Central Limit Theorem

Q1. What does the WLLN say?

A1.

That the average of a series of iid RVs converges to a number, the expected value of the RVs (in probability).

Q2. Assume n Xi iid RVs with mean μ=0 and variance σ^2.

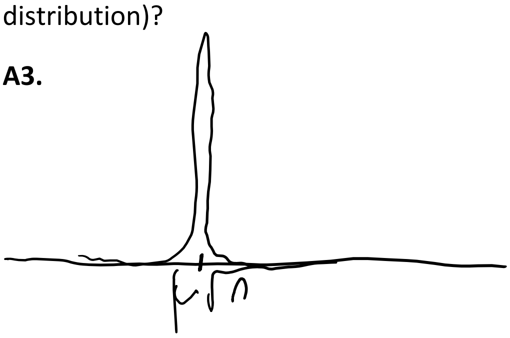
- **a)** Calculate the variance of the sum Sn = X1 + X2 +…+ Xn.
- **b)** Calculate the variance of the mean Mn = Sn/n
- **c)** Find a function of Sn that has constant variance over n

A2.

- **a)** $n\sigma^2$
- **b)** σ^2/n
- **c)** $Var(\frac{S_n}{\sqrt{n}}) = \sigma^2$

Q3. How does the distribution $\frac{S_n}{\sqrt{n}}$ look like as n approaches infinity (what is the limiting shape of the distribution)?

A3.



Q4. The above RV $\frac{S_n}{\sqrt{n}}$ has a mean that increases with sqrt(n). Find a function of Sn with constant mean of 0 and a variance of 1.

A4.

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

The Central Limit Theorem says that:

$$\lim_{n \to \infty} P(Z_n \leq z) = P(Z \leq z)$$

Where Z is the Standard Normal Distribution!

Think what this says. The distribution of Xi doesn't matter!

Q5. What are the assumptions that the CLM works under (that we have been making up to now)?

A5.

1) Independent Xi
2) Xi are drawn from the same distribution
3) 0 < Var(Xi) < Inf. This is important. If it doesn't hold nothing guarantees you convergence.

1 and 2 can be dropped if other assumptions are made.

Q6. How small can n be?

A6.

Let's write some code!

$P(S_n \leq a) \approx b$

In problems you know two of the three variables and try to find the third.

Q7. Using your electrode, you have made a brain helmet that can be used to produce a few mWatts of power from a user wearing it. This, over different users, is a RV drawn from an exponential distribution with λ=1/20 mW^{-1} (i.e. you get almost no power production from most people but some people are like superheroes). Remember for exponential μ = 1/λ, σ = 1/λ. You gather 100 brave test subjects and you try to light up a night light bulb that requires 2.1 Watts to operate. What is the probability you will succeed?

A7.

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

$$P(S_n \geq 2100) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \geq \frac{2100 - n\mu}{\sqrt{n}\sigma}\right) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \geq \frac{2100 - 100 * 20}{10 * 20}\right)$$

$$= P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \geq 0.5\right) = P(Z_n \geq 0.5) = P(Z \geq 0.5) = 1 - P(Z \leq 0.5) = 1 - \Phi(0.5)$$

$$= 1 - 0.6915 = 0.3058$$

Q8. Create two other problems your research efforts might face that would require you to solve for a and for n. Solve them.

A8.

1) You have 100 users and you need to power something with 5% chance of failure every time you try. Choose a device with the appropriate power consumption (a).
2) You have a lamp of 2.1Watts and you need to collect enough people to power it with 5% failure (because the experiment will be on live tv). How many people do you need (n)?

Q9. Realizing Brain Power isn't going to work you go back to the original Brain Computer Interface research, thinking that the CLT will allow you to estimate accurately the percentage of the population (p) your BCI algorithm will work on using fewer test subjects or with a smaller probability you will pass your error threshold every time you run the experiment. So:

A) How many test subjects do we need in order to have a probability of 0.05 that our estimation of p is wrong by no more than 3%?

B) What is the probability you will surpass your 3% error margin if you use 500 test subjects?

A9.

Xi = 1 with p and 0 with p-1 is the Bernoulli RV that the algo will work on person i.

Mn = (X1 + X2 + … + Xn)/n is the sample mean of the experimental subjects population.

We need $|M_n - p| < 0.03$

So the probability for your experiment to fail is $P(|M_n - p| \geq 0.03)$.

$$Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

with $\mu = p \ and \ \sigma = \sqrt{p(1-p)}$ (remember what is the maximum value σ can take and why).

$$|M_n - p| < 0.03$$

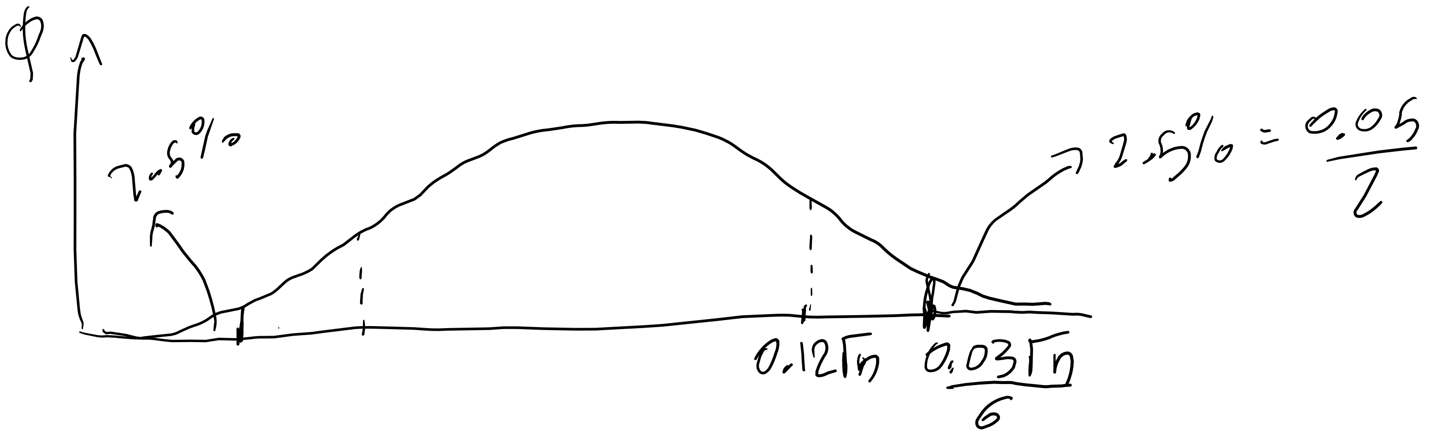
$$\frac{|S_n - np|}{n} < 0.03$$

$$\frac{|S_n - np|}{\sqrt{n}} < 0.03\sqrt{n}$$

$$\frac{|S_n - np|}{\sqrt{n}\sigma} < \frac{0.03\sqrt{n}}{\sigma}$$

$$\frac{|S_n - np|}{\sqrt{n}\sigma} < \frac{0.03\sqrt{n}}{\sigma}$$

$$P(|M_n - p| \geq 0.03) = P\left(|Z_n| \geq \frac{0.03\sqrt{n}}{\sigma}\right) \approx P\left(|Z| \geq \frac{0.03\sqrt{n}}{\sigma}\right)$$



A)

$\sigma = \sqrt{p(1-p)} \leq 0.5$

$P\left(|Z| \geq \frac{0.03\sqrt{n}}{\sigma}\right) \leq P\left(|Z| \geq 0.06\sqrt{n}\right) = 2 * \left(1 - \Phi\left(0.06\sqrt{n}\right)\right) = 0.05$

$1 - \Phi\left(0.06\sqrt{n}\right) = 0.025$

$\Phi\left(0.06\sqrt{n}\right) = 0.975$

$0.06\sqrt{n} = 1.96$

$\mathrm{n} = 1068$

B)

$P\left(|Z| \geq 0.06\sqrt{500}\right) = 2 * \left(1 - \Phi(0.06 * 22.36)\right) = 2 * (1 - 0.90988) = 0.18$