MPO 624 - Assignment 03 (04/14)

Ivenis Pita

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

Pascal - 1654

Laplace - 1812

Bayes - "An essay towards solving a problem in the Doctrine of Chances" was read to the Royal Society in 1763 after Bayes's death (1761).

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

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

3- What defines independence?

Independence is defined by the lack of relation between two or more variables (e.g., when the occurence of one does not affect the probability of occurence of the other).

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

Probability could be obtained only for random samples. Using two samples, the frequentistic probability analyses the number of times an experiment resulted in two different results. As greater the repetitions of the experiments, more precise the probability of getting one or another result.

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

The Bayesian point of view also uses prior knowledge to estimate the probability of getting one or another result. Bayesian probability does not negate a hypothesis, however it states a probability of this hypothesis be truth or false.

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?

Prior is the probability distribution that would express one's beliefs about this quantity before some evidence is taken into account.

Evidence is new data that is to be taken into account (like the result of some experiments).

Posterior s the conditional probabilioty that is assigned after the relevant evidence or background is taken into account

Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$P(A|B)$ -> Likelihooh of event A occuring given that B is true (Conditional probability).

$P(B|A)$ -> Likelihooh of event B occuring given that A is true (Conditional probability).

$P(A)$ -> probability of observing A (Marginal probability).

$P(B)$ -> probability of observing B (Marginal probability).

e.g.,

	Male	Female	Total
BR	.33	.33	.66
US	.33	0	.33
Total	.66	.33	1

$$P(M|BR) = \frac{P(BR|M)P(M)}{P(BR)} = \frac{0.50 * 0.66}{0.66} = 0.50$$

$$P(H_0|Y = y) = \frac{f_Y(y|H_0)P(H_0)}{f_Y(y)}$$

$$P(H_1|Y = y) = \frac{f_Y(y|H_1)P(H_1)}{f_Y(y)}$$

We choose H_0 if and only if $P(H_0|Y = y) \geq P(H_1|Y = y)$

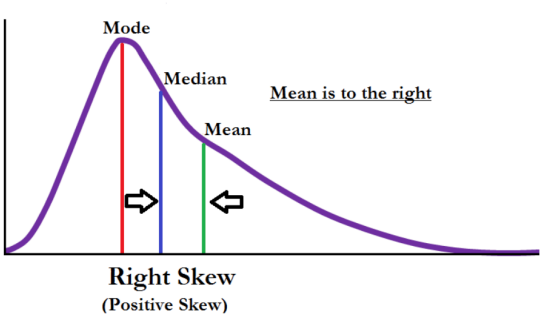
7- What is a probability distribution? a probability density? (hint: discrete vs. continuous). What are the units of p(T), the probability density of temperature of some object at some time?

Probability distribution is a function that describes the likelihood of obtaining the possible values that a discrete random variable can assume.

Probability density is a function whose integral is calculated to find probabilities associated with a continuous random variable

Units of p(T)? $\left[\frac{1}{^oC}\right]$

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*.



$$P(T=\mu)=\frac{1}{n}*(\mu-\mu)=0$$

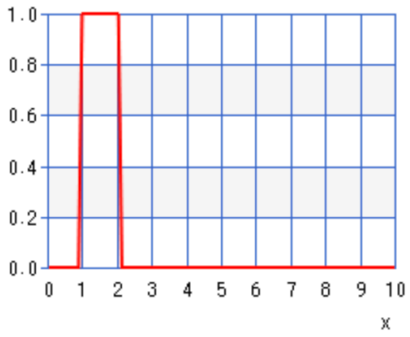
$$P(T=\mu_{median})=\frac{1}{n}*(\mu_{median}-\mu_{median})=0$$

By definition, mode is the term that repeats the most, so

$$P(T=\mu_{mode})=\frac{1}{n}*(\mu_{mode}-\mu_{mode})>0$$

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

$$\mu_i=\int_{-\infty}^{\infty}(x-\mu)^i f(x)dx$$



$$f(x)=\frac{1}{2-1}=1$$

$$\int_1^2 x-\mu dx=\left[\frac{(x-\mu x)^2}{2}\right]|_1^2=0$$

$$\left[\frac{x^2-2x\mu+\mu^2}{2}\right]|_1^2=\left[\frac{4-4\mu+\mu^2}{2}\right]-\left[\frac{1-2\mu+\mu^2}{2}\right]=0$$

$$\frac{3-2\mu}{2}=0,\ \text{so}\ \mu=\frac{3}{2}$$

1^{st} moment is = 0

2^{nd} moment (variance) is = $\int_1^2 (x-1.5)^2 dx == \frac{(x-1.5)^3}{3}|_1^2 = \frac{0.5^3}{3} + \frac{0.5^3}{3} = 0.08333$

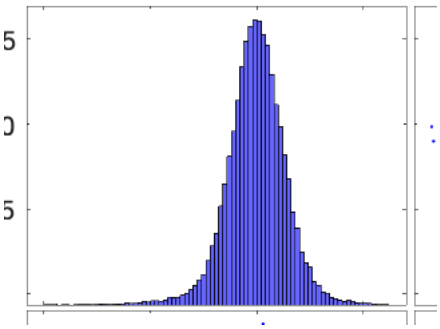
3^{rd} moment is (skewness) = $\int_1^2 (x-1.5)^3 dx = \frac{(x-1.5)^4}{4}|_1^2 = \frac{0.5^4}{4} + \frac{0.5^4}{4} = 0.03125$

4^{th} moment (kurtosis) is = $\int_1^2 (x-1.5)^4 dx = \frac{(x-1.5)^5}{5}|_1^2 = \frac{0.5^5}{5} + \frac{0.5^5}{5} = 0.0125$

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?

Sum. Maybe because the interaction of two continuous independent random variables results in a normal distribution.

The sea surface level obeys a gaussian distribution.



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

Is a test that could shou if your sample/hypothesis is significant/representative or not.

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 a statistical test used to determine whether two population means are different when the variances are known and the sample size is large. normally used for $n \geq 30$ (assuming normal distribution).

https://www.socscistatistics.com/pvalues/normaldistribution.aspx

For a significance level of 0.05,

Z=1; P value $= 0.158655$ (not significant)

Z=2; P value $= 0.02275$ (significant)

Z=3; P value $= 0.00135$ (significant)

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 a statistical test used to determine if there is a significant difference between the means of two groups, which may be related in certain features. Normally used for $n < 30$.

https://www.socscistatistics.com/pvalues/tdistribution.aspx

For a significance level of 0.05,

T=1; P value $= 0.171718$ (not significant)

T=2; P value $= 0.038276$ (significant)

T=3; P value $= 0.007478$ (significant)

Difference could reach values up to 0.013...

Better for a Z-test (probably because has a bigger sample).

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-squared distribution with *k* degrees of freedom is the distribution of a sum of the squares of *k* independent standard normal random variables.

$FALTATERMINAR$!!

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

An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It normally tells whether the model predicts the outcome variable better than using the mean.

We have seen this before in the linear regression, where R^2 is the proportion of variance in dependent variable predicted by independent variable.

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 PDF is a probability distribution that returns the propbability of each random variable to fall in any particular range/discrete set of varles.

Marginal PDF gives the probabilities of various values of the variables in the subset without reference to the values of the other variables.

Conditional PDF gives the probabilities contingent upon the values of the other variables.

	Male	**Female**	**Total**
BR	1	1	2
US	1	0	1
Total	2	1	3

	Male	**Female**	**Total**
BR	.33	.33	.66
US	.33	0	.33
Total	.66	.33	1

Marginal PDF: Born in Brazil? $\frac{2}{3} = 0.66$

Conditional PDF: Probability of be male among those born in Brazil? $\frac{1}{2} = 0.50$

Joint PDF: $P(Female \cap BR) = 0.33$