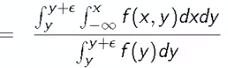
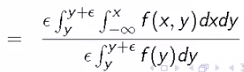
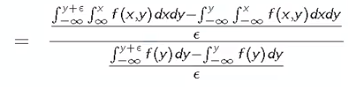
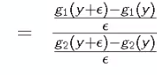
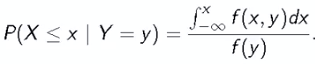
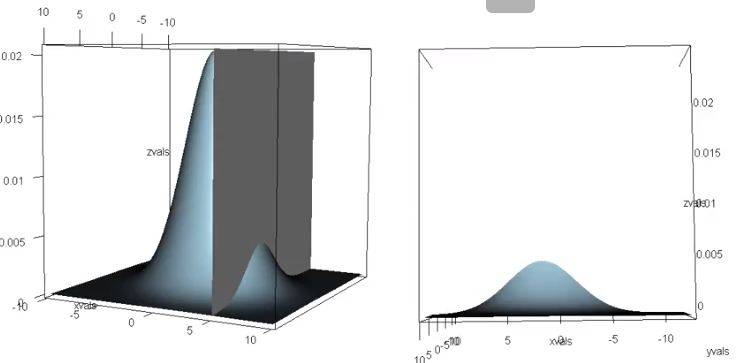
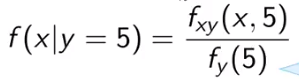
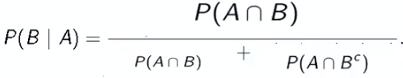
**Conditional Probabilities and Densities**

* P(1) when rolling die = assumed to be 1/6, but suppose we had extra info that the roll ended in an odd
* **Conditional on this new info,** p(1) now = 1/3
* Let B = event w/ P(B) > 0
* Conditional Prob of A given B has occurrent =  (intersection divided by given)
* **If A and B are *independent*, then** (P(B)’s cancel)
* For the die roll, B = {1,2,3}, and A = {1}, so P(A|B) =  🡺 ***A is a subset of B, so the intersection is P(A) by itself*** 🡺 P(A)/P(B) = 1/6 / 1/2 = 2(1/6) = 1/3
* **Conditional Densities/Conditional Mass Functions =** densities + mass functions that govern the behavior of a random variable conditional on the value another random variable took (i.e. **functions of 1 variable conditional on the value of another**)
* f(x,y) = **bivariate (joint) density** or **mass function** for random variables X and Y 🡪 f(x,y) governs the probabilistic behavior of the random variables
* Then let f(x) and f(y) be the associated **marginal mass function** or **densities** that are *disregarding the other variables*
* Then, 
* continuous = integral of joint density f(x,y) over x
* discrete = joint mass function f(x,y) summed over x
* **In other words, to know the marginal probability behavior of random variable y (regardless of what happened w/ respect to x), integrate over all possible values of random variable x**
* Do the flip-side for marginal probability behavior of random variable x
* Then, **conditional density** or **mass function** *given Y = y* is given by  (joint / marginal, from the previous definition)
* Easy to see that (for discrete/finite cases), definition of conditional *probability* is the same as the one for conditional *events* where A == event that X = x and B == event that Y = y
* Continuous definition = bit harder, since events X = x and Y = y have probability = 0
* Useful motivation = take appropriate limits as follow:
* Define:  🡪 random variable X is less than or equal to some real value x, and random variable Y lies in some interval from real value y to y plus some error
* Now both A and B = events w/ probability > 0, and we can apply the standard definition of conditional probability to get P(A|B)
* 
*  🡪 sub in our continous definitions for A and B
*    where 
* We notice that the limit of the numerator and denominator both trend towards g1’ and g2’ as epsilon gets smaller and smaller (*closer to conditioning Y being some specific value y*)
* This gives the **conditional distribution function** associated w/ x as 
* Taking the derivative w/ respect to x gives the **conditional density =** 
* **Densities = derivatives of distribution functions**
* Geometrically, conditional density is obtained by taking the relevant slice of the joint density f(x,y), which itself is a surface (*with volume w/in = 1 to be a valid joint density*) where (x,y) = the plane, f(x,y) = z = the height
* At a particular value for Y, we get some plane (say 5), which slices throught he surface and gives some function

 🡺 

* This slice/function on the right, does not integrate to 1, so it’s not a valid density
* So, we take the slice + appropriately **renormalize** it by something that integrates to 1, say 
* This gives 
* This idea extends to any other line, even non-linear functions)
* Ex: 
* NOTE: **marginal density of y =** integratal joint density function over x 🡺 
* Therefore: **conditional density** ==  == conditional density function/governing behavior of random variable X given y = 3, we have **3\*e^(-3x)**

**Bayes' Rule and DLR’s**

* f(x|y) = conditional density/mass function for X given Y = y, then let f(y) = marginal distribution for y
* If y is continuous 🡺 , if y is discrete = 
* Bayes’ rules related conditional density of f(y|x) to f(x|y) and f(y) (a joint related to another joint and a marginal)
* Special case = 2 sets A and B 🡺 
* Proof:
* Let X = indicator that event A has occurred. Let Y = indicator that event B has occurred
* Plug into discrete version of Bayes’ Rule
* Rules of probability Proof:
* 
* **Diagnostic tests** 🡪 Let + and – be events that the result of a diagnostic test is positive or negative
* **D** and **Dc** = events that the subject does and does not have a disease, respectively
* **Sensitivity** = probability the test is positive given the subject has the disease (TP) = **P(+|D)**
* **Specificity** = probability the test is negative given the subject does not have disease (TN) = **P(-|Dc)**
* **Positive predictive value** = probability the subject *has* the disease, given test is positive, **P(D|+)**
* **Negative predictive value** = probability subject does not have disease, given test is negative, **P(Dc|-)**
* **Prevalence** of disease =*marginal* probability of disease **P(D)**
* **Diagnostic likelihood ratio of a positive test**, **DLR+ = P(+|D) / P(+|Dc) = sensitivity / (1 – specificity)**
* **Diagnostic likelihood ratio of a negative test**, **DLR- = P(-|D) / P(-|Dc) = (1-sensitivity) / specificity**
* *Ex: Study comparing efficacy of HIV test reports on experiment that concluded that HIV antibody test have sensitivity = .997 and specificity = .985*
* *Suppose a subject from a population with prevalence of HIV = .1% receives a positive test.*
* *What is the probability the subject has HIV? 🡪 P(D|+)*
* Bayes’ Rule is convenient in these sorts of settings b/c, in principle, it’s easier to get sensitivity + specificity values by just taking blood samples from a set of people you *know* are HIV+ and seeing what proportion of them had the test comes up positive.
* Then, take a group of people you *know* to be HIV- and seeing the proportion that have a negative test result
* Could get these numbers or *estimates* of these numbers
* very simplistic treatment of how you‘d actually get a sensitivity + specificity
* there's lots of issues 🡺 how do you actually know if you're working in an area where the tests are difficult, how do you actually know if a person has the disease or not, or if you wait so long to where the disease is very clinically relevant, then are we evaluating the test in a stage of the disease where it's not interesting for when we’d be applying the disease?
* There's a lot of issues in development + evaluation of tests + constructing their validity
* Mathematically, we want P(D|+), *given* sensitivity **P(+|D) =** .997, specificity **P(-|Dc)** = .985, and prevalence P(D) = .001
* Use Bayes’ 🡪 
*  =  
* So, *in this specific population*, a positive test result suggests only a 6% probability the subject has the disease 🡺 **Positive predictive value = 6.2%**
* Seems awfully low, and this is due to the very low prevalence and somewhat modest specificity
* Suppose we knew the subject was an IV drug user + routinely had intercourse w/ an HIV-infected partner 🡺 higher **prior** that the subject has HIV, since prevalence among *that* population would be higher, so our **PPV** would be higher
* External info changed our original conclusion 🡪 actual test value does not chance, only the prevalence in the calculation changed
* We notice evidence implied by a positive test does *not* change due to prevalence of the disease in said subject’s population, only our *interpretation* of that evidence changes
* So what component of the calculation does NOT chance, regardless of changes in prevalence? 🡺 **diagnostic likelihood ratios**
* Bayes’ Rule:  and  so  🡺 **DLR+**
* This = a ratio == **post-test odds of having disease = DLR+ \* pre-test odds of having disease**
* So we have some inkling of whether someone has the diseased based on prior knowledge before administrating the test, then we perform the test, which yields data, and DLR+ - the factor by which we multiply pre-test odds to get post-test odds
* So, from before whether a subject is a IV drug user w/ and HIV+ partner is irrelevant to the DLR+, which tells us odds of having disease are increased by x amount given a positive test, regardless of prevalence
* Whereas positive + negative predictive values (LHS of equation above) inherently factor in prevalence (P(D) and P(Dc)) 🡺 higher pre-test odds
* This is basically post-test odds = likelihood ratio (probability model and data combined) times the pretest odds
* This is a very appealing mirror to how we think the scientific process should work.
* i.e. Start out w/ an a-priori set of hypotheses (a hypothesis + its complement) + then collect data
* That **data informs your belief, +** now, we have a *post*-test odds of disease.
* If we were to run *another* test, the starting point for your new **prior** odds would be the post-test odds/**posterior** after this 1st test.
* It all works out just fine if you take 2 tests that’re both positive, so the diagnostic likelihood ratios just multiply.
* But, in terms of Bayesian thinking, this is the idea that we have some **prior**, we **update** it with **data**, we get a **posterior**, + now that posterior is the **new prior**.
* This also codifies a lot of scientific discussion 🡺 if a prior is *absolutely fixed* at a specific point, data is irrelevant, as nothing is going to move it.
* Suppose a subject has a positive test, and DLR+ = .997/(1-.985) = 66
* The result of having positive test == post-test of disease = 66X more than original pre-test odds, regardless of the prior and of the population
* i.e. "hypothesis of disease == 66X more likely (more supported by the data) than the hypothesis of no disease”
* Suppose a subject has a negative test, and DLR\_ = (1- .997/.985 = .003
* The result of having negative test == post-test odds of disease = .3X (.3% of) the original pre-test odds, regardless of the prior and of the population
* i.e. "hypothesis of disease == .003X that of the hypothesis of no disease, given a negative test”
* For a Bayesian, a probability is not objective, but instead is a **quantification of belief**
* Frequentist = person has disease or they don’t
* **NOTE** on the nature of actually collecting data to inform these calculations
* Usually very difficult to know, *conclusively*, whether or not someone has a disease when developing a test
* Very difficult to develop things like *actual, real* prevalence estimates that’re relevant to the person you’re talking about, w/ respect to the disease.
* Very difficult to have whatever samples you're using to develop sensitivity and specificity actually be indicative of the population of samples that the test will be applied to in actual clinical practice.
* So, even though these calculations are very simple + highlight Bayes' rule quite nicely, this is NOT all there is to the world of diagnostic testing and validation, which is a very, very deep subject that involves quite a bit more than Bayes' rule.