Last time

Nomal approximation to binomial

Poisson approximation to binomial.

$k \sim Bin(n,p)$

$$\mathbb{P}(a \le k \le b) \approx \int_{a-0.5}^{b+0.5} Normal(\mu, \sigma)\, dx$$

where $\mu = np$

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

If n is large, p is too small so that $np \le 1$

then, $\mathbb{P}(k \text{ sucesses}) \approx e^{-\mu} \frac{\mu^k}{k!}$ where $\mu = np$.

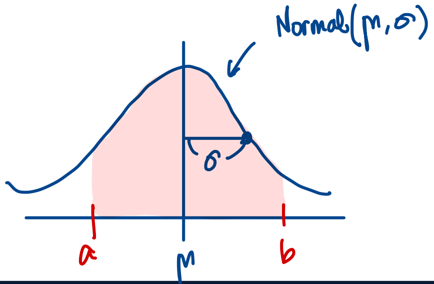
Today

More about Normal distributions

Confidence intervals

Probability density function for the standard normal distribution

PDF of Standard Normal dist.: $f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2}.$



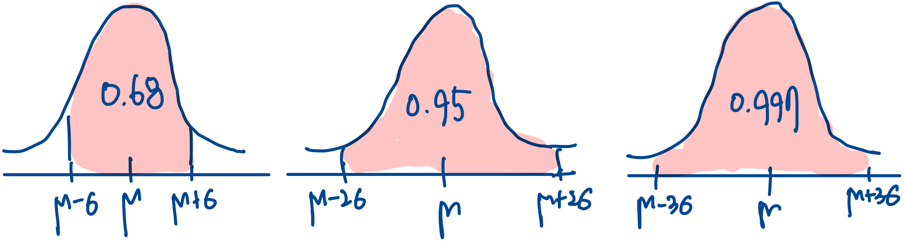
Change of Variable (Normal(μ, σ) $\to$ Normal(0,1))

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\, dx = \int_{\boxed{\frac{a-\mu}{\sigma}}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \cdot e^{-z^2/2}\, dz.$$

- a is raw score
- $\frac{a-\mu}{\sigma}$ is z-score.

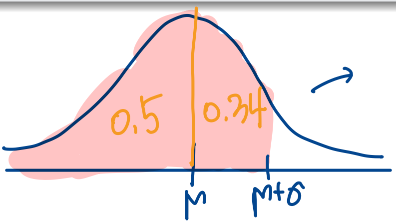
68–95–99.7 rule (a.k.a. Empirical rule)

Whatever μ and σ are, if $X \sim \text{Normal}(\mu, \sigma^2)$



Q. Use the symmetry of normal distribution to evaluate

$$\int_{-\infty}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\ dx.$$

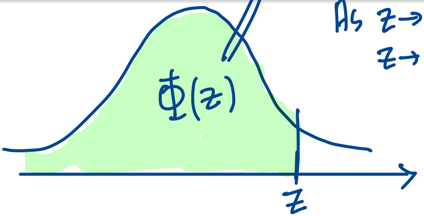


$\rightarrow$ 0.84

Definition

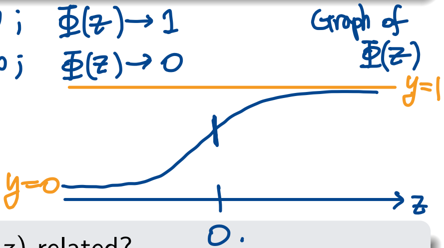
The cumulative distribution function (CDF) for a standard normal distribution

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



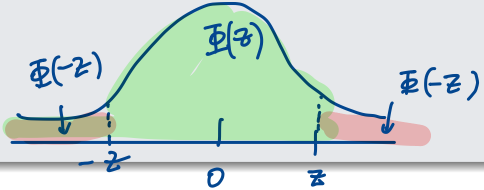
As $z \to \infty$; $\Phi(z) \to 1$

$z \to -\infty$; $\Phi(z) \to 0$



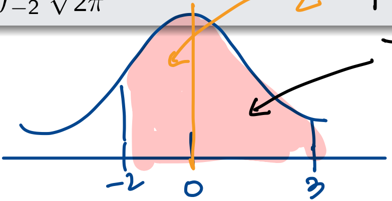
Q. Let $z > 0$. How are $\Phi(z)$ and $\Phi(-z)$ related?

a. $\Phi(z) = \Phi(-z)$
b. $\Phi(z) = -\Phi(-z)$
c. $\Phi(z) - \Phi(-z) = 1$
d. $\Phi(z) + \Phi(-z) = 1$



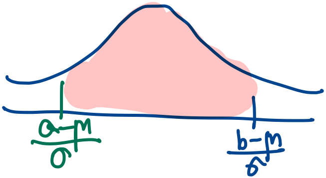
Q. Evaluate $\int_{-2}^{3} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$.

$$\frac{0.95}{2} + \frac{0.997}{2}$$



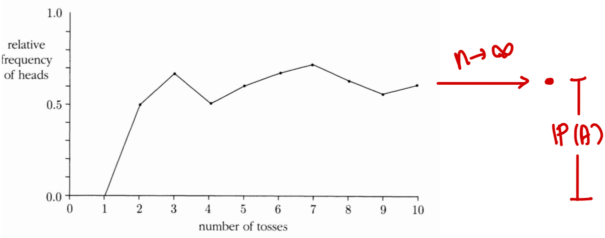
Q. Represent $\int_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\, dx$ in terms of the CDF Φ.

$$= \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$



Big ideas.

1. If the probability of success is p, what are you likely to observe as the proportion of successes $\hat{p}$? What is a likely range of values for $\hat{p}$?
2. Reciprocally, if you observe $\hat{p}$ as your proportion of successes, what are likely values for the actual probability p?



We can say something like: $|P_{100}(A) - \mathbb{P}(A)| \leq \varepsilon$ with probability 95%

i.e. $\mathbb{P}(A) \in [P_{100}(A)-\varepsilon, P_{100}(A)-\varepsilon]$ with 95% prob.

Example 3.2

// p

A biased coin comes up heads 60% of the time. You flip the coin 100 times and observe some frequency of heads $\hat{p}$. What interval centered at p should contain $\hat{p}$ with 99.7% chance?

<u>Goal</u> Find $\varepsilon > 0$ s.t. $\mathbb{P}\left(\hat{p} \in (0.6-\varepsilon, 0.6+\varepsilon)\right) = 99.7\%$

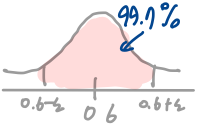
where $\hat{p} = \dfrac{X}{100}$ and X is the number of heads.

$$X \sim \mathrm{Bin}(100, 0.6) \overset{\cdot}{\sim} \mathrm{Normal}(\mu = 60, \sigma = \sqrt{24})$$

$$\hat{p} = \frac{X}{100} \sim \mathrm{Normal}\left(\mu = \frac{60}{100}, \sigma = \frac{\sqrt{24}}{100}\right) \leftarrow$$

Remark (Important)

If $X \sim \mathrm{Normal}(\mu, \sigma)$, $\dfrac{X}{a} \sim \mathrm{Normal}\left(\dfrac{\mu}{a}, \dfrac{\sigma}{a}\right)$



By the empirical rule

$$\varepsilon = 3\sigma = 3 \cdot \frac{\sqrt{24}}{100}$$

Example 3.3

A biased coin comes up heads 60% of the time. How many times do you need to flip the coin in order to be 95% sure that the observed frequency will be between 55% and 65%.?

As n increases, the likelyhood that $\frac{X}{n} = \hat{p}$ is between 0.55 and 0.65 increases.

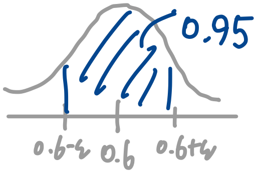
Goal Find n s.t. $\mathbb{P}(\underset{0.6-\varepsilon}{0.55} \le \frac{X}{n} \le \underset{0.6+\varepsilon}{0.65}) = 0.95.$

$\varepsilon = 0.05$

X: # of heads.

$X \sim Bin(n, 0.6) \dot{\approx} Norm(0.6n, 0.24n)$

$\frac{X}{n} \dot{\approx} Norm(0.6, \frac{0.24}{n})$



By the empirical rule.

$\varepsilon = 0.05 = 2 \cdot \sqrt{\frac{0.24}{n}}$

$\Rightarrow \quad n \approx 384.$

Example 3.4 We don't know the true fraction p. We infer p from $\hat{p} = \frac{600}{1000}$

500,000 people live in a city. The pollsters sampled 1000 people and 600 of them preferred candidate X. What is the 99.7% confidence interval of the true fraction of people in the city who prefer candidate X?

X : the people perfering X in the 1000 sample.

Since $X \sim \underset{n=1000}{\mathrm{Bin}(n, p)} \dot{\sim} \mathrm{Normal}\left(np, \sqrt{np(1-p)}\right)$

$$\frac{X}{n} = \hat{p} \sim \mathrm{Normal}\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

Hence, $|p - \hat{p}| \le \sqrt{\frac{p(1-p)}{n}}$ with the chance 68%

$\Leftrightarrow p \in \left[\hat{p} - \sqrt{\frac{p(1-p)}{n}}, \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right]$.

Hence, $\left[\hat{p} - \sqrt{\frac{p(1-p)}{n}}, \hat{p} + \sqrt{\frac{p(1-p)}{n}}\right]$ is the 68% confidence int. of p.

$\left[\quad -2\sqrt{\frac{p(1-p)}{n}}, \quad +2\sqrt{\frac{p(1-p)}{n}}\right]$ 95% "

$\left[\quad -3\sqrt{} - \quad -3\sqrt{-}\right]$ 99.7% int. of p.

Problem! We don't know p ! So we replace p by $\hat{p}$ or $\frac{1}{2}$

After all,

both $\left[\hat{p} - k \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \ \hat{p} + k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$ and

this interval is largest when $\hat{p} = \frac{1}{2}$.

$$\boxed{\left[\hat{p} - k \cdot \frac{1}{2\sqrt{n}}, \ \hat{p} + k\frac{1}{2\sqrt{n}}\right]}$$ are called

the 68%, 95%, 99.7% confidence intervals of p

when $k = 1, 2, 3$, respectively.

Let's use this, for safety.

p belongs to this interval with at least 68, 95, 99.7% chances.

x% confidence interval $\Rightarrow$ Find k s.t.

i.e. $2(\Phi(k) - 0.5) = x\%$