Binomial Distribution

- An experiment with repeated trials.
- Trials are independent.

- Each trial results in SOME EVENT either happening or not happening.

- Occurrence of that event is called SUCCESS and non-occurrence called FALIURE.

- There is the same probability p, of success in each trail.

- Correspondingly, each trial has the probability of failure, $q = 1 - p$.

Binomial distribution

- n independent trials
- probability p of success in each trial
- $0 \leq p \leq 1$, n is a positive integer.

Random variable $\boldsymbol{X}$: number of successes in n independent trials.

Probability Distribution of $\boldsymbol{X}$:

X follows *Binomial Distribution*

$$\boxed{X \sim Bin(n,p)}$$

$P(k\ success\ in\ n\ trials) = P(X = k)$, where

$$\boxed{P(X = k) = \binom{n}{k} p^k (1-p)^{n-k};\ k = 0, 1, \dots, n}$$

Range of $\boldsymbol{X}$: $\{0,1,2,\dots,n\}$

The binomial probabilities $P(X = k)$ are the terms in the binomial expansion:

$$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k},$$

where, $q = 1 - p$.

Exercise:

Check that the binomial distribution is

actually, a distribution by showing that

- $P(X = k) \geq 0 \; for \; all \; k.$
- $\sum_{k=0}^{n} P(X = k) = 1$

Why $\boldsymbol{P(k \; success \; in \; n \; trials) = \binom{n}{k} p^k (1-p)^{n-k}}$???

Exercise:

See Topic 1: Sampling with replacement example.

See Test 1: Q5

Compare the probabilities and events with binomial distribution.

Examples:

Example 1: Draw 5 balls from an urn, with 6 red balls and 4 green balls. Find the probability distribution of number of red balls. what is the probability of drawing 2 red balls?

Example 2: The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted that diseases, what is the probability that

- at least 3 survive.
- from 3 to 8 survive

Example 3: Text book section 2.1 page 83

What is the probability that among five families, each with six children, at least three of the families have four or more girls?

Most Likely Number of Successes/Mode of Binomial Distribution

For $X \sim Bin(n, p)$, $0 < p < 1$, the Mode or the most likely number of successes in n independent trials with probability p of success on each trial is m:

$$m = \lfloor (n+1)p \rfloor \text{ i.e., the greatest integer} \leq (n+1)p.$$

- There may be more than one mode in case of ties.

- The Binomial Distribution is *unimodal.* The probabilities are strictly increasing before they reach the maximum and strictly decreasing after the maximum.

- If $(n+1)p$ is an integer, then there 2 modes/ most likely numbers for the distribution m and $m-1$.

Exercise: Book page 87, For the given p and n, check the mode formula in a few of these cases to see how it works.

Normal Distribution

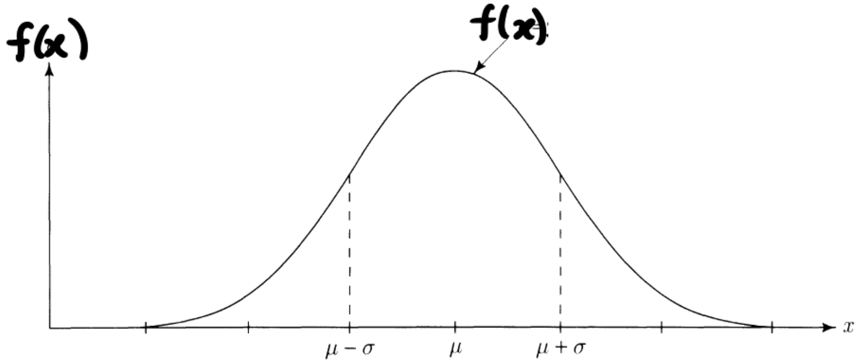
Random variable X has a Normal or Gaussian distribution, with parameters μ and σ^2:

$$\boxed{X \sim N(\mu, \sigma^2)}$$

$$\boxed{\text{Density function: } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

σ > 0 and μ is arbitrary.

- μ is mean of the distribution.
- σ^2 is the variance.
- σ is the standard deviation.
- **Range of X**: $-\infty < x < \infty$
- $N(\mu, \sigma^2)$ is unimodal with **mode** at μ.

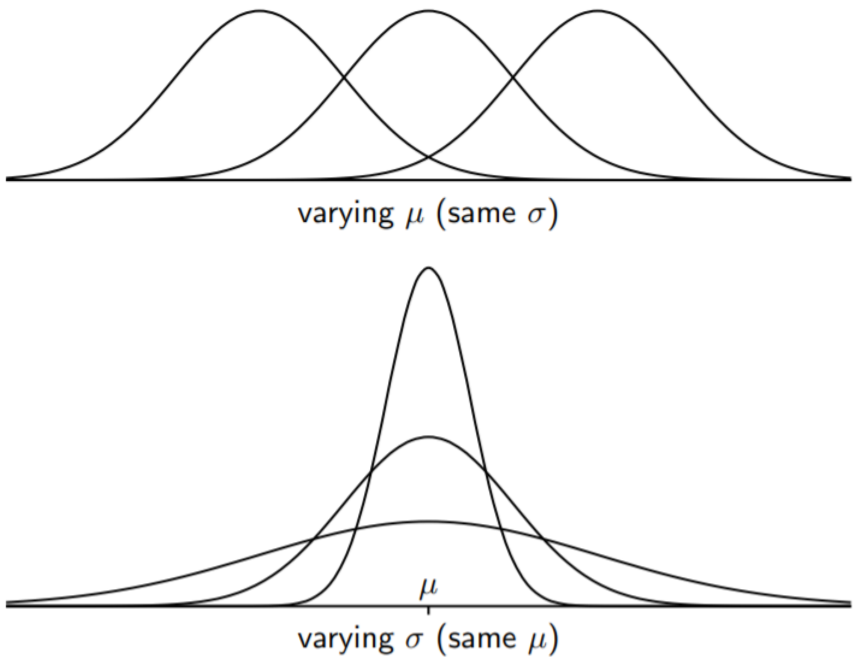


Normal Distribution

Normal Densities:

$$\boxed{\text{Density function: } f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

- μ is a location parameter (changing μ just shifts the distribution without changing its shape)
- σ is a scale parameter (the distribution is concentrated around μ when σ is small, and is spread out when σ is big.)



- Normal densities satisfy the conditions for a probability density function:

1. $f(x) \geq 0, for\ all\ x$
2. $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$

Standard Normal Distribution:

Random variable Z is called Standard Normal distribution if parameters $\mu = 0$ and $\sigma = 1$

$\boxed{Z \sim N(0,1)}$

$\boxed{\text{Density function: } \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}}$

CDF of standard normal $Z \sim N(0,1)$

Let $\Phi(\mathrm{z})$ be the cdf of $Z \sim N(0,1)$

$$\Phi(\mathrm{z}) = \int_{-\infty}^{z} \phi(t)dt$$

- No closed-form expression for Φ, use table of values instead (Appendix 5, for $z \geq 0$).

- $\Phi'(\mathrm{z}) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

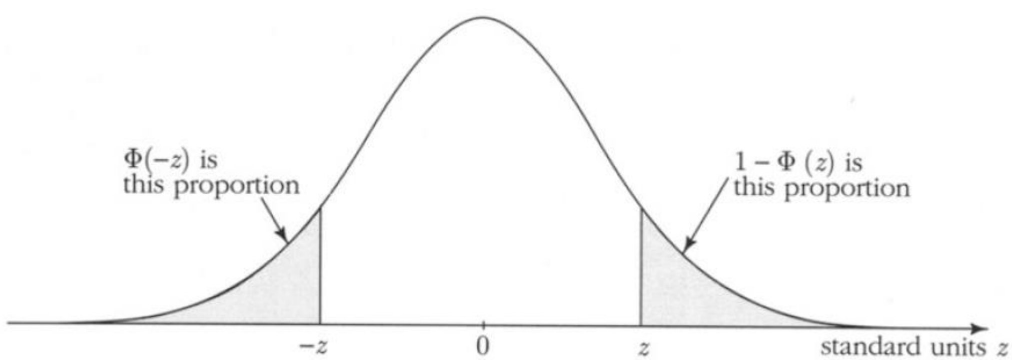
- $\Phi(\mathrm{z})$ is symmetric about mean $\mu = 0$

  $\Phi(\mathrm{z}) = P(Z \leq -z) = P(Z \geq z) = 1 - \Phi(\mathrm{z})$

- $\Phi(\mathrm{a}, \mathrm{b}) = \Phi(\mathrm{b})$ - $\Phi(\mathrm{a})$

- $\Phi(-\mathrm{z}, \mathrm{z}) = 2\Phi(\mathrm{z})$ -1

FIGURE 3. Symmetry of the normal curve.



Cumulative Distribution Function of Normal $X \sim N(\mu, \sigma^2)$

The cdf for normal random variable $N(\mu, \sigma^2)$ is:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- Transformation of Standard Normal Distribution:

  Lemma: If $X = \mu + \sigma Z$ $then$ $X \sim N(\mu, \sigma^2) \Leftrightarrow Z \sim N(0,1)$.

  Exercise:

  Prove the converse. If $X \sim N(\mu, \sigma^2)$ and $X = \mu + \sigma Z$. Show that $Z \sim N(0,1)$

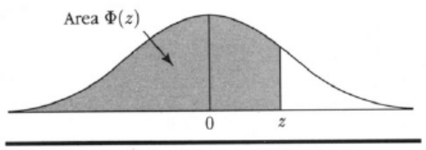
How to find Normal Probabilities?

Calculate probabilities for Normal $X \sim N(\mu, \sigma^2)$ using

1. the general cdf formulae obtained using transformations.
2. Standard normal cdf Φ (table)
3. continuity ($\Phi(z-) = \Phi(z)$);
4. symmetry ($\Phi(\mathrm{z}) = 1 - \Phi(\mathrm{z})$)

Example: Let X ~ $N(1, 4)$. Find $P(0.5 \le X \le 3.46)$.

Solve from notes IV



Appendix 5
Normal Table

Table shows values of $\Phi(z)$ for z from 0 to 3.59 by steps of .01. Example: to find $\Phi(1.23)$, look in row 1.2 and column .03 to find $\Phi(1.2 + .03) = \Phi(1.23) = .8907$. Use $\Phi(z) = 1 - \Phi(-z)$ for negative z.

	.0	.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	.7703	.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
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

What if the z you are looking for is not in the Φ $table$?

- software (The best choice)
  1. *NORMSDIST* function in Excel
  2. *pnorm* function in R
  3. *normcdf* function in Matlab
  4. etc

- crude approximation (round z to 2 decimals and use the corresponding value from the table)

- Linear interpolation for $l \leq z \leq r$

  $z = l + \lambda(r - l)$ then

  $$\Phi(z) \approx \Phi(l) + \lambda\big(\Phi(r) - \Phi(l)\big),$$

  $$\boxed{\Phi(z) \approx \Phi(l) + \frac{z-l}{r-l}(\Phi(r) - \Phi(l))}$$

Note: For $x = r$ or $x = l$, the formula is exact.

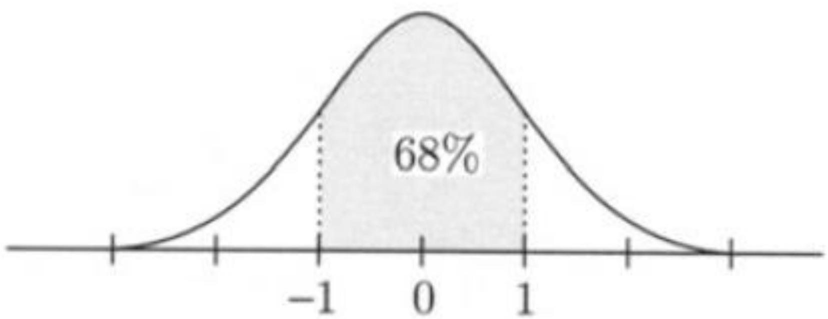
- Program an approximation formula for $\Phi(\mathrm{z})$ [eg. on pg. 95 of the textbook]

Example: $X \sim N(2,3). Find\ P(X \leq 4)$

Rule of Thumb regarding Normal densities:

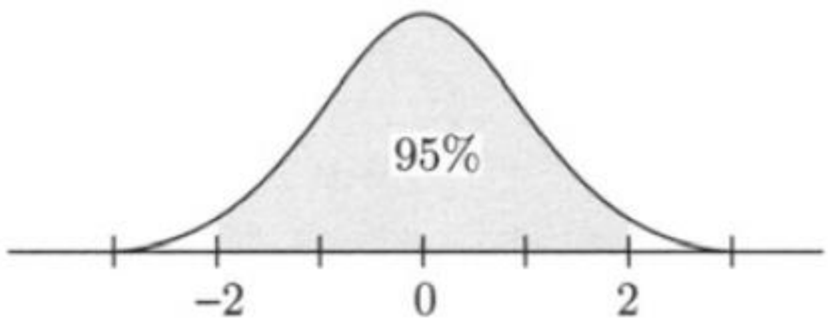
- 70% of the mass lies within 1 standard deviation of the mean.

$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0.70$



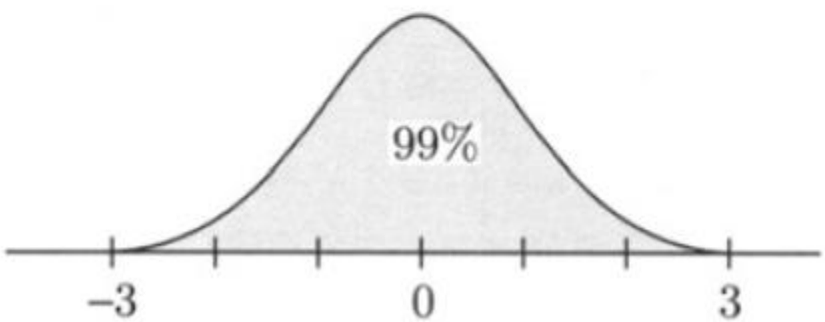
- 95% of the mass lies within 2 standard deviations of the mean.

$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0.95$



- 99% of the mass lies within 3 standard deviations of the mean.

$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) \approx 0.99$$



Normal Approximation to Binomial Distribution:

- $Bin(n, p)$ prob's can be worked out exactly, when n is small.
- For large n, we approximate binomial distribution by normal distribution.

Let n be large and $X \sim Bin(n, p)$. Then,

$X \approx Y$, where $Y \sim N(np, np(1 - p))$.

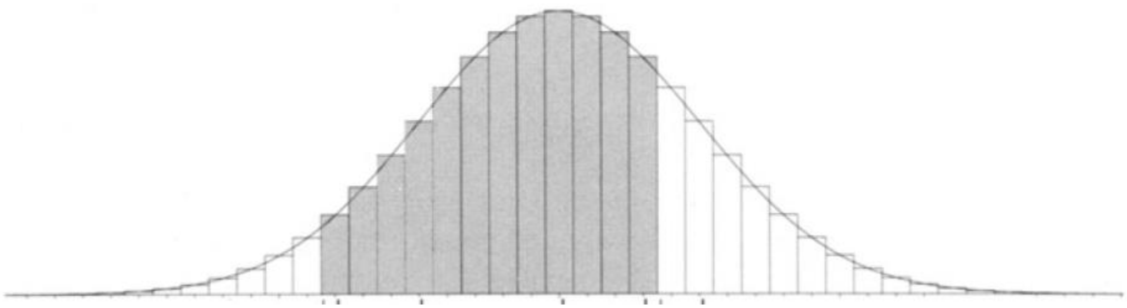
where, $\mu = np$ and $\sigma^2 = np(1 - p)$.

- (Crude) approximation formula for Binomial probabilities:

$X \sim Bin(n, p)$ and n large

$\Rightarrow P(X \leq x) \approx P(Y \leq x)$

where $Y \sim N(np, np(1 - p))$.



Example: $X \sim Bin(1000, 0.5)$. Find $P(X \leq 495)$.

Note: Under this crude approximation, $P(X = 495) \approx P(Y = 495) = 0$ since Y has a continuous distribution.

For discrete Binomial distribution It may be true that $P(X = 495)$ is small. But how small?

Continuity Correction

Need to correct for approximating a discrete distribution by a continuous one.

For a general discrete r.v. X:

- Take possible values $x_1, \ldots, x_n$.
- Let $\delta_i = x_{i+1} - x_i$ be the distance between neighbouring values. (Note: take $\delta_0 = -\infty, \delta_n = +\infty$).

- Split the difference between neighbouring values, we have that x_i is the only possible value for X in the interval $[x_i - \frac{\delta_{i-1}}{2}, x_i + \frac{\delta_i}{2}]$.

- $\boxed{P(X \leq x_i) = P(X \leq x_i + \frac{\delta_i}{2}), P(X \geq x_i) = P(X \geq x_i - \frac{\delta_{i-1}}{2})}$

- $\boxed{P(X = x_i) = P\left(x_i - \frac{\delta_{i-1}}{2} \leq X \leq x_i + \frac{\delta_i}{2}\right).}$

- Approximating X by a r.v. with a continuous distribution, apply the approximation to these expanded events to get more accurate results.

- In the binomial case $\delta_i = 1$, for all i.

Example: $X \sim Bin(1000, 0.5)$. Find

- $P(X = 495)$
- $P(X \leq 495)$

To summarize:

- Can do normal approximation with or without a continuity correction.
- Including the correction gives greater accuracy when approximating binomials
- The normal probabilities can be found using software, crude rounding, or linear interpolation.

Example: Batting averages (§2.2 Problem 11a)

If a player's true batting average is .300, what is the probability of hitting .310 or better over the next 100 at bats?