UPPSALA UNIVERSITET

Föreläsningsanteckningar

Sannolikhetsteori 2

Rami Abou Zahra

Inlämningsdatum
September 6, 2023

Contents

1. Repetition

Anmärkning: Det rekommenderas starkt att läsa igenom anteckningarna från Sannolikhetsteori 1

Definition/Sats 1.1: Random trial

An event is not certain, it usually has a probability associated with it. Taking that "risk" to see what the outcome is, is called a random trial.

Examples of random trials include throwing dice, picking cards, number of people who pass a road

Different possibilities (outcomes).
In the example of the dice, the outcomes are 1-6

Definition/Sats 1.2: Events

An event is something that happens (or does not happen) when you the random trial

You can have an event based on one outcome, or multiple.

Example (one outcome): The dice is 3 after a throw

Example (several outcomes): The card is 7 or lower (1,2,3,4,5,6, all the different colours)

Example (0 outcomes): The card shows both spades and hearts at the same time (impossible)

1.1. Probability measure.

Related to the probability that an event occurs.

Definition/Sats 1.3: Probability measure

A *probability measure* is a function which satisfies Kolmogorovs axioms and for each event gives a number $\in [0,1]$

The number is called the *probability* of the event. Usually denoted $P = P(A)$ where $A \subset \Omega$

Definition/Sats 1.4: Kolmogorovs axioms

Let $P : 2^\Omega \to \mathbb{R}$. P is called a *probability measure* if it satisfies the following

- $P(A) \geq 0 \quad \forall A \in 2^\Omega$
- $P(\Omega) = 1$
- $P(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty P(A_i) \quad$ (A_i disjoint)

1.2. **Conditional probability theory.**

One can think of conditional probability in the term of Venn Diagrams in order to create intuition. Usually, if one event has happened it will affect other events and it is of interest to take this into consideration when calculating the probability of events.

The probability of an event A occuring given that the event B has occured is denoted by

$$P(A|B)$$

Example:
Let A be the event that a person has 2 daughters, let B be the event that a person has 0 daughters, and C be the event that he has at leat 1 daughter.

The probability $P(A|B)$ is of course 0, since given that he has 0 daughters, the probability is 0 for him to have 2 at the same time as he has 0

$P(B|C)$ is also 0, using similar argument as above

$P(A|C) = \dfrac{P(A \cap G)}{P(B)}$ We cannot say much here, other than that the probability is strictly positive since we already have one child

Definition/Sats 1.5: Bayes theorem

Let $F_1, \cdots, F_n$ be disjoint events $\in \Omega$ with $P(F_i) > 0$, and $P(\bigcup F_i) = 1$.

$$P(F_j|E) = \frac{P(E|F_j) \cdot P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

Example:
Suppose we have 3 different cards. The first card is red on both sides (RR), the second card is black on both sides (BB), and the third card is black and red (RB)

We draw a card at random of these three cards such that we only see one side of the card.
Now suppose the side we see is red, what is the probability that the other side is black?

We are interested in the event $P(RB|R)$:

$$\frac{P(RB \cap R)}{P(R)} = \overbrace{=}^{\text{Bayes}} = \frac{P(R|RB)P(RB)}{P(R|RR)P(RR) + P(R|RB)P(RB) + P(R|BB)P(BB)}$$

$$= \frac{(1/2)(1/3)}{1 \cdot (1/3) + (1/2) \cdot (1/3) + (0) \cdot (1/3)} = \frac{1}{3}$$

1.3. **Independent events.**

Definition/Sats 1.6: Independent events

If $P(A|B) = P(A)$, then A and B are independent

Example:
Let A be the event that 2 parents get a daughter, and B be the event that the neighbors child ate an ice cream yesterday.
Since these events do not affect each other, they are independent.

Example:

Let A be the event that the first throw of a dice yields 6, and let B be the event that the second throw is 3. Then A and B are independent since the first throw does not affect the second throw:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

Anmärkning:
There is an equivalent definition for independance through the following:

$$P(A \cap B) = P(A)P(B)$$

Anmärkning:
Independance is a symetric relationship.

1.4. **Random variables.**

Definition/Sats 1.7: Random variable

A *random variable* is a function that for each outcome associates a number with it.

An example is a persons age, or the value of a card drawn. If the outcome is random, the number is also random.

Each random variable has a distribution function associated with it, and is defined as $F(X) = P(X \leq x)$

Anmärkning:

$$\lim_{X \to -\infty} F(X) = 0$$
$$\lim_{X \to \infty} F(X) = 1$$

If $X_1 < X_2 \Rightarrow F(X_1) \leq F(X_2)$

We also have that F is right-continuous, meaning

$$\lim_{X \to a^+} F(X) = F(a)$$

There are 2 types of random variables that will be covered in this course, discrete and continuous (there is also absolutely continuous random variables, but they will not be covered)

1.4.1. *Discrete random variables.*

Definition/Sats 1.8: Discrete random variables

Consists of a finite or countable infinite set of numbers with probabilities:

- $P(X = x_i) = P(x_i) > 0$
- $P(X = \bigcup_{i=1}^{\infty} x_i) = 1$

Anmärkning:
If we have an uncountable infinite set of possibilities, the probability would be 0. Here is where continuous variables come to play

1.4.2. *Continuous random variables.*

For a continuous random variable, $F(x)$ is differentiable so that there exists a function f such that:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

From this comes 2 important things we can derive (both from the discrete and continuous case), namely expected value and the variance

Definition/Sats 1.9: Expected value

For discrete random variables, it is defined as

$$E(X) = \sum xF(x_i)$$

For the continuous case:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Definition/Sats 1.10: Variance

$Var(X) = E(X - E(X))^2$ for both discrete and continuous random variables

An equivalent definition is $E(X^2) - (E(X))^2$

Anmärkning:
$E(X^2)$ is called the second moment

Anmärkning:
If the variance is small, we know that the random variable does not fluctuate a lot from the expected value.

If X is a random variable (r.v) with density function f and g is a function, then we can define a new random variable $g(X) = Y$

Y is a random variable with density function $\hat{f}$.
Then:

$$E(Y) = \int y\hat{f}(y)dy = E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

And for the discrete random variable we have:

$$E(g(X)) = \sum g(x_i)p(x_i)$$

Anmärkning:
It is often better to use the definition of the density function for X rather than Y

Another remark worth noting is that $E(X^2)$ is a special case of $\int g(x)f(x)dx$

Example:
This example considers a distribution with no expected value (∞), and therefore it has no variance.
$P(X = k) = \dfrac{1}{k} - \dfrac{1}{k+1}$, this fulfills Kolmogorovs axioms, and

$$E(X) = \sum_{k=1}^{\infty} \frac{k}{k} - \frac{k}{k+1} = \sum_{k=1}^{\infty} \left(1 - \frac{k}{k+1}\right) = \sum_{k=1}^{\infty} \frac{1}{k+1} = \infty$$

2. Transformations

2.1. Pre-knowledge.

Let X_i be independent random variables with the same mean value (expected value) μ and variance σ^2

Let $S_n = \sum_{i=1}^{n} X_i \quad \overbrace{\Longrightarrow}^{\text{Law of large numbers}} \quad \frac{S_n}{n} \to \mu$ (convergance in probability).

Of course, this is assuming some sort of equal distribution.

The notation for convergance in probability is denoted by $Y_n \xrightarrow{P} a$. This follows from Markovs inequality. It is strongly suggested to look in the notes from the first course here.

From the law of large numbers, $S_n \approx n\mu$. But this does not take into account some erros that may take place in the $\frac{S_n}{n}$ side, as this does not affect the convergance to μ.

This is treated with the *Central Limit Theorem* (CLT), which says that $S_n \sim N(n\mu, n\sigma^2)$

This is equivalent to say that:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \approx N(0,1)$$

When n is large.

Anmärkning:
Here we talked about *convergence in distribution*

3. Multivariate random variables

It is strongly suggested to recall the random n-dimensional vector from Probability theory 1.

It is interesting to look at that the distribution of the vector, but we are often interested in a function $g(X)$

Example:
Starting with 1-dimension, and then working our way up.
Let $Y = g(X)$. Suppose g is strictly increasing of X (larger values of $X \to g(X)$ is larger).

Then

$$Y \leq y \in \mathbb{R} \Leftrightarrow g(X) \leq y \in \mathbb{R} \Leftrightarrow X \leq g^{-1}(y) = h(y)$$

If we look at the distribution function (which gives us everything, the probability the everything):

$$F_Y(y) = P(X \leq g^{-1}(y)) = P(X \leq h(y)) = F_X(h(y))$$

Anmärkning:
The inverse function g^{-1} is denoted by h

From the chain rule for derivates we get:

$$f_Y(y) = f_X(h(y)) \cdot h'(y)$$

We have gotten some information about X from Y. We can of course do the same for strictly decreasing functions:

$$Y \leq y \Leftrightarrow X \geq h(y)$$

$$\Rightarrow F_Y(y) = P(X \geq h(y)) = 1 - F_X(h(y))$$

This is good since we know that density functions are always positive. Taking the derivative gives us:

$$f_Y(y) = -f_X(h(y)) \cdot h'(y)$$

Anmärkning:
We assume g is differentiable with an inverse.

We showed that $f_Y(y) = f_X(h(y)) \cdot |h'(y)|$

3.1. **Multivariate case.**

Think of two n-dimensional space. One for $X = (X_1, X_2, \cdots, X_n)$, and one for $Y = (Y_1, Y_2, \cdots, Y_n)$

Suppose g is a bijective function such that g and g^{-1} are differentiable and let $Y = g(X) = (g_1(X), g_2(X), \cdots, g_n(X))$ where the component $Y_i = g_i(X = (X_1, X$

We have $X = g^{-1}(Y) = h(Y)$ (same as in 1-dimensional case)

Definition/Sats 3.11: Transformation theorem

The density of Y is given by

$$f_Y(y_1, y_2, \cdots, y_n) = f_X(h_1(y), h_2(y), \cdots, h_n(y)) \cdot |J|$$

Where J is the Jacobian matrix

$$J = \left|\frac{d(x)}{d(y)}\right| = \begin{vmatrix} \frac{dx_1}{dy_1} & \cdots & \frac{dx_1}{y_n} \\ \vdots & \vdots & \vdots \\ \frac{dx_n}{dy_1} & \cdots & \frac{dx_n}{dy_n} \end{vmatrix}$$

Anmärkning:
Transformation theorem corresponds to multivariate analysis change of variables

Bevis 3.1: Sketch of Transformation theorem

Let y_0 be a point in the Y-space. Choose an ε-ball C around y_0. Then we can assume that f_Y is constant in C.

The probability that our random vector Y will happen in this region is given by

$$\Delta C \cdot (f_Y(y_0) - \varepsilon) \leq P(Y \in C) \leq \Delta C \cdot (f_Y(y_0) + \varepsilon)$$

Anmärkning: ΔC is the volume/area of C

In the X-space, there then is a region which consists of all x whose $g(x)$ belongs to C.
Since g is bijective $\Rightarrow$ injective, we have that $Y \in C \Leftrightarrow X \in D = g^{-1}(C)$

This means that these probabilities are the same

$$\left| f_Y(y_0) \cdot \Delta C - \int_D f_X(x)dx \right| \leq \Delta C \cdot \varepsilon$$

As C decreases, ΔC decreases as well as ε
Since g is a nice function, D will also decreases
We let $x_0 = g^{-1}(y_0) = h(y_0)$. We can replace the integral by $f_X(x_0) \cdot \Delta D$ and obtain

$$f_Y(y_0) \cdot \Delta C \approx f_X(x_0) \cdot \Delta D \Leftrightarrow f_Y(y_0) \approx f_X(x_0) \frac{\Delta D}{\Delta C}$$

We get equality when $C \to 0$ (choosing a smaller and smaller ε)

Recall the functional determinant (Jacobian) of the matrix $\frac{\Delta x}{\Delta y} = \left|\frac{d(x)}{d(y)}\right|$ (relative volume change)

Thus, we get $f_Y(y_0) = f_X(h(y_0)) \, ||J_n(x_0, y_0)||$
Since this is true for all y, we can take away the index y_0, and we get:

$$f_Y(y) = f_X(h(y)) \cdot |J|$$

$\square$

Example (1-dim case):
Suppose $g(X) = aX + b$. From this it is easy to see what the inverse function $h = g^{-1}$ is, namely $h(y) = \frac{y-b}{a}$.

The Jacobian is just $\frac{1}{a}$. By the transformation theorem we get

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \left|\frac{1}{a}\right|$$

The main thing is that using the density function of X we can get the density function for Y.

Example 2.4:
X, Y are independent normally distributed random variables $N(0,1)$, show that $X+Y$ and $X-Y$ are independent and determine their distribution function.

In order to solve this we do a variable substitution. Let $U = X + Y$ and $V = X - Y$.
Notice that $\frac{U+V}{2} = X$ and $\frac{U-V}{2} = Y$
We have our function $g(x,y) = (u,v) = (x+y, x-y)$
We have our inverse $g^{-1}(u,v) = (x,y) = \left(\frac{u+v}{2}, \frac{u-v}{2}\right)$
We can now use the transformation theorem:

$$f_{U,V}(u,v) = f_{X,Y}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \cdot |J|$$

The Jacobian can be found:

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = \frac{-1}{4} - \frac{1}{4} = \frac{-1}{2}$$

By the transformation theorem:

$$f_{U,V}(u,v) = f_{X,Y}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \cdot \frac{1}{2} \overset{\text{indep.}}{=} f_X\left(\frac{u+v}{2}\right) f_Y\left(\frac{u-v}{2}\right) \cdot \frac{1}{2}$$

We know their density functions since they are normally disitrbuted with $N(0,1)$:

$$= \frac{1}{\sqrt{2\pi}} e^{-(1/2)\left(\frac{u+v}{2}\right)^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(1/2)\left(\frac{u-v}{2}\right)^2} \cdot \frac{1}{2}$$

After simplification we get that $f_{U,V}$ is a product of one function of U and one function of V. This means that U, V are independent since we get them as a product of two different functions.

Example:
Recall the convolution formula from Probability theory 1; if X, Y are independent, we have

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

We can show this from the transformation theorem.

By the independance of X, Y, we have $f_{X,Y}(x,y) = f_X(x) f_Y(y) \forall x, y$
Let $Z = X + Y$. Then $g(X,Y) = (X+Y, X)$
The inverse is given by $Y = Z - X$. We can now use the transformation theorem:

$$f_{Z,Y}(z,x) = f_{X,Y}(h_1(z,x), h_2(z,x)) \cdot |J|$$

The Jacobian is given by

$$\begin{vmatrix} 1 & -1 \\ 1 & 0 \end{vmatrix} = 1$$

Since X, Y are independent, we get:

$$f_{Z,X}(z,x) = f_Y(z-x) f_X(x)$$

The marginal density is given by integrating away x:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,X}(z,x)dx = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)$$

3.2. **Conditional Probabilities.**

It is suggested to do some examples from Probability theory 1.