Contents

Math 148 - Lecture Notes: Selected Topics

Version: **Draft 01**

This write-up contains notes about selected topics given for Math 148 in Fall 2015

1 Lecture Notes for Math 148 - Fall 2015

1.1 Math 148 - Probability

1.1.1 Conditional probabilities

Frequencies and product rule for conditional probabilities

(this was printed) and projected with the overhead projector).

Conditional probabilities $P(B|A) = P(B\textbf{given}A)$ as restricting the context from the overall sample space to A only: We throw away whatever is in that sample space but does not belong to A. This means that only B **and** A of the event of interest B survives within that new sample space A. We abbreviate

Symbol	Description
#(SSpc)	all outcomes for the entire sample space
#(A)	all outcomes for A
#(B)	all outcomes for B
#(A **and** B)	all possible outcomes for both A **and** B

We obtain, because we interpret the probability of an event as its relative frequency $\dfrac{\#(\text{favorable outcomes})}{\#(\text{possible outcomes})}$, the following:

$$
\begin{aligned}
P(A) &= \frac{\#(\text{outcomes in A})}{\#(\text{outcomes in sample space})} = \frac{\#(A)}{\#(\text{SSpc})}\\
P(B) &= \frac{\#(\text{outcomes in B})}{\#(\text{outcomes in sample space})} = \frac{\#(B)}{\#(\text{SSpc})}\\
P(A\textbf{ and }B) &= \frac{\#(\text{outcomes in both A }\textbf{and}\text{ B})}{\#(\text{outcomes in sample space})} = \frac{\#(A\textbf{ and }B)}{\#(\text{SSpc})}\\
P(B|A) &= P(B\textbf{ given }A) = \frac{\#(\text{outcomes in both A }\textbf{and}\text{ B})}{\#(\text{outcomes in new sample space A})} = \frac{\#(A\textbf{ and }B)}{\#(A)}
\end{aligned}
$$

Why that last equation? Answer: The denominator changes from #(SSpc) to #(A) because we restricted our context from the overall sample space to A and the numerator changes from #(B) to #(A **and** B) because we removed the outcomes that are not part of A, i.e., we restricted the favorable outcomes of B to those that also belong to A, i.e., we only consider A**and**B as favorable outcomes.

What now is $P(A)\cdot P(B|A)$? Answer:

$$P(A)\cdot P(B|A) = \frac{\#(A)}{\#(\text{SSpc})}\cdot\frac{\#(A\textbf{ and }B)}{\#(A)} = \frac{\#(A\textbf{ and }B)}{\#(\text{SSpc})} = P(A\textbf{ and }B)$$

Again: $\boxed{P(A\textbf{ and }B) = P(A)\cdot P(B|A)}$

2 Math 148 - Notation used in lecture but not in the FPP text

2.1 Notation - Lists

Given is a list of items $x_1, x_2, x_3, \dots, x_n$. We often write $\vec{x}$ as an abbreviation for such a list [1] , i.e.,

(2.1) $$\vec{x} = (x_1, x_2, x_3, \dots, x_n)$$

It is clear that the last index, here n, denotes for such a list its **size** or **length** There is nothing sacred about the letters x and n. You have seen when we talked about regression that two separate lists $\vec{x} = (x_1, x_2, x_3, \dots, x_n)$ and $\vec{y} = (y_1, y_2, y_3, \dots, y_n)$ of the same length were needed to describe the contents of a scatter diagram. We will also have occasion to look at different lists of different length, so you may see in lecture two separate lists

$$\vec{X} = (X_1, X_2, X_3, \dots, X_n) \quad\text{and}\quad \vec{b} = (b_1, b_2, b_3, \dots, b_k)$$

to describe a chance process $\vec{X}$ of n independent observations $X_1, X_2, \dots, X_n$ each of which will have one of the possible outcomes $b_1, b_2, \dots, b_k$ (the items listed in the "box" [2] $\vec{b}$) with equal probability $1/k$.

[1] That's preferable to writing x because we want to use x only in phrases like "let x be one of the items in the list $\vec{x}$" or as a subscript to an object that is associated with that list (example: SD_x, the standard deviation of the list - see further down in this section).

[2] The notation for chance processes and box models will be described more in detail later in this document.

Examples for lists:

a. If "H" stands for "Heads" and "T" stands for "Tails" then a possible list that describes the outcomes of 5 flips of a coin would be $\vec{t} = (H, H, T, H, T)$: $t_1 = H$, $t_2 = H$, $t_3 = T$, $t_4 = H$, $t_5 = T$. We chose "t" for "toss" as in toss of a coin. In this example the length of the list $\vec{t}$ is 5.

Lists of such qualitative, non-numeric, variables are very limited in terms of the computations that can be done with them.

b. Let us rewrite this list as follows: write 1 instead of H and 0 instead of T.

Now we get the list $\vec{x} = (1, 1, 0, 1, 0)$: $x_1 = 1$, $x_2 = 1$, $x_3 = 0$, $x_4 = 1$, $x_5 = 0$.

For such a list we can compute the sum $x_1 + x_2 + \cdots + x_n$ which is $1 + 1 + 0 + 1 + 0 = 3$ for the list $\vec{x}$ and gives use the number of heads in the original list $\vec{t}$ of heads and tails.

c. If the number of heads is the most important item of interest then we could do the following: create a new list which directly lists for each turn how many heads were obtained until this point in time: We create a new list ("s" for "sum")

$$\begin{aligned}
\vec{s} &= (s_1, s_2, s_3 \ldots, s_n) \quad \text{where} \\
s_1 &= x_1, \\
s_2 &= x_1 + x_2, \\
s_3 &= x_1 + x_2 + s_3, \\
&\ldots \\
s_n &= x_1 + x_2 + s_3 + \cdots + s_n.
\end{aligned}$$

In the example above which was based on five coin tosses: $s_1 = 1$, $s_2 = 1+1 = 2$, $s_3 = 1+1+0 = 2$, $s_4 = 1+1+0+1 = 3$, $s_5 = 1+1+0+1+0 = 3$.

2.2 Notation - "Σ" for Summation

Here is a refresher for the use of "Σ" [3] as a compact means of describing sums of numeric lists.

In section 2.1 on lists we encountered a numbers list

$$\vec{x} = (1, 1, 0, 1, 0): \quad x_1 = 1,\ x_2 = 1,\ x_3 = 0,\ x_4 = 1,\ x_5 = 0$$

and its corresponding list of sums

$$\begin{aligned}
\vec{s} &= (s_1, s_2, s_3, s_4, s_5) \quad \text{where} \\
s_1 &= x_1 \ = \ 1, \\
s_2 &= x_1 + x_2 \ = \ 2, \\
s_3 &= x_1 + x_2 + s_3 \ = \ 2, \\
s_4 &= x_1 + x_2 + s_3 + s_4 \ = \ 3, \\
s_5 &= x_1 + x_2 + s_3 + s_4 + s_5 \ = \ 3.
\end{aligned}$$

[3] For additional background check your highschool math books or tutorials on the internet.

Summation occurs so frequently in statistics that we want to have a shorter way to deal with a sum of 3 or 300 or any k numbers $x_1, x_2, \ldots, x_k$. We write

$$\sum_{j=1}^{k} x_j \text{ as a short for } x_1 + x_2 + x_3 + \cdots + x_k \tag{2.2}$$

We call j the **index variable** and k the **to-variable** or the **end-variable** of the summation. Note that if a variable such as k is used to denote the last item to be included in the sum then it will usually **but not always(!)** be the size of the list.

Let us use this new notation the example above:

$$\begin{aligned}
s_1 &= x_1 = \sum_{j=1}^{1} x_j = 1, \\
s_2 &= x_1 + x_2 = \sum_{j=1}^{2} x_j = 2, \\
s_3 &= x_1 + x_2 + s_3 = \sum_{j=1}^{3} x_j = 2, \\
s_4 &= x_1 + x_2 + s_3 + s_4 = \sum_{j=1}^{4} x_j = 3, \\
.s_5 &= x_1 + x_2 + s_3 + s_4 + s_5 = \sum_{j=1}^{5} x_j = 3.
\end{aligned}$$

Occasionally there are reasons to choose a start index different from 1. For example, we may be interested to sum up the items starting at the 10th turn and ending the 25th turn. We write

$$\sum_{j=10}^{25} x_j \text{ as a short for } x_{10} + x_{11} + x_{12} + \cdots + x_{25}.$$

Sometimes a list may have "start time" 0: $\vec{x} = (x_0, x_1, x_2, \ldots, x_n)$. We then write $\sum_{j=0}^{n} x_j$ for the sum of all its members.

The role of the index variable. In the above the name of the index variable j is irrelevant. You may choose any letter you want as long as it does not match the name of either start variable or end variable. The following all mean the same thing:

$$\sum_{j=4}^{8} x_j = \sum_{i=4}^{8} x_i = \sum_{z=4}^{8} x_z = \sum_{n=4}^{8} x_n = x_4 + x_5 + x_6 + x_7 + x_8$$

and

$$\sum_{j=1}^{n} x_j = \sum_{i=1}^{n} x_i = \sum_{z=1}^{n} x_z,$$

but it is **illegal** to write $\sum_{n=1}^{n} x_n$ because you may not mix up the index variable n with the end variable n in this summation expression.

Lazy ways to write the Σ notation: If there is absolutely no confusion about start index and end index (practically because start index = first index of the list (usually 1) and end index = last index of the list) then these may be dropped and the following each mean the same:

(2.3) For a list $\vec{x} = (x_a, x_{a+1}, x_{a+2}, \dots x_{n-1}, x_{n-1})$, $\sum x_j := \sum_j x_j := \sum_{j=a}^{n} x_j$.

2.3 Basic notation for lists

Given a numeric list $\vec{x} = (x_1, x_2, \dots, x_n)$ we write

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

for its **mean** or **average** and

$$\mathrm{SD}_x = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2}$$

(the root-mean-square of the differences $x_j - \bar{x}$) for the for its **standard deviation**.

Given a numeric list $\vec{x} = (x_1, x_2, \dots, x_n)$, we write

(2.4) $$\vec{\tilde{x}} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) \quad \text{where} \tilde{x}_j = \frac{x_j - \bar{x}}{\mathrm{SD}_x}$$

for the corresponding list of **standard units**.

2.4 Notation for regression

When doing regression for a scatter diagram, note that each observation is not a single number x_j but rather a pair of numbers (x_j, y_j). There are two ways to place the arrows when writing such a scatter diagram as a list:

$$\overrightarrow{(x, y)} := (\vec{x}, \vec{y}) := \big((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\big)$$

The **five point summary**. for such a scatter diagram is $boxed\bar{x}, \mathrm{SD}_x, \bar{y}, \mathrm{SD}_y, r$ where we compute the **correlation coefficient** r as follows:

a. Build associated lists $\vec{\tilde{x}}, \vec{\tilde{y}}$ of standard units:

$$\tilde{x}_j := \frac{x_j - \bar{x}}{\mathrm{SD}_x}, \qquad \tilde{y}_j := \frac{y_j - \bar{y}}{\mathrm{SD}_y}.$$

b. Create a new list $\vec{p}$ from the products $p_j := \tilde{x}_j \cdot \tilde{y}_j$.

c. Take the average of that list:

$$r := r_{x,y} := \bar{p} \ := \ \frac{p_1 + p_2 + \cdots + p_n}{n} \ = \ \frac{1}{n}\sum_{j=1}^{n}\Big(\frac{x_j - \bar{x}}{\mathrm{SD}_x} \cdot \frac{y_j - \bar{y}}{\mathrm{SD}_y}\Big).$$

SD line and regression line for a scatter diagram

Both lines go through the **point of averages**. with coordinates $(\bar{x}, \bar{y})$.

$$\text{The SD line has slope } m = \frac{\mathrm{SD}_y}{\mathrm{SD}_x} \quad \text{if } r > 0, \tag{2.5}$$

$$m = -\frac{\mathrm{SD}_y}{\mathrm{SD}_x} \quad \text{if } r < 0, \tag{2.6}$$

$$\text{The regression line has slope } m = r \cdot \frac{\mathrm{SD}_y}{\mathrm{SD}_x} \quad \text{always.} \tag{2.7}$$

The regression line $\hat{y} = mx + b$ (the slope m was just given above) has as its y-values the best possible estimates for the average of **those** y-values of the scatter diagram $(\vec{x}, \vec{y})$ **only** for which we freeze the x-value to, say, observation j: $x = x_j$. In other words, the y-value $\hat{y}$ on the regression line corresponding to the x -value x_j is the best estimate for $\bar{y}|x = x_j$, an expression which we read as "the mean $\bar{y}$ of the y-values under the condition that the x-values (there could be more than one) are restricted to x_j" or shorter, as **the conditional mean $\bar{y}$ of y given that** $x = x_j$.

Not on any test: The following on total, explained and unexplained variation.

Total variation: $\mathrm{var}_{\mathrm{tot}} = \sum_j (y_j - \bar{y})^2$. Note that $\mathrm{var}_{\mathrm{tot}}/n = \mathrm{SD}_y$.

Explained variation: $\mathrm{var}_{\mathrm{expl}} = \sum_j (\hat{y}_j - \bar{y})^2$.

unexplained variation: $\mathrm{var}_{\mathrm{unexpl}} = \sum_j (y_j - \hat{y}_j)^2$.

$$\boxed{\mathrm{var}_{\mathrm{tot}} = \mathrm{var}_{\mathrm{expl}} + \mathrm{var}_{\mathrm{unexpl}}}$$

What follows might be on a test:

$$s_{est} := \sqrt{\frac{1}{n}\sum_j (y_j - \hat{y}_j)^2} = \sqrt{1 - r^2} \cdot \mathrm{SD}_y = \text{r.m.s. error of the regression line} \tag{2.8}$$

$$= \text{standard error of the estimate} \tag{2.9}$$

2.5 Chance processes and box models

A **box model** is a probability model that behaves like drawing at random **with replacement** from a box with a finite list of tickets $\vec{b} = (b_1, b_2, \ldots, b_k)$. "At random" means that each ticket has the same chance of being drawn and "with replacement" guarantees that the situation is the same before each draw and that the result of the nth draw is independent of all previous and subsequent draws.

Examples:

a. Box for repeatedly tossing a coin: $\vec{b} = (0, 1)$

b. Box for repeatedly rolling a die: $\vec{b} = (1, 2, 3, 4, 5, 6)$

c. Box for counting the occurrence of either a 2 or a 5 when repeatedly rolling a die: $\vec{b} = (0, 1, 0, 0, 1, 0)$

d. Box for counting the net winnings when playing repeatedly Roulette and each time betting 1 on a specific number, say 13: obtain 35 if that number comes up and lose the dollar if it does not. The box has 38 tickets: on for each slot of the Roulette wheel: $\vec{b} = (b_1, b_2, \ldots b_38)$ where 37 of the b_j have a value of -1 (losing a dollar) and only one has a value of $+35$. It does not matter whether #1 or #38 or a ticket inbetween is the winning ticket because remember: the tickets are well shuffled so that each one has the same chance of being drawn.

A box $\vec{b} = (b_1, b_2, \ldots, b_k)$ is just a list of data.

A **chance process** or **random process** is a finite or infinite list ("sequence") of observations which are subject to chance (they are not deterministic and the outcomes may vary if you do the same sequence of actions at another time).

Examples:

a. you toss a coin repeatedly and observe Heads or Tails after each toss.

b. After each such toss (example a) you note how often Heads came up until now.

c. You play Roulette repeatedly, always betting 1 on Red and keep track of your net winnings when Red lets you win a dollar and you otherwise lose a dollar.

d. you randomly select people at an airport and write down their weight.

If you set out tomorrow to do the same set of actions again then it is most unlikely that you end up with exactly the same list of data as you did today: Today your first five tosses may be (H, H, T, H, T), tomorrow you may obtain (H, T, T, T, H) so that you also get a different value in (b) after the fifth toss. Today you may have had a net gain of 4 after 100 games, tomorrow you may loose 9 instead.

We write capital letters such as $\vec{X} = (X_1, X_2, \ldots,)$, $\vec{Y} = (Y_1, Y_2, \ldots,)$, $\vec{S} = (S_1, S_2, \ldots,)$, to describe a chance process.

Do not confuse the chance process $\vec{X} = (X_1, X_2, \ldots,)$ with the specific list of outcomes (data) $\vec{x} = (x_1, x_2, \ldots, x_n)$ you obtain if you let this process happen now and which probably will be quite different if that process happens again tomorrow.

Example c (Roulette) above: do not confuse the process $\vec{S} = (S_1, S_2, \ldots,)$ in which S_j is the difference between dollars won and dollars lost when betting j times a dollar on Red with the list of data $\vec{s} = (s_1, s_2, \dots, s_n)$ that you obtain when actually play Roulette today n times and you may get, e.g., s_1 = net gain after first game S_1 = 1 (Red came up by chance), s_2 = net gain after second game S_2 = 2 (Red came up again by chance), s_n = net gain after nth game S_n = -5 (after the last game in which you particpated today).

In other words, a chance process is a described by a set of rules that describe how a concrete list of data will be obtained by invoking or running that chance process. You might say that it is a potential list of data which will become an actual list of data every time it is run.

Important assumption:
For any chance process $\vec{X} = (X_1, X_2, \dots,)$ we assume that it behaves as if there was a box model: a box $\vec{b} = (b_1, b_2, \dots, b_k)$ of k tickets with value b_i on ticket i ($i = 1, 2, \dots, k$), each of which has the same chance $1/k$ of being drawn and that the possible outcomes of anyone of the X_j ($j = 1, 2, \dots$) is one of those ticket values.

Under this assumption we may expect that any one of the chance process outcomes X_1, X_2 may be expected to equal, on average, the box average $\bar{b}$. and that the expected value of the corresponding summation process $\vec{S} = (S_1, S_2, \dots,)$ ($S_n = X_1 + X_2 + \cdots + X_n$) will be

(2.10) $$E(S_n) = \text{expected value of } S_n = n \cdot \bar{b} = \text{\# of draws} \times \text{average of box}$$

There will also be a spread of the sum $S_n = \sum_{j=1}^n X_j$ about its expected value $E(S_n)$, called the **standard error** SE_S which can becomputed with help of the

square root law:

(2.11) $$\text{SE}(S) = \text{standard error of } S_n = \sqrt{n} \cdot \text{SE}(b) = \sqrt{\text{\# of draws}} \times \text{SD of box}$$

TO BE CONTINUED ....

References

Index