**Probability Distributions**

Now going to consider multivariable probability distributions.

**Linear combination of different Gaussian random variables**

Various combinations of independent (so their covariances are 0) normally distributed random variables have their own special distributions. So I’ll consider a few. Say we have:



where X­j is normally distributed with mean μj and std’s σj. These means and variances can differ from each other. What is the mean and variance of W? Well, using standard formula, we know:



And we’d like to calculate the probability distribution of W. It’s easiest to do this with the moment generating function. So consider the moment generating function of w. We’ll just do example of n = 2.



So,



So we could take the inverse Laplace transform of this to get the probability distribution function for W, but since this is of the form of a Gaussian moment generating function,



We see that the probability distribution of w will be:



As a special case, let’s say we have n identically distributed Gaussian variables, Xi, with mean μ and standard deviation σ (though σ won’t matter). Then let’s form the average:



Of course has average and variance,



And we can say is normally distributed, according to:



Another way to frame this is to form the z-score,



and then we can say that is normally distributed,



Extending our special case, now let’s say we have (k = 1,2) nk identically distributed Gaussian variables, Xki, with mean μk and standard deviations σk. Then let’s form the averages:



Of course k has average and variance,



And let’s introduce a linear combination of these variables (could be average of the averages, or could be difference of the averages, etc.),



The mean and variance of would be:



And we can further say is normally distributed, according to:



Another way to frame this is to form the z-score,



and then we can say that is normally distributed,



**Sum of vector normal variables**

Consider two variables **x** and **y**. We’ll say that each is distributed according to:



and,



and they have moment distribution functions:



What is the probability distribution of, say, **W** = a**x** + b**y**? Since the distributions are independent. The joint probability distribution would just be the product. The moment generating function of the variable **W** would be:



which would be a normally distributed variable with mean aμx + bμy, and covariance a2σx2 + b2σy2. So looks like if we have a general linear combination of such variables, **W** = Σjaj**X**j, then **W** is normally distributed with the following mean and covariance.



**Example**

Two friends, Andrew and Brian, are playing bloody-knuckles. Let’s determine the odds that when they attempt to punch each other’s fist, they get a good hit. For the sake of discussion, let’s say they are both trying to hit a point O on a fictitious x-y coordinate system between them. Andrew has not so great aim so his fist lands at a point normally distributed about point O, with a standard deviation σA = 3cm in each direction. Brian has better aim; his fist lands at a point normally distributed about O with a standard deviation of 2cm in each direction. And to get a good hit, their fists’ centers should be no more than 1cm apart. What are the odds they do?

So we’ll look at the difference in their coordiates. D = A – B. D should be normally distributed with mean μD= μA – μB= 0 and covariance σD2 = σA2 + σB2 = [(9 0), (0 9)] + [(4 0), (0 4)] = [(13 0), (0 13)]. So the probability distribution is:



And now we need to determine the probability distribution of r = √(d12 + d22). This is:



Is this normalized? Let’s check:



So probability the radius is less than 1 is?



**Sum of set of identical squared normal random variables**

Let’s say we have we have the sum of n normalized Gaussian distributed variables (Z has mean 0 and std 1).



where from the previous file, we know that any individual Zj2 has a χ2 distribution with 1 d.o.f. And so <z2> = 1, and <z2>2 = 2. This means we have:



Looking at the previous file, and information regarding the χ2 distribution, this leads us to suspect W has a χ2 distribution with n d.o.f. Let’s see. As usual, it’s best to work this out through the moment generating function,



(recall we worked out E(exp(tz2)) for Gaussian variable in previous file). So we see,



So we see this is indeed a χ2 distributed variable with n d.o.f. So,



Now consider what we get if we look instead at



To get its moment generating function we just replace t by t/n. So we can see that the average of n squared, normally distributed variables would have a moment distribution function of:



Taking the limit that n goes to infinity, we’ll have,



This is the same as the moment generating function for a Gaussian distributed variable.



And we see that the average of n χ2 distributed variables approaches a Guassian distribution with the same mean and variance/n as the original distribution, in the large n limit.



This is a preview of the Central Limit Theorem, that the average of any identically distributed variables will approach a normal distribution.

**Sample variance of Gaussian distributed variables**

Let’s say we form the sample variance of identically distributed variables (actually, they don’t *have* to be Guassian?).



Note is indeed a random variable itself (1/n)ΣjnYj. We can see that this is ultimately just the sum of squares of Gaussian variables. And so we’d suspect this follows a χ2 distribution of some sort. To work this out, we should separate S2 into its independent d.o.f.,



Now let’s look for the moment generating function. So we’ll do:



Ajk has the same value for diagonals and a different same value for off-diagonals.



The determinant of this matrix is, according to Chat GPT:



So our integral is:



which is the same as a χ2 distribution,



So apparently, the variable (n-1)S2/σ2 is χ2 distributed with (n-1) d.o.f. So letting:



we have:



It follows from this that:



The average indicates that S2 is an unbiased estimator for σ2, which we already know, I think. Maybe we’ll see that later. And that’s why we divide by n-1 to get the sample variance, instead of by n.

**Sample T of average of set of identical Gaussian random variables**

Now let’s say we have n identically distributed Gaussian variables, Xi, with mean μ and standard deviation σ (maybe doesn’t have to be Gaussian?). Then let’s form the average:



Of course has average and variance,



But let’s form the sample variance random variable,



We can show this is an unbiased estimator for σ2. Consider,



Let’s then construct the sample variance random variable,



Note the 1/n factor is there because SX2 is just the sample estimated variance of the single random variable Xi, which would be close to σ2. We divide by n to get the sample estimated variance of , which would be close to σ2/n. And now let’s form the sample T-thing.



Breaking this down a bit,



and so we have a Z-variable up top, and in the square root we have a χ2 distributed variable, divided by its d.o.f. So looking back to our single variable pdf file, we see W is Student’s T distributed, with ν = n-1 d.o.f. We’ll recall the Student’s T distribution with ν d.o.f. looks like this:



**Sample T of difference of averages of two sets of Gaussian random variables (different μ’s, same σ’s) (pooled T-test)**

Now let’s say we have (k = 1,2) nk identically distributed Gaussian variables, Xki, with mean μk and standard deviations σ. Then let’s form the averages:



Of course k has average and variance,



Let’s consider the pooled sample variance,



Note this is just a weighted average of the two sample variances. If the two populations have different means, but common variance, then this is an unbiased estimator for that variance. We can show this to be true. Basically, borrowing from above,



Furthermore, we’ll recognize that Sp2 is a χ2 distributed random variable with ν = n1 + n2 – 2 d.o.f. (because it’s the sum of two separate χ2 random variables, and d.o.f. are additive – see below about adding χ2 variables) Now consider the variable:



We know this is unit normally distributed. By analogy, we’ll construct,



and this should be T-distributed. We can see this by breaking it down a little,



which is a Z variable up top, and the square root of a χ2 distributed variable divided by its d.o.f. on bottom. By definition, this is a T-distributed variable. So letting:



we have:



So basic idea is, if have a linear combination of Gaussian variables, then it’s Gaussian distributed with the mean and variance of that linear combination. But a linear combination of Gaussian variables with the value of the variance replaced by the formula for the sample estimated variance, is Student’s T distributed.

**Sample T of difference of averages of two sets of Gaussian random variables (different μ’s, different σ’s) (Welch’s T-test)**

Now let’s say we have (k = 1,2) nk identically distributed Gaussian variables, Xki, with mean μk and standard deviations σk. Then let’s form the averages:



Of course k has average and variance,



Let’s straight-away define,



and work it out a little,



Now (n1 -1)S12/σ12 and (n2-1)S22/σ22 are χ2 distributed variables. And the sum of two χ2’s is a χ2, so it seems plausible that our construction is also a T-distributed variable somehow. Suffice to say it is, but the d.o.f. are crazy (according to ChatGPT). So letting,



we have:



So basic idea is, if have a linear combination of Gaussian variables, then it’s Gaussian distributed with the mean and variance of that linear combination. But a linear combination of Gaussian variables with the value of the variance replaced by the formula for the sample estimated variance, is Student’s T distributed.

**Sample T of average of differences of matched pair of Gaussian random variables (different μ’s, different σ’s) (matched pair T-test)**

Now let’s say we have (k = 1,2) nk identically distributed Gaussian variables, Xki, with mean μk and standard deviations σk. And suppose the sets are correlated, so that cov(X1i, X2j) ≠ 0. The typical scenario is a before/after test, experiment, involving the same sets of people. And so the score/result in the after scenario is correlated with the score/result in the before scenario. Typically mostly higher, or mostly lower. Anyway, if we’re looking to see if there is a meaningfull difference in the after result (X2j) vis a vis the before result (X1j), then the best thing to do is to the form the differences.



The mean of will just be the difference of the two means. But since the Xki’s are correlated, the variance of is not the sum of the variances – see Probability distribution file. Nonetheless, we could work out the variance, and form the variable,



which I think would be normally distributed. But typically, we are not privy to the population variance, though we could approximate it with . So we form the T-variable instead.



where,



and ΔXj = X1j – X2j. This is called a matched pair T-test. And this is T-distributed. So letting,



we have:



**Average of identical exponential random variables**

Let’s say we have the average of n exponential distributed variables with rate parameter λ.



Recalling the poperties of the exponential distribution in the previous file, the first two moments are clearly,



What is its distribution? We can get this by examining the moment generating function:



So,



Now compare to the Γ distribution, and its moment generating function,



and we see that W is Γ distributed with β = λ/n, and α = n. So,



**Sum of different χ2 random variables**

Yeah let’s consider this. Let Xj be χ2 distributed variables, with nj d.o.f. And let’s consider their sum,



And their moment generating function,



So W is also a χ2 distributed variables with ν = Σj=1n nj degrees of freedom. So,



**Sum of Cauchy distributed variables**

Apparently the sum of two Cauchy distributed variables is also Cauchy distributed.