# CS6120/4120 Natural Language Processing – Quiz

## 1. Two events A, B which are disjoint satisfy $P[A \cap B] = P[A] + P[B]$

False.

If two events are disjoint, they are mutually exclusive, they can't occur at the same time, so the probability of their union is zero.

See: https://people.richland.edu/james/lecture/m170/ch05-rul.html

## 2. A, B are two events such that $P[A]=0.3$, $P[B]=0.5$, $P[A \cup B]=0.65$, then A and B are independent.

True.

If A, B independent, then

$$P[A \cap B] = P[A]\,P[B] = 0.3 \times 0.5 = 0.15.$$

and

$$P[A \cup B] = P[A+B] = P[A] + P[B] - P[A \cap B] = 0.3 + 0.5 - 0.15 = 0.65$$

If A depends on B, then

$$P[A \cap B] = P[A|B]\,P[B]$$

If B depends on A, then

$$P[A \cap B] = P[B|A]\,P[A]$$

See: https://www.wyzant.com/resources/lessons/math/statistics_and_probability/probability/further_concepts_in_probability

## 3. The expectation of the random variable X=(1, 2, 6, 7) is 5.

False.

$E[X] = \sum x\, P(X=x) = (1 \times ¼) + (2 \times ¼) + (6 \times ¼) + (7 \times ¼) = 16/4 = 4$

See: https://en.wikipedia.org/wiki/Expected_value

## 4. Suppose (X, Y) is a two dimensional random variable. If the correlation $\rho_{XY}$ equals zero, then X and Y are independent.

False.

The correlation $\rho_{XY}$ refers to the Pearson correlation coefficient which measures linear correlation. If the variables are independent, then $\rho_{XY} = 0$, but converse is not true.

From: https://en.wikipedia.org/wiki/Correlation_and_dependence

If the variables are independent, Pearson's correlation coefficient is 0, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables. For example, suppose the random variable $X$ is symmetrically distributed about zero, and $Y = X^2$. Then $Y$ is completely determined by $X$, so that $X$ and $Y$ are perfectly dependent, but their correlation is zero; they are uncorrelated. However, in the special case when $X$ and $Y$ are jointly normal, uncorrelatedness is equivalent to independence.

## 5. If P[A|B] and P[A] and P[B] are known, we can compute P[B|A]

True.

Using Bayes' rule:

$$P[A|B] = P[B|A]\ P[A] / P[B]$$

so

$$P[B|A] = P[A|B]\ P[B] / P[A]$$

See: https://en.wikipedia.org/wiki/Bayes'_theorem

## 6. If P[A|B] > P[A] then P[B|A] > P[B]

True.

Using Bayes' rule:

$$P[A|B] > P[A]$$

$$P[A|B] > P[A|B]\ P[B] / P[B|A]$$

$$P[B|A]\ P[A|B] > P[A|B]\ P[B]$$

$$P[B|A] > P[B]$$

Since all of the probabilities are positive, multiplying or dividing both sides does not change the inequality.

See: https://en.wikipedia.org/wiki/Inequality_(mathematics), **Multiplication and Division**.

## 7. Suppose X and Y are independent random variables. For a given value x, we can estimate E[X+Y | X=x ] if we know E[Y].

True.

If X and Y are independent, E[X+Y] = E[X] + E[Y]. The expression E[X+Y | X=x] means the value of E[X+Y] when X=x, so E[X+Y] = E[X=x] + E[Y] = x + E[Y]. If X is a constant, E[X] is the constant.

See: https://en.wikipedia.org/wiki/Expected_value#Expected_value_of_a_constant

## 8. The average of a list of numbers cannot be smaller than the standard deviation.

False.

The standard deviation for discrete samples is:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}, \text{ where } \mu = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

If the numbers are centered on zero, the standard deviation can be large while the average is zero.

Example:

samples = { -100, -75, -50, -25, 25, 50, 75, 100 }

μ = (1/8) * ∑ ( -100, -75, -50, -25, 25, 50, 75, 100 ) = 0

σ = sqrt( (1/8) * ∑ ($-100^2$, $-75^2$, $-50^2$, $-25^2$, $25^2$, $50^2$, $75^2$, $100^2$) )

= sqrt( (1/8) * ∑ (10000, 5625, 2500, 625, 625, 2500, 5625, 10000) )

= sqrt( (1/8) * 37500 )

= sqrt( 4687.5 )

= 68.4653

See: https://www.mathsisfun.com/data/standard-deviation-formulas.html

See: http://mathworld.wolfram.com/StandardDeviation.html

## 9. Suppose X is a normal random variable with $\sigma^2 = 10$, then 2X is a normal random variable with $\sigma^2 = 20$.

False.

From: https://en.wikipedia.org/wiki/Normal_distribution#Operations on normal deviates

The family of normal distributions is closed under linear transformations: if $X$ is normally distributed with mean $\mu$ and standard deviation $\sigma$, then the variable $Y = aX + b$, for any real numbers $a$ and $b$, is also normally distributed, with mean $a\mu + b$ and standard deviation $|a|\sigma$.

So for 2X, $\sigma^2_{2X} = (|2|\sigma_X)^2 = 4\sigma^2_X = 40$, not 20.

## 10. If a pair of dice is rolled simultaneously, the probability that the sum of the dots is an even number equals 0.5.

True.

P(sum is even)

= P(both even) + P(both odd)

= P(die1 even) P(die2 even) + P(die1 odd) P(die2 odd) *because the dice are independent*

= (0.5)(0.5) + (0.5)(0.5)

= 0.25 + 0.25

= 0.5

## 11. If you roll three dice, the expected value of the sum of the rolls is 10.

False.

E[roll of a die] = ( 1 + 2 + 3 +4 + 5 + 6 ) / 6 = 3.5

E[X+Y+Z] = E[X] + E[Y] + E[Z] = 3.5 + 3.5 + 3.5 = 10.5

where X, Y, and Z are the rolls of the three dice.

See: https://en.wikipedia.org/wiki/Expected_value, **Linearity**.

## 12. The vectors [ 1 1 1 ]$^T$, [ 1 1 0 ]$^T$, [ 1 0 0 ]$^T$ are linearly independent.

True.

A set of vectors are linearly independent if there is no way to represent any of them as a linear combination of the others.

Multiplying any of the vectors by a constant will not change the zero elements. If you compute a linear combination of the second and third vectors, $a_1$ [ 1 1 0 ]$^T$ + $a_2$ [ 1 0 0 ]$^T$ with $a_1$ and $a_2$ not zero, the result is [ $a_1+a_2$, $a_1$, 0 ]$^T$ which cannot equal [ 1 1 1 ]$^T$. Similarly, including a [1 1 1 ]$^T$ in a linear combination will result in a sum with a non-zero entry in the last element, so the sum cannot be equal to either of the other two vectors. None of the vectors can be expressed as a linear combination of the other two, so they are linearly independent.

See: https://en.wikipedia.org/wiki/Linear_independence

## 13. The least square solution of the linear system Ax=y is a real, unique solution of the system if A is invertible.

False? True?

The approximate least squares solution of the linear system Ax=y exists if $A^TA$ is invertible:

$$x = (A^TA)^{-1}A^Ty$$

If A is invertible, then the least squares solution becomes:

$$x = (A^TA)^{-1}A^Ty = A^{-1}(A^T)^{-1}A^Ty = A^{-1}y$$

which is the unique solution to the original equation, with no approximations.

Whether x is real depends, I think, on whether the components of A and y are real.

See: *Linear Algebra Review and Reference*, **4.4 Least Squares**

See: https://en.wikipedia.org/wiki/Least_squares#Solving_the_least_squares_problem

## 14. There is a system of linear equations Ax=y which has exactly 2 solutions.

False.

A system of linear equations can have …

- no solutions
- exactly 1 unique solution
- an infinite number of solutions

See: https://en.wikipedia.org/wiki/System_of_linear_equations

## 15. In general, a normal distribution N(0, 1) has a fatter tail than a Student’s t-distribution.

False.

A Student’s t-distribution has a fatter tail than a normal distribution though it approaches the normal distribution as the sample size increases.

See: https://en.wikipedia.org/wiki/Student's_t-distribution

See: http://rpsychologist.com/d3/tdist/

## 16. A 95% confidence interval for the population mean of (90, 110) implies that 95% of the data points lie between 90 and 110.

False.

A 95% confidence interval of (90, 110) for the population mean says that there is a 95% probability that the actual population mean is between 90 and 110.  A stricter interpretation is that if such an interval were calculated for many independent experiments, 95% of the intervals would contain the mean.

The data points (measured values or values of the underlying population) could lie anywhere in or out of the interval provided the mean ended up in it in 95% of the experiments.

It does not mean that 95% of the data points are within the interval, though they may be.

See: https://en.wikipedia.org/wiki/Confidence_interval

See: http://www.stat.yale.edu/Courses/1997-98/101/confint.htm

## 17. An experiment consists of taking the average of 100 rolls of a fair die. Let X be the average of 100 rolls. X must have a normal distribution.

True?

I interpret the question as meaning that if I performed the experiment many, many times, the averages X would have a normal distribution.

I found it hard to find a straight answer to this question.

The variable X is a discrete variable, because the average of 100 throws of a die = sum / 100. The sums are integers from 100 to 1200, so the average is one of a set of fractions.

Also, throwing one die 100 times is equivalent to throwing 100 dice at once. The distribution of the average is the same as the distribution of the sum, when the trials are all of the same number of rolls.

The best answer I found implies that the average approaches a normal distribution for a large number of experiments: the sum of 50 dice thrown at once approaches a normal distribution as the number of trials increases. The explanation is rather complicated.

See: https://math.stackexchange.com/questions/406192/probability-distribution-of-rolling-multiple-dice

See: https://math.stackexchange.com/questions/397689/why-convolution-regularize-functions/398146#398146, **Summing Dice**

## 18. If a hypothesis is rejected at a significance level of 0.001, it is possible that the hypothesis is not rejected if the test was done at significance level 0.01 (with everything else staying the same).

True.

In statistical hypothesis testing, the significance level (α) is the probability of mistakenly rejecting the null hypothesis if the null hypothesis is true. So a smaller significance level is a stricter standard for rejecting the null hypothesis (loosely speaking, accepting the hypothesis being tested, the opposite of the null hypothesis).

See: https://en.wikipedia.org/wiki/Statistical_significance

## 19. Suppose you are performing a hypothesis test. All other things being equal, if the value of α increases, the value of p increases.

False.

The value of p is an experimental result. The value of α is an experimental parameter. Increasing α means that p can be larger without rejecting the experiment hypothesis (or rather, accepting the null hypothesis), but the value of p comes from the experiment and does not change if a different α is used to test the hypothesis.

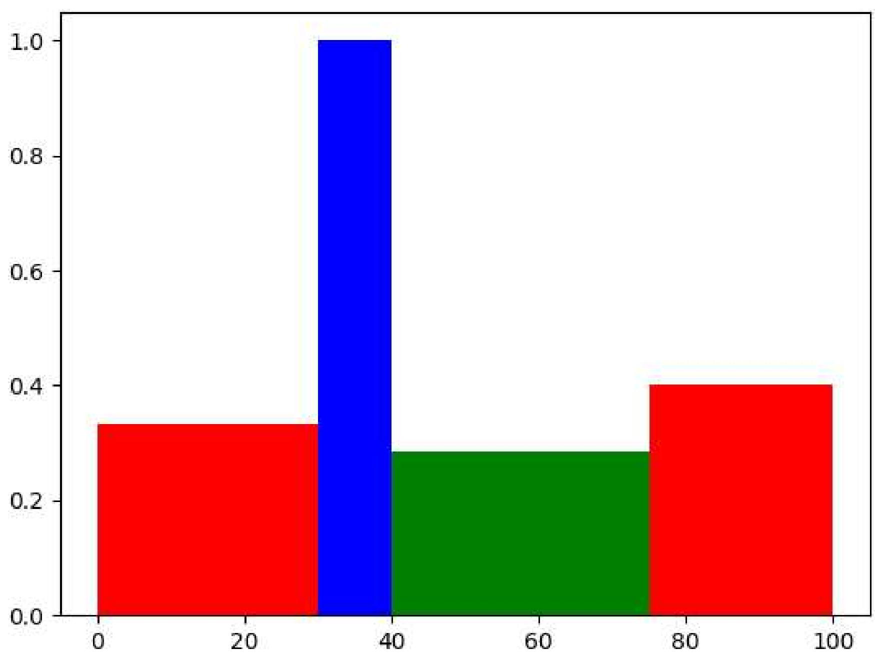
See: https://en.wikipedia.org/wiki/Statistical_significance

See: https://en.wikipedia.org/wiki/Statistical_hypothesis_testing

### 20. A set of data has quartiles of 30 and 75 and a median 40. We can conclude that the shape of the data distribution is skewed to the left.

True.

The median divides the data set so half the samples are below it and half above it. The quartiles divide each half in the same way. So in this example, ¼ of the values are less than 30, ¼ are between 30 and 40, ¼ are between 40 and 75, and ¼ are greater than 75. Since 40 is closer to 30 than to 75, the data points are crowded to the left of their range. The example below has 10 samples in each quartile.




See: https://en.wikipedia.org/wiki/Quartile

See: http://www.cvgs.k12.va.us/digstats/main/descriptv/d_skewd.html