This lemma was used in my old clunky proof of chi squared tests from level 6 so it has be moved to this section.

Theorem: for a fixed vector a, if a sequence of random vectors Y_n converges in distribution to a random vector Y, then $a.Y_n$ converges in distribution to $a.Y$, and that if $a.Y_n$ converges in distribution to $a.Y$ for all vectors a, then Y_n converges in distribution to Y. The first part is because a linear combination of the components of the Y_n converges in distribution to the linear combination of the components of Y since Y_n converges to Y. The other direction is the hard part: Note that it is not as obvious as it seems: Just because each component converges to the right thing does not mean the whole thing does, since the components are not independent in general.

In fact, by having a distribution Y and setting the sequence to be Y, Y, Y,… the above theorem actually says that every probability distribution in higher dimensional space is uniquely determined by its 1D projections, or that scanning something in every direction determines its structure, which is really interesting. (This is for distributions which are part discrete part continuous – which is more than what we need – since our characteristic function theory only works for those types of distributions)

Levels recommended for proof: 6 (A good understanding of all the hard stuff in level 6 stats and technical results is expected)

Proof:

Lemma: Convergence in distribution implies pointwise convergence of characteristic function

Proof:

Suppose we have a sequence of probability distributions X_n converging in distribution to a probability distribution X. For any ε>0, we can pick a point on the cdf of X with a value strictly less than $\frac{\varepsilon}{2}$, and another point with a value strictly greater than $1-\frac{\varepsilon}{2}$. This is always possible since in the extremes, the cdf approaches 0 and 1, so we can find points as close as we want to 0 and 1. Let M be at least the maximum absolute value of these points we've picked and ensure we pick M such that at M and -M, the cdf of X is continuous and does not jump. Then $P(|X|>M)<\varepsilon$. But since X_n converges in distribution to X, the cdfs converge pointwise, so at -M and M where the cdf of X is within $\frac{\varepsilon}{2}$ of 0 and 1 respectively, by definition of convergence we can find an N such that for all X_n with n>N, the cdf of X_n at M and -M get as close as we want to the cdf of X at M and -M. Specifically, make it so close that it is still within $\frac{\varepsilon}{2}$ of 0 and 1 respectively. Then we have shown that for any ε>0 there exists M and N such that if n>N then $P(|X_n|>M)<\varepsilon$.

Now for any ε>0 pick M and N_1 such that if $n>N_1$ we have that $P(|X_n|>M)<\frac{\varepsilon}{4}$. Now let 1_S be the indicator function of the set S, ie the function that returns 1 for points inside the set S and 0 everywhere else. Then $E\left(e^{itX_n}\right)-E\left(e^{itX}\right)$

$$=E\left(e^{itX_n}1_{|X_n|\leq M}+e^{itX_n}1_{|X_n|>M}\right)-E\left(e^{itX}1_{|X|\leq M}+e^{itX}1_{|X|>M}\right)$$

We simply have that the expectation overall is the expectation when inside a set plus the expectation when outside that set, this is obvious. We split this even further to get the following:

$$=E\left(e^{itX_n}1_{|X_n|\leq M}\right)+E\left(e^{itX_n}1_{|X_n|>M}\right)-E\left(e^{itX}1_{|X|\leq M}\right)-E\left(e^{itX}1_{|X|>M}\right)$$

Now, recall that we have the triangle inequality for both real numbers and integrals. Since expectations are actually defined in terms of integrals, we do have the inequality |E(x)|≤E(|x|) directly from the integral inequality. So, we have:

$$\left|E\left(e^{itX_n}\right) - E\left(e^{itX}\right)\right|$$

$$= \left|E\left(e^{itX_n}1_{|X_n|\leq M}\right) + E\left(e^{itX_n}1_{|X_n|>M}\right) - E\left(e^{itX}1_{|X|\leq M}\right) - E\left(e^{itX}1_{|X|>M}\right)\right|$$

By what we just did

$$\leq \left|E\left(e^{itX_n}1_{|X_n|\leq M}\right) - E\left(e^{itX}1_{|X|\leq M}\right)\right| + \left|E\left(e^{itX_n}1_{|X_n|>M}\right)\right| + \left|E\left(e^{itX}1_{|X|>M}\right)\right|$$

By the normal triangle inequality

$$\leq \left|E\left(e^{itX_n}1_{|X_n|\leq M}\right) - E\left(e^{itX}1_{|X|\leq M}\right)\right| + E\left(|e^{itX_n}|1_{|X_n|>M}\right) + E\left(|e^{itX}|1_{|X|>M}\right)$$

By the triangle inequality for expectations. However, $\left|e^{itX_n}\right| = |e^{itX}| = 1$ so since this is only when $|X_n| > M$, we have that

$$\left|E\left(e^{itX_n}\right) - E\left(e^{itX}\right)\right| \leq \left|E\left(e^{itX_n}1_{|X_n|\leq M}\right) - E\left(e^{itX}1_{|X|\leq M}\right)\right| + \frac{\varepsilon}{4}$$

Which helps a lot, since the goal is to make this term less than a full epsilon so we have that the characteristic functions get as close together as we want so we have the desired result.

Now, e^{itx} on [-M,M] is a continuous function on a closed bounded interval so it is uniformly continuous. Therefore we can find a δ small enough that on any interval of length at most δ, e^{itx} is within a ball in the complex plane of diameter $\frac{\varepsilon}{4}$, which we can do by considering real and imaginary parts separately and making sure they are in an interval of length $\frac{\varepsilon}{4\sqrt{2}}$, ensuring the whole thing is in a square within the circle that we need. Since this δ works everywhere by uniform continuity, we can find a finite partition $-M = a_0 < a_1 < \cdots < a_m = M$ where the distance between each x is at most δ. We take care to make sure that each a is a point where the cdf is continuous, and this will have the property that if $g(x) := e^{itx}$, then for any x and y between a_{j-1} and a_j we have that $|g(x) - g(y)| < \frac{\varepsilon}{4}$. In particular, for any ξ_j with $a_{j-1} < \xi_j \leq a$, which I will now pick arbitrarily for each interval j, we have that the difference between $c_j := g(\xi_j)$ and g(anything else in that interval) is bounded above in absolute value by $\frac{\varepsilon}{4}$. I will now define the simple function $s(x) := \sum_{j=1}^{m} c_j 1_{(a_{j-1},a_j]}$, then since $|c_j|$ is 1 since it is equal to $e^{i*stuff}$, we have that it is always the case that $|g(x) - s(x)| < \frac{\varepsilon}{4}$. Now lets work on the $\left|E\left(e^{itX_n}1_{|X_n|\leq M}\right) - E\left(e^{itX}1_{|X|\leq M}\right)\right|$ term which we hope to bound by $\frac{3\varepsilon}{4}$ in order to be done: This term is equal to $\left|\int_{-M}^{M} g(x)f_n(x)dx - \int_{-M}^{M} g(x)f(x)dx\right|$ because of the definition of expected value in terms of integrals and the definition of g(x) and the fact that the indicator function makes this be on the interval [-M,M]. Here f_n is the pdf of X_n and f is the pdf of x. I can write the term as

$$\left|\int_{-M}^{M} (g-s)(x)f_n(x)dx - \int_{-M}^{M} (g-s)(x)f(x)dx + \int_{-M}^{M} s(x)f_n(x)dx - \int_{-M}^{M} s(x)f(x)dx\right|$$

Then both versions of the triangle inequality mean that this is

$$\leq \int_{-M}^{M} |(g-s)(x)|f_n(x)dx + \int_{-M}^{M} |(g-s)(x)|f(x)dx + \left|\int_{-M}^{M} s(x)f_n(x)dx - \int_{-M}^{M} s(x)f(x)dx\right|$$

Since f is positive and real so we can pull it out of the absolute value (this also justifies the inequality for expectations in general). Now using the bounds we got earlier we can simplify this even further:

$$\leq \int_{-M}^{M} \frac{\varepsilon}{4} f_n(x)dx + \int_{-M}^{M} \frac{\varepsilon}{4} f(x)dx + \left| \int_{-M}^{M} s(x)f_n(x)dx - \int_{-M}^{M} s(x)f(x)dx \right|$$

$$\leq \frac{\varepsilon}{4} \int_{-M}^{M} f_n(x)dx + \frac{\varepsilon}{4} \int_{-M}^{M} f(x)dx + \left| \int_{-M}^{M} s(x)f_n(x)dx - \int_{-M}^{M} s(x)f(x)dx \right|$$

$$\leq \frac{\varepsilon}{2} + \left| \int_{-M}^{M} s(x)f_n(x)dx - \int_{-M}^{M} s(x)f(x)dx \right|$$

Since those integrals cannot be greater than 1. So we just have to make sure that

$\left| \int_{-M}^{M} s(x)f_n(x)dx - \int_{-M}^{M} s(x)f(x)dx \right| \leq \frac{\varepsilon}{4}$

Then we're done, since we will have that the characteristic function converges for every fixed t and therefore converges pointwise.

To do this, we note that since s is equal to c_j on each interval (a_{j-1}, a_j), we have that the term in question can be written as

$$\left| \sum_{j=1}^{m} \int_{a_{j-1}}^{a_j} c_j f_n(x)dx - \sum_{j=1}^{m} \int_{a_{j-1}}^{a_j} c_j f(x)dx \right|$$

But the integral of f and f_n is the cdf, which I'll call F and F_n respecively, so the term becomes

$$\left| \sum_{j=1}^{m} c_j [F_n(a_j) - F_n(a_{j-1})] - \sum_{j=1}^{m} c_j [F(a_j) - F(a_{j-1})] \right|$$

$$= \left| \sum_{j=1}^{m} c_j \left[\left(F_n(a_j) - F_n(a_{j-1}) \right) - \left(F(a_j) - F(a_{j-1}) \right) \right] \right|$$

$$\leq \sum_{j=1}^{m} \left| c_j \left[\left(F_n(a_j) - F_n(a_{j-1}) \right) - \left(F(a_j) - F(a_{j-1}) \right) \right] \right|$$

By the triangle inequality

$$= \sum_{j=1}^{m} \left| \left[\left(F_n(a_j) - F_n(a_{j-1}) \right) - \left(F(a_j) - F(a_{j-1}) \right) \right] \right|$$

Since $|c_j| = 1$

$$\leq \sum_{j=1}^{m} \left[|F_n(a_j) - F(a_j)| + |F_n(a_{j-1}) - F(a_{j-1})| \right]$$

By the triangle inequality.

Now, since each a_j is a point where F is continuous, it means F_n converges to F there. So if n is large enough $F_n(a_j)$ will be within $\frac{\varepsilon}{8m}$ of $F(a_j)$ if we pick $n > N_j$. Since there are finitely many (m, the number of intervals in our partition) N_j's, simply pick the largest one, then we have that n is large enough so that our sum is

$$\leq \sum_{j=1}^{m}\left[\frac{\varepsilon}{8m}+\frac{\varepsilon}{8m}\right] = \sum_{j=1}^{m}\left[\frac{\varepsilon}{4m}\right] = \frac{\varepsilon}{4}$$

So done (with the lemma).

Lemma 2: We have the other direction (Characteristic functions determine the distribution and convergence in cf implies convergence in distirbution) for random vectors.

Proof:

A random vector is defined by a probability density function in $\mathbb{R}^k$, ie sets of k real numbers. We define the cf of a random vector Y to be a function that takes in a vector t and outputs $E\left(e^{i(t.Y)}\right)$. If the space is d dimensional, we define

$$I_\varepsilon := \frac{1}{(2\pi)^d}\int_{R^d} \phi_y(t) e^{-\varepsilon|t|^2} \prod_{j=1}^{d} \frac{e^{-ia_jt_j} - e^{-ib_jt_j}}{it_j} dt$$

Where that big scary looking thing just means product.

$$= \frac{1}{(2\pi)^d}\int_{R^d}\int_{R^d} f(y) e^{i(t.y)} \prod_{j=1}^{d} \frac{e^{-ia_jt_j} - e^{-ib_jt_j}}{it_j} dy e^{-\varepsilon|t|^2} dt$$

Note: Each $\frac{e^{-ia_jt_j}-e^{-ib_jt_j}}{it_j}$ is bounded for the same reasons as bsfore, so we have the product of bounded things times a thing which integrates to 1 in the inner dy integral, so that's bounded. Now in total we have a bounded thing times the integral of $e^{-\varepsilon|t|^2}$, which is finite, so we have the conditions to swap the integrals around.

Since $e^{-\varepsilon|t|^2}$ is just the product of $e^{-\varepsilon(t_j^2)}$, we can simplify I_ε as follows:

$$= \frac{1}{(2\pi)^d}\int_{R^d} f(y) \int_{R^d} e^{i(t.y)} \prod_{j=1}^{d} \frac{e^{-ia_jt_j} - e^{-ib_jt_j}}{it_j} e^{-\varepsilon|t|^2} dt dy$$

$$= \int_{R^d} f(y) \int_{R^d} \prod_{j=1}^{d} \frac{1}{2\pi} e^{i(t_jy_j)} \frac{e^{-ia_jt_j} - e^{-ib_jt_j}}{it_j} e^{-\varepsilon(t_j)^2} dt dy$$

We know what each term looks like from earlier, so we end up with

$$= \int_{R^d} f(y) \int_{R^d} \prod_{j=1}^{d} \frac{1}{2}\left(\operatorname{erf}\left(\frac{y_j - a_j}{2\sqrt{\varepsilon}}\right) - \operatorname{erf}\left(\frac{y_j - b_j}{2\sqrt{\varepsilon}}\right)\right) dt dy$$

Which in the limit vanishes exactly when we are outside the (a,b) high dimensional rectangle and goes to 1 when we are inside it, for the same reasons – Each term goes to 1 or 0 and the product is 1 only when all terms go to 1.

Now at last we have, because of dominated convergence again, the same result for random vectors: the characteristic function determines the distribution.

We also need the result for random vectors that convergence in cf implies convergence in distribution.

To do this, we will define yet another function. Let $L_\varepsilon(t) \coloneqq \frac{e^{-\varepsilon|t|^2}}{2\pi}\prod_{j=1}^{d}\frac{e^{-ia_jt_j}-e^{-ib_jt_j}}{it_j}$, then $I_\varepsilon = \int_{R^d}\phi(t)L_\varepsilon(t)dt$. With exactly the same proof as the univariate case, all the hypotheses for the dominated convergence theorem apply, so the limit of $\int_{-\infty}^{\infty}\phi_n(t)L_\varepsilon(t)dt$ is indeed I_ε. Therefore, letting ε approach 0, we have that the probability our distribution lands between a and b approaches that probability for a pdf with cf ϕ if the cf's converge to ϕ. So done.

The proof of the main theorem is very short now.

Suppose we have a vector t, and $Z_n \coloneqq t.Y_n$ and $Z \coloneqq t.Y$. Suppose also that Z_n converges in distribution to Z. Then we will show that Y_n converges in distribution to Y. $\phi_{Y_n}(t) = E\left(e^{i(t.Y_n)}\right) = E\left(e^{i*1*(Z_n)}\right) = \phi_{Z_n}(1)$. Since Z converges in distribution, we have that as n goes to infinity $\phi_{Y_n}(t) = \phi_{Z_n}(1) \to \phi_Z(1) = E\left(e^{iZ}\right) = E\left(e^{i(t.Y)}\right) = \phi_Y(t)$. This is true for every fixed vector t, so since we know that convergence in characteristic functions implies convergence in distribution even for vectors, the result follows.