Questions

1. When did Pascal and Laplace write their great works on probability? How about Bayes?

-Pascal wrote about probability in the seventeenth century 1654; correspondence with Fermat on the probability of points.

-Laplace wrote T*héorie analytique des probabilités* in 1812.

-Bayes wrote *An Essay towards solving a Problem in the Doctrine of Chances* in 1763.

2. What formula relates the probability of two events to the probability of each event, if they are **independent**?

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

3. What defines **independence**?

-two random variables are independent if the realization of one does not affect the probability distribution of the other (i.e rolling a die does not affect the probability of drawing an ace from a deck of cards)

4. In your own words, what is the frequentist interpretation of probability?

-There's a show on Netflix called Ozark and it's about a money laundering scheme and season 3 is about laundering through a riverboat casino. The main character says a quote about predicting the gambling habits of people and he says something along the lines of "predicting the odds of a single person making a single decision is nearly impossible to predict, but you get thousands of people doing the same thing and suddenly it's much clearer to develop a sense of the pattern" That's what came to my mind when I was reading about frequentist probability. "For example, the probability of rolling a dice (having 1 to 6 number) and getting a number 3 can be said to be Frequentist probability. Consider another example of head occurring as a result of tossing a coin. Note that the Frequentist frequencies can be calculated by conducting the experiment in a repetitive manner for possibly a large number of times and calculating the probability by counting the number of times an of particular type occurred." https://vitalflux.com/difference-frequentist-vs-bayesian-probability/

5. In your own words, what is the Bayesian interpretation of probability?

-I understood this as the probability of an event that is not repeated. Something that doesn't make sense to calculate based on chance.

6. In Bayesian statistics, what are *prior* and *evidence* and *posterior*? Write Bayes' theorem and define the symbols you use, perhaps with an example. How does it relate to the scientific method of a cycle from hypothesis to experiment and on to better hypothesis?

-A prior is the probability distribution that would express one's beliefs about this quantity before some evidence is taken into account. A prior can be determined from past information, such as previous experiments.

-Evidence corresponds to new data that were not used in computing the prior probability

-Posterior is what we want to know: the probability of a hypothesis *given* the observed evidence.

$-P(A \mid B)=(P(B \mid A)*P(A))/(P(B))$

- $P(A \mid B)$ is the likelihood of event A occurring given that B is true

-$P(B \mid A)$ is the likelihood of event B occurring given that A is true

-P(A) and P(B) are the probabilities of observing A and B respectively

OR …

$-P(H \mid E)=(P(E \mid H)*P(H))/(P(E))$

- $P(H \mid E)$ the *posterior probability*, is the probability of observing E given H. It indicates the compatibility of hypothesis.

-$P(E \mid H)$ is the probability of the evidence with the given

P(H) is the prior probability while P(E) is the evidence or posterior probability

https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/

-Serena Williams beat Venus Williams 5/6 times between 1999 and 2003. If you asked me to guess who would win in 2004 I would guess Serena.

-I think it helps develop a better hypothesis because you usually go into a hypothesis with prior knowledge. You run the experiment and gather evidence. Your evidence either does or doesn't support the hypothesis. If it does not support it, you know have more prior knowledge to reform your hypothesis and try again with new evidence. I like this as "rolling hypothesis". Which is how I think science should be performed. We shouldn't be so set on one idea or one outcome. This type of method allows us to keep learning and keep applying what we learned to new ideas and hypotheses.

7. What is a probability distribution? a probability density? (hint: discrete vs. continuous). What are the units of p(T), the probability density of temperature of some object at some time?

-A probability distribution is a statistical function that describes all the possible values and likelihoods that a random variable can take within a given range.

-Probability density function (PDF) is a statistical expression that defines a probability distribution (the likelihood of an outcome) for a discrete random variable (i.e. a stock) as opposed to a continuous random variable.

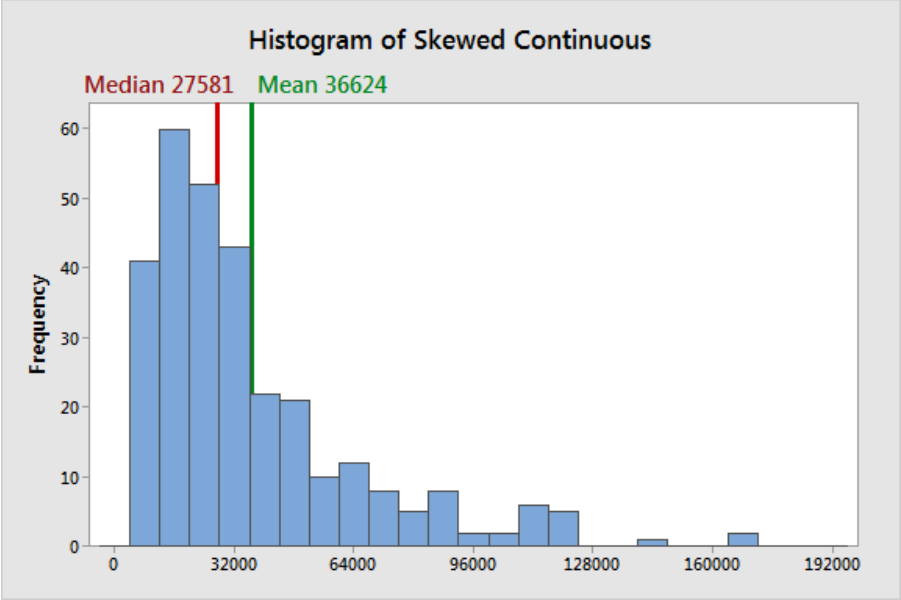
- This probability is given by the integral of this variable's PDF over that range—that is, it is given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the range. The probability density function is nonnegative everywhere, and its integral over the entire space is equal to 1.

8. Using the term *likelihood* for probability density, what is the *maximum likelihood* value of T if p(T) is not symmetric -- the mean, the median, or the mode? Sketch a nonsymmetric distribution and indicate these 3 different measures of its *central tendency*.

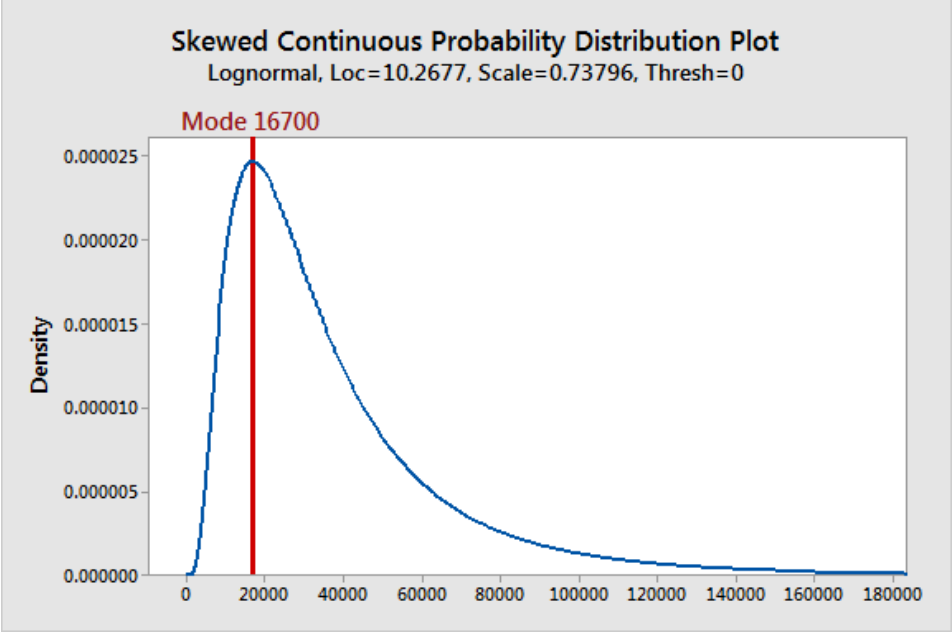
- likelihoods attach to hypotheses

-maximum likelihood estimation is a method that will find the values of mean (μ) and standard deviation (σ) that result in the curve that best fits the data.

-central tendency : central tendency (or measure of central tendency) is a central or typical value for a probability distribution. It may also be called a center or location of the distribution. Colloquially, measures of central tendency are often called averages. The most common measures of central tendency are the arithmetic mean, the median and the mode. A middle tendency can be calculated for either a finite set of values or for a theoretical distribution, such as the normal distribution.



In a skewed distribution, the outliers in the tail pull the mean away from the center towards the longer tail. For this example, the mean and median differ by over 9000, and the median better represents the central tendency for the distribution.



The mode is the value that occurs the most frequently in your data set.
https://statisticsbyjim.com/basics/measures-central-tendency-mean-median-mode/

9. Consider a uniform distribution over [1,2]. What is its first *moment*? What are its first four *central moments*?

-The *s*th moment of the data set with values x_1, x_2, x_3, ... , x_n is given by the formula: $(x_1s + x_2s + x_3s + ... + x_n^s)/n$

-The 1st moment is the mean set s=1

-The 2nd moment is the variance set s=2

-The 3rd moment is skewness set s=3

The 4th moment is kurtosis set s=4

Significance of moments (raw, central, normalised) and cumulants (raw, normalised), in connection with named properties of distributions

Moment ordinal	Moment			Cumulant	
	Raw	Central	Standardized	Raw	Normalized
1	Mean	0	0	Mean	N/A
2	–	Variance	1	Variance	1
3	–	–	Skewness	–	Skewness
4	–	–	(Non-excess or historical) kurtosis	–	Excess kurtosis
5	–	–	Hyperskewness	–	–
6	–	–	Hypertailedness	–	–
7+	–	–	–	–	–

10. What fundamental mathematical operation (addition, subtraction, multiplication, division) creates the *Normal distribution* according to the *Central Limit Theorem*? Can you think of reasons it is so commonly observed? That is, can you name some natural processes that mimic that mathematical operation?

- central limit theorem (CLT) establishes that, in some situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution (informally a *bell curve*) even if the original variables themselves are not normally distributed.

-the mean of all samples from the same population will be approximately equal to the mean of the population. Furthermore, all the samples will follow an approximate normal

distribution pattern, with all variances being approximately equal to the variance of the population, divided by each sample's size.

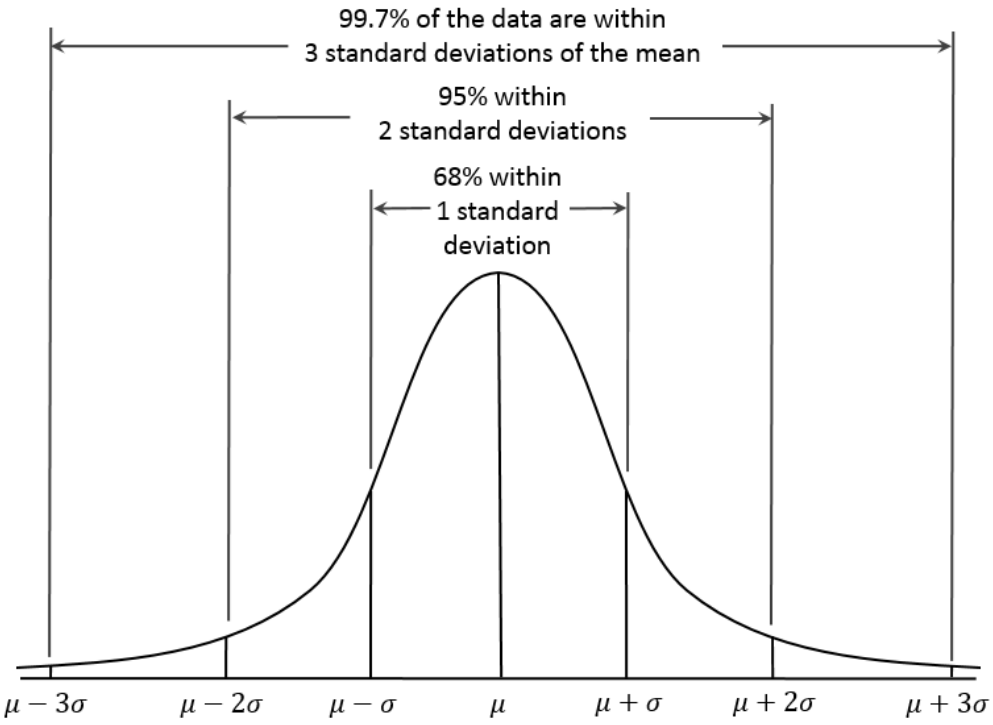
- voter demographics, mean household incomes for a city, weights of students in a school ? I'm not exactly sure this is what you meant about natural processes?

11. In your own words, what is a *test statistic*?

-A test statistic is used to determine if a hypothesis will be accepted or rejected? To me it seems like this what what we use to accept or reject the hypothesis or null hypothesis. Examples include t-test, chi test and it seems like these are specific tests statistics to look for a certain property.

12. What is the Z-test (based on the Z-statistic)? What are the one-tailed (or one-sided) p-values for Z values of 1, 2, 3? (sometimes called *one-sigma*, *two-sigma*, *three-sigma* events or excursions of a variable away from its mean).

- A *Z*-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. It is used to determine whether two population means are different when the variances are known and the sample size is large. A z-statistic, or z-score, is a number representing how many standard deviations above or below the mean population a score derived from a z-test is.



Z value	1	2	3
P value (1tail)	0.159	0.023	0.00135
P value (2 tail)	0.317	0.046	0.0027

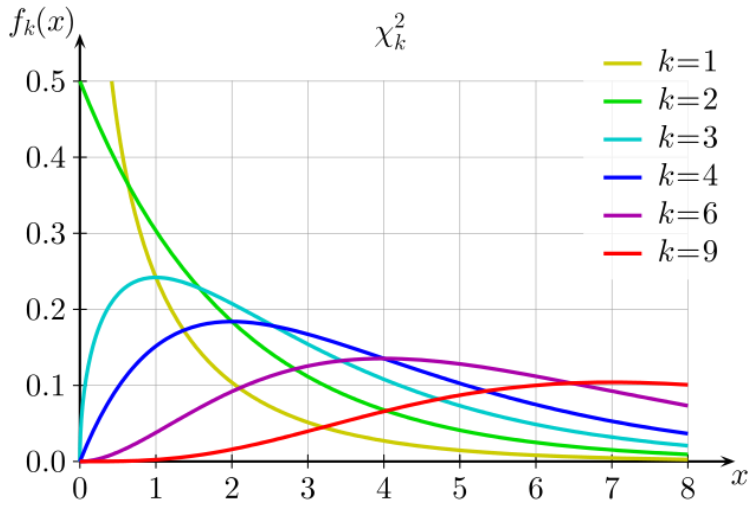
13. What is the t-test, based on the t-distribution? Find a table or Web page to answer the p-value question above, for a t-test with sample number N=10. How small is the difference from a Z-test?

-A t-test is used to determine if there is a significant difference between the means of two groups, which may be related in certain features. A t-test looks at the t-statistic, the t-distribution values, and the degrees of freedom to determine the statistical significance. Normal distributions are used when the population distribution is assumed to be normal. The T distribution is similar to the normal distribution, just with fatter tails

T value	1	2	3
P value (1tail)	0.171	0.0382	0.00747
P value (2 tail)	0.343	0.0765	0.01496
Δ P value for T and Z (1 tail)	.012	.0152	.00612
Δ P value for T and Z (2 tail)	.026	.0305	.0126

14. What is the chi-squared distribution? When might you use it in a statistical test? (hint: where in science do we see a sum of squares involving just two or three variables?)

-Chi-square is a statistical test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis.
-The chi square distribution is the distribution of a sum of the squares of *k* independent standard normal random variables.
-The chi-square test is always testing the null hypothesis, which states that there is no significant difference between the expected and observed result.
-Do we see sum of squares in linear regressions?



15. What is the F-test? Where have we seen a *ratio of variances* before? (hint: what was r-squared in linear regression)?

-The F statistic is defined as the ratio between the two independent chi square variates that are divided by their respective degree of freedom.

-R^2 is the proportion of the variance for a dependent variable that's explained by an independent variable. This is like the F-test in the sense that the F-test is used to test for the equality of two variances. Test whether or not two independent samples are drawn from a normal population with the same variability

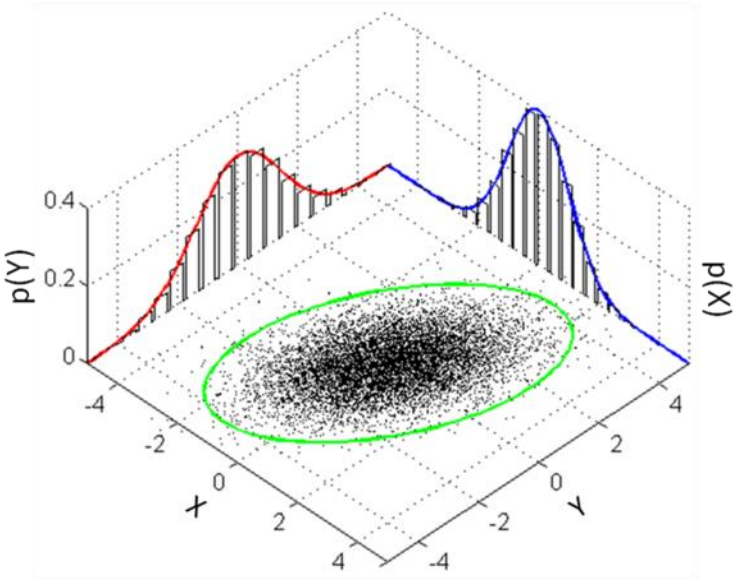
16. What are the *joint* PDF (probability density or distribution function)? *Marginal*? *Conditional*? Illustrate these 3 quantities with a sketch or annotation on an example like at https://en.wikipedia.org/wiki/Multivariate_normal_distribution.

-Joint probability distribution for is a probability distribution that gives the probability that each falls in any particular range or discrete set of values specified for that variable.

- marginal distribution giving the probabilities for any one of the variables with no reference to any specific ranges of values for the other variables.

-conditional probability distribution giving the probabilities for any subset of the variables conditional on particular values of the remaining variables.

A multivariate distribution is a distribution where there are multiple random variables. If the variables are independent, then they become more trivial, because the joint pdf is just the product of the individual pdfs.



I do not understand the illustrations / figures that represent multivariate distributions.