Chapter 4 and beyond: Learning about distributions from finite data

Part 2, Wed 13 July

Jess Kunke

MATH/STAT 394: Probability I (Summer 2022 A-term)

Outline

Weak law of large numbers (WLLN)

Normal (Gaussian) distribution

Normal Approximation

Additional details

Outline

Weak law of large numbers (WLLN)

Normal (Gaussian) distribution

Normal Approximation

Additional details

Weak law of large numbers

- Setting: an experiment consisting of a series of iid trials
- Recall the definition of the empirical mean:

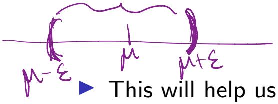
$$\overline{X_n} = \frac{X_1 + \ldots + X_n}{n}$$

for $X_i \overset{\text{iid}}{\sim} X$ where X is a RV that models a single trial in our experiment

→ coin flips: $\overline{X_n}$ = proportion of flips that came up heads (tails) ⇒ estimate of p

- If μ is the mean of X and σ^2 is its variance, then $\mathrm{E}[\overline{X_n}] = \mu$ and $\mathrm{Var}(\overline{X_n}) = \sigma^2/n$.
- For any number $\varepsilon > 0$, by Chebyshev's inequality,

$P(|\overline{X_n} - \mu| \leq \varepsilon) \to 1$



$$P(|\overline{X_n} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \underset{n\to\infty}{\to} 0$$

- This will help us formalize the idea that $\overline{X_n}$ converges to μ (next slide)

Weak law of large numbers

Theorem

Let $X_1, \ldots, X_n$ be iid RVs with finite variance σ^2 and finite mean μ.
For any fixed $\varepsilon > 0$,

$$\lim_{n \to +\infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| < \varepsilon \right) = 1.$$

We say the sample average converges in probability *towards the expected value.*

Interpretation:

No matter how small an interval $[\mu - \varepsilon, \mu + \varepsilon]$ you choose around μ,
as n becomes large, the observed empirical mean will lie inside this interval
with overwhelming probability.

Strong law of large numbers

It turns out, actually, that an even stronger type of convergence holds:

Theorem (Strong Law of Large Numbers)

Let $X_1, \ldots, X_n$ be iid RV with finite mean μ.

$$P\left(\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu\right) = 1$$

We say the sample average converges almost surely *towards the expected value.*

Strong law of large numbers

It turns out, actually, that an even stronger type of convergence holds:

Theorem (Strong Law of Large Numbers)

Let $X_1, \ldots, X_n$ be iid RV with finite mean μ.

$$P\left(\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu\right) = 1$$

We say the sample average converges almost surely *towards the expected value.*

The proof is much more complex; we will not cover it. In this course, we will focus on the WLLN instead.

Note:

- Almost sure convergence is similar to pointwise convergence in real analysis
- MATH/STAT 395 will cover more about different types of convergence

WLLN: example

let's say $c = 2$.

Example

Suppose we want to estimate the population mean of a RV X using the sample mean $\overline{X}_n$ over a finite number of independent samples or data points $X_1, ..., X_n$. Suppose that we know that $\text{Var}(X) \leq c$ for some value c. How large does our sample need to be (how many data points, or how large does n need to be) in order for us to be 99% sure that our estimate (the sample mean) is within 0.05 of the correct value?

Let $\mu = E[X_i]$, $\sigma^2 = Var(X_i) \leq c$. $\Rightarrow E[\overline{X}_n] = \mu$, $Var(\overline{X}_n) = \frac{\sigma^2}{n} \leq \frac{c}{n}$.

Cheby.: $P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$

need n s.t. $P(|\overline{X}_n - \mu| < 0.05) \geq 0.99$

$\Rightarrow$ n s.t. $P(|\overline{X}_n - \mu| \geq 0.05) \leq 0.01$ $\leftarrow$ here $\varepsilon = 0.05$

sufficient: $\frac{\sigma^2}{n\varepsilon^2} \leq 0.01 \Rightarrow n \geq \frac{\sigma^2}{(0.05)^2(0.01)} = 4\times10^4 \cdot \sigma^2$.

$2 = c \geq \sigma^2$, so $4\times10^4 \cdot 2 \geq 4\times10^4 \cdot \sigma^2$. Suppose $\sigma^2 = 1$

$\therefore$, if $n \geq 4\times10^4 \cdot 2$, then n meets the condition $n \geq 4\times10^4 \cdot \sigma^2$.

Outline

Weak law of large numbers (WLLN)

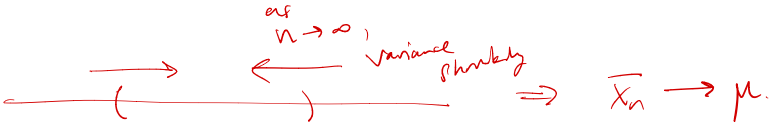
Normal (Gaussian) distribution

Normal Approximation

Additional details

Motivation from the law of large numbers

- When you flip a coin, eventually the frequency of tails you observe will be the actual probability to get a tail
- We want to quantify the error in our estimate of that probability, or quantify the number of flips we need to do to ensure the error is below some amount
- We know that the Chebyshev bound can be loose/uninformative
- How can we model the distribution of the sample mean $\overline{X_n}$ around the true mean as $n \to +\infty$?
- This is given by the Gaussian or normal distribution



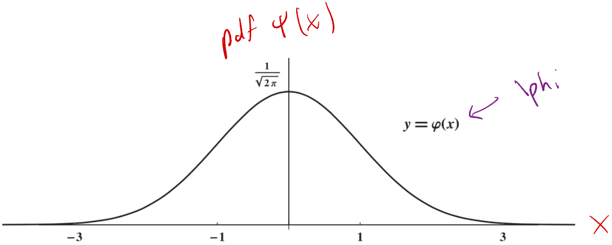
Gaussian distribution

Definition

A RV Z has the **standard normal distribution** (or **standard Gaussian distribution**) if Z has density function

$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{for } x \in \mathbb{R}.$$

We denote it $Z \sim \mathcal{N}(0, 1)$ since it has mean 0 and variance 1.



Gaussian distribution

Sanity check:

- Is the pdf of the Gaussian distribution a valid pdf? (properties?)

Gaussian distribution

Sanity check:

- Is the pdf of the Gaussian distribution a valid pdf? (properties?)

Lemma

$$\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}.$$

Derivation: The trick is to compute the square of the integral as a double integral and switch to polar coordinates

$$\begin{aligned}
\left(\int_{-\infty}^{+\infty} e^{-x^2/2} dx\right)^2 &= \left(\int_{-\infty}^{+\infty} e^{-x^2/2} dx\right) \cdot \left(\int_{-\infty}^{+\infty} e^{-y^2/2} dy\right) \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x^2/2 - y^2/2} dxdy \\
&= \int_{0}^{2\pi} \int_{0}^{+\infty} e^{-r^2/2} r dr d\theta \\
&= \int_{0}^{2\pi} \left[-e^{-r^2/2}\right]_{0}^{+\infty} d\theta = \int_{0}^{2\pi} d\theta = 2\pi
\end{aligned}$$

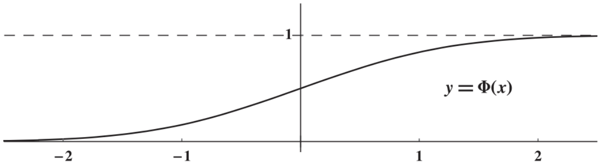
where we used the change of variable $x = r\cos(\theta), y = r\sin(\theta)$ (i.e. we used polar coordinates), such that $x^2 + y^2 = r^2$, $dxdy = rdrd\theta$ and the bounds go to 0 to $+\infty$ for the radius r and 0 to 2π for the angle θ.

cdf of Gaussian distribution

~~$F_Z(z) = z^2 - z$~~

- There is no closed form expression for the standard normal cdf!
- We'll denote the cdf by

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx.$$



cdf of Gaussian distribution

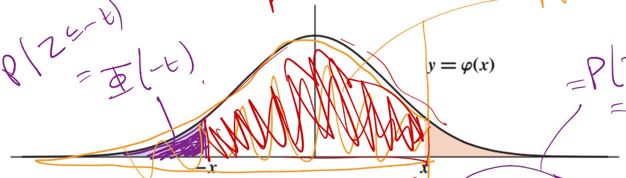
- There is no closed form expression for the standard normal cdf!
- We'll denote the cdf by

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx.$$

- So how do we compute the cdf $\Phi(t)$ for a value t? Lookup tables (e.g. textbook) or statistical software (e.g. `pnorm` in R)
- By symmetry of the distribution, for any t,

$$\Phi(-t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-t} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{t}^{+\infty} e^{-x^2/2} dx = 1 - \Phi(t)$$

- Example: What is the probability that $|Z| \leq 1$?



Lookup table (see also the back of your textbook)

remember $\Phi(-t) = 1 - \Phi(t)$.

\Phi in LaTeX

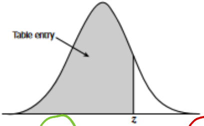
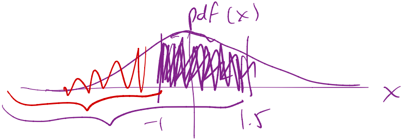


Table entry for z is the area under the standard normal curve to the left of z.

$\Phi(1) = 0.8413$

$\Phi(1.73) = 0.9582$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990



Example

Let $Z \sim \mathcal{N}(0,1)$. Find $P(-1 \leq Z \leq 1.5)$.

$$P(-1 \leq Z \leq 1.5) = \Phi(1.5) - \Phi(-1)$$

$$= \Phi(1.5) - (1 - \Phi(1)) = \Phi(1.5) + \Phi(1) - 1$$

$$= 0.9332 + 0.8413 - 1$$

$$= 0.7745.$$

77% prob.

or 0.7745

Lemma

If $Z \sim \mathcal{N}(0,1)$, then $E[Z] = 0$ and $Var(Z) = 1$.

- First check that $\mathrm{E}[Z]$ is well defined, which means showing that $\mathrm{E}[|Z|] < +\infty$.

  For that one shows that $\int_{-\infty}^{+\infty} |x| e^{-x^2/2} dx = 2 \int_0^{+\infty} x e^{-x^2/2} dx = 2$ is finite
- Then since the pdf of Z satisfies $\psi(x) = \psi(-x)$, we have that (ψ is the pdf of Z)

$$\int_{-a}^{a} x\psi(x)dx = \int_{-a}^{0} x\psi(x)dx + \int_{0}^{a} x\psi(x)dx = -\int_{0}^{a} x\psi(-x)dx + \int_{0}^{a} x\psi(x)dx = 0$$

- Therefore $\mathrm{E}[Z] = 0$
- On the other hand by integration by parts, i.e., $\int_a^b f'g = [fg]_a^b - \int_a^b fg'$ for $f(x) = -e^{-x^2/2}$ and $g(x) = x$.

$$\begin{aligned}
\mathrm{E}[Z^2] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx \\
&= -\frac{1}{\sqrt{2\pi}} \left(\left[x e^{-x^2/2} \right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-x^2/2} dx \right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = 1
\end{aligned}$$

Gaussian distribution

Can we generalize the standard normal distribution?

Example

Let $Z \sim \mathcal{N}(0,1)$ and let $X = \sigma Z + \mu$ for $\sigma > 0, \mu \in \mathbb{R}$.

1. Compute $E[X]$, $Var(X)$. — remember expectation is linear, variance is not
   $E(aX+b) = aE[X]+b$, $Var(aX+b) = a^2 Var(X)$
2. Compute the pdf of X.

$$E[X] = E[\sigma Z + \mu] = \sigma E[Z] + \mu = \sigma \cdot 0 + \mu = \mu.$$

$$Var(X) = Var(\sigma Z + \mu) = \sigma^2 \underbrace{Var(Z)}_{1} = \sigma^2.$$

$$F_X(x) = P(X \le x) = P(\sigma Z + \mu \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

$g(x)$

$$f_X(x) = F_X'(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)^2 \cdot \frac{1}{2}}$$

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

(general) normal distribution

chain rule:

$$\frac{d}{dx}\left[\Phi(g(x))\right] = \underbrace{g'(x)}_{1/\sigma} \cdot \underbrace{\Phi'}_{\varphi}(\underbrace{g(x)}_{\frac{x-\mu}{\sigma}}) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right).$$

Generic Gaussian distribution

- From the standard normal distribution, we can define a whole family of normal distributions as $X = \sigma Z + \mu$
- These distributions are entirely characterized (parameterized) by their mean and their variance

Definition

Let $\mu \in \mathbb{R}$ and $\sigma > 0$, a RV X has **the normal/Gaussian distribution with mean μ and variance σ^2** if X has the pdf

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \quad \text{for } x \in \mathbb{R}.$$

We denote it $X \sim \mathcal{N}(\mu, \sigma^2)$.

e.g. $X \sim N(99, 16)$

Some more cdf method practice!

Example

Let $\mu \in \mathbb{R}$, $\sigma > 0$ and $X \sim \mathcal{N}(\mu, \sigma^2)$
Let $a \neq 0$ and $b \in \mathbb{R}$, show that $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
In particular what is the dist. of $Z = \frac{X-\mu}{\sigma}$?

First consider $a > 0$.

$$F_Y(y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

$$f_Y(y) = F_Y'(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi}\, a\sigma} e^{-(y-(a\mu+b))^2 / 2a^2\sigma^2}$$

($a\sigma$: sd; $a\mu+b$: new mean; $a^2\sigma^2$: new variance)

follow similar steps for $a < 0$.

$$Z = \frac{X-\mu}{\sigma} = \frac{1}{\sigma} X - \frac{\mu}{\sigma} \Rightarrow Z \sim N\left(\frac{1}{\sigma}\mu - \frac{\mu}{\sigma}, \frac{1}{\sigma^2} \cdot \sigma^2\right)$$

($\frac{1}{\sigma}$: a; $-\frac{\mu}{\sigma}$: b)

$$= N(0, 1).$$

standardization

Generic Gaussian distribution

From generic to standard normal

- Computing prob. of $X \sim \mathcal{N}(\mu, \sigma^2)$ can be done by using the cdf of the standard normal dist.

$$P(X \in [a,b]) = P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right),$$

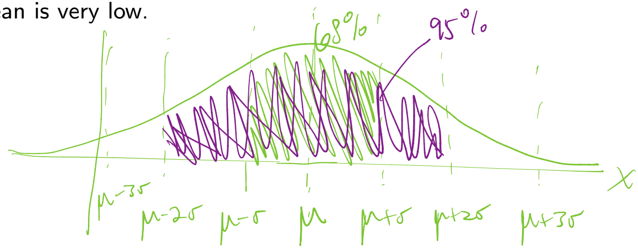
$$\implies \boxed{P(X \in [a,b]) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

If $X \sim N(\mu = 100, \sigma^2 = 4)$, then $\sigma = 2$

$\Rightarrow P(X \text{ is within } 1\ \sigma \text{ of } \mu) = P(|X-\mu| \le \sigma)$

$= P(|Z| \le 1)$

- if X has a normal distribution,
  - about 90% probability X falls within 1.645 SD of the mean,
  - about 95% probability X falls within **1.96 SD** of the mean,
  - about 99% probability X falls within 2.576 SD of the mean.
  - How does this compare to what you found with the standard normal?
- The probability of X falls 4, 5, or more standard deviations away from the mean is very low.



Summary

Weak law of large numbers (WLLN)

Let $X_1, \dots, X_n$ be iid RVs with finite variance σ^2 and finite mean μ.
For any fixed $\varepsilon > 0$,

$$\lim_{n \to +\infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| < \varepsilon \right) = 1$$

Standard normal/Gaussian distribution

- $Z \sim \mathcal{N}(0, 1)$ has pdf
$$\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$
- cdf $\Phi(x)$ not available in closed form but given by tables
- $\mathrm{E}[Z] = 0$, $\mathrm{Var}(Z) = 1$

Normal/Gaussian distribution

- $X \sim \mathcal{N}(\mu, \sigma^2)$ has $\mathrm{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$, and pdf
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$
- X can be constructed from Z via $X = \sigma Z + \mu$, and $Z = \frac{X-\mu}{\sigma}$

Outline

Weak law of large numbers (WLLN)

Normal (Gaussian) distribution

Normal Approximation

Additional details

Motivation

- We turn back to our original motivation:
  How close is the empirical mean to the true mean of a RV?
- We know that for n iid observations $X_i \overset{\text{iid}}{\sim} X$ of X,
  1. $\overline{X_n} = \frac{X_1 + \ldots + X_n}{n}$ converges to $\mathrm{E}[X]$ (law of large numbers)
  2. the variance of $\overline{X_n}$ is $\mathrm{Var}(X)/n$
- We can use concentration inequalities such as Chebyshev to bound

$$P(|\overline{X_n} - \mathrm{E}[X]| \geq \varepsilon)$$

  for any $\varepsilon > 0$
- Could we know more than that?
- Could we know the whole distribution of $\overline{X_n}$ around $\mathrm{E}[X]$ as $n \to +\infty$?

Normal approximation

Idea

- Isolate the unknown information about $\overline{X_n}$ by *standardizing* $\overline{X_n}$

Definition

Let X be a RV with finite mean $\mu = \mathrm{E}[X]$, **centering** X consists in considering

$$Y = X - \mu$$

s.t. $\mathrm{E}[Y] = 0$.
For X with finite standard deviation σ, **standardizing** X consists in considering

$$Z = \frac{X - \mu}{\sigma}$$

s.t. $\mathrm{E}[Z] = 0$ and $\mathrm{Var}(Z) = 1$.

Example
We already saw that standardizing $X \sim \mathcal{N}(\mu, \sigma^2)$ yields $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Normal approximation

Standardizing the empirical mean

For n iid observations of X, i.e., $X_i \overset{\text{iid}}{\sim} X$, with $\mu = \mathrm{E}[X]$, $\mathrm{Var}(X) = \sigma^2$ the standardized empirical mean is

$$Z_n = \frac{\overline{X_n} - \mathrm{E}[\overline{X_n}]}{\sqrt{\mathrm{Var}(\overline{X_n})}} = \frac{\overline{X_n} - \mu}{\sigma/\sqrt{n}}$$

$\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$

such that

$$\overline{X_n} = \mu + \frac{\sqrt{n}}{\sigma} Z_n.$$

Question

Now, what could be the distribution of Z_n as $n \to +\infty$?

Normal Approximation

Bernoulli case

- Let's look at a simple example: $X_i \overset{\text{iid}}{\sim} \text{Ber}(p)$, in that case,

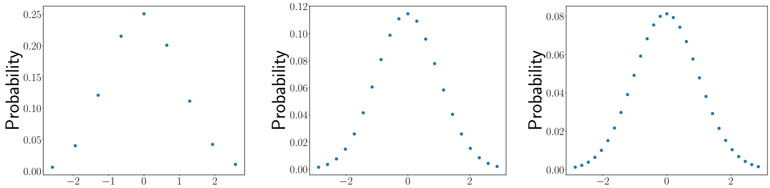
$$Z_n = \frac{S_n/n - p}{\sqrt{p(1-p)/n}} = \frac{S_n - np}{\sqrt{np(1-p)}}$$

$\frac{\sigma}{\sqrt{n}}$

with $S_n = X_1 + \ldots + X_n \sim \text{Bin}(n, p)$.

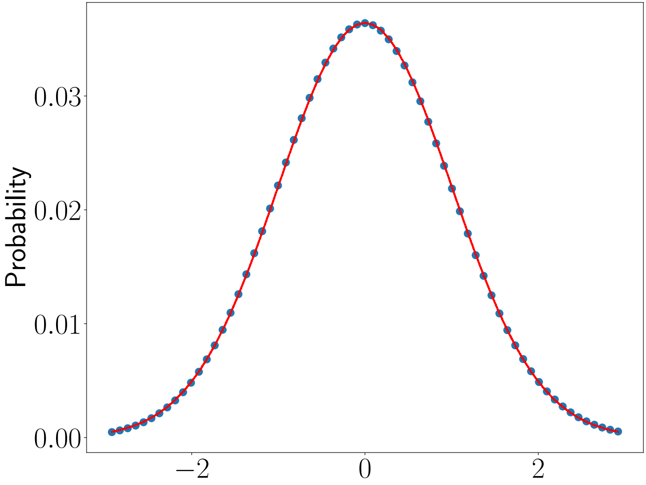
- The pmf of Z_n is then given by

$$P\left(Z_n = \frac{k - np}{\sqrt{np(1-p)}}\right) = P(S_n = k) \quad \text{for } k \in \{0, \ldots n\}$$



Plots of pmf of Z_n for $p = 0.4$ and $n = 10, 50, 100$
$\rightarrow$ Looks like the bell of a Gaussian distribution!

Normal Approximation



Bullets: pmf of Z_n for $p = 0.4$ and $n = 500$
Red curve[1]: $X \sim \mathcal{N}(0, 1)$

[1] See additional slides for more details on how the plot is done

Normal approximation

Theorem (Central Limit Theorem (CLT) for binomial random variables)

Let $0 < p < 1$, consider n iid observations $X_i \overset{iid}{\sim} Ber(p)$ of a Bernoulli RV. The distribution of the standardized empirical mean

$$Z_n = \frac{\overline{X_n} - E[\overline{X_n}]}{\sqrt{Var(\overline{X_n})}} = \frac{S_n - np}{\sqrt{np(1-p)}},$$

where $S_n = X_1 + \ldots + X_n \sim Bin(n,p)$ and $\overline{X_n} = S_n/n$, converges to the distribution of a standard normal distribution, i.e., for any $-\infty \leq a \leq b \leq +\infty$,

$$\lim_{n\to+\infty} P\left(a \leq Z_n \leq b\right) = P(a \leq Z \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

for $Z \sim \mathcal{N}(0,1)$.

Notes:

- Compared to the law of large numbers, this is a limit in distribution, i.e., as $n \to +\infty$, we get a formulation of the prob. in terms of a fixed pdf

Normal approximation

Application

- Previous theorem can be used to approx. the distribution of a binomial (which could be hard to compute as $n \to +\infty$ because of the choose numbers)
- Previous theorem is still only valid for a limit, below is a practical rule

Normal approximation

Application

- Previous theorem can be used to approx. the distribution of a binomial (which could be hard to compute as $n \to +\infty$ because of the choose numbers)
- Previous theorem is still only valid for a limit, below is a practical rule

Lemma

Suppose that $S_n \sim Bin(n,p)$ with n large and p not too close to 0 and 1, then

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \approx \Phi(b) - \Phi(a)$$

with Φ the cdf of $Z \sim \mathcal{N}(0,1)$.
As a rule of thumb the approx. is good if $np(1-p) > 10$.

Normal approximation

Application

- Previous theorem can be used to approx. the distribution of a binomial (which could be hard to compute as $n \to +\infty$ because of the choose numbers)
- Previous theorem is still only valid for a limit, below is a practical rule

Lemma

Suppose that $S_n \sim Bin(n,p)$ with n large and p not too close to 0 and 1, then

$$P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) \approx \Phi(b) - \Phi(a)$$

with Φ the cdf of $Z \sim \mathcal{N}(0,1)$.
As a rule of thumb the approx. is good if $np(1-p) > 10$.

Note:

- We will see that if p is too small even for large n the normal distribution is not the right approximation of the binomial.

Tip: frame this as a series of Bernoulli trials (i.e. frame X as a Binomial)

Example

Suppose we roll a pair of fair 6-sided dice 10,000 times. Estimate the probability that the number of times we get snake eyes (two ones) is between 280 and 300.

what are n & p?

X = # snake eyes = sum of Bern(1/36) trials $\Rightarrow X \sim \text{Bin}(n = 10^4, p = 1/36)$

$$P(280 \le X \le 300)$$

$$= P\left(\frac{280-np}{\sqrt{np(1-p)}} \le \frac{X-np}{\sqrt{np(1-p)}} \le \frac{300-np}{\sqrt{np(1-p)}}\right)$$

normal approx.

$$\approx \Phi\left(\frac{300-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{280-np}{\sqrt{np(1-p)}}\right)$$

plug in n, p

$$\approx \Phi(1.352) - \Phi(0.1352) \approx 0.358$$

pbinom(300, 10000, 1/36) − pbinom(280, 10000, 1/36) = 0.3459

Consider $X \sim \text{Binom}(10, 1/3)$

$$np(1-p) = 10 \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{20}{9} < 10.$$

$$P(3 \le X \le 5) \approx \Phi\left(\frac{5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{3-np}{\sqrt{np(1-p)}}\right)$$

$$\approx 0.457$$

using binom. pmf,

$$P(3 \le X \le 5) \approx 0.364$$

Preview: the Central Limit Theorem

We can generalize this beyond the binomial distribution:

Theorem (Central Limit Theorem)

Suppose that we have iid RVs $X_1, ..., X_n$ *with finite mean* $E[X_i] = \mu$ *and finite variance* $Var(X_i) = \sigma^2$*. Let* $S_n = \sum_i X_i$*. Then*

$$P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \approx \Phi(b) - \Phi(a).$$

The CLT will be covered in greater detail in MATH/STAT 395.

Outline

Weak law of large numbers (WLLN)

Normal (Gaussian) distribution

Normal Approximation

Additional details

Details on the plots

Notes

An attentive reader may have noticed that the plot of slide 11 is not the plot of the pdf of a standard normal distribution, since on 0 the pdf of $X \sim \mathcal{N}(0,1)$ should be approx. 0.4

Indeed a continuity correction has been used (see lecture 26).

Namely, we have with the notations of slide 10

$$\begin{aligned} P\left(Z_n = \frac{k-np}{\sqrt{np(1-p)}}\right) &= P\left(\frac{k-1/2-np}{\sqrt{np(1-p)}} \le Z_n \le \frac{k+1/2-np}{\sqrt{np(1-p)}}\right) \\ &\approx \Phi\left(\frac{k+1/2-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-1/2-np}{\sqrt{np(1-p)}}\right) \\ &\approx \frac{\psi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)}{\sqrt{np(1-p)}} \end{aligned}$$

where $\psi(x) = e^{-x^2/2}/\sqrt{2\pi}$ is the pdf of $X \sim \mathcal{N}(0,1)$

and I used in the last line that for a function f,

$$f(x+1/2) - f(x-1/2) \approx f'(x)$$

with $f(x) = \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$.

So the red curve on slide 11 is the plot of a scaled version of the normal distribution

Namely it is the plot of $\frac{\psi(x)}{\sqrt{np(1-p)}}$ for $x \in \left\{\frac{k-np}{\sqrt{np(1-p)}},\ k \in \{0,\dots n\}\right\}$

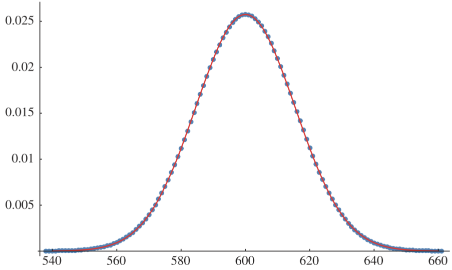
Details on the plots

Direct visualization

Another way to visualize how much a binomial is close to some normal distribution is to consider that since $S_n = np + \sqrt{np(1-p)}Z_n$ with $Z_n \approx \mathcal{N}(0,1)$ (Z_n is the standardized empirical mean), then we should have

$$S_n \approx \mathcal{N}(np, np(1-p))$$

The pmf of S_n and the pdf of its normal approx. are given below (without any scaling)
Though these plots are more natural, they hide the general reasoning of "standardizing the empirical mean" which can be applied for any empirical mean (not only the empirical mean of Bernoulli RV)



Bullets: pmf of $S_{1000} \sim \text{Bin}(1000, 0.6)$
Red curve: pdf of $X \sim \mathcal{N}(600, 240)$