STAT1301 - Lecture Notes

Ismael Khan

1 Lecture 1

2 Lecture 2

3 Lecture 3

4 Lecture 4

5 Lecture 5

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum\limits_{i=1}^{n} x_i}{n}$$

The value $b = \bar{x}$ minimizes the squared deviations from the data:

$$\sum_{i=1}^{n} (x_i - b)^2$$

Prove this Observe that

$$g(b) = (170 - b)^2 + (182 - b)^2 + (160 - b)^2$$
$$g(b) = 3b^2 - \beta b - \gamma b -$$

We can take the derivative to find the minimum

$$\begin{aligned} g(b) &= \sum_{i=1}^{n} (x_i - b)^2 \\ g'(b) &= \sum_{i=1}^{n} 2(x_i - b) \cdot -1 = 0 \\ &= test \end{aligned}$$

5.1 Robustness

What would happen if the third student had given their value in metres? Outliers can havea big effect on the sample mean. The median is also more informative for skewed data. Thus visualising is important before using means.

5.2 Sample Variance

Consider the three height values again: 170cm, 185cm, 162cm. How do we measure the spread of thse values?

One way to measure spread is via the sample variance:

$$\beta = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

5.3 Sample standard deviation

Note that the sample variance is on the squared scale of the original observations. To get a measure of spread on the original scale, we use the function

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Take an average height of people to be 170cm. A person of height 182cm leaves the room. The mean height will decrease. We can show this.

$$\frac{x_1 + x_2 + x_3 + ... + x_n}{n} = 170$$

Without loss of generality, that $x_1 = 180$, then

$$\frac{x_2 + ... + x_n}{n} = \frac{x_1 + x_2 + ... + x_n - x_1}{n-1}$$

$$\frac{n \times 170 - 180}{n-1} < \frac{n \times 170 - 170}{n-1} = \frac{170(n-1)}{n-1} = 170$$

5.4 Uniform random numbers

6 Lecture 6

6.1 Modelling Relationships

We often model relationships as a tend in the mean response plus variability about that trend.

$$\gamma(x) = g(x) + \varepsilon$$

For γ being the response, x being the explnation variable and εbeing the random error

6.2 Describing Relations

How much would descrie the relationship between the father and son?

6.3 Pearson Correlation

The Pearson correlation coefficient r, measures the strength of a linear relationship. If the points in the scatter plot are (x_1, y_1), (x_2, y_2), (x_n, y_n) then

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

It is important to remember that r is only appropriate for linear relationships between variables.

6.4 Least-Squared Lines

Once we have deterimend that a straight line relationship may be appropriate we can go ahead and fit a line of best fit to the data. For any given line $b_0 + b_1 x$, how to judge how well it fits the data? We can look at the sum of the squared prediction errosr (or sum of squared deviations)

$$\sum_{i=1}^{n}(y_i - [b_0 + b_1 x_i])^2$$

The line $b_0 + b_1 x_1$, that minimises this is called the least-squares line.

7 Lecture 7

8 Lecture 8

8.1 Simple Probability Models

The simplest probability model is when the sample spaces is a finite collection of outcomes, all being equally likely.

Example 1 (Rolling a Die)**.** Consider rolling a fair 6 sided die. What is Ω?,

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Then what is the probability $\mathbb{P}$. For a given $A \subset \Omega$, we assign the probability $\mathbb{P}(A)$ as

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},\ A \subset \Omega$$

Thus for $|\Omega| = 6$,

$$\mathbb{P}(A) = \frac{|A|}{6},\ A \subset \Omega$$

Drawing balls from an urn. The Urn can be counted in 4 possible urn experiments

1. Ordered, with replacement (Notes missing)
2. Ordered, without replacement
3. Unordered, without replacement

   Consider a horse race with 8 horses. How many ways are there to gamble on placings (1st,2nd,3rd).
4. Unordered, with replacement

Theorem 1. *Let $X \sim U[a, b]$. Then,*

1. $\mathbb{E}(X) = \dfrac{(a+b)}{2}$
2. $Var(X) = \dfrac{(b-a)^2}{12}$

9 Lecture 9 - Conditional Probability and Independence

9.1 Conditional Probability

How do we update the probability of an event A when we know that some other event B has occured? If we are given that B has occured, then A will occur if and only if $A \cap B$ occurs.

The relative chance of A occuring given B has occured is therefore $\dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}}$. Which is the *conditional probability* of A given B.

Definition 1. The **condtional probability** of A given B is defined as.

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

Example 2. Suppose we have rolled a fair six-sided die. What is the conditional probability that we get a "4" given we know that we rolled an *even number.*

Intuition implies that the probability is $\dfrac{1}{3}$. We can formally show this however.

Let B be getting an even number $= \{2, 4, 6\}$ and A to be getting a $4 = \{4\}$. We already know that $\mathbb{P}(B) = \frac{3}{6}$. Now $A \cap B = \{4\}$, so that $\mathbb{P}(A \cap B) = \frac{1}{6}$. Thus,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$$

9.2 Product Rule

Rearranging the definition of the conditional probability gives us an expression for the product rule.

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A \mid B) = \mathbb{P}(A)\mathbb{P}(B \mid A)$$

More generally,

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) \ldots \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1})$$

The crucial usage of this is that in many cases conditional probabilities are easy to figure out.

Example 3. We draw consecutively 3 balls rom an urn with 5 white and 5 black balls, without putting them back. What is the probability that all drawn balls will be black? Let A_i be the event that the i-th ball is black. We wish to find the probability of $A_1 \cap A_2 \cap A_3$ (also written as $A_1 A_2 A_3$), which by the product rule is

$$\mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) = \frac{5}{10}\frac{4}{9}\frac{3}{8} \approx 0.083$$

9.3 Independence

If the join probability of two events A and B happens to factorise into the product of the two individual probabilities, i.e.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Then we say that the events A and B are independent. In other words, knowledge of A gives no additional knowledge to B and vice versa. This implies that

$$\mathbb{P}(A \mid B) = \mathbb{P}(A)$$

and

$$\mathbb{P}(B \mid A) = \mathbb{P}(B)$$

Example 4. Example 3 is used however 3 balls are chosen **with** putting them back. So by independence,

$$\mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) = \frac{5}{10}\frac{5}{10}\frac{5}{10} \approx 0.125$$

10 Lecture 10 - Random Variables

A random variable canbe viewed as measurement of a random experiemnt that becomes available **tommorow**. However all the thinking about the measurements can be carried out **today**. We denote random variables with capital letters X, X_1, X_2, Y, Z etc.

Example 5. Some random variables may be

1. THe number of defective transistors out of 100 inspected ones.
2. The number of bugs in a computer program
3. The amount of r ain in a certain location in June.
4. The amount of time needed for an operation.

We distinguish between discrete and continuous random variables:

- **Discrete** random variables can only take *countably many* values.
- **Continuous** random variables can take a *continuous range* of values, for example, any value on the positive real line $\mathbb{R}_+$.

10.1 Probability distrubition

Let X be a random variable. We would like to designate the probabilities of events such as $\{X = x\}$ and $\{a \leq X \leq b\}$. If we can specify all probabilities involving X, we say that we have determined the probability distrubution of X.

A way to specifiy the P.D is to give all probabilties of the form $\{X \leq X\}$.

Definition 2. The **cumulative distribution function** (cdf) of a random variable X is the funciton F defined by

$$F(x) = \mathbb{P}(X \leq x),\ x \in \mathbb{R}$$

10.2 Cumulative distribution function (cdf)

Note that any cdf is increasing (if $x \leq y$ then $F(x) \leq F(y)$ and lies inbetween 0 and 1. We can use any function F with these properties to specify the distrubution of a random variable X

10.3 Probability mass function

Definition 3. A random variable X is said to have a discrete distrubution if $\mathbb{P}(X = x_i) > 0,\ i = 1, 2...$ for some finite or countable set of values $x_1, x_2...$ such that $\sum_i \mathbb{P}(X = x_i) = 1$. The **probability mass function (pmf)** of X is the function f defined by $f(x) = \mathbb{P}(X = x)$.

By the sum rule, if we know $f(x)\ \forall x$ then we can calculate all possible probabilties involving X

$$\mathbb{P}(X \in B) = \sum_{x \in B} f(x)$$

Note. $\{X \in B\}$ should be read as X is an elemnt of B

Definition 4. A random variable X with cdf F is said to have a continuous distrubution if there exists a positive function f with total integral I such that $\forall a < b$

$$\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_a^b f(u)\ du$$

Function f is called the **probability distribution function (pdf)** of X

10.4 Calculating probabilities

Once we know the pdf, we can calculate any probability taht X lies in some set B by means of integration

$$\mathbb{P}(X \in B) = \int_B f(x)\ dx$$

10.5 Relationship between cdf and pdf

Suppose f and F are the pdf and cdf of a continuous random variable X respectively. Then F is an anti-derivative of f

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(u)\ du$$

Conversely, f is the derivative of the cdf F

$$f(x) = \frac{d}{dx} F(x) = F'(x)$$

10.6 Density

In the continuous case, $f(x) \neq \mathbb{P}(X = x)$ because the latter is 0 $\forall x$. Instead we intepret $f(x)$ as the density of the probability at distrubution x, in the sense that for any small h.

$$\mathbb{P}(x \leq X \leq x + h) = \int_x^{x+h} f(u)\ du \approx h f(x)$$

Note. $\mathbb{P}(x \leq X \leq x + h) = \mathbb{P}(x < X \leq x + h)$ in this case.

Example 6. Draw a random number X from the interval of real numbers $[0, 2]$, where each number is equally likely to be drawn. What are the pdf and cdf F of X

We have

$$\mathbb{P}(X \leq x) = F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{x}{2} & \text{if } 0 \leq x \leq 2 \\ 1 & \text{if } x > 2 \end{cases}$$

By differentiating F we find

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Note that theis density is constant on the interval $[0, 2]$ and zero elsewhere. Reflecting the fact that each point in $[0, 2]$ are equally **equally likely to be drawn**.

10.7 Expectation (discrete case)

Although all probability info about a random variable is contained in its cdf or pmf/pdf, it is often useful to consider various numerical characteristics of a random variable.

Definition 5. Let X be a *discrete* random variable with pmf f. The **expectation** (or expected value) of X, denoted as $\mathbb{E}(X)$ is defined as

$$\mathbb{E}(X) = \sum_x x\mathbb{P}(X = x) = \sum_x xf(x)$$

The expectation is thus a "weighted average" of the values that X can take.

10.8 Expectation (continuous case)

Definition 6. Let X be a *continuous* random variable with pdf f, The **expectation** (or expected value) of X, denoted $\mathbb{E}X$, is defined as

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

10.9 Expectation of a function of a random variable

If X is a random variable, then a function of X such as X^2 or $\sin(X)$ is also a random variable.

Theorem 2 (Expectation of a Function of a Random Variable)**.** *If X is discrete with pdf f, then for any real values function g*

$$\mathbb{E}(g(X)) = \sum_x g(x)f(x)$$

Replace the integral with a sum for the discrete case

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\, dx$$

10.10 Linearity of expectation

An important consequence of Theorem 2 is that the expectation is "linear".

Theorem 3. *For any real numbers a and b, and functions g and h.*

1. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
2. $\mathbb{E}(g(X) + h(X)) = \mathbb{E}(g(X)) + \mathbb{E}(h(X))$

Proof of Theorem 3. We show it for the discrete case. Suppose X has pmf f. The first statement follows from

$$\mathbb{E}(aX + b) = \sum_x (ax + b)f(x) = a\sum_x xf(x) + b\sum_x f(x) = a\mathbb{E}(X) + b$$

Similarly, the second statement follows from

$$\mathbb{E}(g(X)+h(X)) = \sum_x (g(x)+h(x))f(x) = \sum_x h(x)f(x)$$

$$\mathbb{E}(g(X)) + \mathbb{E}(h(X))$$

□

10.11 Variance

Another useful characteristic of the distrubution of X is the variance of X. This number is sometimes denoted as σ^2 and measures the *spread* of the distrubution of X.

Definition 7. The **variance** of a random variable X, denoted $\mathrm{Var}(X)$, is defined as

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

where $\mu = \mathbb{E}(X)$. The square root of the variance is called the **standard deviation**. The number $\mathbb{E}X'$ is called the r-th **moment** of X

10.12 Properties of Variance

Theorem 4. *For any random variable X, the following properties hold for variance*

1. $Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$
2. $Var(a+bX) = b^2\, Var(X)$

10.13 Valuable Conclusions

- The probability distrubution of a random variable is completely specified by its cdf (cumulative distrubution function). For discrete random variables, it is more useful to specifiy distrubution with pmf (probability mass function). For continuous random variables, use pdf (probability density function)
- The expectation (expected value) of a random variable is the weighted average of the values that random variable can take. It measures the locality of the distrubution of a random variable.
- The variance is the expected squared distance from the random variable to its expected value. As a consequence, it is a measure of the spread of distrubution of a random variable.

Note. The lecture 11 (scheduled on a Wednesday) was cancelled due to the Ekka holiday. So all lectures proceeding from this point are one less then their *true* lecture number.

11 Lecture 11 - Bernoulli and binomial distribution

Test test test

12 Lecture 12 - Uniform and normal distribution

The simplest continuous distribution is the uniform distribution

Definition 8. A random variable X is said to have **uniform** distribution on the interval $[a,b]$ if its pdf is given by

$$f(x) = \frac{1}{b-a},\ a \le x \le b$$

(and $f(x) = 0$ otherwise). We write $X \sim U[a,b]$

12.1 Intepreteation

The random variable X $U[a,b]$ can model a randomly chosen point from the interval $[a,b]$, where each choice is equally likely.

Theorem 5. *Let $X \sim U[a,b]$. Then,*

1. $\mathbb{E}(X) = \frac{a+b}{2}$
2. $Var(X) = \frac{(b-a)^2}{12}$

12.2 Normal Distribution

We introduce the most important distribution in the study statistics: the normal (or Gaussian) distribution.

Definition 9. A random variable X is said to have a **normal** distribution with parameters μ (expectation) and σ^2 (variance) if its pdf is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \; x \in \mathbb{R}$$

Theorem 6. *Let $X \sim N(\mu, \sigma^2)$. Then*

1. $\mathbb{E}(X) = \sigma^2$
2. $Var(X) = \sigma^2$

if $\mu = 0$ and $\sigma = 1$, the distribution is called the **standard normal distribution**. Its pdf is denoted with φ

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}$$

The respective cdf is denoted with capital phi Φ

12.3 Standardisation

Theorem 7. *If Z has standard normal distribution, then $X = \mu + \sigma Z$ has a $N(\mu, \sigma^2)$ distribution. Consequently, if $X \sim N(\mu, \sigma^2)$ then the standardised random variable*

$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution.