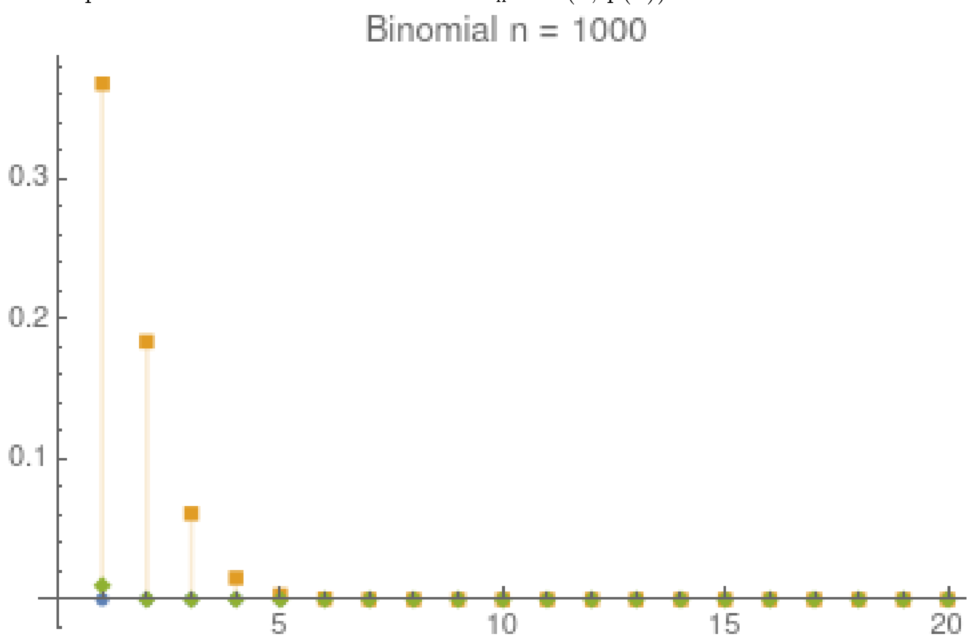
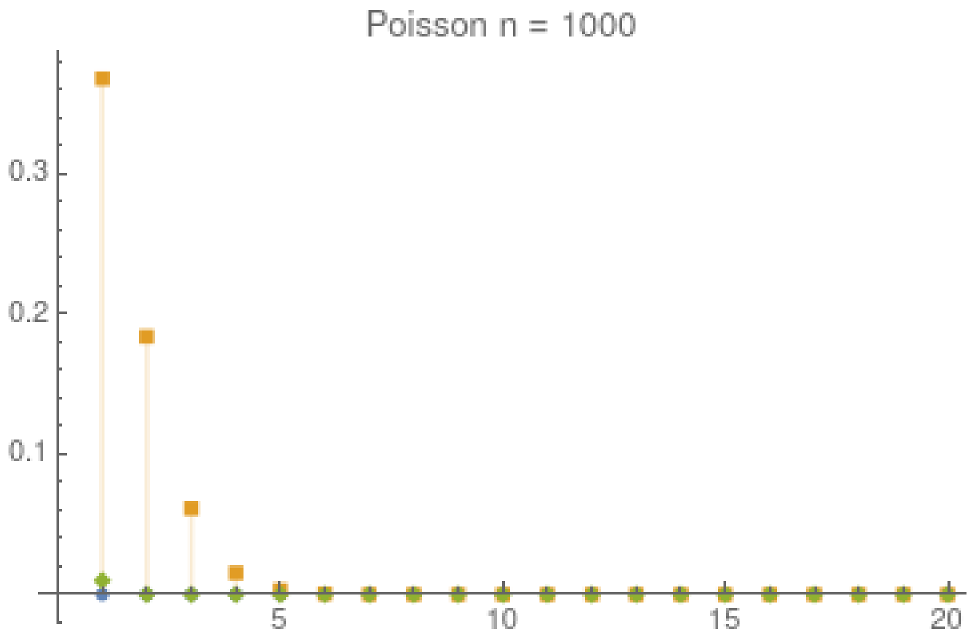
1 Binomial (with proof) [Theor 4.30] and multinomial [Theor 4.31] Poisson theorems. Theorem Poisson's limit (with proof) [Theor 4.32] and Gaussian convergence of Poisson's laws (with proof) [Theor 4.33]

1.1 Binomial Poisson theorem:

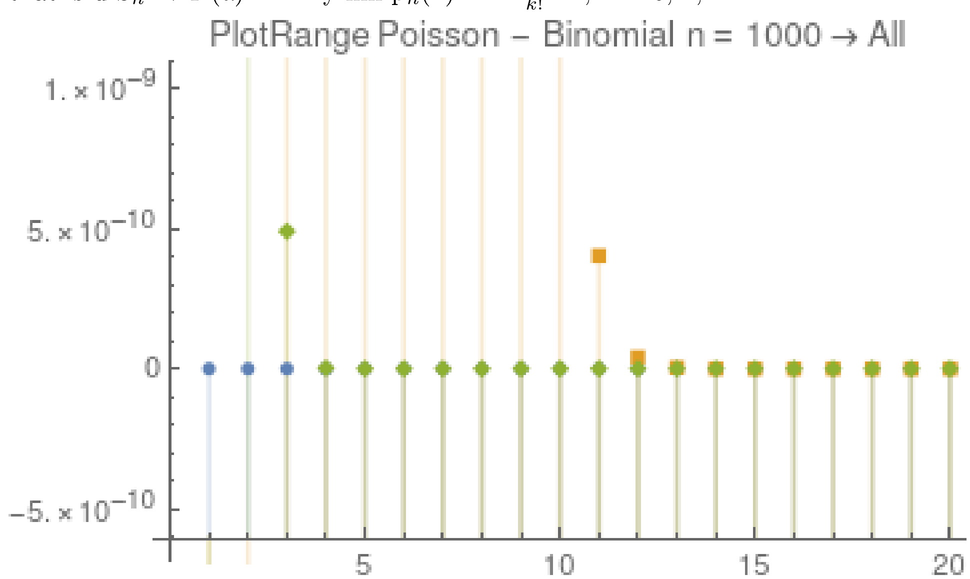
Take a sequence of binomial rv 's binomiali $S_n \sim \mathcal{B}(n; p(n))$:



if it exists a number $\alpha > 0$ such that:
$p(n) \to 0$
$q(n) = 1 - p(n) \to 1$
$np(n) \to \alpha$
$n \to \infty$
then S_n converges in distribution to the Poisson law $P(\alpha)$,



that is d $S_n \xrightarrow{d} P(\alpha)$ namely lim $p_n(k) = \frac{\alpha^k k e^{-\alpha}}{k!}$, k = 0, 1, . . .



1.1.1 Proof

Since for every $\alpha > 0$, from a certain n onward we have $\alpha/n < 1$, starting from there our hypotheses empower us to write:

$p(n) = \frac{\alpha}{n} + o(n^{-1})$

so that for k = 0, 1, . . . , n we will get:

$p_n(k) = \frac{n(n-1)\ldots(n-k+1)}{k!}[\frac{\alpha}{n} + o(n^{-1})]^k[1 - \frac{\alpha}{n} + o(n^{-1})]^{n-k}$

Dealing with the three factors separately:

$n(n-1)\ldots(n-k+1)[\frac{\alpha}{n} + o(n^{-1})]^k = \frac{n(n-1)\ldots(n-k+1)}{n^k}[\alpha + o(1)]^k = (1-\frac{1}{n})\ldots(1-\frac{k-1}{n})[\alpha + o(1)]^k \xrightarrow{n} \alpha^k$

$[1 - \frac{\alpha}{n} + o(n^{-1})]^{n-k} = {}^k[1 - \frac{\alpha}{n} + o(n^{-1})]^n[1 - \frac{\alpha}{n} + o(n^{-1})]^{-k} \xrightarrow{n} e^{-\alpha}$

we get:

$p_n(k) = \frac{\alpha^k e^{-\alpha}}{k!}$

1.2 Multinomial Poisson theorem:

Take a sequence of multinomial r-vec's$S_n = (X_1 \cdots X_r \mathcal{B}(n; p_1, \cdots, p_r)$ with

$P\{X_1 = k_1, \cdots, X_r = k_r\} = \frac{n!}{k_0!k_1!\cdots k_r!}p_0^{k_0}p_1^{k_1}\cdots p_r^{k_r}$

$$\begin{cases} p_0 + p_1 + \ldots + p_r = 1 \\ k_0 + k_1 + \ldots + k_r = n \end{cases}$$

If for $j = 1, \ldots, r$ and$n \to \infty$ there exist $\alpha_j > 0$ such that

$p_j = p_j(n) \to 0$

$p_0 = p_0(n) \to 1$

then we have $np_j(n) \to \alpha_j$

$S_n = (X_1, \ldots, X_r) \xrightarrow{d} \mathcal{B}(\alpha_1)\mathcal{B}(\alpha_2)\ldots\mathcal{B}(\alpha_r)$

1.3 Theorem Poisson's limit

For every $n \in N$ and $k = 1, \ldots, n$ take the independent rv 's X_k^n with:

- $P\{X_k^n = 1\} = p_k^n$
- $P\{X_k^n = 0\} = q_k^n$
- $p_k^n + q_k^n = 1$
- and positive $S_n = X_1 + \cdots + X_n$

if

$$\max_{1\le k\le n} p_k^n \xrightarrow{n} 0$$

$$\sum_{k=1}^{n} p_k^n \xrightarrow{n} \alpha > 0$$

then we have

$$S_n \xrightarrow{d} P(\alpha)(n)$$

1.3.1 Proof

From the independence of the X_k, we have

$$\varphi_{Sn}(u) = E[e^{iuS_n}] = \prod_{k=1}^{n} [p_k^n e^{iu} + q_k^n] = \prod_{k=1}^{n} [1 + p_k^n(e^{iu} - 1)]$$

Since by hypothesis$p_k^n \to 0$, from the series expansion of the logarithm we have

$$ln\varphi_{Sn}(u) = \sum_{k=1}^{n} ln[1 + p_k^n(e^{iu} - 1)] = \sum_{k=1}^{n} ln[p_k^n(e^{iu} - 1) + o(p_k^n)] \xrightarrow{n} \alpha(e^{iu} - 1)$$

and given the continuity of the logarithm

$$\varphi_{Sn}(u) \xrightarrow{n} e^{\alpha(e^{iu}-1)}$$

so

$$S_n \xrightarrow{d} P(\alpha)(n)$$

1.4 Gaussian convergence of Poisson's laws

If $S \sim P(\alpha)$ is a Poisson rv , then

$$S^* = \frac{S - \alpha}{\sqrt{\alpha}} \xrightarrow{d} N(0, 1)$$

1.4.1 Proof

If φ α is the chf of S^*, from the series expansion of an exponential we find for$\alpha \to +\infty$

$\varphi_\alpha(u) = E[e^{iuS^*}] = e^{-iu\sqrt{\alpha}}E[e^{iuS/\sqrt{\alpha}}] = exp[-iu\sqrt{\alpha} + \alpha(e^{iuS/\sqrt{\alpha}} - 1)] = exp[-iu\sqrt{\alpha} - \alpha + \alpha(1 + \frac{iu}{\sqrt{-\alpha}} - \frac{u^2}{2\alpha} + o(\frac{1}{\alpha}))] \to e^{-u^2/2}$

2 Definition of conditional densities with respect to events of zero measure, and its justification [Sect 3.4.1]. Conditional expectation value, even with respect to a random variable [Section 3.4.2]. Property of the conditional expectations [Prop 3.42] with proof only of E [E [X | Y]] = E [X]

2.1 Conditional densities

2.1.1 CDF

If X, Y are two rv 's with a joint cdf $F_{XY}(x, y)$ which is y-differentiable, and if Y is ac with pdf f Y (y), we will call cdf of X conditioned by the event {Y = y} the function:

$$F_{X|Y}(x|y) = \begin{cases} \frac{\partial_y F_{X,Y}(x,y)}{f_Y(y)} & f_Y(y) \neq 0 \\ arbitrary, possibly = 0 & f_Y(y) = 0 \end{cases}$$

2.1.2 PDF

If X is ac and the joint pdf is $f_{XY}(x, y)$, then the pdf of X conditioned by the event {Y = y} is:

$$f_{X|Y}(x|y) = \begin{cases} \frac{f_{XY}(x,y)}{f_Y(y)} & f_Y(y) \neq 0 \\ 0 & f_Y(y) = 0 \end{cases}$$

2.2 Conditional expectation

Given the rv 's X, Y and a Borel function g(x), we will call conditional expectation of g(X) w.r.t. {Y = y} the y-function:

$$m(y) \equiv E[g(X|Y = y)] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) dx$$

We will call instead conditional expectation of g(X) w.r.t. the rv Y the rv:

$$E[g(X)|Y] \equiv m(Y)$$

2.2.1 Properties

- E [E [X|Y]] = E [X]
- E [X|Y] = E [X] P -a.s. if X and Y are independent
- E [φ(X, Y)|Y = y] = E [φ(X, y)|Y = y]

- E [φ(X, Y)|Y = y] = E [φ(X, y)]
- E [X g(Y)|Y] = g(Y) E [X|Y] P Y -as P -a.s.

2.2.2 Proof for E[E[X|Y]]=E[X]

$E[E[X|Y]] = E[m(Y)] = \int_R m(y)F_Y(y)dy = \int_R E[X|Y=y]f_Y(y)dy = \int_R[\int_R xf_{X|Y}(x|y)dx]f_Y(y)dy = \int_R[\int_R x\frac{f_{XY}(x,y)}{f_Y(y)}dx]f_Y(y)dy = \int_R x[\int_R f_{XY}(x,y)dy]dx = \int_R xf_X(x)dx = E[X]$

3 Definition of expectation value and moments of a random variable [Section 3.3.1]. Procedure for the calculation of the expectation value with change of variables [Theor 3.22, Coroll. 3.23]. Main ownership of expectation values [Theor 3.26]

3.1 Expectation value

The expectation of a rv X is a numerical indicator specifying the location of the barycenter of a distribution P_X

For every $A \in \mathcal{F}$ we always have E [I_A] = P{A}

For a rv X expectation value can be calculated using Lebesgue integral:

$$E[X] = \int_\Omega XdP = \int_\Omega X(\omega)P\{d\omega\}$$

3.2 Moments

We will call moment of order k of a rv X the expectation (if it exists)

$E[X^k] = \int_\Omega X^k dP$

$k = 0, 1, 2$

and absolute moment of order r the expectation (if it exists)

$E[|X^r|] = \int_\Omega |X|^r dP$

$r \geq 0$

3.3 Calculation of the expectation value with change of variables

Take the r-vec$X = (X_1, \cdots, X_n)$ on (Ω, F, P) with joint distribution P_X , and the Borel function $g : (R_n, B(R_n)) \to (R, B(R))$; if $Y = g(X)$ we have:

$E[Y] = \int_\Omega YdP = \int_\Omega Y(\omega)P\{d\omega\} = E[g(X)] = \int_\Omega g(X)dP = \int_\Omega g(X(\omega))P\{d\omega\} = \int_{R^n} g(x)P_x\{dx\} = \int_{R^n} g(x_1, x_2, \ldots, x_n)P_x\{dx_1, dx_2, \ldots, dx_n\}$

3.4 Main ownership of expectation

Because expectation value is defined with the Lebesgue integral it is sharing it's properties:

- $E[aX + bY] = aE[X] + bE[Y]$with $a, b \in R$
- $E[X] \leq E[|X|]$
- if $X = 0$, P -a.s., then$E[X] = 0$; if moreover X is an arbitrary rv and A an event - then $E[XI_A] = \int_A XdP = 0$ if $P\{A\} = 0$
- if $X \leq Y$, P -a.s. then$E[X] \leq E[Y]$, and if$X = Y$ P -a.s., then $E[X] = E[Y]$
- if $X \geq 0$ and $E[X] = 0$, then $X = 0$, P -a.s., namely X is degenerate δ_0
- if $E[XI_A] \leq E[YI_A], \forall A \in F$, then$X \leq Y$, P -a.s., and in particular if $E[XI_A] = E[YI_A], \forall A \in F$, then $X = Y$, P -a.s.
- if X and Y are independent, then also XY is integrable and $E[XY] = E[X] \cdot E[Y]$

4 Probability density f(x) of a law [Sect 2.2.3]: conditions of existence (Radon Nikodym's theorem) [Theor 2.15] and at least two examples. Concept of singular distribution [Section 2.2.4]. Blends laws and Lebesgue-Nikodym Theorem [Section 2.2.5]

4.1 Radon Nikodym's theorem

A cdf $F(x)$ on R is ac iff there exists a non negative function $f(x)$ defined on R such that:

$\int_{-\infty}^{\infty} f(x)dx = 1$
$F(x) = \int_{-\infty}^{x} f(z)dz$
$f(x) = F'(x)$
The function f(x) is called a pdf (probability density function) of F(x).

4.1.1 Example of Uniform distribution

The cdf:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

The pdf can be calculated by derivation of cdf, so

$$f(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \le x \le b \\ 0 & x > b \end{cases}$$

4.1.2 Example of Gaussian distribution

The cdf:

$F(x) = \frac{1}{a\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(z-b)^2/2a^2} dz$

The pdf can be calculated by derivation of cdf, so

$f(x) = \frac{1}{a\sqrt{2\pi}} e^{-(-b+x)^2/2a^2}$

4.2 Singular distribution

We say that P is a singular distribution when its cdf F (x) is continuous, but not ac.

Namely, it's neither degenerate nor absolutely continuous.

The example could be a Cauchy distribution with "a" aproaching 0. At some point the pdf function would take the infinite value, ceasing to be a function.

5 Sums of independent random variables, convolutions with at least one example [Section 3.5.2] and them characteristic functions [Prop 4.9]

5.1 Sums of independent random variables

Given two independent rv 's X and Y with pdf 's $f_X(x)$ and $f_Y(y)$, the pdf of their sum Z = X + Y is the convolution of the respective pdf 's.

$$f_Z(x) = (f_X * f_Y)(x) = (f_X * f_Y)(x)$$

5.2 Sum of uniformly distributed rv's

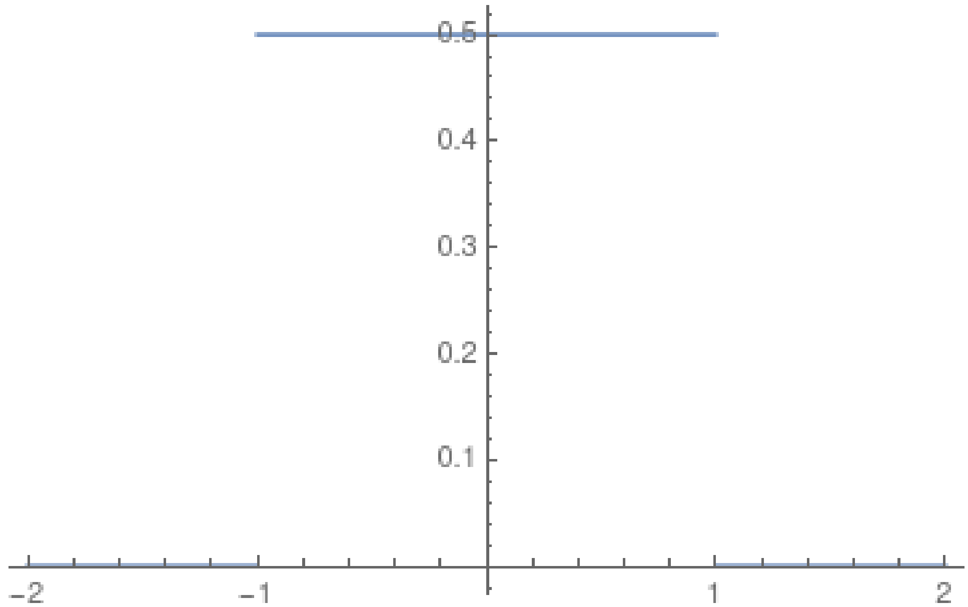
Given $X_{1,} \ldots, X_n$with pdf's described by:

$f_{X_k}(x) = \frac{1}{2}\upsilon(1 - |x|)$

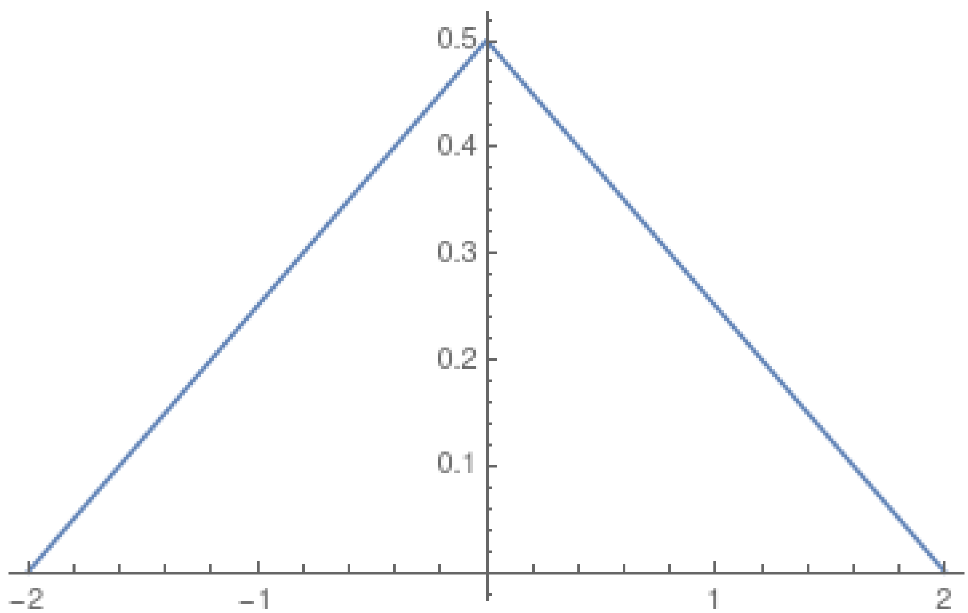
we can construct $Y = X_1 + X_2 + \ldots + X_n$by convoluting it's pdf's

$f_Y(x) = (f_{X_1} * f_{X_2} * \ldots * f_{X_n})(x)$

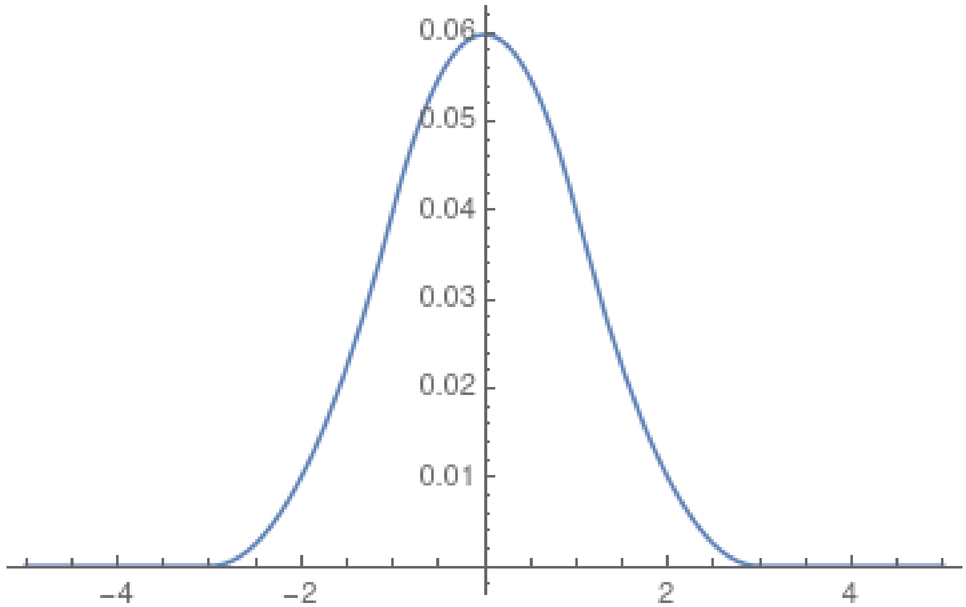
5.2.1 Example for plot n=1



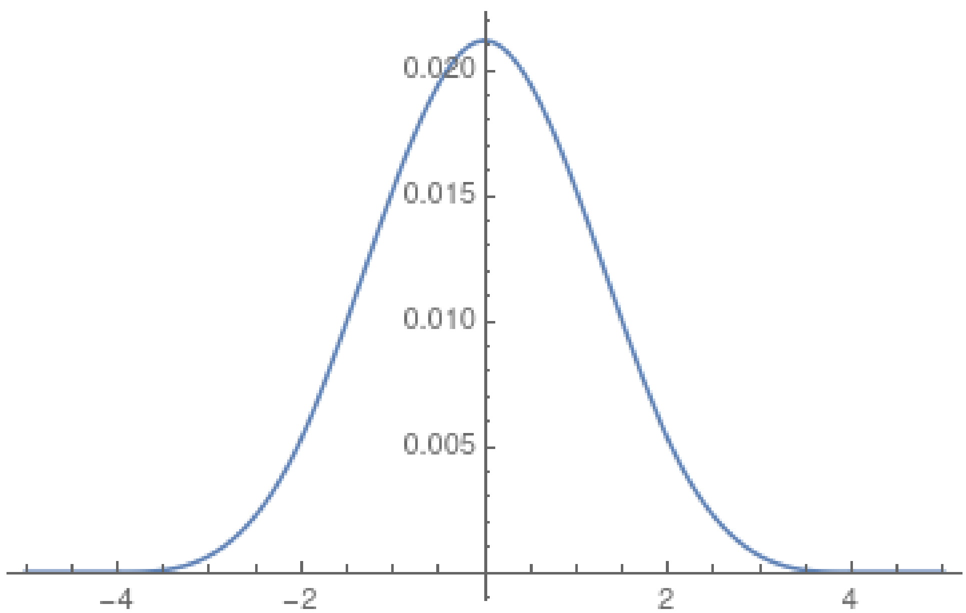
5.2.2 Example for plot n=2



5.2.3 Example for plot n=3



5.2.4 Example for plot n=4



5.3 Characteristic functions

By calculating a Fourier Transform of pdf we can obtain a chf (characteristic function).

$$\varphi_X(u) = \varphi_X(u_1, ..., u_n) = E[e^{iu \cdot X}]; u \in R^n$$

It's useful because the chf of sum of rv's is a product of chf and not a convolution.

```
f[x_] := 1/2 HeavisideTheta[1 - Abs[x]]
h = FourierTransform[f[x], x, y]
f2 = InverseFourierTransform[h*h, y, x]
f3 = InverseFourierTransform[h*h*h, y, x]
f4 = InverseFourierTransform[h*h*h*h, y, x]
Plot[f3, {x, -5, 5}]
Plot[f4, {x, -5, 5}]
```

The plot's for n=3 and n=4 in previous section were created with the code above, which was faster than running convolution and I could do that thanks to the properties of chf.

6 Law of large numbers weak (with proof) [Theor 4.23] and strong [Theor 4.25] with applications to the calculation of an integral with the Monte Carlo method [Ex 4.26]

6.1 Weak law of large numbers

In short, law of large numbers claims that avarege probability of n iid rv. will approach it's expected value with n aproaching infinity.

Given a sequence$(X_n)n \in N$ of rv 's iid with $E[|X_n|] < +\infty$, and taken $S_n = X_1 + \cdots + X_n$ and $E[X_n] = m$, it turns out that:

$$\frac{S_n}{n} \xrightarrow{P} m$$

6.1.1 Proof

First lets calculate chf of the given random sequence

$\varphi_n(u) = E[e^{iuS_n/n}] = \prod_{k=1}^{n} E[e^{iuX_k/n}] = [\varphi_n(\frac{u}{n})]^n$

We will write the chf of the rv as:

$\varphi(u) = 1 + ium + o(u)$ for $u \to 0 \Longrightarrow \varphi(\frac{u}{n}) = 1 + i\frac{u}{n}m + o(\frac{1}{n})$ for $n \to \infty$

Then we can just plug it in to the previous equation and get:

$\varphi_n(u) = [1 + i\frac{u}{n}m + o(\frac{1}{n})]^n \to e^{imu}$

6.2 Strong law of large numbers

The only difference between weak and strong law of large numbers is that the strong law claims that the sequence converges to the expected value almost surely, so the probabilty that it doesn't converge is 0.

6.2.1 Monte Carlo

The Monte Carlo method is using strong law of large numbers to estimate a numerical value of an integral, by randomly selecting points on the plot and checking if the point is below or above a function.

By having enough points and knowing the area of the whole plot, one can use a ratio of point below to point above, multiplied by the area of the plot, to get the area under the function - namely the integral.

This can be done, because the probability to get below the line, will aproach almost surely the expectation value of this event, when we will try enough number of times.

7 Central Limit Theorem (with proof) [Theor 4.27] and Lyapunov conditions to replace the hypothesis of identical distribution [Theor 4.28]

7.1 Central Limit Theorem

Take a sequence$(X_n)n \in N$ of iid rv 's with$E[X_n^2] < +\infty$ and $V[X_n] > 0$, and define $S_n = X_1 + ... + X_n$then:

$$\frac{S_n - E[S_n]}{\sqrt{V[S_n]}} \xrightarrow{d} N(0,1)$$

What it means, is that for big enough random sequences and enough sampling points, with well defined expectation value and variance, it's mean distribution will approach the Normal Gauss distribution.

7.1.1 Proof

We will take:

$m = E[X_n]$

$\sigma^2 = V[X_n]$

$\varphi(u) = E[e^{iu(X_n - m)}]$

$E[S_n] = nm$

$V[S_n] = n\sigma^2$

$S_n^* = \frac{1}{\sigma\sqrt{n}} \sum\limits_{k=1}^{n} (X_k - m)$

r.vs are iid, so we can use the sam trick as for a law of large numbers and get

$\varphi_n(u) = E[e^{iuS_n^*}] = \prod_{k=1}^{n} E[e^{iu(X_k-m)/\sigma\sqrt{n}}] = [\varphi_n(\frac{u}{\sigma\sqrt{n}})]^n$
$\varphi(u) = 1-\frac{\sigma^2 u^2}{2} + o(u^2)$with $u \to 0$
and pluging $\varphi(u)$ into $\varphi_n(u)$:
$\varphi_n(u) = [1-\frac{\sigma^2 u^2}{2} + o(u^2)]^n \Rightarrow \varphi_n(u) = [1 - \frac{u^2}{2n} + o(\frac{1}{n})]^n \xrightarrow{e} e^{-u^2/2}$
Where $e^{-u^2/2}$is the chf of $N(0,1)$

7.2 Lyapunov conditions

If the distribution meets Lyapunov conditions:
$\frac{1}{V_n^{2+\delta}} \sum_{k=1}^{n} E[|X_k - m_k|^{2+\delta}] \xrightarrow{n} 0$ for $m_n = E[X_k]$ and $V_n = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}$
then it holds also for independent r.v (not necesserily identically distrubuted)