



Normal Distributions

Arulalan Rajan¹

February 6, 2024

National Institute of Technology Karnataka

Dept. of Electronics & Communication Engg.,

¹The instructor acknowledges various authors, their articles and blogs which gave wonderful insight to some of the concepts discussed in the lecture slides

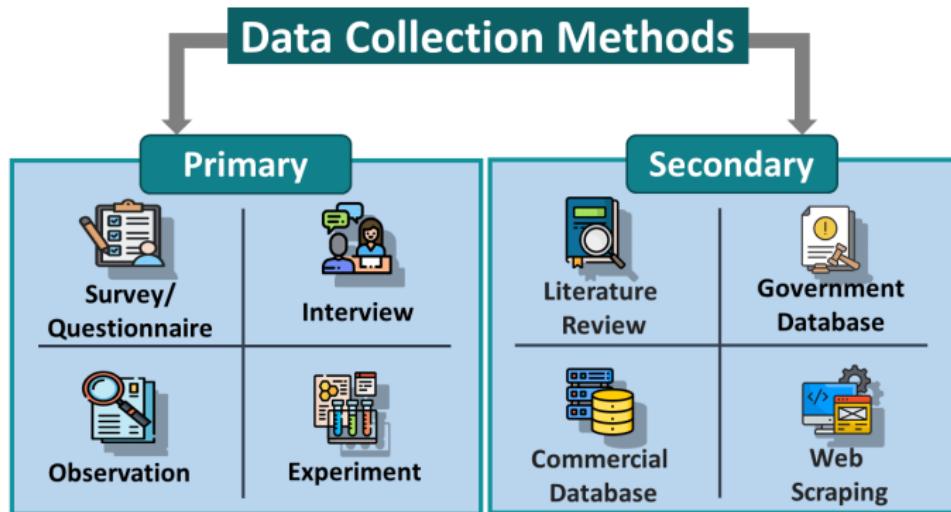
Module Contents

- ▶ Sampling and Types of sampling
- ▶ Sample Mean
- ▶ Sample Variance
- ▶ Sampling distributions from a normal population
- ▶ Sampling from a finite population
- ▶ Normal Distribution
- ▶ Approximating Binomial, Poisson distributions using normal distribution

Sampling

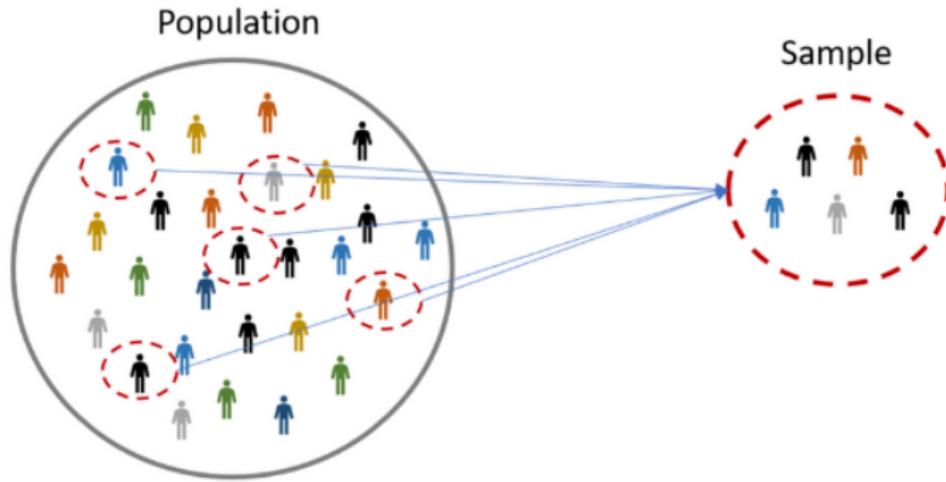
- ▶ Statistics - Converting Data to information
- ▶ Where does the data come from?
- ▶ How is it gathered?
- ▶ How do we ensure its accuracy?
- ▶ Does it represent the population from which it was drawn?

Data Collection Techniques²



²<https://www.educba.com/data-collection-methods/>

Sampling³



Sampling

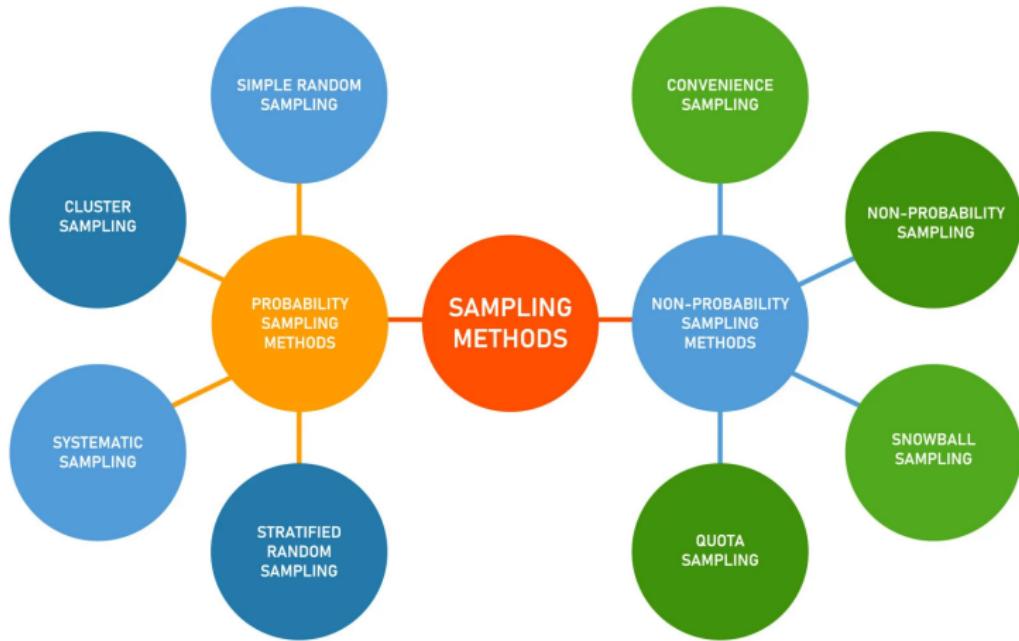


³<https://hotcubator.com.au/research/a-complete-guide-to-sampling-techniques/>

Sampling

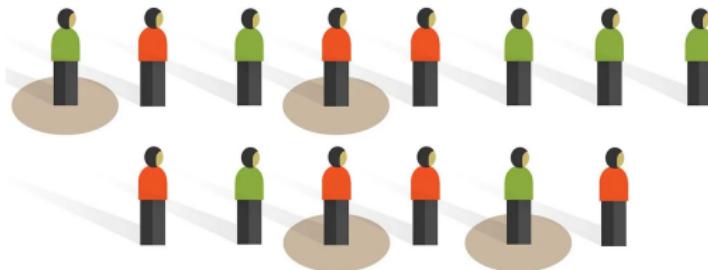
- ▶ Selecting representative subset of individuals from larger population
- ▶ Sampling Goal: Sample large enough and represent the target population
- ▶ Employs Probability ideas

Sampling Methods⁴



⁴<https://www.simplypsychology.org/sampling.html>

Random Sampling



- ▶ Equal chance of being chosen
- ▶ Unbiased representation of a population
- ▶ **Expensive and time consuming**
- ▶ **Access to respondents**

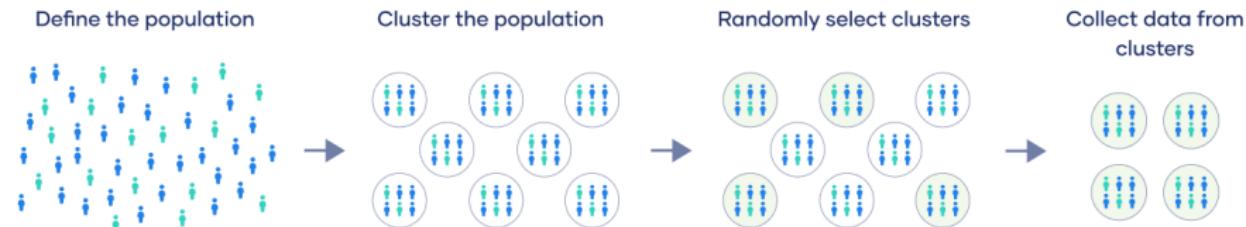
Stratified Sampling⁵



- ▶ Divide a population into homogeneous subpopulations or strata (gender, location etc)
- ▶ Strata must be mutually exclusive
- ▶ Use random sampling to choose from strata

⁵<https://www.scribbr.com/methodology/stratified-sampling/>

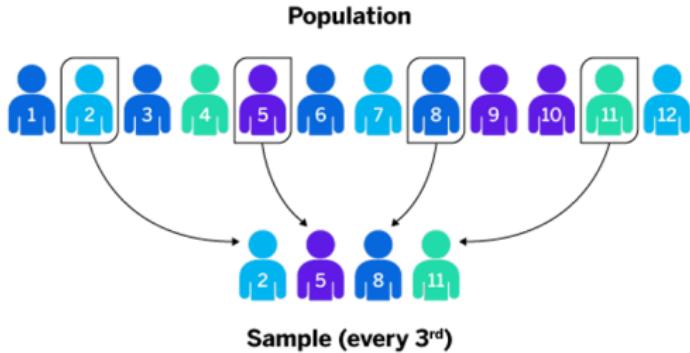
Cluster Sampling⁶



- ▶ Divide a population into small groups or clusters
- ▶ Schools, colleges etc
- ▶ Make the individual clusters as diverse as possible
- ▶ Clusters must be non-overlapping

⁶<https://www.scribbr.com/methodology/cluster-sampling/>

Systematic Sampling⁷

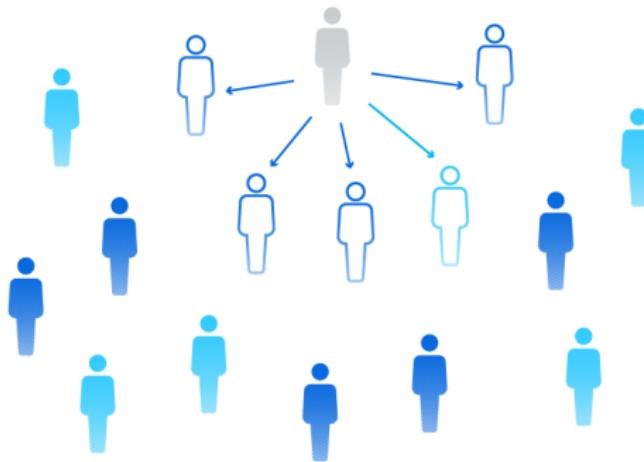


- ▶ Assign a desired sample size of the population
- ▶ Assign a regular interval number to decide who in the target population will be sampled
- ▶ Sampling Interval = Population Size / Sample Size
- ▶ **Data Manipulation - Sample collection could lead to higher chance of achieving a predetermined result**

⁷<https://www.scribbr.com/methodology/cluster-sampling/>

Convenience Sampling

Convenience sample



- ▶ Choose participants based on their convenience and availability
- ▶ Eg: Malls, school campuses

Snowball Sampling

- ▶ Mimics a pyramid in its selection pattern
- ▶ Known as referral / Respondent Driven or Chain referral sampling
- ▶ Early participants recruit further participants and so on till sample size is reached
- ▶ Recruiting sample members thru media channels to promote our work in their network

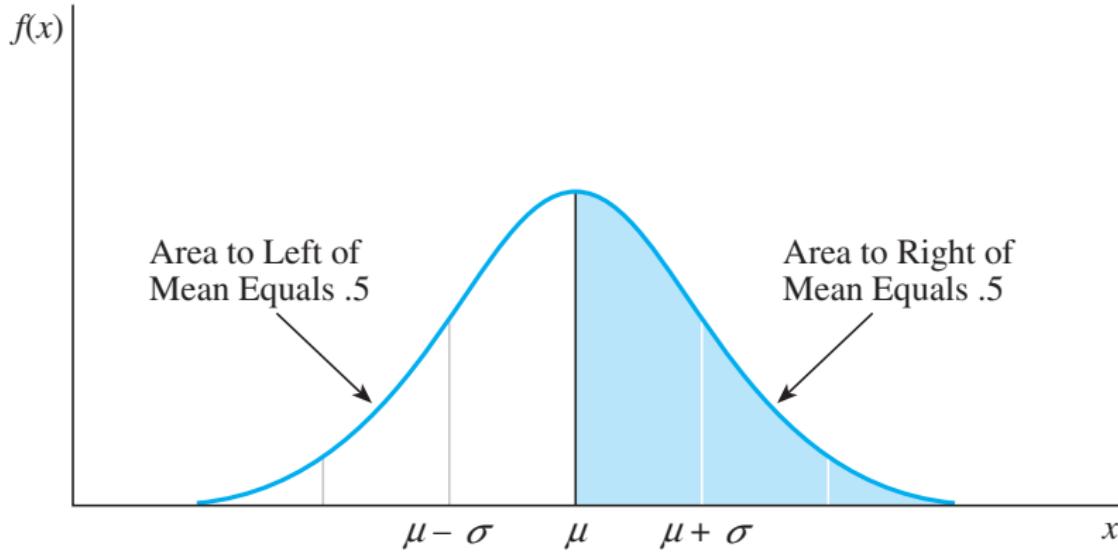
Normal Distribution

- ▶ Continuous probability distribution
- ▶ Large number of random variables observed in nature possess a mound-shaped or a bell-shaped profile

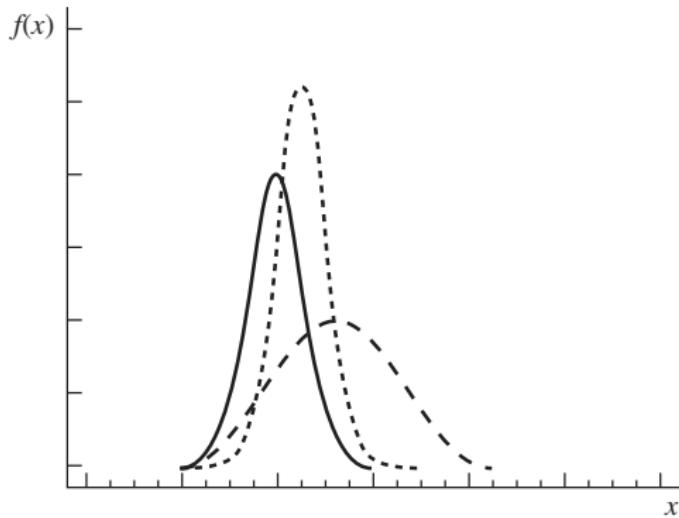
$$\text{▶ } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

- ▶ μ and σ , ($\sigma > 0$) - Parameters that represent the mean and standard deviation respectively
- ▶ μ locates the center of the distribution, with the distribution being symmetric about the mean
- ▶ Area to the left of μ = Area to the right of μ = 0.5
- ▶ Shape of the distribution - Determined by the population standard deviation

Normal Distribution



Normal Probability Distribution with different μ and σ



Larger (smaller) values of σ reduce (increase) the height of the curve and increase (decrease) the spread. Almost all values of a random variable lie in the interval $\mu \pm 3\sigma$

Normal distribution - *Jus leave me alone!!!*

For any random variable X , we define $F_X(x) = P(X \leq x)$

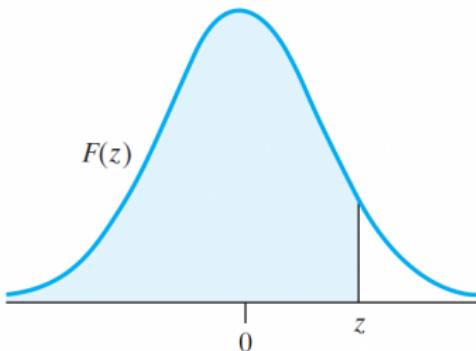


Figure: Cumulative Distribution Function $F_X(x) = P(X \leq x)$

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

All'z well !!! Nothing Complicated!!! Dont Panic!!!

Standard Normal Random Variable, Z

- ▶ X - Normal random variable
- ▶ Express its value as the number of standard deviations (σ) it lies to the right of its mean, μ

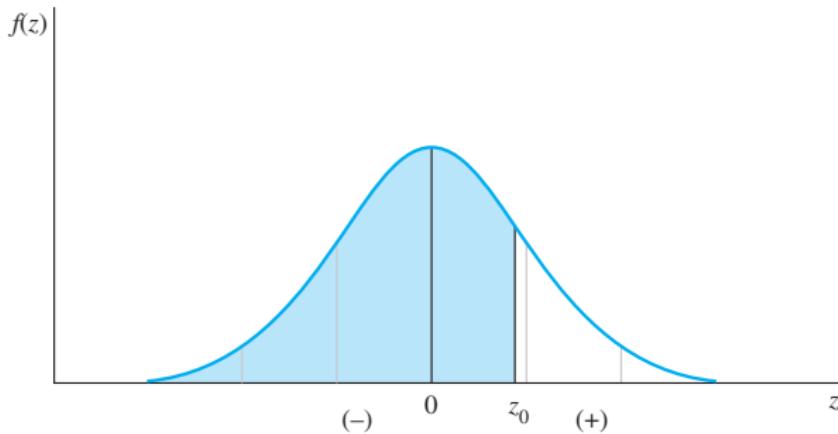
$$Z = \frac{x - \mu}{\sigma} \Rightarrow X = \mu + Z\sigma$$

Observe the following

- ▶ $X < \mu \Rightarrow Z$ is negative
- ▶ $X > \mu \Rightarrow Z$ is positive
- ▶ $X = \mu \Rightarrow Z$ is 0

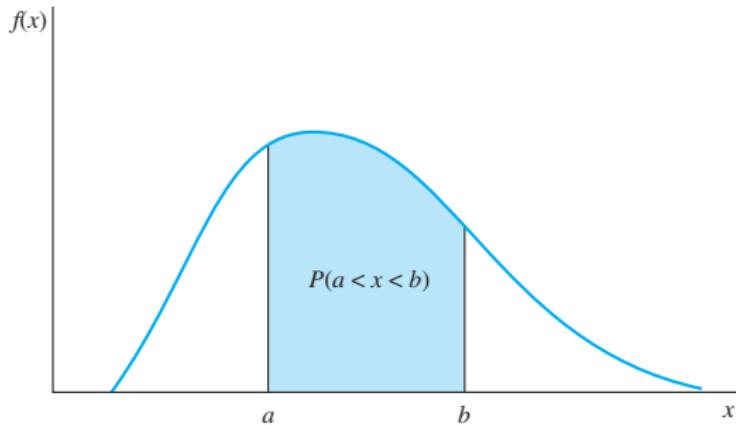
Probability Distribution for Z

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



- ▶ Mean 0, Standard Deviation 1
- ▶ Values on the left (**right**) side of the curve are negative (**positive**)
- ▶ Area under the standard normal curve to the left of a specified value $z_0 = P(Z \leq z_0)$

Tabulated Areas of the Normal Distribution



What is the probability that a normal random variable x lies in the interval from a to b ?

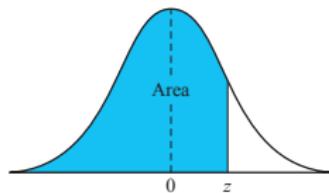
Infinitely large number of normal distributions - one for each different mean and standard deviation

Impractical to have separate tables for each of these curves

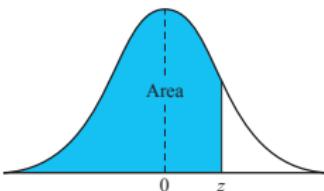
Standardise and use same table for all normal distributions

Probability Distribution for Z, Christian Kramp

<i>z</i>	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010	.0010
-2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0022	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

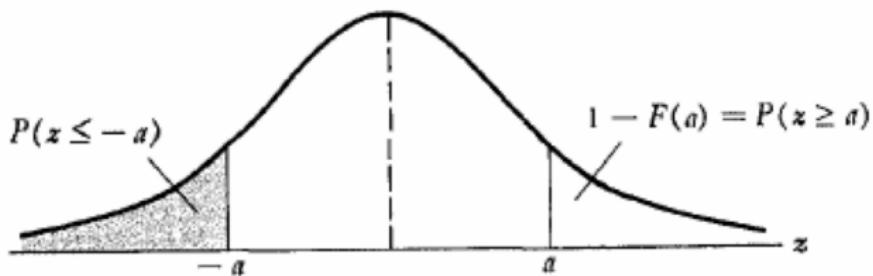
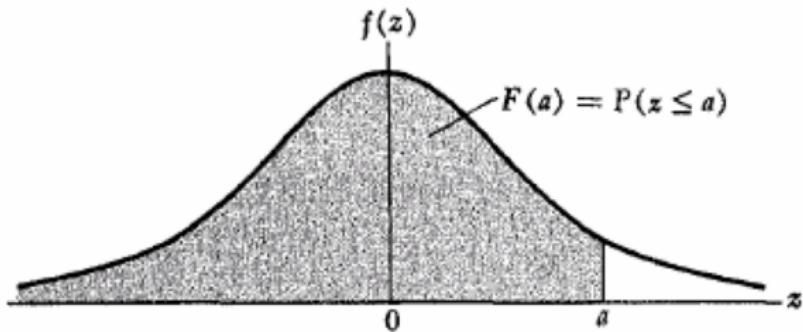


Probability Distribution for Z , Christian Kramp



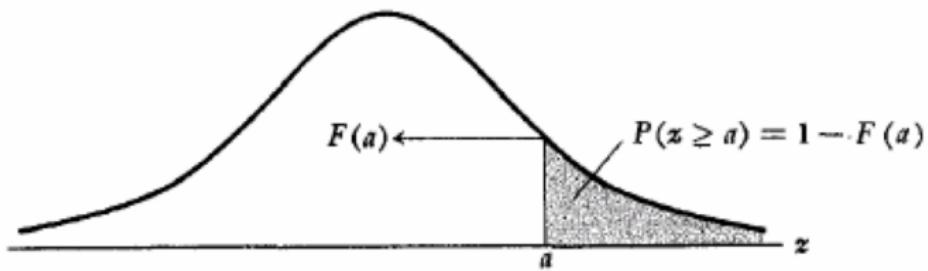
Rules for using standardized normal distribution

$$P(Z \leq a) = \begin{cases} F_Z(a), & a > 0 \\ 1 - F_Z(-a), & a < 0 \end{cases}$$



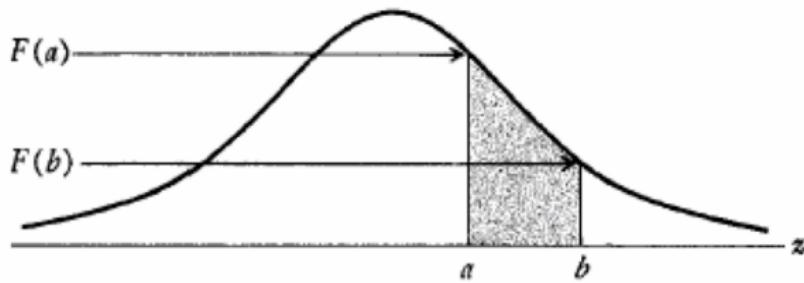
Rules for using standardized normal distribution

$$P(Z \geq a) = \begin{cases} 1 - F_Z(a), & a > 0 \\ F_Z(-a), & a < 0 \end{cases}$$



Rules for using standardized normal distribution

$$P(a < Z \leq b) = F_Z(b) - F_Z(a)$$



Rules for using standardized normal distribution

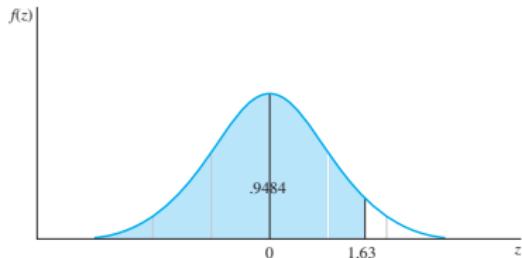
For a positive,

$$P(-a \leq Z \leq a) = F_Z(a) - F_Z(-a) = F_Z(a) - (1 - F_Z(a)) = 2F_Z(a) - 1$$

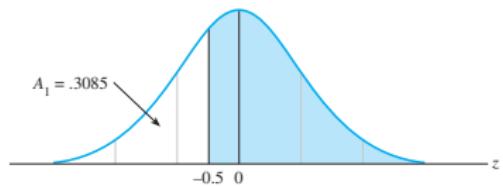
For a positive,

$$F_Z(a) = [1 + P(-a \leq Z \leq a)]/2$$

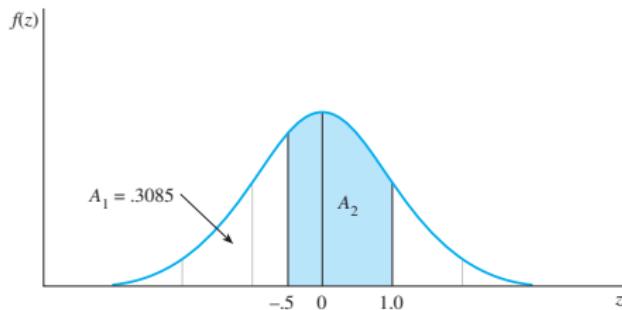
Try the following



$$\text{Find } P(Z \leq 1.63)$$



$$\text{Find } P(Z \geq -0.5)$$



$$\text{Find } P(-0.5 \leq Z \leq 1.0)$$

Try the following

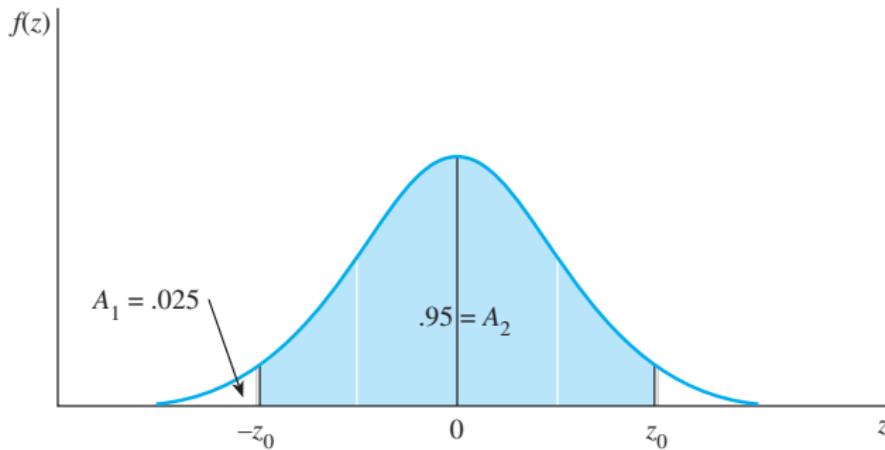
1. Find the probability that a normally distributed random variable will fall within **one** standard deviation of its mean
2. Find the probability that a normally distributed random variable will fall within **two** standard deviation of its mean

Given a distribution of measurements that is approximately mound shaped

- The interval $\mu \pm \sigma$ contains 68% of the measurements
- The interval $\mu \pm 2\sigma$ contains 95% of the measurements
- The interval $\mu \pm 3\sigma$ contains 99.7% of the measurements

Try the following

- Find the value of Z , say z_0 such that 95% of the area is within $\pm z_0$ standard deviations of the mean



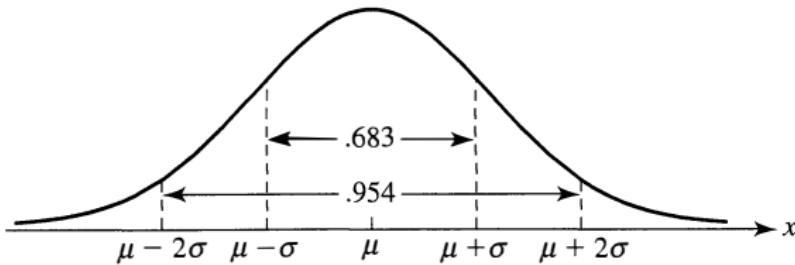
- Let X be a normally distributed random variable with mean 10 and standard deviation 2. Find the probability that X lies between 11 and 13.6

Summary - Normal Distribution

Generalization of bell-shaped normal density to several dimensions -
Vital role in multivariate analysis

Recall the univariate normal distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$



$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &\approx 0.68 \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &\approx 0.95 \end{aligned}$$

Try the following!!!

1. The top 5% of applicants as measured by a competitive exam (CE) scores will get scholarships. If $CE \sim N(500, 100^2)$, how high should your competitive exam score be to get scholarship?
2. Suppose the duration of trouble-free operation of a new vacuum cleaner is normally distributed with a mean of 530 days and standard deviation of 100 days. What is the probability that a randomly selected vacuum cleaner will work for at least 730 days without trouble? What is the sampling distribution of the sample mean for a random sample of 25 vacuums. If 25 random vacuum cleaners are selected, within what limits do you expect the sample average to lie with probability 0.95?

Jointly Gaussian Random Variables

Used for modelling communication systems dealing with signals in the presence of noise

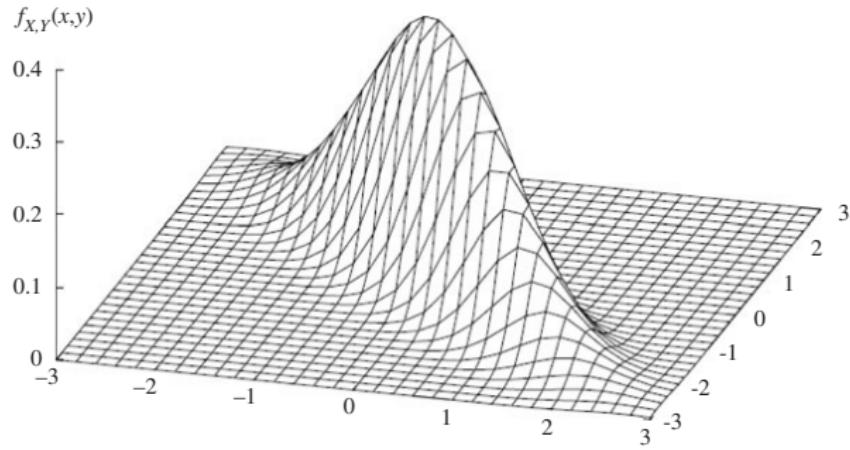
$$f_{X,Y}(x,y) = \frac{\exp \left\{ k \left[\left(\frac{x - \mu_1}{\sigma_1} \right)^2 - 2\rho_{X,Y} \left(\frac{x - \mu_1}{\sigma_1} \right) \left(\frac{y - \mu_2}{\sigma_2} \right) + \left(\frac{y - \mu_2}{\sigma_2} \right)^2 \right] \right\}}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho_{X,Y}^2}}$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$ and where $k = \frac{-1}{2(1 - \rho_{X,Y}^2)}$

The pdf is centered at (μ_1, μ_2)

Has a bell shape depending on $\sigma_1\sigma_2, \rho_{X,Y}$

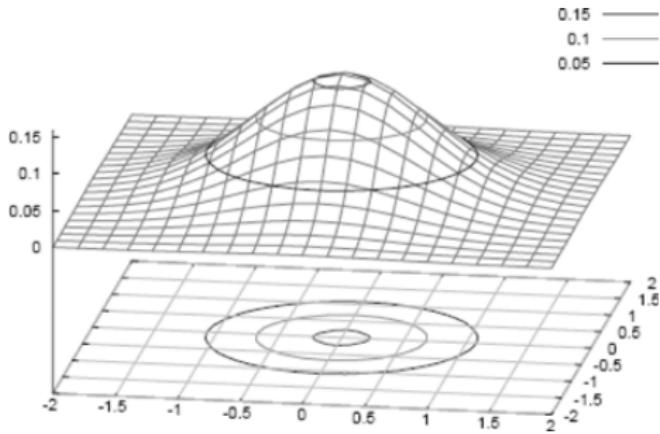
Jointly Gaussian Random Variables



pdf is constant for values x and y for which the argument of the exponent is constant

$$\left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho_{X,Y} \left(\frac{x-\mu_1}{\sigma_1} \right) \left(\frac{y-\mu_2}{\sigma_2} \right) + \left(\frac{y-\mu_2}{\sigma_2} \right)^2 \right] = \text{constant}$$

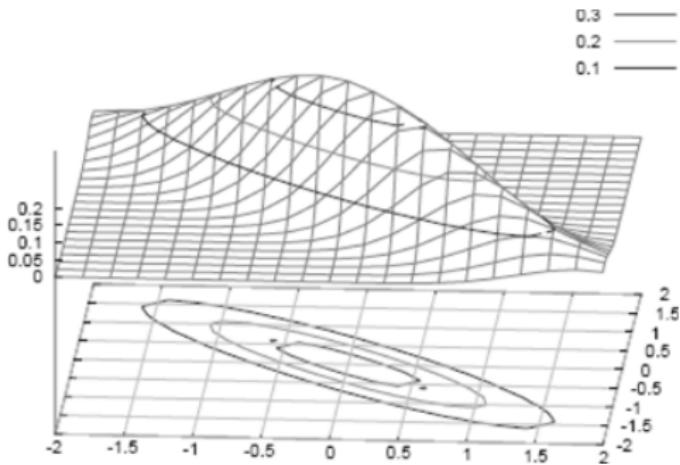
Jointly Gaussian Random Variables



$$\rho_{X,Y} = 0$$

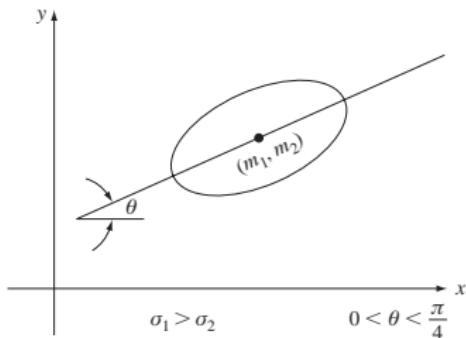
X and Y are independent, PDF Contour is an ellipse with principal axes aligned with the $x-$ and $y-$ axes.

Jointly Gaussian Random Variables

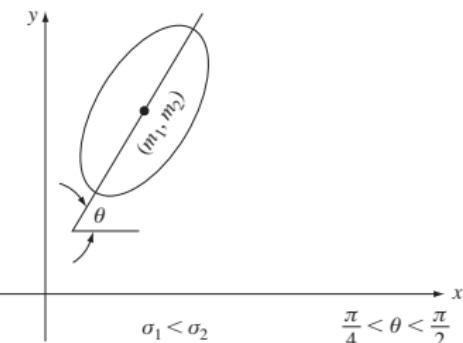


$$\rho_{X,Y} = -0.9$$

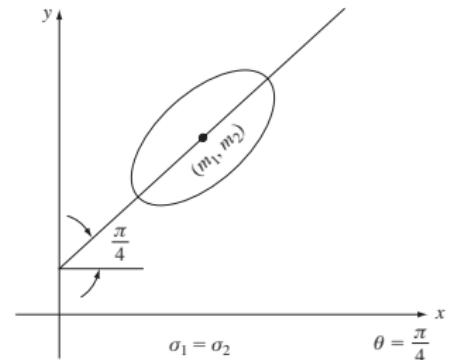
Jointly Gaussian Random Variables



$$\sigma_1 > \sigma_2$$



$$\sigma_1 < \sigma_2$$



$$\sigma_1 = \sigma_2$$

When $\rho_{X,Y} \neq 0$, major axis of the ellipse is oriented along the angle

$$\theta = \frac{1}{2} \arctan \left(\frac{2\rho_{X,Y}\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2} \right)$$

Here you go!!! A Toughie!!!

An engineering professional body estimates that 75% of the students taking undergraduate engineering courses are in favour of studying of statistics as part of their studies. If this estimate is correct, what is the probability that more than 780 undergraduate engineers out of a random sample of 1000 will be in favour of studying statistics?

- ▶ Binomial distribution with large n
- ▶ $X \sim \text{Binomial}(1000, 0.75)$
- ▶ To find the probability that X is greater than 780, i.e.,
 $P(X > 780)$
- ▶ $P(X > 780) = P(X = 781) + P(X = 782) + \dots + P(X = 1000)$
- ▶ Tedious computations

Normal Approximation to Binomial

- ▶ Approximate very closely by normal distribution with the same mean and standard deviation as $X \approx \text{Binomial}(1000, 0.75)$
- ▶ $\mu = np = 1000 \times 0.75 = 750$
- ▶ $\sigma = \sqrt{np(1-p)} = \sqrt{1000 \times 0.75 \times 0.25} = 13.7$
- ▶ $X \sim \text{Binomial}(1000, 0.75)$ and $Y \sim N(750, 13.7^2)$
- ▶ $P(X = 781) \approx P(780.5 \leq Y \leq 781.5)$
- ▶ Similarly compute for the rest of the values
- ▶ *Oooh Boiii!!!! Thats tedious again!!!!*

- ▶ $P(X > 780) = \sum_{i=781}^{1000} P(X = i)$
- ▶ $P(X > 780) \approx P(780.5 < X < 1000.5)$
- ▶ $P(X > 780) = P\left(\frac{780.5 - 750}{13.7} \leq Z \leq \frac{1000.5 - 750}{13.7}\right) = P(2.23 \leq Z \leq 18.28) = 0.013$

Normal Approximation to Binomial

- ▶ Can we do what we did in the problem we just solved?
- ▶ Isn't Binomial discrete and Normal continuous?
- ▶ Doesn't that mean there exists no values between any two integers??
- ▶ **Continuity correction: Accounts for approximating a discrete random variable with a continuous one**
- ▶ Use this correction only for binomial probabilities and not with a random variable which is already continuous
- ▶ Normal Approximation to binomial probabilities if $np > 5$ and $np(1 - p) > 5$

Normal Approximation to Binomial

- Find the necessary values of n and p . Calculate $\mu = np$ and $\sigma = \sqrt{npq}$.
- Write the probability you need in terms of x and locate the appropriate area on the curve.
- Correct the value of x by $\pm .5$ to include the entire block of probability for that value. This is the *continuity correction*.
- Convert the necessary x -values to z -values using

$$z = \frac{x \pm .5 - np}{\sqrt{npq}}$$

Figure: Normal Approximation to Binomial

My Dearest Students!!! - You will reach stars and beyond!!!

*When things go wrong, as they sometimes will,
When the road you're trudging seems all uphill,
When the funds are low and the debts are high,
And you want to smile, but you have to sigh,
When care is pressing you down a bit-
Rest if you must, but don't you quit.*

*Often the goal is nearer than
It seems to a faint and faltering man;
Often the struggler has given up
When he might have captured the victor's cup;
And he learned too late when the night came down,
How close he was to the golden crown.*

*Success is failure turned inside out -
The silver tint in the clouds of doubt,
And you never can tell how close you are,
It might be near when it seems afar;
So stick to the fight when you're hardest hit -
It's when things seem worst that you must not quit.*

Dont Quit !!!

Poem Courtesy: www

*Life is queer with its twists and turns,
As every one of us sometimes learns,
And many a fellow turns about
When he might have won had he stuck it out.
Don't give up though the pace seems slow -
You may succeed with another blow.*

**Thanks for putting up with
my pranks and tantrums and
advices!!!
And of course my lectures! 😊**

**My best of best wishes to
all of u for grand success
in all ur endeavours!**

What a wonderful bunch of kids you are! Will miss you all!!!

Never doubt yourself!

I have fullest faith in all of you.. and am sure you will all glitter like diamond ... It is jus that you are getting polished now to shine amazingly well soon...

Don't get stressed out. Life is really beautiful and beyond!

If in case you feel down, jus feel free to call me! *The fact that you will have to listen to me over call should be sole reason for you to feel cheerful always!!!*

You are MY students and You are the BEST !!!

Loads of best wishes, blessings to you all....

signing off - Alan!!!