

Lecture 32: Central limit theorem

The central limit theorem explains why the normal distribution

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

is prevalent. If we add independent random variables and normalize them so that the mean is zero and the standard deviation is 1, then the distribution of the sum converges to the normal distribution.

Given a random variable X with expectation m and standard deviation σ define the **normalized random variable** $X^* = (X - m)/\sigma$.

The normalized random variable has the mean 0 and the standard deviation 1. The standard normal distribution mentioned above is an example. We have seen that S_n/n converges to a definite number if the random variables are uncorrelated. We have also seen that the standard deviation of S_n/n goes to zero.

A sequence X_n of random variables **converges in distribution** to a random variable X if for every trigonometric polynomial f , we have $E[f(X_n)] \rightarrow E[f(X)]$.

This means that $E[\cos(tX_n)] \rightarrow E[\cos(tX)]$ or $E[\sin(tX_n)] \rightarrow E[\sin(tX)]$ converge for every t . We can combine the cos and sin to $\exp(itx) = \cos(tx) + i \sin(tx)$ and cover both at once by showing $E[e^{itX_n}] \rightarrow E[e^{itX}]$ for $n \rightarrow \infty$. So, checking the last statement for every t is equivalent to check the convergence in distribution.

The function $\phi_X(t) = E[e^{itX}]$ is called the **characteristic function** of a random variable X .

Convergence in distribution is equivalent to the statement that the cumulative distribution functions $F_n(c) = P[X_n \leq c]$ converge to $F(c) = P[X \leq c]$ at every point c at which F is continuous. An other statement which is intuitive is that if the distribution is such that all moments $E[X^m]$ exist, it is enough to check that the moments $E[X_n^m]$ converge to $E[X^m]$ for all m . Trigonometric polynomials are preferred because they do not require the boundedness of the moments. "Convergence in distribution" is also called "convergence in law" or "weak convergence".

The following result is one of the most important theorems in probability theory. It explains why the standard normal distribution is so important.

Central limit theorem: Given a sequence of IID random variables with finite mean and variance and finite $E[X^3]$. Then S_n^* converges in distribution to the standard normal distribution.

Proof. Let X be a $N(0, 1)$ distributed random variable. We show $E[e^{itS_n^*}] \rightarrow E[e^{itX}]$ for any fixed t . Since any of the two random variables X_k, X_l are independent,

$$E[\exp(i(X_k + X_l))] = E[\exp(iX_k)]E[\exp(iX_l)].$$

More generally

$$E[\exp(itS_n)] = E[\exp(it(X_1 + \dots + X_n))] = E[\exp(itX_1)] \cdots E[\exp(itX_n)].$$

Since we normalize the random variables, we can assume that each X_k has zero expectation and variance 1. If $E[X_k] = 0, E[X_k^2] = 1$, we have for each of the n factors

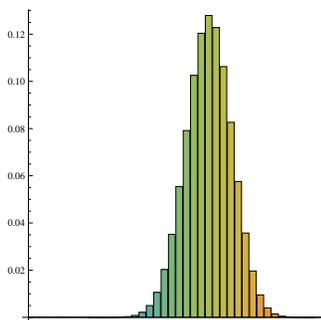
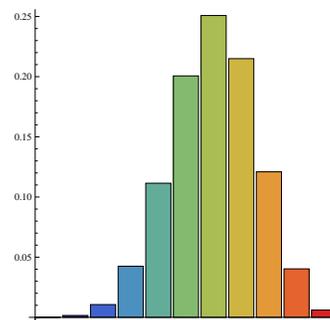
$$E[e^{itX_k/\sqrt{n}}] = (1 - \frac{t^2}{2n} - \frac{it^3 E[X^3]}{3n^{3/2}} + \dots).$$

Using $e^{-t^2/2} = 1 - t^2/2 + \dots$, we get

$$E[e^{itS_n/\sqrt{n}}] = (1 - \frac{t^2}{2n} + R_n/(n^{3/2}))^n \rightarrow e^{-t^2/2}.$$

The last step uses a Taylor remainder term $R_n/n^{3/2}$ term. It is here that the $E[X^3] < \infty$ assumption has been used. The statement now follows from

$$E[e^{itX}] = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx = e^{-t^2/2}.$$



- 1 We throw a fair dice n times. The distribution of S_n is a binomial distribution. The mean is np and the standard deviation is $\sqrt{np(1-p)}$. The distribution of $(S_n - np)/\sqrt{np(1-p)}$ looks close to the normal distribution. This special case of the central limit theorem is called the **de Moivre-Laplace theorem**. It was proven by de Moivre in 1730 in the case $p = 1/2$ and in 1812 for general $0 < p < 1$ by Laplace.

Statistical inference

For the following see also Cliffs notes page 89-93 (section 14.6): what is the probability that the average S_n/n is within ϵ to the mean $E[X]$?

The probability that S_n/n deviates more than $R\sigma/\sqrt{n}$ from $E[X]$ can for large n be estimated by

$$\frac{1}{\sqrt{2\pi}} \int_R^\infty e^{-x^2/2} dx .$$

Proof. Let $p = E[X]$ denote the mean of X_k and σ the standard deviation. Denote by X a random variable which has the standard normal distribution $N(0, 1)$. We use the notation $X \sim Y$ if X and Y are close in distribution. By the central limit theorem

$$\frac{S_n - np}{\sqrt{n}\sigma} \sim X .$$

Dividing nominator and denominator by n gives $\frac{\sqrt{n}}{\sigma}(\frac{S_n}{n} - p) \sim X$ so that

$$\frac{S_n}{n} - p \sim X \frac{\sigma}{\sqrt{n}} .$$

The term σ/\sqrt{n} is called the **standard error**. The central limit theorem gives some insight why the standard error is important.

In scientific publications, the standard error should be displayed rather than the standard deviation.¹

A squirrel accidentally drank from liquor leaking from a garbage bag. Tipsy, it walks forth and back on a telephone cable, randomly taking a step of one foot forward or backwards each second. How far do we expect him to be drifted off after 3 minutes

2 **Answer:** The mean is zero, the standard deviation is 1. By the central limit theorem we expect the drift off from $p = 0$ to be \sqrt{n} because

$$S_n \sim X\sigma\sqrt{n} .$$

That means we can expect the squirrel to be within $\sqrt{180} = 13$ feet. The chance to see it in this neighborhood is about 2/3 because the probability to be within the standard deviation interval is about 2/3 (see homework).



A probability space and a random variable X define a **null hypothesis**, a model for your experiment. Assume you measure $X = c$. Assuming c is larger than the expectation, the **P-value** of this experiment is defined as $P[X \geq c]$. If c is smaller than the expectation, we would define the P-value as $P[X \leq c]$.

The **P-value** is the probability that the test statistics is at least as extreme as the experiment.

3 The assumption in the previous problem that our squirrel is completely drunk the **null hypothesis**. Assume we observe the squirrel after 3 minutes at 20 feet from the original

place. What is the **P-value** of this observation? The **P-value** is $P[S_{180} \geq 20]$. Since $S_{180}/\sqrt{180}$ is close to normal and $c = 20/\sqrt{180} = 1.49$, we can estimate the **P** value as

$$\int_c^\infty f_{normal}(x) dx = 0.068 .$$

Since we know that the actual distribution is a Binomial distribution, we could have computed the P-value exactly as $\sum_{k=100}^{180} B(180, k)p^k(1-p)^{n-k} = 0.078$ with $p = 1/2$. A P-value smaller than 5 percent is called **significant**. We would have to reject the null hypothesis and the squirrel is not drunk. In our case, the experiment was not significant.

The central limit theorem is so important because it gives us a tool to estimate the **P-value**. It is much better in general than the estimate given by Chebyshev.

¹Geoff Cumming, Fiona Fidler, 1 and David L. Vaux: Error bars in experimental biology, Journal of Cell biology, 177, 1, 2007 7-11

Homework due April 20, 2011

- 1 Look at the standard normal probability distribution function

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

Verify that for this distribution, there is a probability of 68 percent to land within the interval $[-\sigma, \sigma]$ is slightly larger than $2/3$.

You have verified the rule of thumb:

For the normal distribution $N(m, \sigma)$, the chance to be in the interval $[m - \sigma, m + \sigma]$ is about two thirds.

- 2 In many popular children games in which one has to throw a dice and then move forward by the number of eyes seen. After n rounds, we are at position S_n . If you play such a game, find an interval of positions so that you expect to be with at least $2/3$ percent chance in that interval after n rounds.



Remark: An example is the game "Ladder" which kids play a lot in Switzerland: it is a random walk with drift 3.5. What makes the game exciting are occasional accelerations or setbacks. Mathematically it is a **Markov process**. If you hit certain fields, you get pushed ahead (sometimes significantly) but for other fields, you can almost lose everything. The game stays interesting because even if you are ahead you can still end up last or you trail behind all the game and win in the end.

- 3 We play in a Casino with the following version of the martingale strategy. We play all evening and bet one dollar on black until we reach our goal of winning 10 dollars. Assume each game lasts a minute. How long do we expect to wait until we can go home? (You can assume that the game is fair and that you win or lose with probability $1/2$ in each game.)

Remark: this strategy appears frequently in movies, usually when characters are desperate. Examples are "Run Lola Run", "Casino Royale" or "Hangover".

