Now that we have showed that "Sum of two independent Poissons is still a Poisson", and "Sum of two independent Binomials with same p is still a Binomial". One might ask the following question:

Does the sum of two independent random variables within same distribution family stay in the same family?

or even

If I want the distribution of $X + Y$ do I just sum up the paramters?

The answer is **No**.
How so? Let us look at the simplest example.

Example. *Colin rolls two fair dice, and let X_1 and X_2 denote the numbers he gets respectively. Clearly X_1 and X_2 are independent and have the uniform distribution within $\{1, 2, 3, 4, 5, 6\}$. What about $X_1 + X_2$, is it still a uniform distribution?*

We write down the possible outcomes of $X_1 + X_2$ in the following table:

$X_1 + X_2$	$X_2 = 1$	2	3	4	5	6
$X_1 = 1$	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

What do you think? Is this a uniform distribution? Does number 8 have the same chance of appearing compared to number 12?

We see in the above example that even for the simplest uniform distribution the statement does not hold. In the later chapter we will also see that sum of two normal random variables is still normal. But that along with the theorems we see in this chapter are very special examples, one cannot generalise them based on intuition.

Fun fact: A direct application of the above counter example is playing games. For example *Dungeons and Dragons*, where sometimes one needs to throw a 20-sided die (d20). Is it ok to substitute it with two 10-sided die (d10) if you don't have a d20 at hand? The answer is no (unless you want to cheat), because with two d10's you will not have a uniform distribution from 1 to 20. (Another way to see this is that with two d10's it is impossible to get a 1). However, you can use a coin (d2) and a d10 to substitute a d20. How does it work?