Lecture 13 Worksheet

Chrysafis Vogiatzis

Every worksheet will work as follows.

1. You will be asked to form a group with other students in the class: you can make this as big or as small as you'd like, but groups of 4-5 work best.
2. Read through the worksheet, discussing any questions with the other members of your group.
   - You can call me at any time for help!
   - I will also be interrupting you for general guidance and announcements at random points during the class time.
3. Answer each question (preferably in the order provided) to the best of your knowledge.
4. While collaboration between students is highly encouraged and expected, each student has to submit their own version.
5. You will have 24 hours (see gradescope) to submit your work.

Activity 1: The multinomial distribution

A product is manufactured in one of three factories located in Fargo, ND, Atlanta, GA, and Portland, OR. The product can be of high or low quality (assume these are the only two possibilities). The three factories have slightly different quality outputs, that are provided in Table 1.

	Fargo, ND	Atlanta, GA	Portland, OR
High quality	95%	92%	97.5%
Low quality	5%	8%	2.5%

Table 1: The high and low quality output rates of each factory. For example, Fargo, ND will have 95% of their products be of high quality, whereas Atlanta, GA will only have 92% of their products be of high quality.

Last, assume that Fargo, ND is responsible for 40% of the production in the United States, with Atlanta, GA and Portland, OR sharing the remaining production.

Problem 1: Law of total probability

Remember the law of total probability? Let's apply it here (in its discrete form). What is the probability a random product you pick up is of low quality?

Answer to Problem 1.

Problem 2: An application for the multinomial

You pick 10 products at random. What is the probability that exactly 4 of them were produced in Fargo, ND, exactly 3 of them were produced in Atlanta, GA, and the other 3 of them were produced in Portland, OR?

Answer to Problem 2.

Problem 3: Marginal and conditional distributions

You pick 10 products at random again. Answer the two probability questions here:

1. What is the probability that at most one of them was produced in Atlanta, GA? [1]
2. What is the probability 4 of them were produced in Fargo and the other 4 of them were produced in Portland given that 2 of them were manufactured in Atlanta? [2]

[1] Use the **marginal** distribution for this.

[2] Use the **conditional** distribution for this.

Answer to Problem 3.

In general:

1. The marginal distribution of a multinomial distribution is the binomial.
2. The conditional distribution of a multinomial distribution is another multinomial.

Activity 2: The bivariate normal distribution

We have asked a set of people about their views on climate change and associated sustainability policies proposed to curb its effects. Specifically, we want to measure the net favorability of each question (measured in %) defined as the % of respondents agreeing minus the % of respondents disagreeing. The survey in several different states has shown that the net favorability in the climate change question is $X \sim \mathcal{N}(4, 4)$, while the net favorability in the policies question is $Y \sim \mathcal{N}(2, 9)$. In English, we assume the net favorability in agreeing that climate change is a real threat is (continuous) random variable X normally distributed with mean $\mu_X = 4$ and variance $\sigma_X^2 = 4$; while the net favorability in agreeing with policies proposed is another normally distributed random variable called Y with mean $\mu_Y = 2$ and $\sigma_Y^2 = 9$.

Problem 4: Simple normal

Let's jog our memories with the normal distribution. Answer the two next probability questions. [3]

[3] A z-table is provided in the last page of the worksheet.

- What is the probability that there is more than a 6% net favorability difference for climate change? That is, what is $P(X > 6)$?
- What is the probability that there is less than or equal to a 1% net favorability difference adopting sustainability policies? That is, what is $P(Y \leq 1)$?

Answer to Problem 4.

$P(X > 6) =$

$P(Y \leq 1) =$

Problem 5: Independent?

Let's now look at the problem from a "data analyst" perspective. Do you think it is fair to assume that these two answers are independent? Or would they be correlated? And is that correlation positive or negative? [4] Simply a sentence of explaining why yes/no would suffice here.

[4] Think of the following setup: would a person agreeing that climate change is a threat also be more prone to agreeing with policies meant to curb it? Would the opposite be true?

Answer to Problem 5.

Hopefully you have agreed that the two are indeed correlated. From now on, please assume that we have a positive correlation equal to $\rho_{XY} = 0.5$.

Problem 6: The case of NC

We are analyzing the results from North Carolina. It appears that in NC respondents agree that climate changes is an important threat. Specifically, we observe that the net favorability of this question is $X = 3\%$. What is the probability that respondents disagree with policies to promote sustainability given that $X = 3\%$? That is, what is $P(Y \leq 0|X = 3)$? [5] Recall that we have $\rho_{XY} = 0.5$.

[5] What is the **conditional** distribution of a bivariate normal distribution? Is it.. another normal distribution?

Answer to Problem 6.

Problem 7: Overwhelming support for one; so-and-so on the other

What is the probability of getting a state that both has a higher than 6% net favorability for the first question but is very close on the second question? Let's define "very close" as being between -1% and +1% of net favorability. In mathematical terms, what is $P\left((X > 6) \cap -1 \leq (Y \leq 1)\right)$? [6]

Answer to Problem 7.

[6] Also written as $P(X > 6, -1 \leq Y \leq 1)$ or $P(X > 6 \text{ and } -1 \leq Y \leq 1)$. Recall that to answer questions like these we need to integrate properly the pdf of the bivariate normal distribution. Thankfully there are online calculators!

Activity 3: The velocity of a particle

A particle's velocity (measured in m/s) in a gas is a continuous random variable (let it be V) with pdf $f_V(v) = av^2e^{-bv}, v > 0$, where b is a constant that depends on the temperature of the gas and the mass of the particle. While we could make the case that the speed of anything can be no larger than the speed of light, we do allow speed to go up to infinity here, as it makes the math a little easier :) so please consider that $v \in (0, +\infty)$.

Problem 8: Back to basics

Once again, what should a be in order for $f_V(v)$ to be a valid pdf? Note that your result will be a function fo b. [7]

Answer to Problem 8.

[7] As a reminder, we want $\int\limits_{-\infty}^{+\infty} f(x)dx$ to be equal to 1. In our case, random variable velocity V cannot be negative, and we assume it goes to $+\infty$, so we would like $\int\limits_{0}^{+\infty} f_V(v)dv = 1$.

Problem 9: The expectation of a function

Back to expectation properties now. The kinetic energy of a particle is $W = \frac{1}{2}mV^2$, where V is the velocity (the random variable from earlier). What is the expected kinetic energy of the particle? [8] Of course, as in Problem 1, your final answer will be a function of b.

Answer to Problem 9.

[8] Check Lecture 9: the expectation of a function of a *continuous* random variable X, $g(X)$, is calculated as $E[g(X)] = \int\limits_{-\infty}^{+\infty} g(x)f(x)dx$, where $f(x)$ is the pdf of X itself. As a reminder, that would be slightly different in the discrete case (the integration becomes a summation).

Problem 10: The probability density function of a function

Assuming again that the kinetic energy of a particle is $W = \frac{1}{2}mV^2$, where V is the velocity random variable: what is the pdf of W?

Answer to Problem 10.

This is a very useful derivation. Let X be a random variable distributed with pdf $f_X(X)$. Also consider Y a random variable that is a function of X, as in $Y = g(X)$. Let $u(y)$ be the inverse function, i.e., $u(y) = g^{-1}(y)$. Then, Y is distributed with pdf:

$$\boxed{f_Y(y) = f_X(u(y)) \cdot |u'(y)|}$$

NORMAL CUMULATIVE DISTRIBUTION FUNCTION ($\Phi(z)$)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441