CS395T
Computational Statistics with Application to Bioinformatics

Prof. William H. Press
Spring Term, 2011
The University of Texas at Austin

Lecture 5

The Central Limit Theorem is the reason that the Normal (Gaussian) distribution is uniquely important. We need to understand where it does <u>and doesn't</u> apply.

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.)

$$\phi_X(0) = 1$$
$$\phi'_X(0) = \int ixp_X(x)dx = i\langle X \rangle$$
$$-\phi''_X(0) = \int x^2 p_X(x)dx = \mathrm{Var}(X) + \langle X \rangle^2$$

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

"The c.f. of the sum of independent r.v.'s is the product of their individual c.f.'s"

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

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

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

Last line follows immediately from the Fourier convolution theorem. (In fact, it <u>is</u> the Fourier convolution theorem!)

Proof:

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

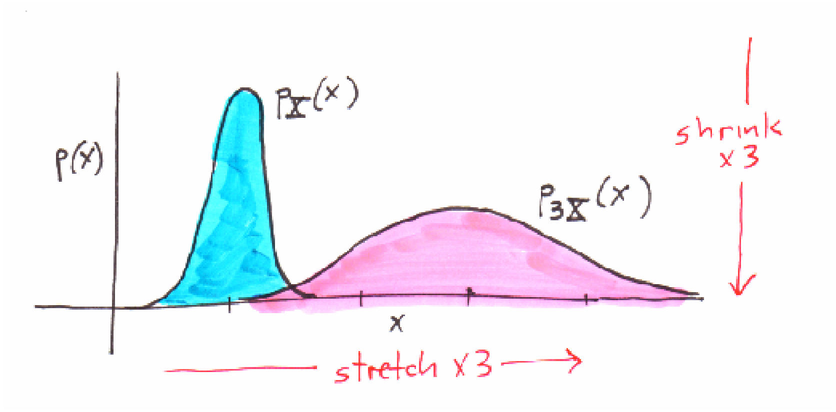
$$p_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(t)e^{-itx}dt$$

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

Scaling law for r.v.'s:



Scaling law for characteristic functions:

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

What's the characteristic function of a Gaussian?

```
syms x mu pi t sigma
p = exp(-(x-mu)^2 / (2*sigma^2)) / (sqrt(2*pi)*sigma)
p =
1/2*exp(-1/2*(x-mu)^2/sigma^2)*2^(1/2)/pi^(1/2)/sigma
norm = int(p,x,-Inf,Inf)
norm =
1
cf = simplify(int(p*exp(i*t*x),x,-Inf,Inf))
cf =
exp(1/2*i*t*(2*mu+i*t*sigma^2))
```

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

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

Out[15]=

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

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

Out[16]=

1

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

Out[17]=

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

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

So the CF of a Gaussian is itself a Gaussian:

$$\phi_{\text{Normal}}(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$$

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

Recall:

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

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

Matlab and Mathematica both (sadly) fail at computing the characteristic function of the Cauchy distribution, but you can use old-fashioned wetware methods* (see proof posted on forum) and get:

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

note non-analytic at t=0

*Or social networking! My co-author Saul says: "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 $x-\mu$."

Central Limit Theorem

Let $S = \frac{1}{N}\sum X_i = \sum \frac{X_i}{N}$ with $\langle X_i \rangle \equiv 0$

Can always subtract off the means, then add back later.

Then

$$\phi_S(t) = \prod_i \phi_{X_i/N}(t) = \prod_i \phi_{X_i}\left(\frac{t}{N}\right)$$

$$= \prod_i \left(1 - \tfrac{1}{2}\sigma_i^2 \frac{t^2}{N^2} + \cdots \right)$$

Whoa! It better have a convergent Taylor series around zero! (Cauchy doesn't, e.g.)

$$= \exp\left[\sum_i \ln\left(1 - \tfrac{1}{2}\sigma_i^2 \frac{t^2}{N^2} + \cdots \right)\right]$$

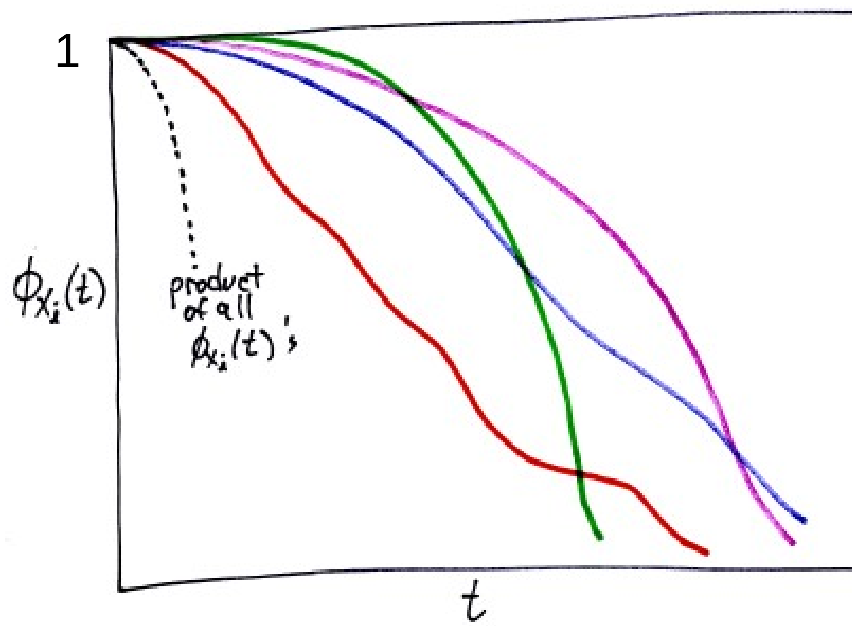
These terms decrease with N, but how fast?

$$\approx \exp\left[-\frac{1}{2}\left(\frac{1}{N^2}\sum_i \sigma_i^2\right)t^2 + \cdots\right]$$

So, S is normally distributed

$$p_S(\cdot) \sim \mathrm{Normal}(0, \tfrac{1}{N^2}\sum \sigma_i^2)$$

Intuitively, the product of a lot of arbitrary functions that all start at 1 and have zero derivative looks like this:



Because the product falls off so fast, it loses all memory of the details of its factors except the starting value 1 and fact of zero derivative. In characteristic function space that's basically the CLT.

CLT is usually stated about the sum of RVs, not the average, so

$$p_S(\cdot) \sim \mathrm{Normal}(0, \tfrac{1}{N^2}\sum \sigma_i^2)$$

Now, since

$NS = \sum X_i$ and $\mathrm{Var}(NS) = N^2\mathrm{Var}(S)$

it follows that the simple sum of a large number of r.v.'s is normally distributed, with variance equal to the sum of the variances:

$$p_{\sum X_\mathrm{i}}(\cdot) \sim \mathrm{Normal}(0, \sum \sigma_i^2)$$

If N is large enough, and if the higher moments are well-enough behaved, and if the Taylor series expansion exists!

Also beware of borderline cases where the assumptions technically hold, but convergence to Normal is slow and/or highly nonuniform. (This can affect p-values for tail tests, as we will soon see.)

Since Gaussians are so universal, let's learn estimate the parameters μ and σ of a Gaussian from a set of points drawn from it:

For now, we'll just find the maximum of the posterior distribution of (μ,σ), given some data, for a uniform prior. This is called "maximum a posteriori (MAP)" by Bayesians, and "maximum likelihood (MLE)" by frequentists.

The data is: $x_i,\ i = 1, \ldots, N$

The statistical model is: $$P(\mathbf{x}|\mu,\sigma) = \prod_i \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$

The posterior estimate is: $$P(\mu,\sigma|\mathbf{x}) \propto \frac{1}{\sqrt{2\pi}\sigma^N} e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2} \times \cancel{P(\mu,\sigma)}$$ uniform

Now find the MAP (MLE):

$$0 = \frac{\partial P}{\partial \mu} = \frac{P}{\sigma^2}\left(\sum_i x_i - N\mu\right) \;\Rightarrow\; \mu = \frac{1}{N}\sum_i x_i$$

Ha! The MAP mean is the sample mean, the MAP variance is the sample variance!

$$0 = \frac{\partial P}{\partial \sigma} = \frac{P}{\sigma^3}\left[-N\sigma^2 + \sum_i (x_i-\mu)^2\right] \;\Rightarrow\; \sigma^2 = \frac{1}{N}\sum_i (x_i-\mu)^2$$

It won't surprise you that I did the algebra by computer, in Mathematica:

```
p =
 (1 / s ^N)
  Exp[-(1 / (2 s ^2)) Sum[(x[i] - mu) ^2, {i, 1, N}]]
```

$$e^{-\frac{\sum_{i=1}^{N}(-mu+x[i])^2}{2 s^2}} s^{-N}$$

```
Simplify[D[p, mu]]
```

$$-\frac{1}{2} e^{-\frac{\sum_{i=1}^{N}(-mu+x[i])^2}{2 s^2}} s^{-2-N} \sum_{i=1}^{N} -2\ (-mu + x[i])$$

```
Simplify[D[p, s]]
```

$$e^{-\frac{\sum_{i=1}^{N}(-mu+x[i])^2}{2 s^2}} s^{-3-N} \left(-N s^2 + \sum_{i=1}^{N} (-mu + x[i])^2\right)$$

(I don't know if MATLAB can deal with symbolic sums. Could someone find out?)