Mathematics for Informatics 4a

José Figueroa-O'Farrill

Lecture 14
9 March 2012

The story of the film so far...

- X, Y independent random variables and $Z = X + Y$: $f_Z = f_X \star f_Y$, where $\star$ is the **convolution**
- X, Y with joint density $f(x, y)$ and $Z = g(X, Y)$:

$$\mathbb{E}(Z) = \iint g(x, y) f(x, y)\, dx\, dy$$

- X, Y independent:
  - $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
  - $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
  - $M_{X+Y}(t) = M_X(t)M_Y(t)$, where $M_X(t) = \mathbb{E}(e^{tX})$
- We defined **covariance** and **correlation** of two r.v.s
- Proved **Markov's** and **Chebyshev's** inequalities
- Proved the **(weak) law of large numbers** and the **Chernoff bound**
- Waiting times of Poisson processes are exponentially distributed

More approximations

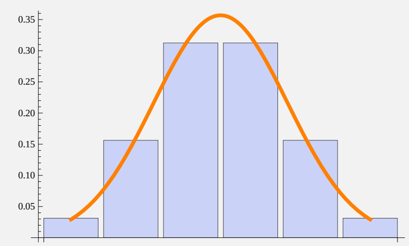
In Lecture 7 we saw that the binomial distribution with parameters n, p can be approximated by a Poisson distribution with parameter λ in the limit as $n \to \infty$, $p \to 0$ but $np \to \lambda$:

$$\binom{n}{k} p^k (1-p)^{n-k} \sim e^{-\lambda} \frac{\lambda^k}{k!}$$

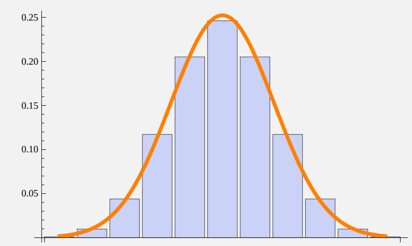
But what about if $n \to \infty$ *but* $p \not\to 0$*?*
For example, consider flipping a fair coin n times and let X denote the discrete random variable which counts the number of heads. Then

$$\mathbb{P}(X = k) = \binom{n}{k} \frac{1}{2^n}$$

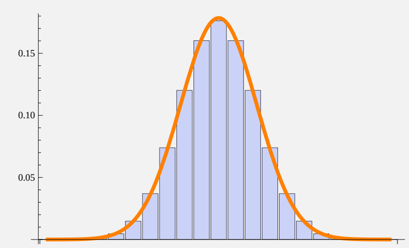
Lectures 6 and 7: this distribution has $\mu = n/2$ and $\sigma^2 = n/4$.



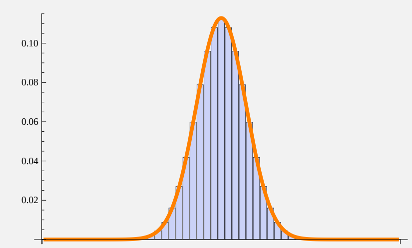
$n = 5$



$n = 10$



$n = 20$



$n = 50$

Normal limit of (symmetric) binomial distribution

Theorem

Let X be binomial with parameter n and $p = \frac{1}{2}$. Then for n large and $k - n/2$ not too large,

$$\binom{n}{k}\frac{1}{2^n} \simeq \frac{1}{\sqrt{2\pi}\sigma} e^{-(k-\mu)^2/2\sigma^2} = \sqrt{\frac{2}{n\pi}} e^{-2(k-n/2)^2/n}$$

for $\mu = n/2$ and $\sigma^2 = n/4$.

The proof rests on the de Moivre/Stirling formula for the factorial of a large number:

$$n! \simeq \sqrt{2\pi} n^n \sqrt{n} e^{-n}$$

which implies that

$$\binom{n}{n/2} = \frac{n!}{(n/2)!(n/2)!} \simeq 2^n \sqrt{\frac{2}{\pi n}}$$

Proof

Let $k = \frac{n}{2} + x$. Then

$$\begin{aligned}\binom{n}{k} 2^{-n} &= \binom{n}{\frac{n}{2}+x} 2^{-n} = \frac{n! 2^{-n}}{\left(\frac{n}{2}+x\right)! \left(\frac{n}{2}-x\right)!} \\ &= \frac{n! 2^{-n}}{\left(\frac{n}{2}\right)! \left(\frac{n}{2}\right)!} \times \frac{\frac{n}{2}\left(\frac{n}{2}-1\right)\cdots\left(\frac{n}{2}-(x-1)\right)}{\left(\frac{n}{2}+1\right)\left(\frac{n}{2}+2\right)\cdots\left(\frac{n}{2}+x\right)} \\ &\simeq \sqrt{\frac{2}{n\pi}} \times \frac{1\left(1-\frac{2}{n}\right)\cdots\left(1-(x-1)\frac{2}{n}\right)\left(\frac{n}{2}\right)^x}{\left(1+\frac{2}{n}\right)\left(1+2\frac{2}{n}\right)\cdots\left(1+x\frac{2}{n}\right)\left(\frac{n}{2}\right)^x}\end{aligned}$$

Now we use the exponential approximation

$$1 - z \simeq e^{-z} \quad \text{and} \quad \frac{1}{1+z} \simeq e^{-z}$$

(valid for z small) to rewrite the big fraction in the RHS.

Proof – continued.

$$\begin{aligned}\binom{n}{k} 2^{-n} &\simeq \sqrt{\frac{2}{n\pi}} \exp\left[-\frac{4}{n} - \frac{8}{n} - \cdots - \frac{2(x-1)}{n} - \frac{2x}{n}\right] \\ &= \sqrt{\frac{2}{n\pi}} \exp\left[-\frac{4}{n}(1 + 2 + \cdots + (x-1)) - \frac{2x}{n}\right] \\ &= \sqrt{\frac{2}{n\pi}} \exp\left[-\frac{4}{n}\frac{x(x-1)}{2} - \frac{2x}{n}\right] \\ &= \sqrt{\frac{2}{n\pi}} e^{-2x^2/n}\end{aligned}$$

which is indeed a normal distribution with $\sigma^2 = \frac{n}{4}$. □

A similar proof shows that the general binomial distribution with $\mu = np$ and $\sigma^2 = np(1-p)$ is also approximated by a normal distribution with the same μ and σ^2.

Example (Rolling a die *ad nauseam*)

It's raining outside, you are bored and you roll a fair die 12000 times. Let X be the number of sixes. *What is* $\mathbb{P}(1900 < X < 2200)$*?*

The variable X is the sum $X_1 + \cdots + X_{12000}$, where X_i is the number of sixes on the ith roll. This means that X is binomially distributed with parameter $n = 12000$ and $p = \frac{1}{6}$, so $\mu = pn = 2000$ and $\sigma^2 = np(1-p) = \frac{5000}{3}$.

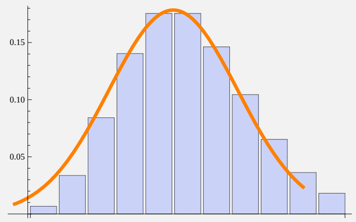
$X \in (1900, 2200)$ iff $\frac{X-2000}{\sigma} \in (-\sqrt{6}, 2\sqrt{6})$, whence

$$\mathbb{P}(1900 < X < 2200) \simeq \Phi(2\sqrt{6}) - \Phi(-\sqrt{6}) \simeq 0.992847$$

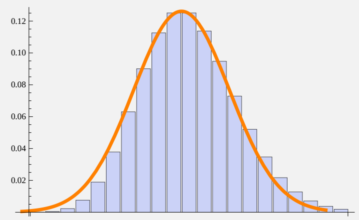
The exact result is

$$\sum_{k=1901}^{2199} \binom{12000}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{12000-k} \simeq 0.992877$$

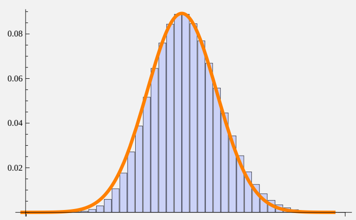
Normal limit of Poisson distribution



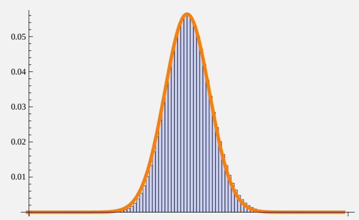
$\lambda = 5$



$\lambda = 10$

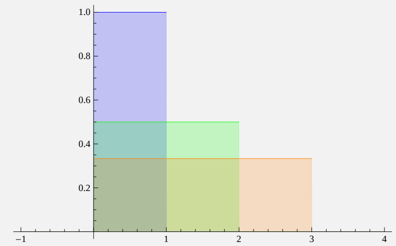


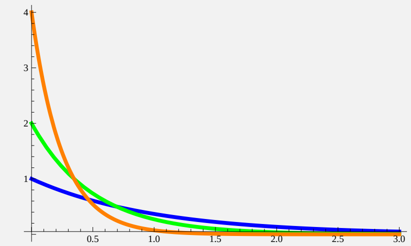
$\lambda = 20$



$\lambda = 50$

- We have just shown that in certain limits of the defining parameters, two discrete probability distributions tend to normal distributions:
  - the binomial distribution in the limit $n \to \infty$,
  - the Poisson distribution in the limit $\lambda \to \infty$
- What about continuous probability distributions?
- We could try with the uniform or exponential distributions:



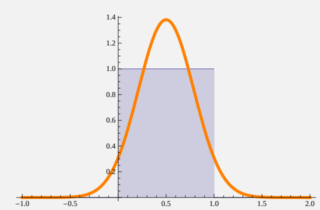


- No amount of rescaling is going to work. Why?

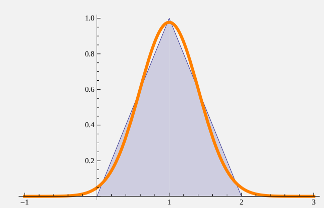
- The binomial and Poisson distributions have the following property:
  - if X, Y are binomially distributed with parameters (n, p) and (m, p), $X + Y$ is binomially distributed with parameter $(n + m, p)$
  - if X, Y are Poisson distributed with parameters λ and μ, $X + Y$ is Poisson distributed with parameter $\lambda + \mu$
- It follows that if $X_1, X_2, \dots$ are i.i.d. with binomial distribution with parameters (m, p), $X_1 + \cdots + X_n$ is binomial with parameter (nm, p). Therefore m large is equivalent to adding many of the X_i.
- It also follows that if $X_1, X_2, \dots$ are i.i.d. with Poisson distribution with parameter λ, $X_1 + \cdots + X_n$ is Poisson distributed with parameter $n\lambda$ and again λ large is equivalent to adding a large number of the X_i.
- The situation with the uniform and exponential distributions is different.

Sum of uniformly distributed variables

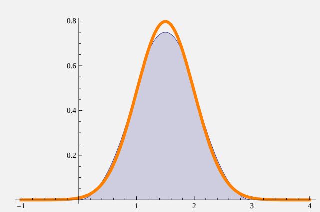
X_i i.i.d. uniformly distributed on $[0, 1]$: then $X_1 + \cdots + X_n$ is



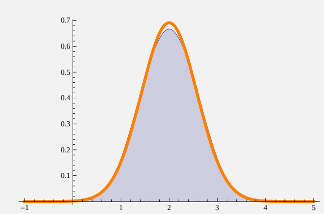
$n = 1$



$n = 2$



$n = 3$



$n = 4$

Sum of exponentially distributed variables

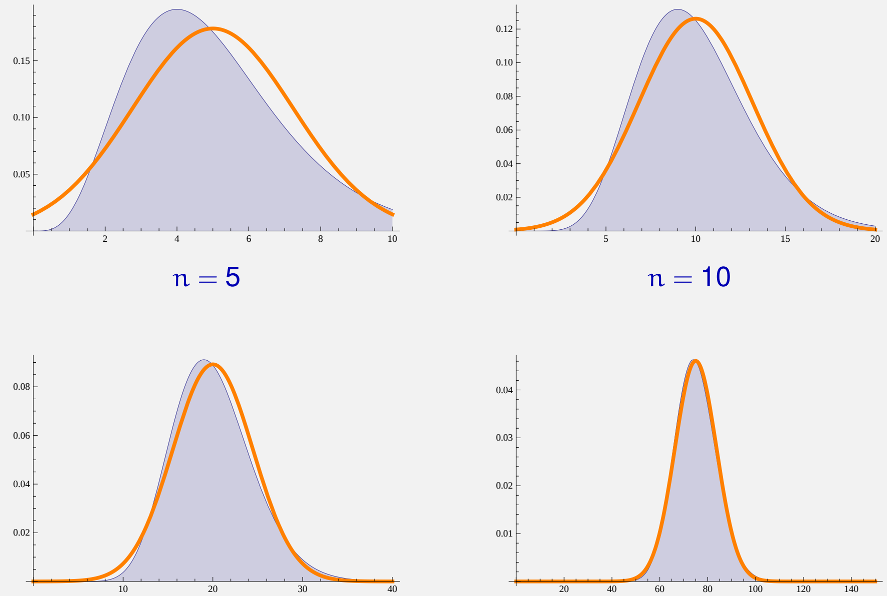
- If X_i are i.i.d. exponentially distributed with parameter λ, we already saw that $Z_2 = X_1 + X_2$ has a "gamma" probability density function:

$$f_{Z_2}(z) = \lambda^2 z e^{-\lambda z}$$

- It is not hard to show that $Z_n = X_1 + \cdots + X_n$ has probability density function

$$f_{Z_n}(z) = \lambda^n \frac{z^{n-1}}{(n-1)!} e^{-\lambda z}$$

- What happens when we take n large?



$n = 5$ $n = 10$

$n = 20$ $n = 75$

The Central Limit Theorem

- Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean μ and variance σ^2.
- Let $Z_n = X_1 + \cdots + X_n$.
- Then Z_n has mean $n\mu$ and variance $n\sigma^2$, but in addition we have

Theorem (Central Limit Theorem)

In the limit as $n \to \infty$,

$$\mathbb{P}\left(\frac{Z_n - n\mu}{\sqrt{n}\sigma} \leqslant x\right) \to \Phi(x)$$

with Φ the c.d.f. of the standard normal distribution.

In other words, for n large, Z_n is normally distributed.

- Our 4-line proof of the CLT rests on **Lévy's continuity law**, which we will not prove.
- Paraphrasing: "the m.g.f. determines the c.d.f."
- It is then enough to show that the limit $n \to \infty$ of the m.g.f. of $\frac{Z_n - n\mu}{\sqrt{n}\sigma}$ is the m.g.f. of the standard normal distribution.

Proof of CLT.

- We shift the mean: the variables $Y_i = X_i - \mu$ are i.i.d. with mean 0 and variance σ^2, and $Z_n - n\mu = Y_1 + \cdots + Y_n$.
- $M_{Z_n - n\mu}(t) = M_{Y_1}(t) \cdots M_{Y_n}(t) = M_{Y_1}(t)^n$, by i.i.d.
- $M_{\frac{Z_n - n\mu}{\sqrt{n}\sigma}}(t) = M_{Z_n - n\mu}\left(\frac{t}{\sqrt{n}\sigma}\right) = M_{Y_1}\left(\frac{t}{\sqrt{n}\sigma}\right)^n$
- $M_{\frac{Z_n - n\mu}{\sqrt{n}\sigma}}(t) = \left(1 + \frac{\sigma^2 t^2}{2n\sigma^2} + \cdots\right)^n \to e^{t^2/2}$, which is the m.g.f. of a standard normal variable.

□

Crucial observation

The CLT holds regardless of how the X_i are distributed!
The sum of **any** large number of i.i.d. normal variables **always** tends to a normal distribution.

This also explains why normal distributions are so popular in probabilistic modelling.
Let us look at a few examples.

Example (Rounding errors)

Suppose that you round off 108 numbers to the nearest integer, and then add them to get the total S. Assume that the rounding errors are independent and uniform on $[-\frac{1}{2}, \frac{1}{2}]$. *What is the probability that S is wrong by more than 3? more than 6?*
Let $Z = X_1 + \cdots + X_{108}$. We may approximate it by a normal distribution with $\mu = 0$ and $\sigma^2 = \frac{108}{12} = 9$, whence $\sigma = 3$.
S is wrong by more than 3 iff $|Z| > 3$ or $\frac{|Z-\mu|}{\sigma} > 1$ and hence

$$\begin{aligned}\mathbb{P}(|Z-\mu| > \sigma) &= 1 - \mathbb{P}(|Z-\mu| \leqslant \sigma)\\ &= 1 - (2\Phi(1) - 1) = 2(1-\Phi(1)) \simeq 0.3174\end{aligned}$$

S is wrong by more than 6 iff $\frac{|Z-\mu|}{\sigma} > 2$ and hence

$$\begin{aligned}\mathbb{P}(|Z-\mu| > 2\sigma) &= 1 - \mathbb{P}(|Z-\mu| \leqslant 2\sigma)\\ &= 1 - (2\Phi(2) - 1) = 2(1-\Phi(2)) \simeq 0.0456\end{aligned}$$

Place your bets!

Example (Roulette)

A roulette wheel has 38 slots: the numbers 1 to 36 (18 black, 18 red) and the numbers 0 and 00 in green.

You place a $\pounds 1$ bet on whether the ball will land on a red or black slot and win $\pounds 1$ if it does. Otherwise you lose the bet. Therefore you win $\pounds 1$ with probability $\frac{18}{38} = \frac{9}{19}$ and you "win" $-\pounds 1$ with probability $\frac{20}{38} = \frac{10}{19}$.

After 361 spins of the wheel, what is the probability that you are ahead? (Notice that $361 = 19^2$.)

Example (Roulette – continued)

Let X_i denote your winnings on the ith spin of the wheel. Then $\mathbb{P}(X_i = 1) = \frac{9}{19}$ and $\mathbb{P}(X_i = -1) = \frac{10}{19}$. The mean is therefore

$$\mu = \mathbb{P}(X_i = 1) - \mathbb{P}(X_i = -1) = -\frac{1}{19}$$

and the variance is

$$\sigma^2 = \mathbb{P}(X_i = 1) + \mathbb{P}(X_i = -1) - \mu^2 = 1 - \frac{1}{361} = \frac{360}{361}$$

Then $Z = X_1 + \cdots + X_{361}$ has mean -19 and variance 360. This means that after 361 spins you are down $\pounds 19$ on average. We are after the probability $\mathbb{P}(Z \geqslant 0)$:

$$\begin{aligned}\mathbb{P}(Z \geqslant 0) &= \mathbb{P}\left(\frac{Z+19}{\sqrt{360}} \geqslant \frac{19}{\sqrt{360}}\right) = 1 - \mathbb{P}\left(\frac{Z+19}{\sqrt{360}} \leqslant \frac{19}{\sqrt{360}}\right)\\ &\simeq 1 - \Phi(1) \simeq 0.1587\end{aligned}$$

So there is about a 16% chance that you are ahead.

Example (Measurements in astronomy)

Astronomical measurements are subject to the vagaries of weather conditions and other sources of errors. Hence in order to estimate, say, the distance to a star one takes the average of many measurements. Let us assume that different measurements are i.i.d. with mean d (the distance to the star) and variance 4 (light-years^2). *How many measurements should we take to be "reasonably sure" that the estimated distance is accurate to within half a light-year?*

Let X_i denote the measurements and $Z_n = X_1 + \cdots + X_n$. Let's say that "reasonably sure" means 95%, which is 2σ in the standard normal distribution. (In Particle Physics, "reasonably sure" means 5σ, but this is Astronomy.) Then we are after n such that

$$\mathbb{P}\left(|\tfrac{Z_n}{n} - d| \leqslant 0.5\right) \simeq 0.95$$

Example (Measurements in astronomy – continued)

By the CLT we can assume that $\frac{Z_n - nd}{2\sqrt{n}}$ is standard normal, so we are after

$$\mathbb{P}\left(|\tfrac{Z_n - nd}{2\sqrt{n}}| \leqslant \tfrac{\sqrt{n}}{4}\right) \simeq 0.95$$

or, equivalently, $\frac{\sqrt{n}}{4} = 2$, so that $n = 64$.

A question remains: *is $n = 64$ large enough for the CLT?* To answer it, we need to know more about the distribution of the X_i. However Chebyshev's inequality can be used to provide a safe n. Since $\mathbb{E}\left(\frac{Z_n}{n}\right) = d$ and $\mathrm{Var}\left(\frac{Z_n}{n}\right) = \frac{4}{n}$, Chebyshev's inequality says

$$\mathbb{P}\left(|\tfrac{Z_n}{n} - d| > 0.5\right) \leqslant \frac{4}{n(0.5)^2} = \frac{16}{n}$$

so choosing $n = 320$ gives $\mathbb{P}\left(|\frac{Z_n}{n} - d| > 0.5\right) \leqslant 0.05$ or $\mathbb{P}\left(|\frac{Z_n}{n} - d| \leqslant 0.5\right) \geqslant 0.95$ as desired.

Summary

- The binomial distribution with parameters n, p can be approximated by a normal distribution (with same mean and variance) for n large.
- Similarly for the Poisson distribution with parameter λ as $\lambda \to \infty$
- These are special cases of the **Central Limit Theorem**: if X_i are i.i.d. with mean μ and (nonzero) variance σ^2, the sum $Z_n = X_1 + \cdots + X_n$ for n large is normally distributed.
- We saw some examples on the use of the CLT: rounding errors, roulette game, astronomical measurements.