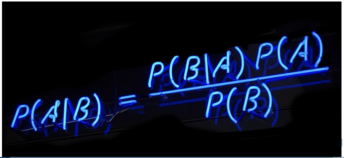
Statistics in Data Analysis

All you ever wanted to know about statistics but never dared to ask

part 3

Pawel Brückman de Renstrom
(`pawel.bruckman@ifj.edu.pl`)

March 20, 2024



Question from the previous lecture

1. Suppose two independent measurements of the same quantity gave the following results:

$$x_1 \pm \sigma_1 \quad \text{and} \quad x_2 \pm \sigma_2$$

Take the weighted mean to be $\bar{x} = wx_1 + (1-w)x_2$. Find the w which minimizes the error on the mean, hence provide expressions for the weighted mean $\bar{x}$ and its variance $\sigma^2_{\bar{x}}$.

Solution to be sent to me before the next lecture

Solution

We have to express the variance of the weighted mean

$$\bar{x} = wx_1 + (1-w)x_2$$

using the recipe for error propagation:

$$\begin{aligned} Var(\bar{x}) &= \left(\frac{\partial \bar{x}}{\partial x_1}\right)^2 \sigma_1^2 + \left(\frac{\partial \bar{x}}{\partial x_2}\right)^2 \sigma_2^2 \\ &= w^2\sigma_1^2 + (1-w)^2\sigma_2^2 \end{aligned}$$

and minimise it w.r.t. the weight w.

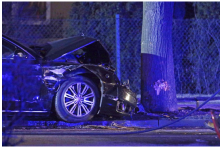
$$\begin{aligned} \frac{\partial Var(\bar{x})}{\partial w} &= 2w\sigma_1^2 - 2(1-w)\sigma_2^2 = 0 \\ \Longrightarrow w &= \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \end{aligned}$$

Hence we get:

$$\bar{x} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2} \quad \text{and} \quad Var(\bar{x}) = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \qquad \therefore$$

Accidents happen...

Exponential distribution



- Imagine a fleet of governmental limousines circulating daily. For any of them there is a probability λ to be crashed in an accident in a day. We start with N_0 limousines. What is the time p.d.f. of the accidents?
- For many circulating cars, accident rate is simply proportional to their number:

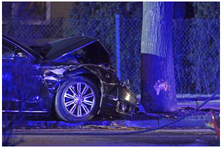
$$\frac{dN}{dt} = -\lambda N \quad \Rightarrow \quad \frac{dN}{N} = -\lambda dt \qquad \Big/ \int$$

$$\ln N = -\lambda t + C \quad \Rightarrow \quad N(t) = N_0 e^{-\lambda t} \quad \Rightarrow \quad \frac{dN(t)}{dt} = -\lambda N_0 e^{-\lambda t} \tag{1}$$

...so we observe an exponential decay of the fleet.

Accidents happen...

Exponential distribution



- Now consider just a single limousine of the PM. What is the time p.d.f. for its accident?
  Let $t_{1/2}$ (half-life) be the time of 50% survival probability:

$$F_{\mathrm{s}}(t_{1/2}) = (1-\varepsilon)^n = 0.5, \quad n\delta = t_{1/2}, \quad k\delta = t, \quad \delta \text{ is an infinitesimal time interval.}$$

$$n = \frac{\ln(0.5)}{\ln(1-\varepsilon)} \simeq \frac{-\ln(0.5)(1-\varepsilon)}{\varepsilon} \overset{\varepsilon\to 0}{\to} \frac{\ln(2)}{\varepsilon}$$

$$F_{\mathrm{s}}(t) = (1-\varepsilon)^k = (1-\varepsilon)^{\frac{1}{\varepsilon}\frac{t}{t_{1/2}}\ln(2)} = \left| \lim_{\varepsilon\to 0}(1-\varepsilon)^{\frac{\alpha}{\varepsilon}} = e^{-\alpha} \right| =$$

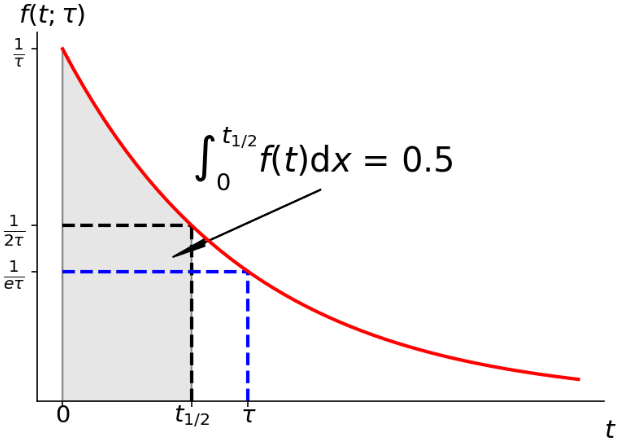
$$= e^{-\frac{t}{t_{1/2}}\ln(2)} \quad \Longrightarrow \quad F_{\mathrm{a}}(t) = 1 - e^{-\frac{t}{t_{1/2}}\ln(2)} \tag{2}$$

- F_{a} is the cumulative accident probability. Hence the the p.d.f.:

$$f_{\mathrm{a}}(t) = F_{\mathrm{a}}'(t) = \frac{1}{\tau}e^{-\frac{t}{\tau}}, \quad \text{with} \quad \tau = \frac{t_{1/2}}{\ln 2} \approx 1.44\ t_{1/2} \tag{3}$$

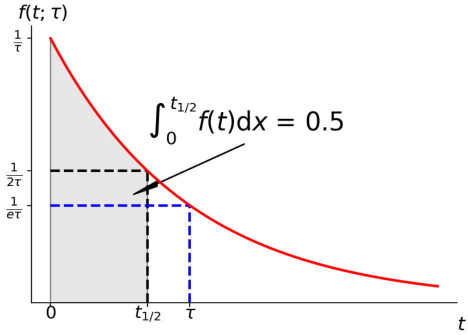
$$E[t] = \tau = \text{mean lifetime}, \qquad V[t] = \tau^2. \quad \text{show these!} \tag{4}$$

Exponential distribution



You are most likely to damage a brand new limousine!!!

Exponential distribution



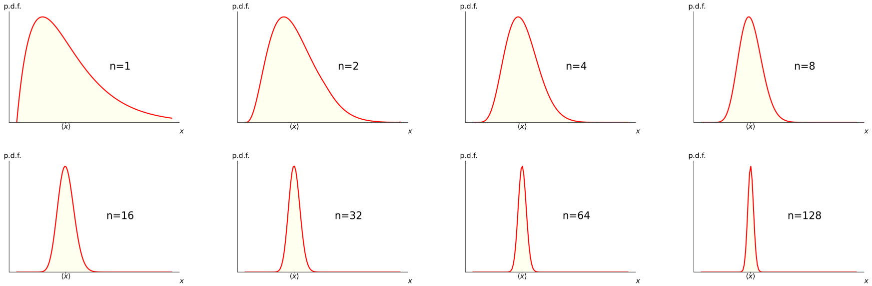
$$f_{\mathrm{a}}(t|t_0) = f_{\mathrm{a}}(t)/F_s(t_0) =$$
$$\frac{1}{\tau}e^{-\frac{t}{\tau}}/e^{-\frac{t_0}{\tau}} =$$
$$\frac{1}{\tau}e^{-\frac{t-t_0}{\tau}} = f_{\mathrm{a}}(t-t_0).$$

Do not be fooled! Probability of crashing a limo any day remains constant provided it has survied this far (conditional probability!).

Mean of a random variable ensamble

Central Limit Theorem

Imagine a measurement being a sum of of many n independent ones, or an average of n random numbers drawn from an **arbitrary distribution** (sampling distribution).



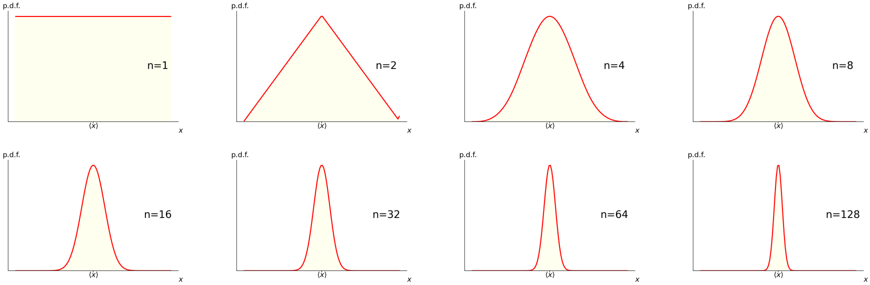
The mean $< x >$ converges on the initial distribution mean while the shape tends to a...

...**Gaussian** with ever decreasing width as $n \nearrow$.

Mean of a random variable ensamble

Central Limit Theorem

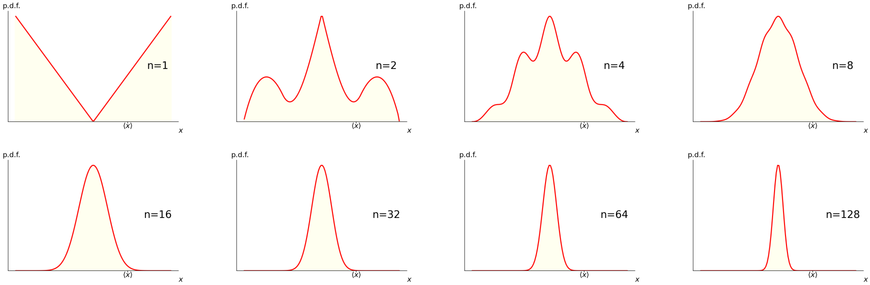
Ok, that was a well behaved distribution. Let's try something a bit less "gaussian" to start with:



The mean $< x >$ converges on the initial distribution mean while the shape tends to a...

...**Gaussian** with ever decreasing width as $n \nearrow$.

Ok, that was not austere enough. Let's try being bolder:



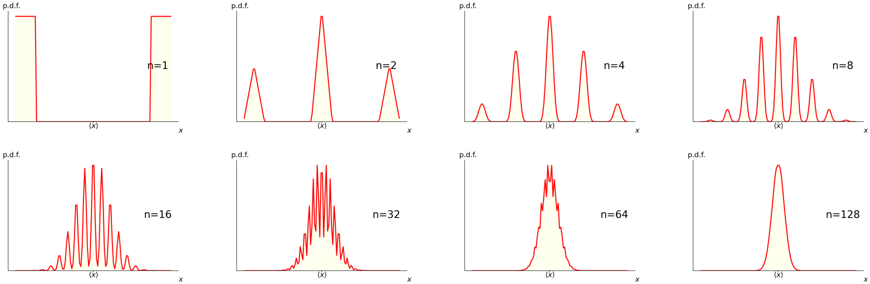
The mean $< x >$ converges on the initial distribution mean while the shape tends to a...

...**Gaussian** with ever decreasing width as $n \nearrow$.

Mean of a random variable ensamble

Central Limit Theorem

And again. Something manifestly non-Gaussian:

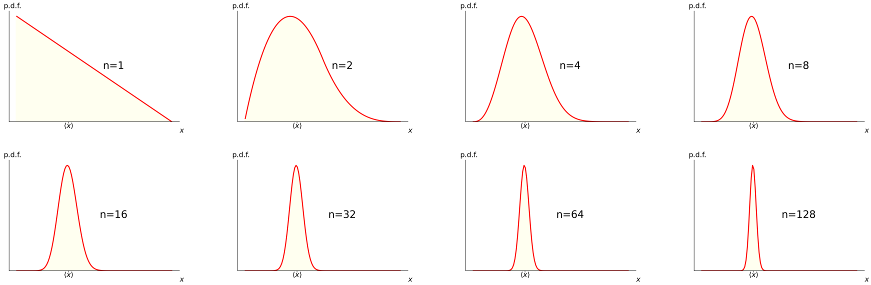


The mean $< x >$ converges on the initial distribution mean while the shape tends to a...
...**Gaussian** with ever decreasing width as $n \nearrow$.

Mean of a random variable ensamble

Central Limit Theorem

Finally, give up the symmetry:



The mean $< x >$ converges on the initial distribution mean while the shape tends to a...
...**Gaussian** with ever decreasing width as $n \nearrow$.

Central Limit Theorem

Sum of n random variables drawn from a probability distribution function of a finite variance, σ^2, tends to be Gaussian distributed about the expectation value for the sum, with variance $n\sigma^2$.

Consequently, the mean of the same n random values will have the expectation value of the initial p.d.f. and varaince $\frac{1}{n}\sigma^2$.

Ex: What is the probability that the mean salary of 50 randomly chosen emploies of our institute exceeds 6000 pln?

NOTE: We don't need to know the actual distribution of salaries in the institute. All we need to know is the average and the varaiance (or standard dev.).

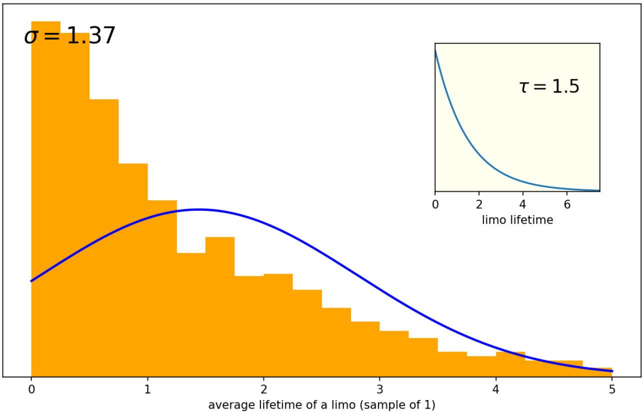
Careful: The *finite variance* is an important (and the only) requirement. A notable exception is the Cauchy (Breit-Wigner) distribution describing resonant states:

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

You can trivially show that the $E[x^2]$ is divergent!

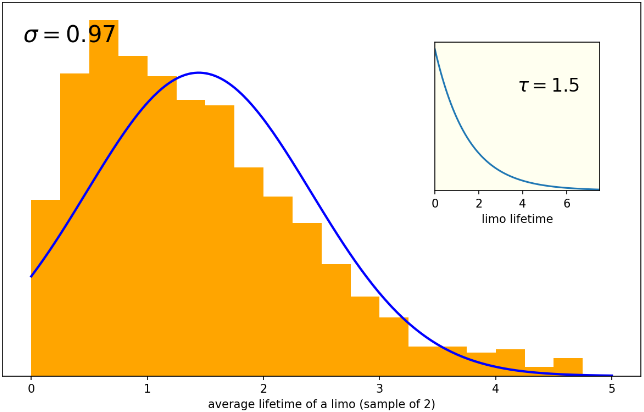
Back to fleet of limousines...

a single limo



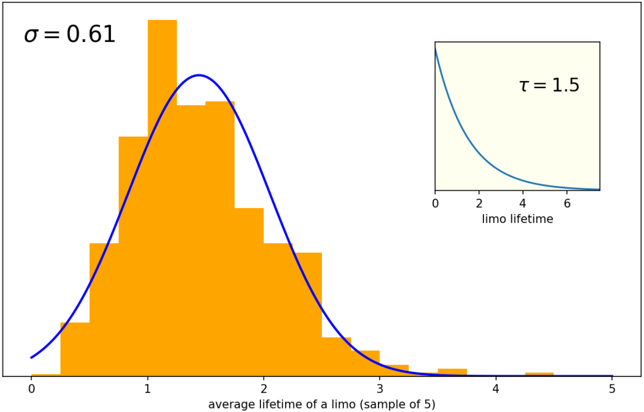
Back to fleet of limousines...

2 limo's



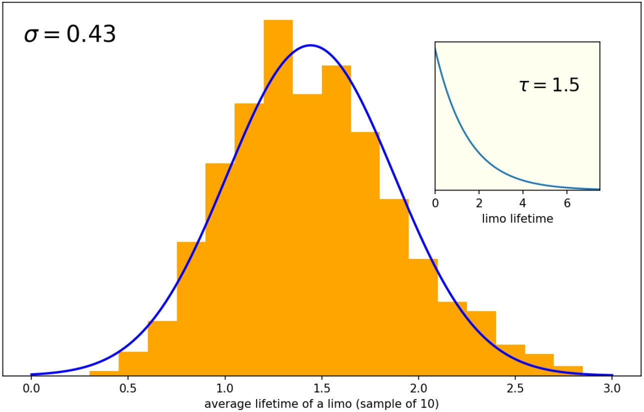
Back to fleet of limousines...

5 limo's



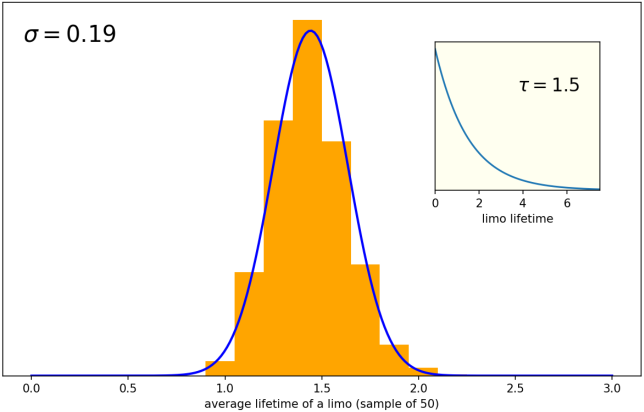
Back to fleet of limousines...

10 limo's



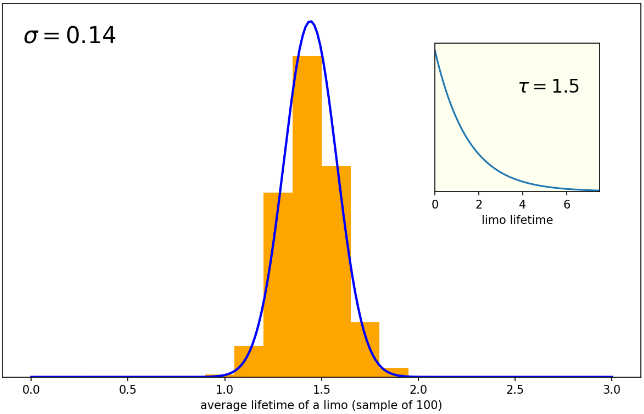
Back to fleet of limousines...

50 limo's



Back to fleet of limousines...

100 limo's



Gaussian distribution

The **Gaussian** p.d.f. of the continuous random variable x with $-\infty < x < \infty$ is defined by:

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \tag{5}$$

The term **normal** distribution is used when $\mu = 0 \ \ \& \ \ \sigma = 1$.

Gaussian p.d.f.: normalisation, mean & variance

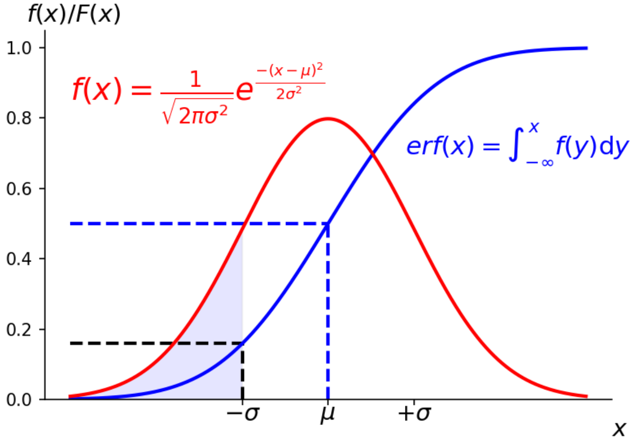
$$\int_{-\infty}^{\infty} f(x;\mu,\sigma^2) = 1 \tag{6}$$

$$E[x] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx = \mu, \tag{7}$$

$$V[x] = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx = \sigma^2. \tag{8}$$

Can you prove the above?

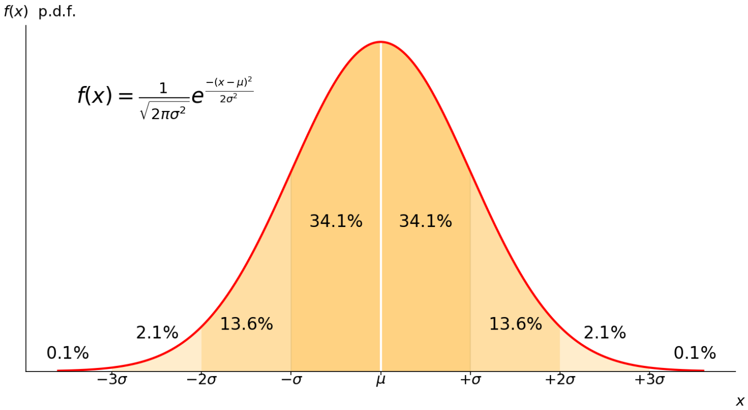
Gaussian distribution



The cumulative distribution of the Gaussian p.d.f. is not analitically calculable.
Nonetheless, quantiles of the normal distribution are of paramount importance for statistics!

Gaussian distribution

Quantiles



Standard deviation (σ) of a Gaussian distribution has central importance for error analysis:

$$\mu \pm 1\sigma : 68.27\%, \qquad \mu \pm 2\sigma : 95.45\%, \qquad \mu \pm 3\sigma : 99.73\%.$$

Characteristic function

Fourier Transform of a p.d.f.: the **characteristic function**

$$\phi(k) = E[e^{ikx}] = \int_{-\infty}^{\infty} dx\; f(x) e^{ikx} \quad \Rightarrow \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\; \phi(k) e^{-ikx} \tag{9}$$

- m'th algebraic moment of $f(x)$ is obtained by simple diferentiation of $\phi(k)$:

$$(-i)^m \frac{d^m}{dk^m} \phi(k)\Big|_{k=0} = (-i)^m \frac{d^m}{dk^m} \int_{-\infty}^{\infty} dx\; f(x) e^{ikx}\Big|_{k=0} =$$

$$= (-i^2)^m \int_{-\infty}^{\infty} dx\; x^m f(x) = E[x^m] \tag{10}$$

- Let $z = \sum_i x_i$, where $x_1, ..., x_n$ are n independent random variables:

$$\phi_z(k) = \int ... \int e^{ik \sum_i x_i} f_1(x_1)...f_n(x_n) dx_1...dx_n = \tag{11}$$

$$= \int e^{ikx_1} f_1(x_1) dx_1 ... \int e^{ikx_n} f_n(x_n) dx_n = \phi_1(k)...\phi_n(k). \tag{12}$$

Central Limit Theorem

Derivation of...

Let $z = \frac{1}{\sqrt{n}}(x_1 + ... + x_n) = \sum_{j=1}^{n} \frac{x_j}{\sqrt{n}}$. For a single variable $u \equiv x/\sqrt{n}$, the characteristic function is given by:

$$\begin{aligned}\phi_u(k) = \int_{-\infty}^{\infty} du\ f(u) e^{iku} = 1 + iE[u]k - \frac{1}{2}E[u^2]k^2 + O(k^3) = \\ = 1 + iE[x]\frac{k}{\sqrt{n}} - \frac{1}{2}E[x^2]\frac{k^2}{n} + O(\frac{k}{\sqrt{n}}^3)\end{aligned} \tag{13}$$

Without any loss of generality, we can assume that $E[x] = 0$ and $E[x^2] = \sigma^2$ (otherwise use $\bar{x} \equiv x - E[x]$):

$$\begin{aligned}\lim_{n\to\infty} \phi_z(k) = \lim_{n\to\infty} \prod_{j=1}^{n} \phi_{u_j}(k) = \lim_{n\to\infty} \prod_{j=1}^{n} \left(1 - E[x^2]\frac{k^2}{2n} + O(\frac{k^3}{n^{3/2}})\right) \simeq \\ \simeq \lim_{n\to\infty} \left(1 - \frac{\sigma^2 k^2}{2n}\right)^n = e^{-\sigma^2 k^2/2}\end{aligned} \tag{14}$$

Central Limit Theorem

... and the Gaussian distribution

So far we have found the characteristic function of the z. The p.d.f. is given by its inverse Fourier transform:

$$f_z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ \phi_z(k) e^{-ikz} = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ e^{-\sigma^2 k^2/2} e^{-ikz} = \\ = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\ e^{-\left(\sigma k/\sqrt{2} + iz/(\sigma\sqrt{2})\right)^2 - z^2/(2\sigma^2)} = \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/(2\sigma^2)} \tag{15}$$

We have derived the **Central Limit Theorem**

The sum of independent random variables, sampled from the same distribution, will tend towards a **Gaussian** distribution, independently of the initial distribution.

Note: In the proof we used the strong assumption that all moments were finite. In fact, it is sufficient that the second moment (σ^2) is finite, but we shall leave it without a proof. This holds for most well-behaved p.d.f.'s, but not all!

Central Limit Theorem

concequences

For the above derivation we used particularly normalised sum ($z = \sum_{j=1}^{n} \frac{x_j}{\sqrt{n}}$) which led to the variance of the z being equal to the variance of x_i.
It is easy to see that:

1. For the algebraic sum $z = \sum_{j=1}^{n} x_j$ we obtain $\sigma_z = \sqrt{n}\sigma$, or more generally $\sigma_z^2 = \sum_{j=1}^{n} \sigma_j^2, \quad (< z >= \sum_{j=1}^{n} < x_j >)$.
2. For the algebraic mean $z = \frac{1}{n} \sum_{j=1}^{n} x_j$ we obtain $\sigma_z = \frac{1}{\sqrt{n}}\sigma$, or more generally $\sigma_z^2 = \frac{1}{n} \sum_{j=1}^{n} \sigma_j^2, \quad (< z >= \frac{1}{n} \sum_{j=1}^{n} < x_j >)$.

What does it mean?

- If we estimate the mean from a sample, we will always tend towards the true mean,
- The uncertainty in our estimate of the mean will decrease as the sample gets bigger.

Gaussian distribution

... generalisation

Let $\mathbf{x} = (x_1, x_2, ..., x_n)$ be a n-dimensional sample space.

n-dimensional Gaussian distribution

$$f(\mathbf{x}; \mu, V) = \frac{1}{(2\pi)^{n/2}|V|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T V^{-1}(\mathbf{x}-\mu)\right) \tag{16}$$

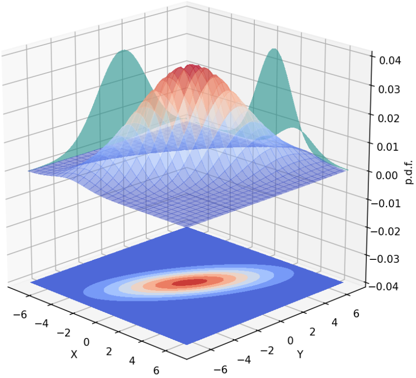
V is the covariance matrix of $\mathbf{x}$ and V^{-1} is its inverse, called the *weight* matrix.
$|V|$ is the determinant of V.

What does the above give for independent random variables?

Gaussian distribution

... 2D case

- $\sigma_1 = 2$
- $\sigma_2 = 3$
- $\rho = 0.7$



$$V = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

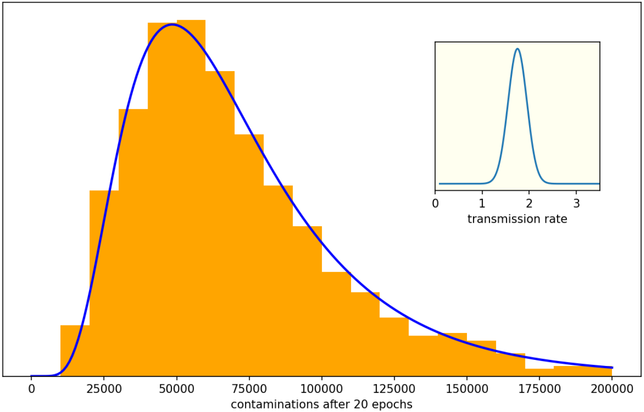
$$V^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

$$f(x_1, x_2; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$
$$\exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right) \quad (17)$$

Spread of a pandemic

multiplicative Gaussian

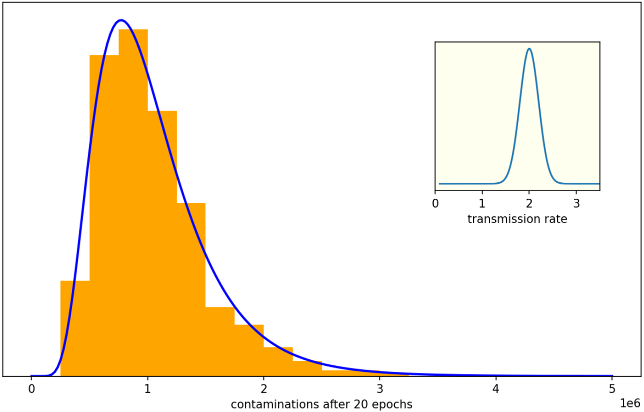
Average transmission rate: 1.75 with standard deviation of 0.2.
Number of infected after 20 epochs:



Spread of a pandemic

multiplicative Gaussian

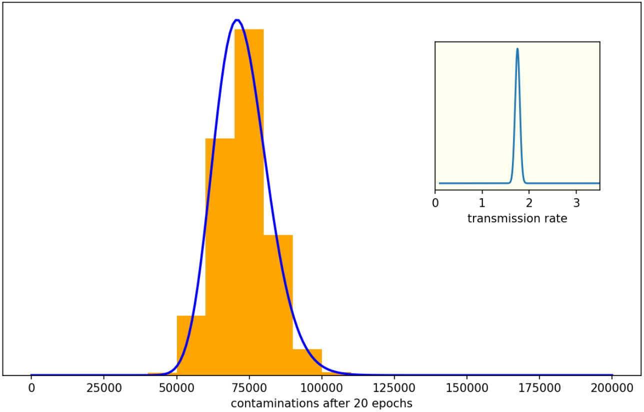
Average transmission rate: 2.0 with standard deviation of 0.2.
Number of infected after 20 epochs:



Spread of a pandemic

multiplicative Gaussian

Average transmission rate: 1.75 with standard deviation of 0.05.
Number of infected after 20 epochs:



Log-normal distribution

Let y be a Gaussian-distributed random variable with mean and variance μ, σ^2. What is the p.d.f. of $x = e^y$?

$$g(x) = f(y(x); \mu, \sigma^2) \left| \frac{dy}{dx} \right| = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2} \right) \frac{d(\ln x)}{dx} \tag{18}$$

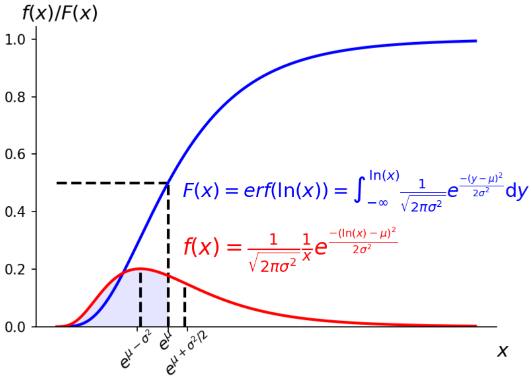
log-normal p.d.f.

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2} \right) \tag{19}$$

$$E[x] = e^{\mu + \frac{1}{2}\sigma^2} \tag{20}$$

$$V[x] = e^{2\mu + \sigma^2} \left[e^{\sigma^2} - 1 \right] \tag{21}$$

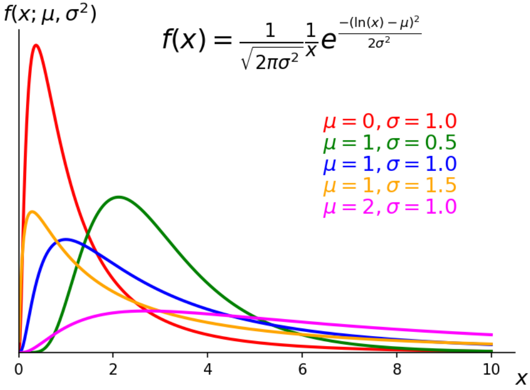
Log-normal distribution



$$\int_0^X \frac{1}{x} e^{\frac{-(\ln(x)-\mu)^2}{2\sigma^2}} dx = \left| \ln(x) = y, \frac{1}{x} dx = dy \right| = \int_{-\infty}^{\ln(X)} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy = \sqrt{2\pi\sigma^2} erf\,(\ln(X))$$

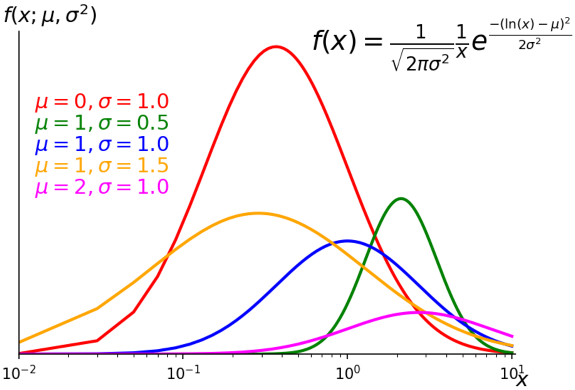
$$\int_0^\infty x \frac{1}{x} e^{\frac{-(\ln(x)-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} e^{\frac{-(y-\mu)^2}{2\sigma^2}} e^y dy = \int_{-\infty}^{\infty} e^{\frac{-(y-(\mu+\sigma^2))^2}{2\sigma^2}} e^{\mu+\frac{1}{2}\sigma^2} dy = \sqrt{2\pi\sigma^2} e^{\mu+\frac{1}{2}\sigma^2}$$

mode: $e^{\mu-\sigma^2}$, median: e^{μ}, mean: $e^{\mu+\frac{1}{2}\sigma^2}$, $F(X) = erf(\ln(X))$



It becomes apparent that if $z = \prod_{j=1}^{n} x_j = e^{\sum_{j=1}^{n} y_j}$, the product of many random variables tends to a log-normal distribution with $\mu = \sum_{j=1}^{n} \mu_j$ and $\sigma^2 = \sum_{j=1}^{n} \sigma_j^2$. Here, $\mu_j = E[\ln x]$ and $\sigma_j^2 = E[\ln^2 x] - E[\ln x]^2$. Certainly, $\forall_j x_j > 0$.

Log-normal distribution

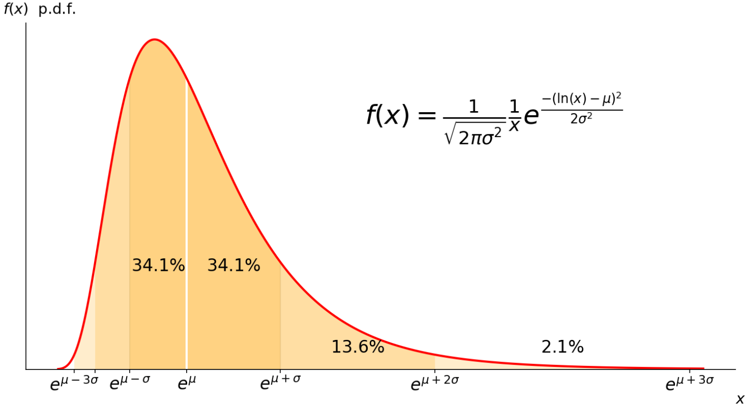


In logarythmic scale, log-norm distributions appears as Gaussian (normal).

$$y = \ln(x):\ \frac{1}{x}e^{\frac{-(\ln(x)-\mu)^2}{2\sigma^2}} = e^{\frac{-(y^2-2\mu y+\mu^2)-2\sigma^2 y}{2\sigma^2}} = e^{-\mu+2\sigma^2}e^{\frac{-(y-(\mu-\sigma^2))^2}{2\sigma^2}}$$

Log-normal distribution

Quantiles



$e^{\mu} \times / e^{\sigma} : 68.27\%,\quad e^{\mu} \times / e^{2\sigma} : 95.45\%,\quad e^{\mu} \times / e^{3\sigma} : 99.73\%.$

χ^2 test statistic

Let x be a Gaussian-distributed randon variable with known μ and σ. We can make a simple linear transformation of this variable such, that the distribution becomes so-called *standard normal* ($\mu = 0$, $\sigma = 1$):

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad x \to z = \frac{x-\mu}{\sigma}, \quad f(z;0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \tag{22}$$

What is the distribution of $u \equiv z^2$ ($E[u] = E[z^2] = V[z] = 1$)?

$$\chi_1^2(u) = 2f(z(u)) \left|\frac{dz}{du}\right| = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{u}} \exp\left(-\frac{u}{2}\right) \tag{23}$$

Recall: $z \in (-\infty, \infty) \longrightarrow u = z^2 \in (0, \infty)$.

χ_1^2: mean & variance

$$E[u] = \int_0^\infty u\chi_1^2(u)du = 1 \tag{24}$$

$$V[u] = \int_0^\infty u^2\chi_1^2(u)du = 2 \tag{25}$$

χ^2 test statistic

χ^2_1 can be extended to distribution of two independent normal-distributed random variables $u = z_1^2 + z_2^2$ by means of Fourier convolution. The operation executed recurrently provides the expression for any value of n ($u = \sum_{i=1}^{n} z_i^2$):

$$\chi^2_n(u) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} u^{\frac{n}{2}-1} \exp\left(-\frac{u}{2}\right) \tag{26}$$

Recall: $\Gamma(n) = (n-1)!, \qquad \Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$

χ^2_n: mean & variance

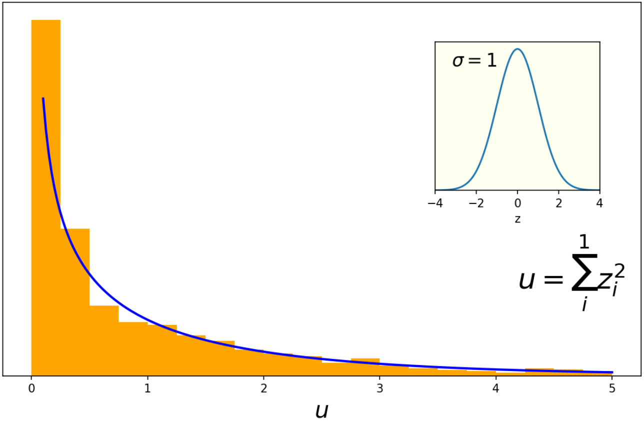
$$E[u] = \int_0^\infty u\chi^2_n(u)du = n \tag{27}$$

$$V[u] = \int_0^\infty u^2\chi^2_n(u)du = 2n \tag{28}$$

Note: χ^2 distribution has only one parameter, n, called *number of degrees of freedom* (nDoF).

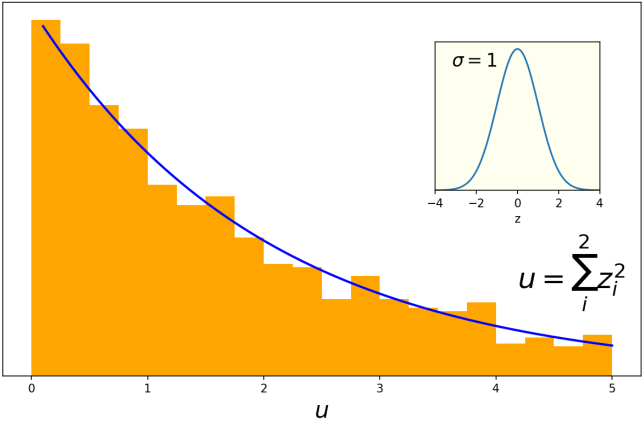
χ^2 test statistic

nDoF=1



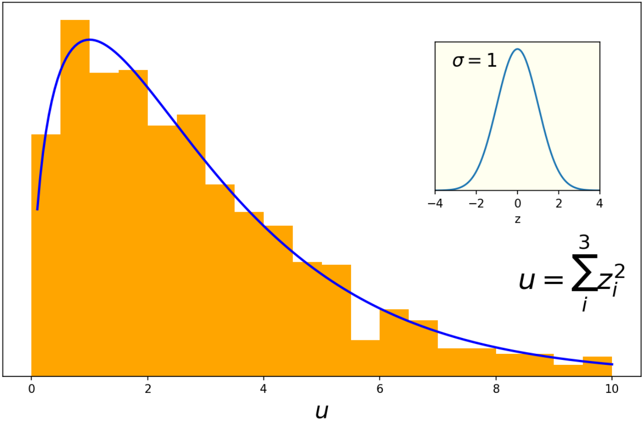
χ^2 test statistic

nDoF=2



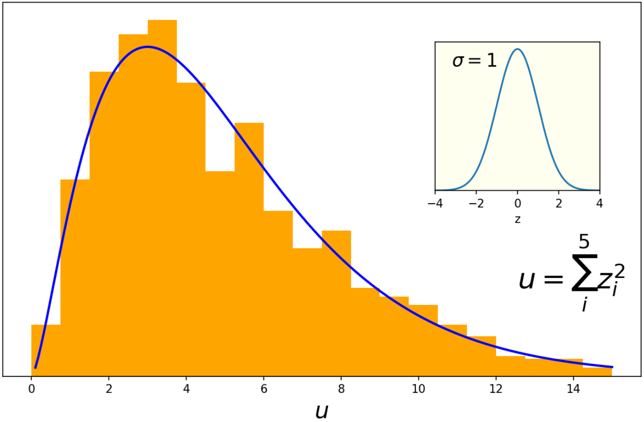
χ^2 test statistic

nDoF=3



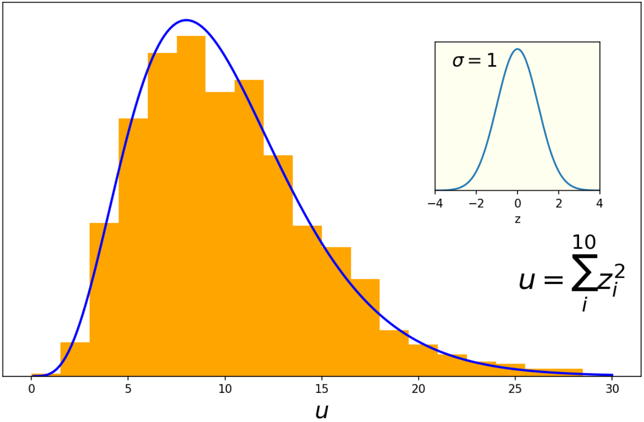
χ^2 test statistic

nDoF=5



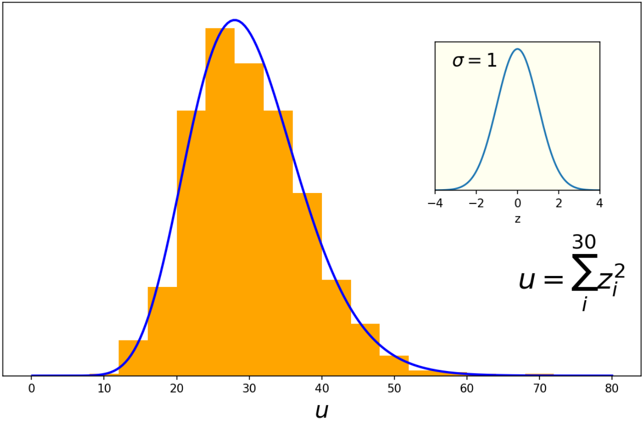
χ^2 test statistic

nDoF=10



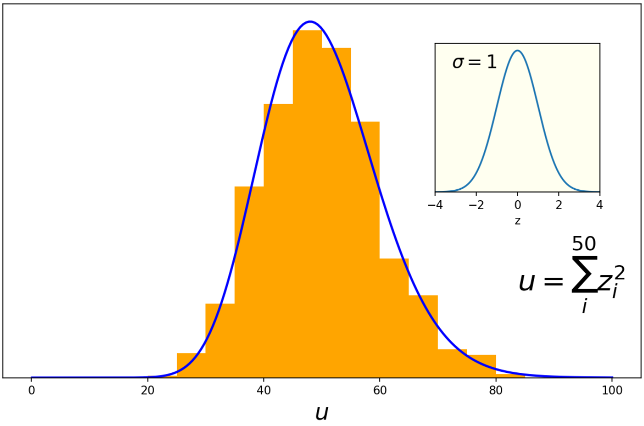
χ^2 test statistic

nDoF=30



χ^2 test statistic

nDoF=50



χ^2 test statistic

general n-dimensional case

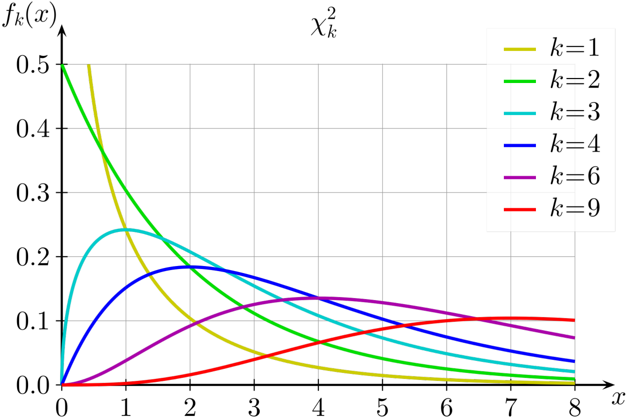
So far independence of the normal-distributed variables was as assumed. This can be generalised to n-dimensional Gaussian distribution with an arbitrary covariance matrix $\mathbf{V}$.

χ^2-distributed n-dimensional Gaussian

$$z = (\mathbf{x} - \mu)^T \mathbf{V}^{-1} (\mathbf{x} - \mu) \tag{29}$$

is a χ^2_n random variable with n DoF's.

χ^2 distribution



The χ_k^2 distribution approaches a Gaussian (recall CLT!) for $k \to \infty$. For practical applications, it can be considered Gaussian for $n > O(50)$ ($\mu = k$, $\sigma = \sqrt{2k}$).

mode: $k - 2$, median: $\approx k \left(1 - \frac{2}{9k}\right)^3$, mean: k, $F(X, k) = \frac{1}{\Gamma\left(\frac{k}{2}\right)} \gamma\left(\frac{k}{2}, \frac{X}{2}\right)$

$$\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$$

Questions

Consider the exponential p.d.f.,

$$f(x;\tau) = \frac{1}{\tau} e^{-x/\tau}, \qquad x \geq 0.$$

1. Show that the corresponding cumulative distribution is given by

$$F(x;\tau) = 1 - e^{-x/\tau}$$

2. Show that the conditional probability to find a value $x < x_0 + x'$ given that $x > x_0$ is equal to the (unconditional) probability to find x less than x', i.e.

$$P(x < x_0 + x' | x \geq x_0) = P(x \leq x').$$

Solutions to be sent to me before the next lecture

Thank you

Back-up

Fourier convolution - revisited

$z = x + y$, find $f_z(z)$ given $f_{x,y}(x,y)$

$$P(z \leq z_1) = \int_{-\infty}^{z_1} d\kappa f_z(\kappa) =$$

$$= \int_{-\infty}^{\infty} dy \int_{-\infty}^{z_1 - y} dx \, \underbrace{f_{x,y}(x,y)}_{\text{joint p.d.f.}} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z_1 - x} dy f_{x,y}(x,y) \tag{30}$$

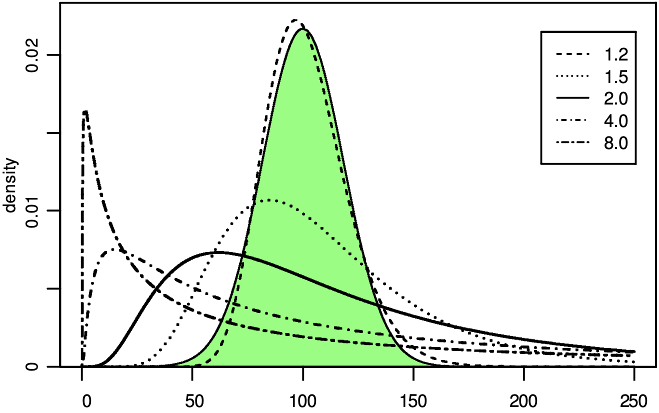
$$f_z(z) = \frac{dP}{dz} = \int_{-\infty}^{\infty} dx f_{x,y}(x, z-x) = \int_{-\infty}^{\infty} dy f_{x,y}(z-y, y) \tag{31}$$

Hence for independent variables ($f_{x,y}(x,y) = f_x(x) * f_y(y)$) we obtain:

$z = x + y$: Fourier convolution

$$f(z) = \int_{-\infty}^{+\infty} g(x)h(z-x)dx = \int_{-\infty}^{+\infty} g(z-y)h(y)dy. \tag{32}$$

Log-normal distribution



Gaussian μ, σ^2 are additive, log-normal are multiplicative.
The log-normal distribution approaches a Gaussian for $\sigma \to 0$.