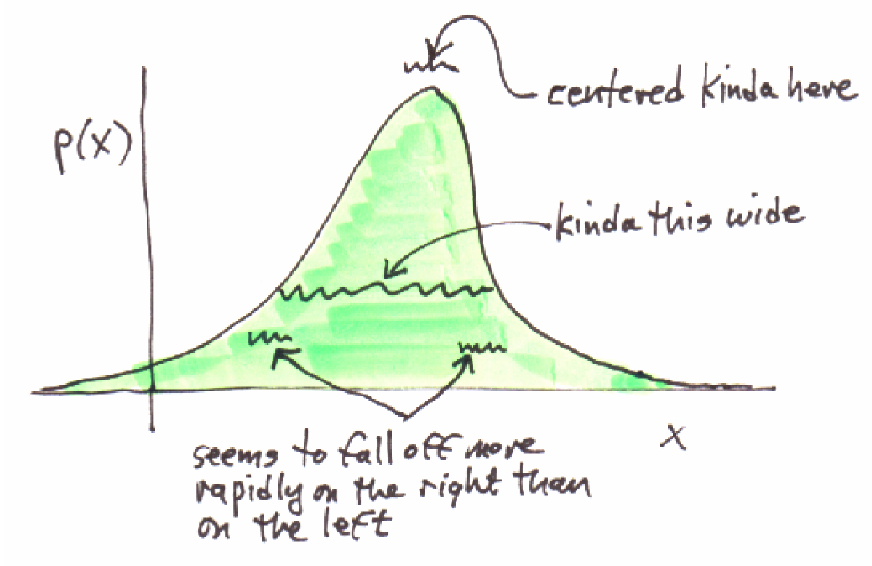
4th IMPRS Astronomy Summer School

Drawing Astrophysical Inferences from Data Sets

William H. Press
The University of Texas at Austin

Lecture 2

Many, though not all, common distributions look sort-of like this:



We already saw the beta distribution with $\alpha, \beta > 0$ as an example on the interval [0,1]. We'll see more examples soon.

Suppose we want to <u>summarize</u> *p(x)* by a single number *a*, its "center". Let's find the value *a* that minimizes the mean-square distance of the "typical" value *x*:

expectation notation:
$\langle \text{anything} \rangle \equiv \int_x (\text{anything})\, p(x) dx$
expectation is linear, etc.

$$\text{minimize:} \quad \Delta^2 \equiv \left\langle (x-a)^2 \right\rangle = \left\langle x^2 - 2ax + a^2 \right\rangle$$
$$= \left(\left\langle x^2 \right\rangle - \langle x \rangle^2 \right) + \left(\langle x \rangle - a \right)^2$$

This is the variance Var(x), but all we care about here is that it doesn't depend on a.

(in physics this is called the "parallel axis theorem")

The minimum is obviously $a = \langle x \rangle$. (Take derivative wrt a and set to zero if you like mechanical calculations.)

Higher moments, centered moments defined by

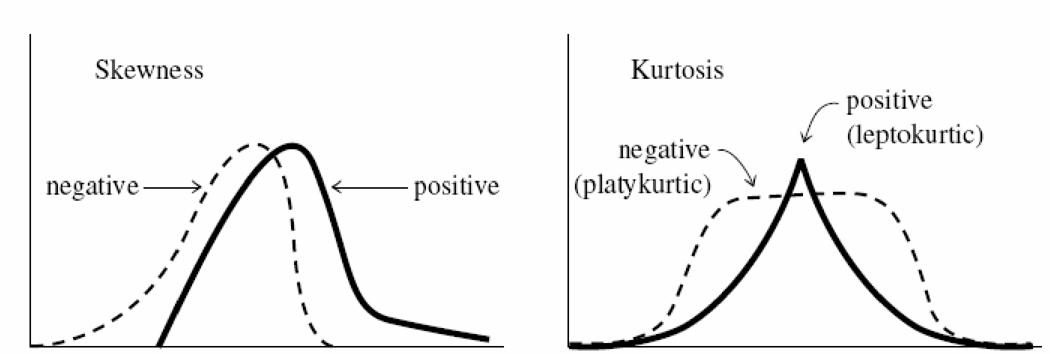
$$\mu_i \equiv \langle x^i \rangle = \int x^i \, p(x) dx$$
$$M_i \equiv \left\langle (x - \langle x \rangle)^i \right\rangle = \int (x - \langle x \rangle)^i \, p(x) dx$$

The centered second moment M_2, the variance, is by far most useful

$$M_2 \equiv \mathrm{Var}(x) \equiv \left\langle (x - \langle x \rangle)^2 \right\rangle = \langle x^2 \rangle - \langle x \rangle^2$$
$$\sigma(x) \equiv \sqrt{\mathrm{Var}(x)}$$

"standard deviation" summarizes a distribution's half-width (r.m.s. deviation from the mean)

Third and fourth moments also have "names"



But generally wise to be cautious about using high moments. Otherwise perfectly good distributions don't have them at all (divergent). And (related) it can take a <u>lot</u> of data to measure them accurately.

Mean and variance are additive over independent random variables:

$$\overline{(x+y)} = \bar{x} + \bar{y} \qquad \mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(x)$$

note “bar” notation, equivalent to < >

Certain combinations of higher moments are also additive. These are called semi-invariants.

$$I_2 = M_2 \qquad I_3 = M_3 \qquad I_4 = M_4 - 3M_2^2$$

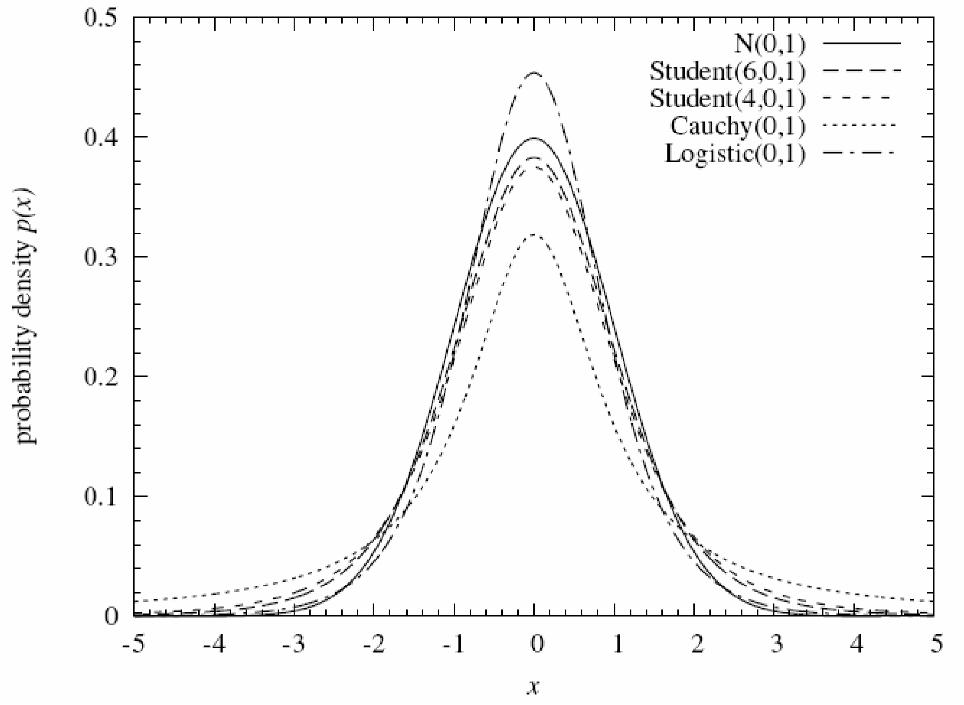
$$I_5 = M_5 - 10M_2M_3 \qquad I_6 = M_6 - 15M_2M_4 - 10M_3^2 + 30M_2^3$$

Skew and kurtosis are dimensionless combinations of semi-invariants

$$\mathrm{Skew}(x) = I_3/I_2^{3/2} \qquad \mathrm{Kurt}(x) = I_4/I_2^2$$

A Gaussian has all of its semi-invariants higher than I_2 equal to zero.
A Poisson distribution has all of its semi-invariants equal to its mean.

This is a good time to review some standard (i.e., frequently occurring) distributions:



Normal (Gaussian):

$$x \sim \mathrm{N}(\mu, \sigma), \qquad \sigma > 0$$
$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)$$

tails fall off "as fast as possible"

Cauchy (Lorentzian):

$$x \sim \mathrm{Cauchy}(\mu, \sigma), \qquad \sigma > 0$$
$$p(x) = \frac{1}{\pi\sigma}\left(1 + \left[\frac{x-\mu}{\sigma}\right]^2\right)^{-1}$$

tails fall off "as slowly as possible"

Student:

$$t \sim \text{Student}(\nu, \mu, \sigma), \qquad \nu > 0,\ \sigma > 0$$

$$p(t) = \frac{\Gamma(\frac{1}{2}[\nu+1])}{\Gamma(\frac{1}{2}\nu)\sqrt{\nu\pi}\sigma}\left(1 + \frac{1}{\nu}\left[\frac{t-\mu}{\sigma}\right]^2\right)^{-\frac{1}{2}(\nu+1)}$$

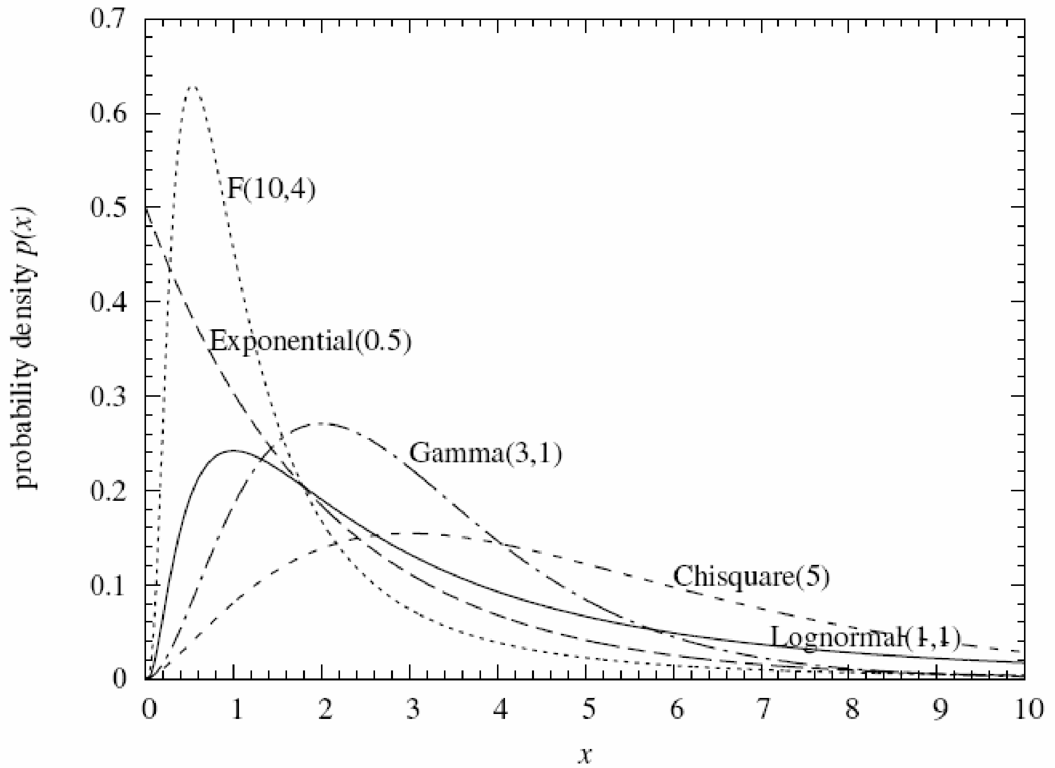
"bell shaped" but you get to specify the power with which the tails fall off. Normal and Cauchy are limiting cases. (Also occurs in some statistical tests.)

note that σ is not (quite) the standard deviation!

we'll see uses for "heavy-tailed" distributions later

"Student" was actually William Sealy Gosset (1876-1937), who spent his entire career at the Guinness brewery in Dublin, where he rose to become the company's Master Brewer.

Common distributions on positive real line:



Exponential:

$$x \sim \text{Exponential}(\beta), \qquad \beta > 0$$
$$p(x) = \beta \exp(-\beta x), \qquad x > 0$$

Lognormal:

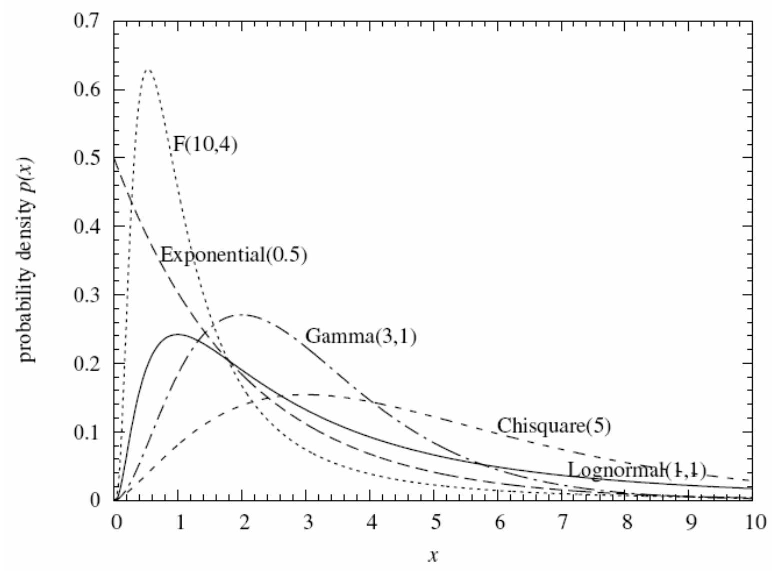
$$x \sim \text{Lognormal}(\mu, \sigma), \qquad \sigma > 0$$

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{1}{2}\left[\frac{\log(x) - \mu}{\sigma}\right]^2\right), \qquad x > 0 \tag{6.14.31}$$

Note the required extra factor of x^{-1} in front of the exponential: The density that is "normal" is $p(\log x) d \log x$.

While μ and σ are the mean and standard deviation in $\log x$ space, they are *not* so in x space. Rather,

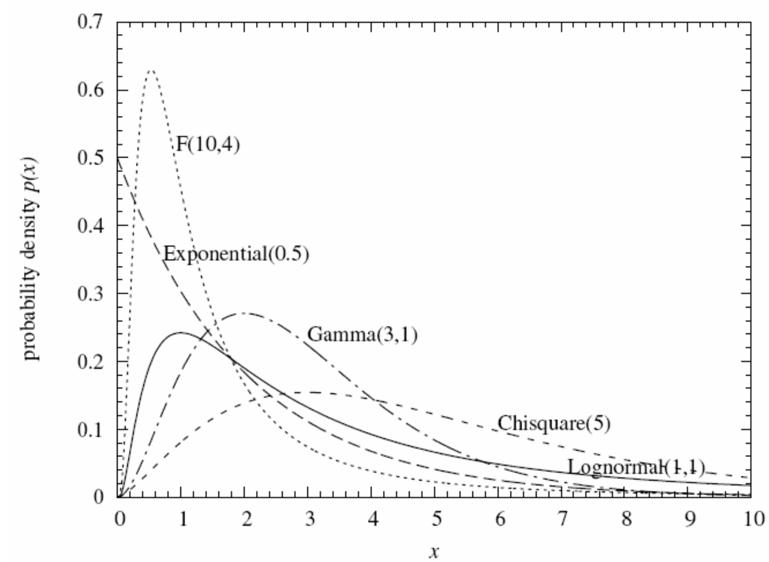
$$\begin{aligned} \text{Mean}\{\text{Lognormal}(\mu, \sigma)\} &= e^{\mu + \frac{1}{2}\sigma^2} \\ \text{Var}\{\text{Lognormal}(\mu, \sigma)\} &= e^{2\mu}\, e^{\sigma^2} (e^{\sigma^2} - 1) \end{aligned} \tag{6.14.32}$$



Gamma distribution:

$$x \sim \text{Gamma}(\alpha, \beta), \qquad \alpha > 0,\ \beta > 0$$
$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad x > 0$$

$$\text{Mean}\{\text{Gamma}(\alpha, \beta)\} = \alpha/\beta$$
$$\text{Var}\{\text{Gamma}(\alpha, \beta)\} = \alpha/\beta^2$$



When $\alpha \geq 1$ there is a single mode at $x = (\alpha - 1)/\beta$

- Gamma and Lognormal are both commonly used as convenient 2-parameter fitting functions for “peak with tail” positive distributions.
- Both have parameters for peak location and width.
- Neither has a separate parameter for how the tail decays.
  - Gamma: exponential decay
  - Lognormal: long-tailed (exponential of square of log)

Chi-square distribution (we'll use this a lot!)

Has only one parameter ν that determines both peak location and width.
ν is often an integer, called "number of degrees of freedom" or "DF"

the independent variable is χ^2, not χ

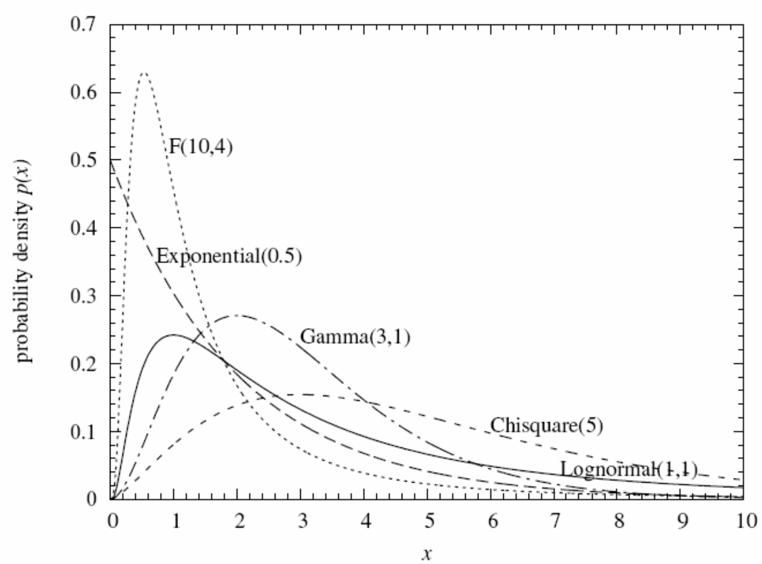
$$\chi^2 \sim \text{Chisquare}(\nu), \qquad \nu > 0$$

$$p(\chi^2)d\chi^2 = \frac{1}{2^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu)}(\chi^2)^{\frac{1}{2}\nu-1}\exp\left(-\tfrac{1}{2}\chi^2\right)d\chi^2, \qquad \chi^2 > 0$$

$$\text{Mean}\{\text{Chisquare}(\nu)\} = \nu$$
$$\text{Var}\{\text{Chisquare}(\nu)\} = 2\nu$$

When $\nu \geq 2$ there is a single mode at $\chi^2 = \nu - 2$

It's actually just a special case of Gamma, namely Gamma(ν/2,1/2)

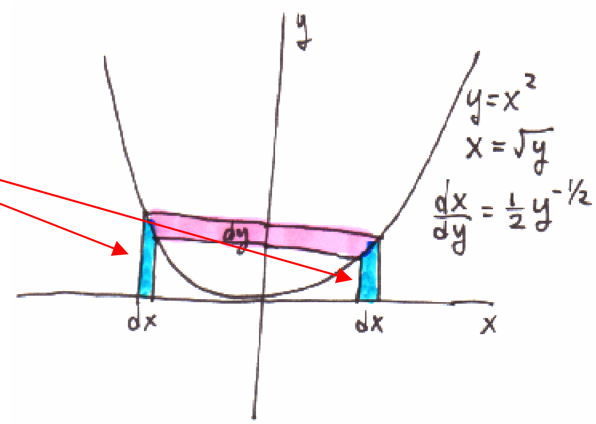


A deviate from $\mathrm{N}(0,1)$ is called a t-value.

$\mathrm{Chisquare}(\nu)$ is exactly the distribution of the sum of the squares of ν t-values.

Let's prove the case of ν=1:

$$p_Y(y)\,dy = 2p_X(x)\,dx$$



$$\text{So, } p_Y(y) = y^{-1/2} p_X(y^{1/2}) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}$$
$$\sim \mathrm{Chisquare}(1)$$

Why this will be important: We will know the distribution of any "statistic" that is the sum of a known number of t^2-values.

Characteristic function of a distribution

The characteristic function of a distribution is its Fourier transform.

$$\phi_X(t) \equiv \int_{-\infty}^{\infty} e^{itx} p_X(x) dx$$

(Statisticians often use notational convention that *X* is a random variable, *x* its value, $p_X(x)$ its distribution.)

$$\phi_X(0) = 1$$

$$\phi'_X(0) = \int ix p_X(x) dx = i\mu$$

$$-\phi''_X(0) = \int x^2 p_X(x) dx = \sigma^2 + \mu^2$$

So, the coefficients of the Taylor series expansion of the characteristic function are the (uncentered) moments.

Properties of characteristic functions:

Addition of independent r.v.'s:

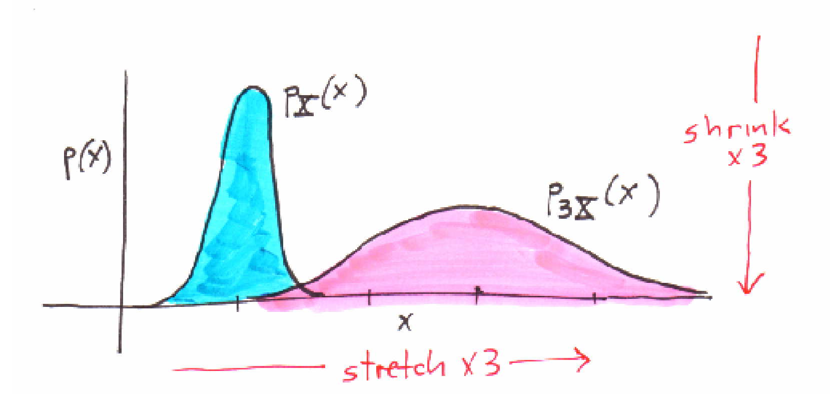
$$\text{let } S = X + Y$$

$$p_S(s) = \int p_X(u)p_Y(s-u)du$$

$$\phi_S(t) = \phi_X(t)\phi_Y(t)$$

(Fourier convolution theorem.)

Scaling law for r.v.'s:



Scaling law for characteristic functions:

$$\begin{aligned}
\phi_{aX}(t) &= \int e^{itx}\underline{p_{aX}(x)}dx \\
&= \int e^{itx}\underline{\frac{1}{a}p_X\left(\frac{x}{a}\right)}dx \\
&= \int e^{i(at)(x/a)}p_X\left(\frac{x}{a}\right)\frac{dx}{a} \\
&= \phi_X(at)
\end{aligned}$$

Proof of convolution theorem:

$$\left.\begin{aligned}\phi_X(t) &\equiv \int_{-\infty}^{\infty} e^{itx} p_X(x)\,dx \\ p_X(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(t) e^{-itx}\,dt\end{aligned}\right\}$$

Fourier transform pair

$$\begin{aligned}p_S(s) &= \int_{-\infty}^{\infty} p_X(u) p_Y(s-u)\,du \\ &= \int_{-\infty}^{\infty} p_X(u)\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_Y(t) e^{-it(s-u)}\,dt\right] du \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_Y(t) e^{-its}\left[\int_{-\infty}^{\infty} p_X(u) e^{itu}\,du\right] dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_Y(t)\phi_X(t) e^{-its}\,dt\end{aligned}$$

So, $\quad \phi_S(t) = \phi_Y(t)\phi_X(t)$

What's the characteristic function of a Gaussian?

Tell Mathematica that sig is positive. Otherwise it gives "cases" when taking the square root of sig^2

In[14]:=
```
$Assumptions = $Assumptions && (sig > 0)
```

In[15]:=
```
p = (1 / (Sqrt[2 Pi] sig)) Exp[-(1 / 2) ((x - mu) / sig) ^2]
```

Out[15]=

$$\frac{e^{-\frac{(-\mathrm{mu}+x)^2}{2\,\mathrm{sig}^2}}}{\sqrt{2\pi}\,\mathrm{sig}}$$

In[16]:=
```
Integrate[p, {x, -Infinity, Infinity}]
```

Out[16]=

$$1$$

In[17]:=
```
Integrate[p Exp[I t x], {x, -Infinity, Infinity}]
```

Out[17]=

$$e^{i\,\mathrm{mu}\,t-\frac{\mathrm{sig}^2\,t^2}{2}}$$

What's the characteristic function of $\text{Chisquare}(\nu)$?

```
In[9]:= pchi2 = (1 / (2 ^ (nu / 2) Gamma[nu / 2])) y ^ (nu / 2 - 1) Exp[-y / 2]
```

Out[9]= $\dfrac{2^{-nu/2}\, e^{-y/2}\, y^{-1+\frac{nu}{2}}}{\text{Gamma}\left[\frac{nu}{2}\right]}$

```
In[10]:= Integrate[pchi2, {y, 0, Infinity}, GenerateConditions → False]
```

Out[10]= 1

```
In[11]:= Integrate[pchi2 Exp[I t y], {y, 0, Infinity},
          GenerateConditions → False]
```

Out[11]= $(1 - 2\, i\, t)^{-nu/2}$

Since we already proved that $\nu=1$ is the distribution of a single t^2-value, this proves that the general ν case is the sum of ν t^2-values.

Cauchy distribution has ill-defined mean and infinite variance, but it has a perfectly good characteristic function:

$$x \sim \text{Cauchy}(\mu, \sigma), \qquad \sigma > 0$$
$$p(x) = \frac{1}{\pi\sigma}\left(1 + \left[\frac{x-\mu}{\sigma}\right]^2\right)^{-1}$$

Matlab and Mathematica both sadly fails at computing the characteristic function of the Cauchy distribution, but you can use fancier methods* and get:

$$\phi_{\text{Cauchy}}(t) = e^{i\mu t - \sigma|t|}$$

note non-analytic at t=0

*If t>0, close the contour in the upper 1/2-plane with a big semi-circle, which adds nothing. So the integral is just the residue at the pole $(x-\mu)/\sigma=i$, which gives $\exp(-\sigma t)$. Similarly, close the contour in the lower 1/2-plane for t<0, giving $\exp(\sigma t)$. So answer is $\exp(-|\sigma t|)$. The factor $\exp(i\mu t)$ comes from the change of x variable to $x-\mu$.