Foundation of Statistics Sec. 1 & 2 under Shirin Cheat Sheet

by Dylan (dylablo) via cheatography.com/68322/cs/17214/

Basic Mathematical Symbols and Explanations

$\sum$: Sum of the following set/set of values returned from a function. There's ususally a variable name and assignment underneath it, and a limit above - this means you are summing the values returned by the function with the value range from the bottom to to the top.

```
sum=0;
for(int i =0; i++; i<=10){
sum += do_function(i); }
```

$\prod$: Product of the following set/set of values returned from a function. There's ususally a variable name and assignment underneath it, and a limit above - this means you are getting the product of all the values returned by the function with the value range from the bottom to to the top.

```
prod=0;
for(int i =0; i++; i<=10){
prod *= do_function(i); }
```

$\forall$: For all/For every instance. Universal quantifier in predicate logic. ie. The stated holds true for every situation. Can be further expanded with $\forall i$ (assuming i is defined in the previously stated function) followed by a subset or function which would read as "The stated holds true for all i in the following set/function"

$\mathbb{R}$: Real numbers. ie. Not imaginary numbers (sqrt of a negative number) and not infinity. Integers, negatives, floats, doubles etc. are all considered "Real numbers"

$\int$: Integral. Used for finding areas, volumes, central points etc. Not confident in my own summary, please follow this link https://www.mathsisfun.com/calculus/integration-introduction.html

$\lim_{a->x}$: This is the function to find/define the limit of possible values returned when a is fed into the following function ie.

$\lim_{a->-\infty} F(a)=0$

Which means that the lowest value for F(a) where a gets as close to -∞ as possible (approaching from 0) is limited to 0. ie. Lower limit is 0. Can be used to define upper limits with +∞ and limits for discrete variables by specifying their unique upper and lower bounds ie.

$\lim_{a->6} f(a)=1$

where $0<=a<=6$

Definitions, Properties, Rules and Laws

The Additivity Property	If $A \cap B = \varnothing$ then $P(A \cup B) = P(A)+P(B)$ If $A \cap B != \varnothing$ then $P(A \cup B) = P(A)+P(B)-P(A \cap B)$ $P(A^c)=1-P(A)$
The Multiplication Rule	$P(A \cap B) = P(A\|B)P(B)$

Definitions, Properties, Rules and Laws (cont)

The Law of Total Probability	Given disjoint events B1,B2,...,Bm such that $\cup^{m}_{i=1} Bi = \Omega$ (ie. The union of all events B1 through Bm is the same as the entire sample space) Then the probability of a random/arbitrary event A is expressed as... $P(A) = \sum^{m}_{i=1} P(A\|Bi)P(Bi)$ (ie. The sum of probabilities of all events Bi where A occurs)
Bayes' Rule	Given disjoint events B1,B2,...,Bm and $\cup^{m}_{i=1} Bi = \Omega$ (ie. The union of all events B1 through Bm is the same as the entire sample space) Then the conditional probability of Bi, given that a random/arbitrary event A occurs is... $P(Bi\|A) = P(A\|Bi)P(Bi)/\sum^{m}_{j=1}P(A\|Bj)P(Bj)$ (ie. !!!VERIFY!!!THe probability of Bi given that A occurs is the calculated by dividing the the probabiltiy $P(A \cap B)$ <according to the multiplication rule> by the sum of the probabilities of A intersecting all other events in the sample space <according to the multiplication rules)>

Definitions, Properties, Rules and Laws (cont)

Properties of the Probability Mass Function aka pmf	All probabilities are positive: $fx(x) \geq 0$. Any event in the distribution (e.g. “scoring between 20 and 30”) has a probability of happening of between 0 and 1 (e.g. 0% and 100%). The sum of all probabilities is 100% (i.e. 1 as a decimal): $\Sigma fx(x) = 1$. An individual probability is found by adding up the x-values in event A. $P(X \in A)=\Sigma_{x \in A} f(X)$
Properties of the Cumulative Distribution Function aka cdf	1. For a<=b then F(A)<=F(b) ie. if a<=b then the cdf of a<=the cdf of b. 2. F(a) is a probability 0<=F(a)<=1, and $\lim_{a \to +\infty} F(a)=1$ $\lim_{a \to -\infty} F(a)=0$ ie. F(a) will never return a result bigger than 1 or smaller than 0. 3. F is right-continuous $\lim_{b \to 0} F(a+b) = F(a)$. *a<=b implies that the event {X<=a} is contained(subset of) the event {X<=b}
Properties of Expectation aka E(X)	$E(aX) = aE(X)$ ∀ a is a constant $E(XY) = E(X)E(Y)$ when X and Y are independent $E(a+bX) = a+bE(X)$ linearity $E(X+Y) = E(X)+E(Y)$ linearity $E[\Sigma_{i=1}^{n} X_i] = \Sigma_{i=1}^{n} E[X_i]$
Properties of Variance aka V(X)	$Var(aX) = a^2 Var(X)$ ∀ a is a constant $Var(a+X) = Var(X)$ ∀ a is a constant

Probability/Statistics & Set Notation

P(A)	Probability of event A occuring. (Number of ways event A can occur / Number of total outcomes possible))
Ω	Sample Space/Universe. P(Ω)=1
∅	Empty/Null set
P(A∩B)	Probability of A Intersection B
Disjoint/Independent/Mutually Exclusive	If A∩B = ∅ then disjoint/independent of each other
P(A∪B)	If disjoint/independent of one another P(A∪B) = P(A) + P(B) If not disjoint P(A∪B) = P(A) + P(B) - P(A∩B)
A^c	A complement. Everything outside A. $P(A^c) = 1 - P(A)$
A∈B / A∉B	A is an element of B / A is not an element of B
A: A ∈ B	A such that A is an element of B
n! aka Permutations	Counting method where ORDER matters. n! = n(n-1)(n-2)...(n-k+1) where k = sample size
$(^n k)$ aka Combinations	Counting method where order does not matter. $(^n k) = n!/k!(n-k)!)$
P(A|B) aka Conditional Probability	The Probability of A happening, given that B occurs. If A and B are disjoint/independent/mutually exclusive then P(A|B)=P(A) as B has no effect on A. If A and B are dependent ie. B has an effect on the chances of A the P(A|B) = P(A∩B)/P(B) $P(A\|B)+P(A^c\|B)=1$

Probability/Statistics & Set Notation (cont)

P(Bi|A)P(A) = P(A|Bi)P(Bi)	Proven by the combination of Bayes' rule and Law of total probability applied to P(A)
Independence of more than 2 events	Events A1,A2,...,Am are independent if $P(\cap^{m}{i=1}Ai) = \prod^{m}{i=1}P(Ai)$ (ie. They are independent events if the probability of all of their intersections are equal to the product of all of their individual probabilities) A and B are independent. B and C are independent. This does not mean that A and C are independent, nor does it mean they must be dependent.
Random variable aka. rv	Any variable whose value is not known prior to the experiment and are subject to chance aka. Variability aka. Change. Has an associated probability aka. mass An rv is a type of mapping function over the whole sample space and is associated with measure theory. ie. An rv can transform the sample space.
Discrete	There is a set number of outcomes

Probability/Statistics & Set Notation (cont)

Discrete Random Variable	Any function X: $\Omega \rightarrow \mathbb{R}$ that takes on some value. eg. X could be S=sum or M=max ran on a sample space, getting the sum/max of each experiment outcome and constructing a new sample space out of it.
Probability Mass Function aka pmf	The pmf of some discrete rv X. Essentially creating a table/graph displaying all the probabilities of all possible values our discrete rv can be. Please refer to "Properties of the Probability Mass Function aka PMF for more details."Explained here http://www.statisticshowto.com/probability-mass-function-pmf/
Cumulative Distribution Function aka cdf	The cdf of some discrete rv can be used to determine the probability above, below and between values occuring. Please refer to "Properties of the Cumulative Distribution Function aka CDF for more details." Explained here http://www.statisticshowto.com/cumulative-distribution-function/
Continuous	An infinite number of possible values.

Probability/Statistics & Set Notation (cont)

Continuous Random Variables	Is a function X: $\Omega \to \mathbb{R}$ that takes on any value $a \in \mathbb{R}$ Mass/Associated probability no longer considered for each possible value of X instead consider the likelihood that $X \in (a,b)$ for $a<b$.
Probability Density Function aka pdf	Pdf on a continuous rv f(x) of X is an integrable function such that... $P(a<=X<=b) = \int^b_a f(x)dx$ ie. it is the area under the cure between points a and b. Therefore it is the probability of a range of values occuring s.t. conditions on f $f(x)>=0 \ \forall x \in \Omega$ $\int^{\infty}_{-\infty} f(x)dx=1$ ie. the complete area under the curve contains all outcomes. This is defined by the formula... $F(x)=\int^x_{-\infty} f(u)du = P(X<=x)$
Expectation aka. E(x)	The expected value of a random variable This is found using the formula when our rv is discrete $E(X) = \Sigma_{xi \in \Omega}$ xi p(xi) and the following formula when the rv is continuous $E(X)=\int_\Omega x f(x)dx$ To make this easier to understand The expected value is simply the mean and is calculated as the sum of (each possible value muiltiplied by it's independent probability) ie The sum of weighted values to probabilities

Probability/Statistics & Set Notation (cont)

Variance aka Var(X)	A method of measuring how far the actual value of a rv may be from the expected value. Given a discrete variable X the formula is.. $Var(X)=\Sigma_{xi \in \Omega} x^2 i\ p(xi) - (\Sigma_{xi \in \Omega} xi\ p(xi))^2$ Or given a continuous rv use the formula $Var(X)=\int_\Omega x^2 f(x)dx - (\int_\Omega x f(x)dx)^2$ In other words we sum up (the squared value's multiplied by their individual probabilities) and finally deduct the the squared expected value.
Standard Deviation	Another method similar to variance about looking at how far distribution goes from the mean ie. The actual value vs the expected value. Simply calculated with the sqrt(Var(X)). Benefit of this is that it is expressed in the same unit that X is expressed in rather than the squared as variance is.