S&DS 241 Lecture 16

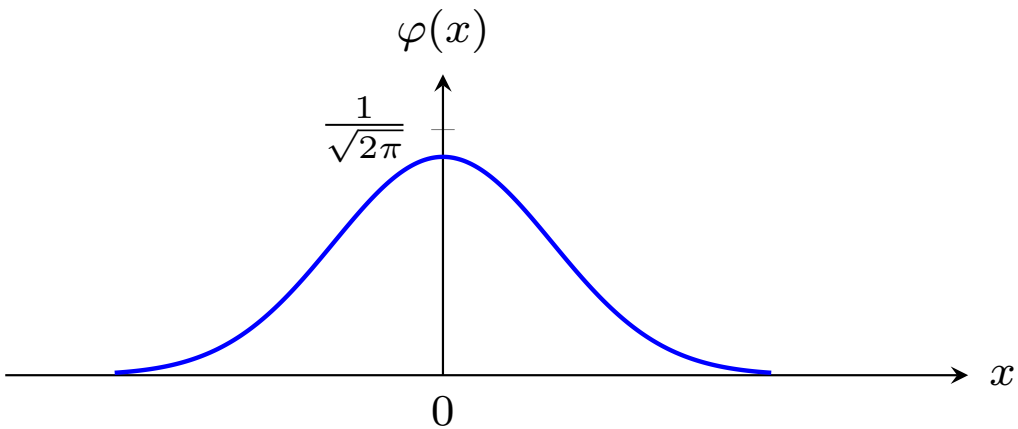
Normal distributions, CLT for binomials

B-H: 5.4, 10.3

Standard normal distribution

A continuous random variable X is said to have the standard normal (Gaussian) distribution, if it has the following PDF:

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \triangleq \varphi(x)}$$

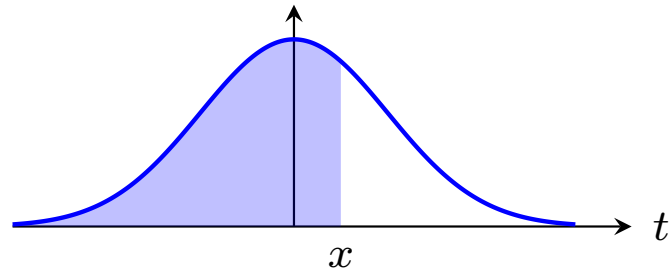


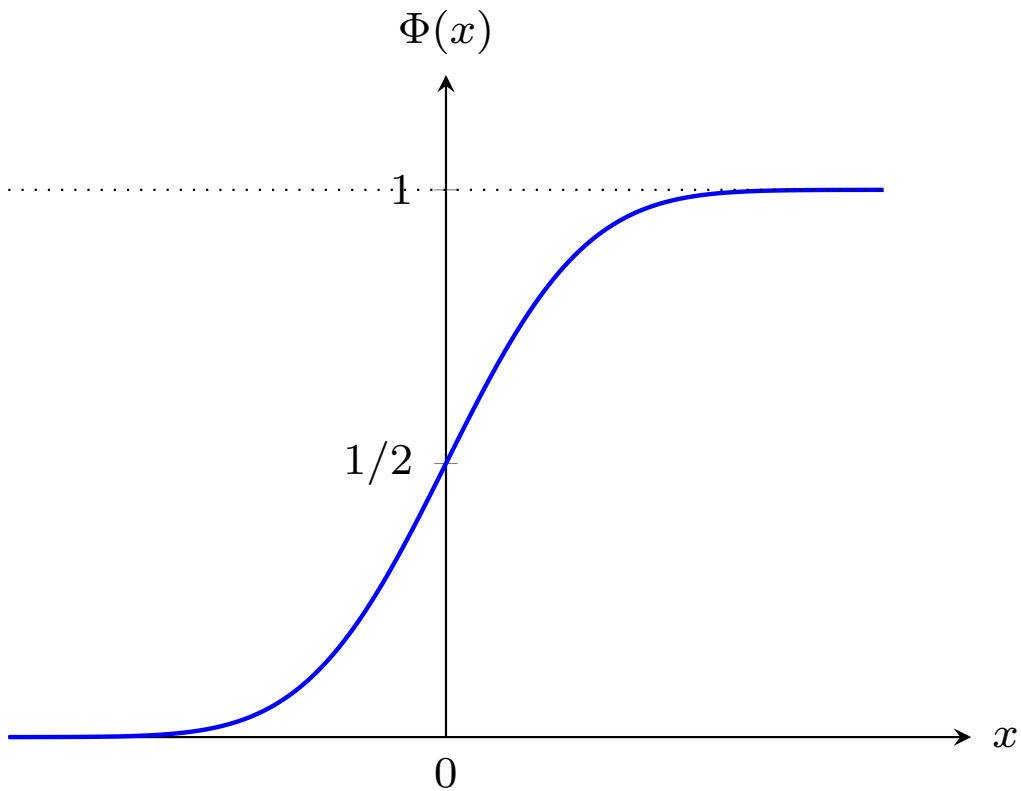
Carl Friedrich Gauss

- Symmetric and bell-shaped
- $E(X) = 0, \mathrm{Var}(X) = \mathrm{SD}(X) = 1$

CDF of standard normal

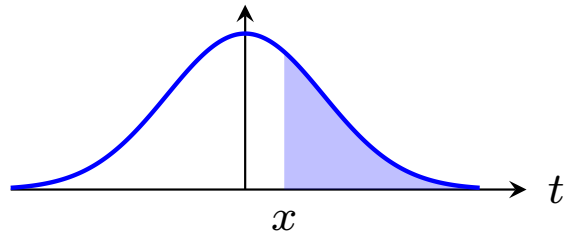
$$P(X \leq x) = \Phi(x) \triangleq \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt =$$

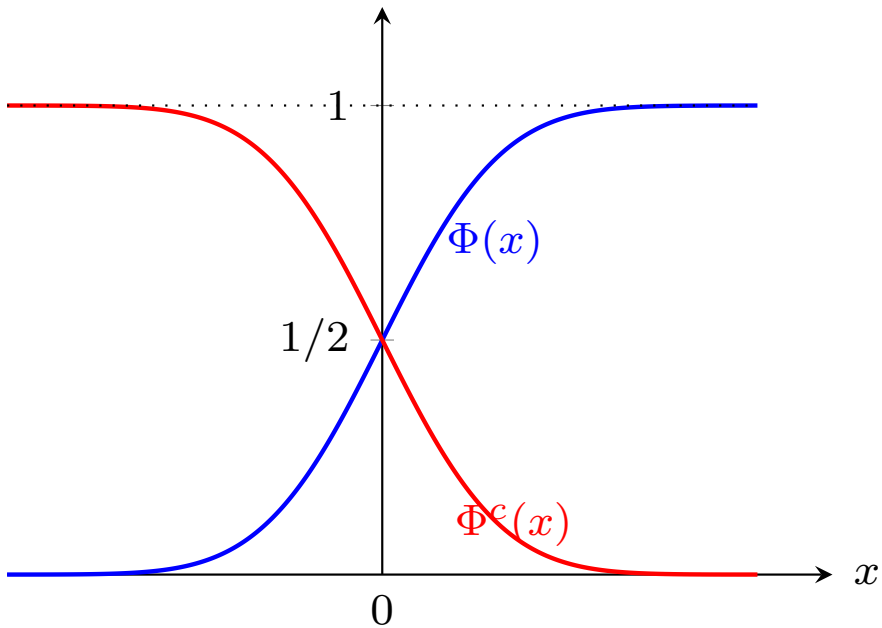




Complementary CDF (tail)

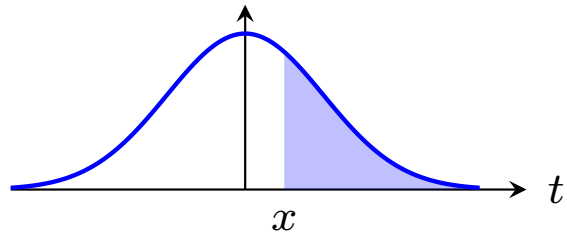
$$\Phi^c(x) \triangleq 1-\Phi(x) = P(X > x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt =$$

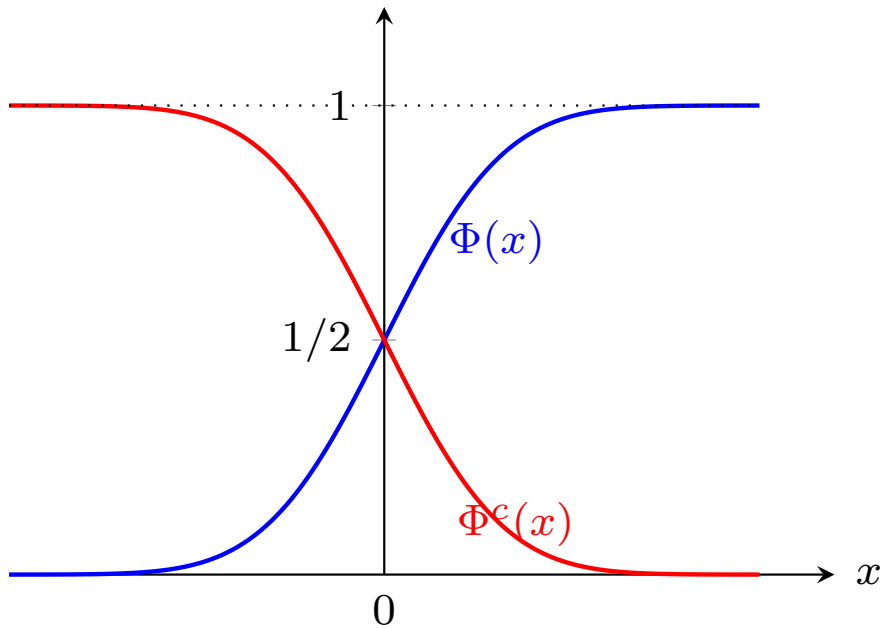




Complementary CDF (tail)

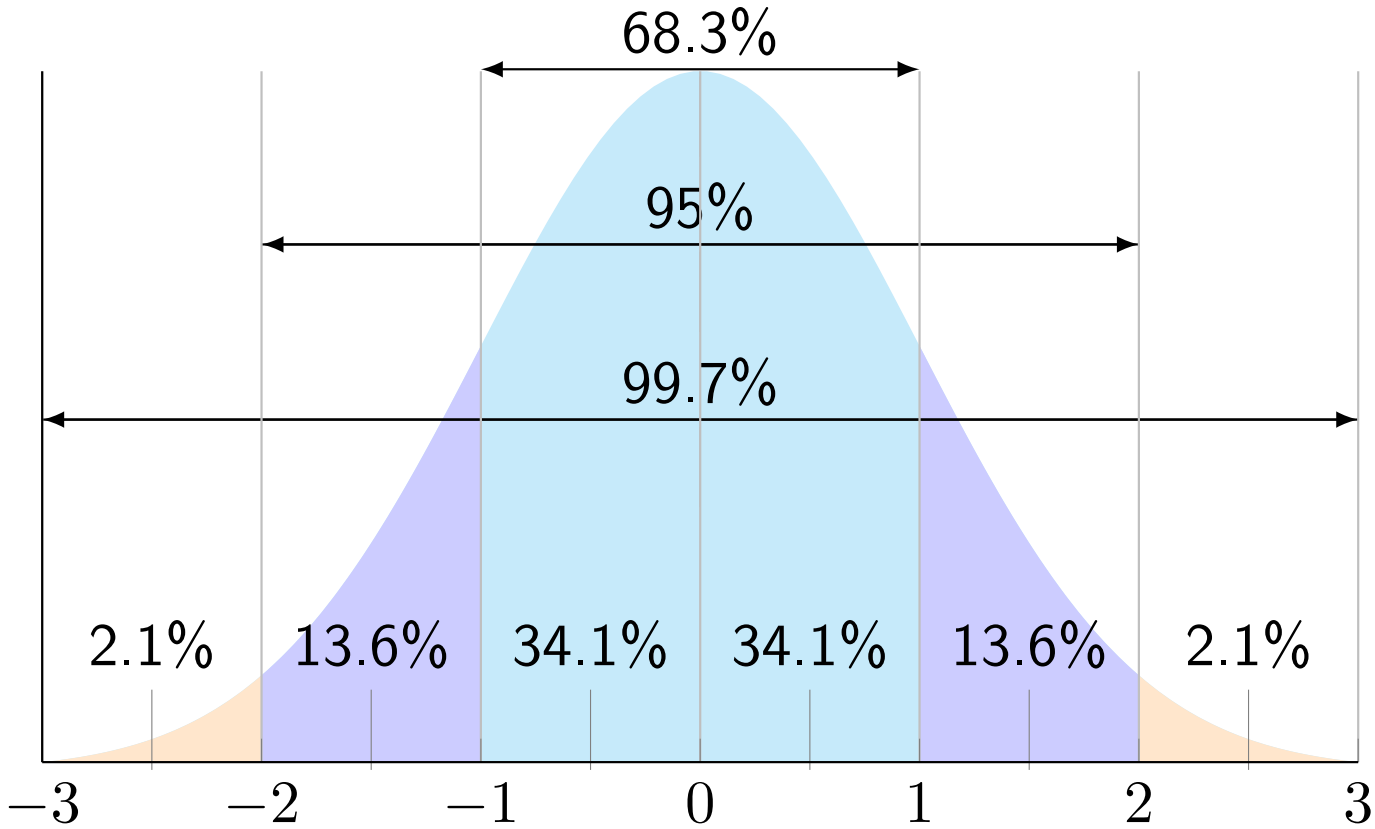
$$\Phi^c(x) \triangleq 1-\Phi(x) = P(X > x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt =$$





Consequences of symmetry:

- $P(X \leq -x) = \Phi(-x) = \Phi^c(x) = P(X \geq x)$,
- $P(|X| \leq x) = 2\Phi(x) - 1$

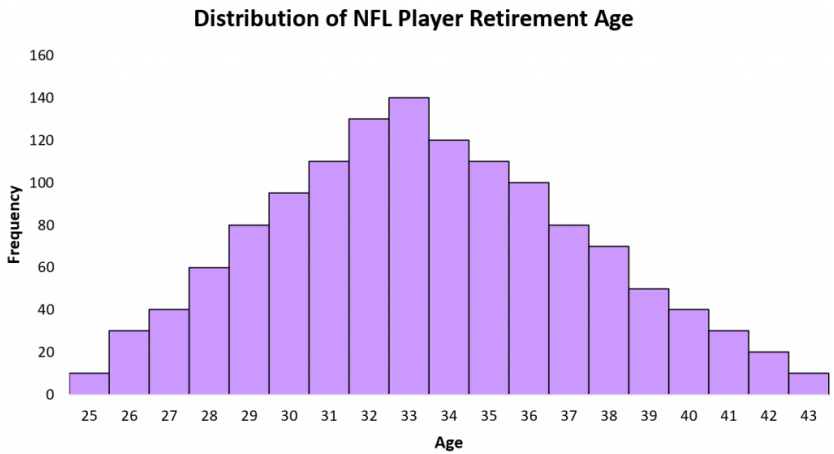


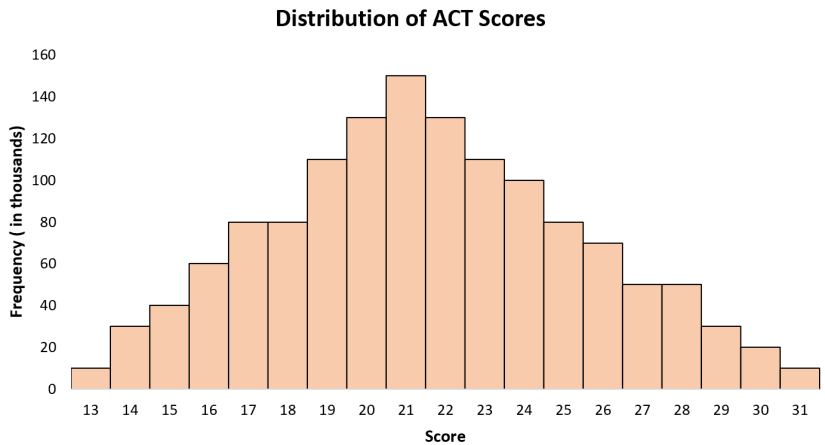
$$P(-1 < X < 1) = \Phi(1) - \Phi(-1) \approx 0.683$$

$$P(-2 < X < 2) = \Phi(2) - \Phi(-2) \approx 0.95$$

$$P(-3 < X < 3) = \Phi(3) - \Phi(-3) \approx 0.997$$

Normal distribution in real life

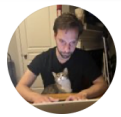




For more examples:

https://www.statology.org/example-of-normal-distribution/

Normal distribution in real life



Nuno Clara
@NunoClara_

Behavioral biases. Distribution is pretty normal, but 95 always rounds to 100.

Normalization

To show $\int_{-\infty}^{\infty} \varphi(x)dx = 1$, we need to verify:

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$$

Normalization

To show $\int_{-\infty}^{\infty} \varphi(x)dx = 1$, we need to verify:

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$$

Let's prove

$$\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right)^2 = 2\pi$$

The magic of Gauss

$$\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy\right)$$

The magic of Gauss

$$\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy\right)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}}\,dxdy$$

The magic of Gauss

$$\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)^2 = \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy\right)$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\,dx\,dy$$

The magic of Gauss

$$\begin{aligned}\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)^2 &= \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy\right)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\,dx\,dy\\ &= \int_{0}^{2\pi}\int_{0}^{\infty} e^{-\frac{r^2}{2}}\,r\,dr\,d\theta \qquad \text{(polar coordinates cf. A.7.2)}\end{aligned}$$

The magic of Gauss

$$
\begin{aligned}
\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)^2 &= \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\,dx\,dy \\
&= \int_{0}^{2\pi}\int_{0}^{\infty} e^{-\frac{r^2}{2}}\,r\,dr\,d\theta \qquad \text{(polar coordinates cf. A.7.2)} \\
&= \left(\int_{0}^{2\pi} d\theta\right)\left(\int_{0}^{\infty} e^{-\frac{r^2}{2}}\,r\,dr\right)
\end{aligned}
$$

The magic of Gauss

$$
\begin{aligned}
\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)^2 &= \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\,dx\,dy \\
&= \int_{0}^{2\pi}\int_{0}^{\infty} e^{-\frac{r^2}{2}}\,r\,dr\,d\theta \qquad \text{(polar coordinates cf. A.7.2)} \\
&= \left(\int_{0}^{2\pi} d\theta\right)\left(\int_{0}^{\infty} e^{-\frac{r^2}{2}}\,r\,dr\right) \\
&= 2\pi\left(-e^{-\frac{r^2}{2}}\Big|_0^{\infty}\right) = 2\pi
\end{aligned}
$$

Mean and variance

- Expectation: $E(X) = 0$, by symmetry.

Mean and variance

- Expectation: $E(X) = 0$, by symmetry.
- Variance:

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}} d(-e^{-\frac{x^2}{2}}) \\
&= \underbrace{\frac{-x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty}}_{0} + \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx}_{1} \qquad \text{(IBP)} \\
&= 1
\end{aligned}$$

Shifting and scaling a standard normal

Let Z be standard normal. Let $\mu \in \mathbb{R}$ and $\sigma > 0$. Let

$$X = \mu + \sigma Z$$

Then

- $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$

Shifting and scaling a standard normal

Let Z be standard normal. Let $\mu \in \mathbb{R}$ and $\sigma > 0$. Let

$$X = \mu + \sigma Z$$

Then

- $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$
- PDF: from Lec 15:

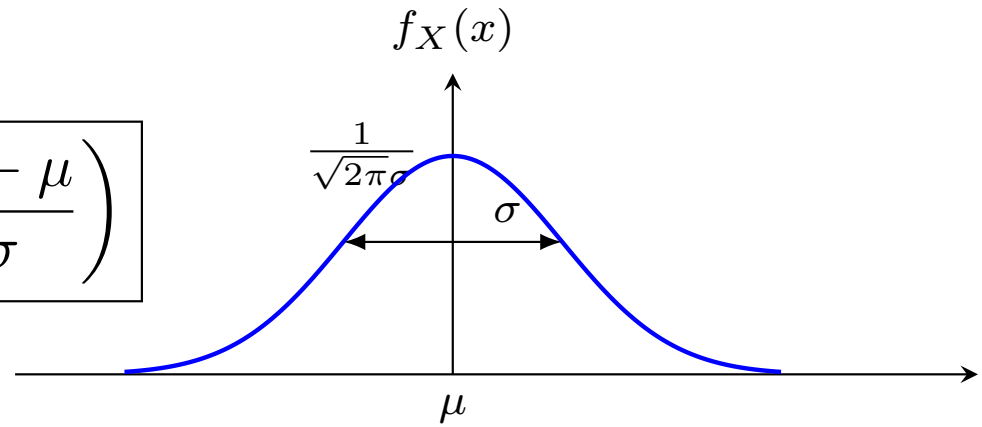
$$f_X(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

  This is the general definition of normal distribution.

Normal distribution

A continuous random variable X is said to have the normal (Gaussian) distribution with mean μ and variance σ^2, denoted by $X \sim N(\mu, \sigma^2)$, if it has the following PDF:

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



- CDF:

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

- Standard normal: $N(0, 1)$
- "68-95-99.7 rule"

$$P(X \in (\mu - \sigma, \mu + \sigma)) \approx 68.3\%$$
$$P(X \in (\mu - 2\sigma, \mu + 2\sigma)) \approx 95\%$$
$$P(X \in (\mu - 3\sigma, \mu + 3\sigma)) \approx 99.7\%$$

Higgs boson discovery[1]


SCIENTIFIC AMERICAN

S MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODC

5 Sigma What's That?

STAFF | By Evelyn Lamb on July 17, 2012 10

Chances are, you heard this month about the discovery of a tiny fundamental physics particle that may be the long-sought Higgs boson. The phrase five-

LATEST

- Particle physics: 5σ-rule is the norm, corresponding to "p-value" 6×10^{-7} – probability such a discovery is due to sheer chance
- Other subjects, e.g., social sciences: 2σ corresponding to 5%

[1] https://blogs.scientificamerican.com/observations/five-sigmawhats-that/

Why is normal distribution so important

CLT and normal approximation

Central limit theorem (CLT)

The cumulative effect of many small independent effects is approximately normal

CLT and normal approximation

Central limit theorem (CLT)

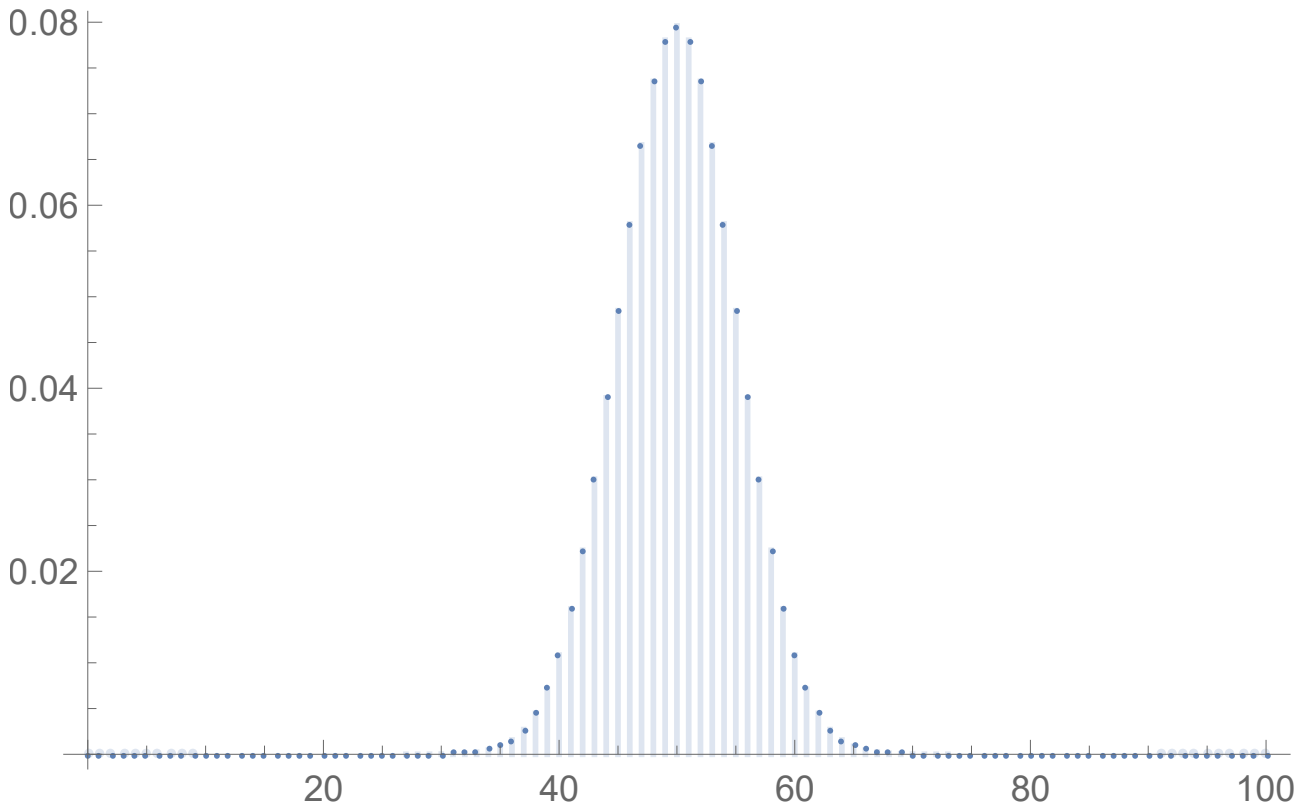
The cumulative effect of many small independent effects is approximately normal

- Normal approximation: approximate the distribution of X by the normal distribution with matching mean μ_X and variance σ_X^2, i.e.,

$$\tilde{X} \sim N(\mu_X, \sigma_X^2)$$

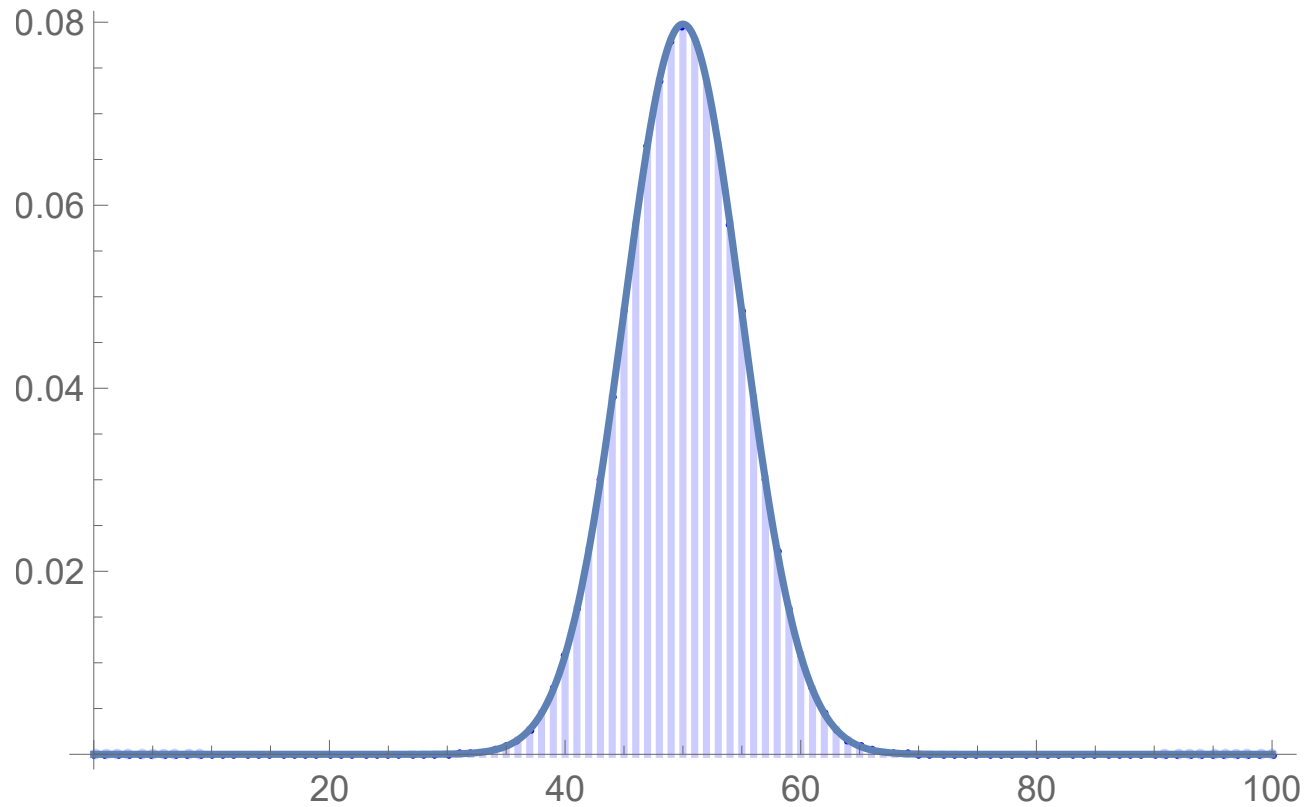
  Then we approximate the CDF $P(X < x)$ by $P(\tilde{X} < x) = \Phi(\frac{x-\mu_X}{\sigma_X})$
- CLT provides a theoretical justification for normal approximation

Normal approximation of binomials



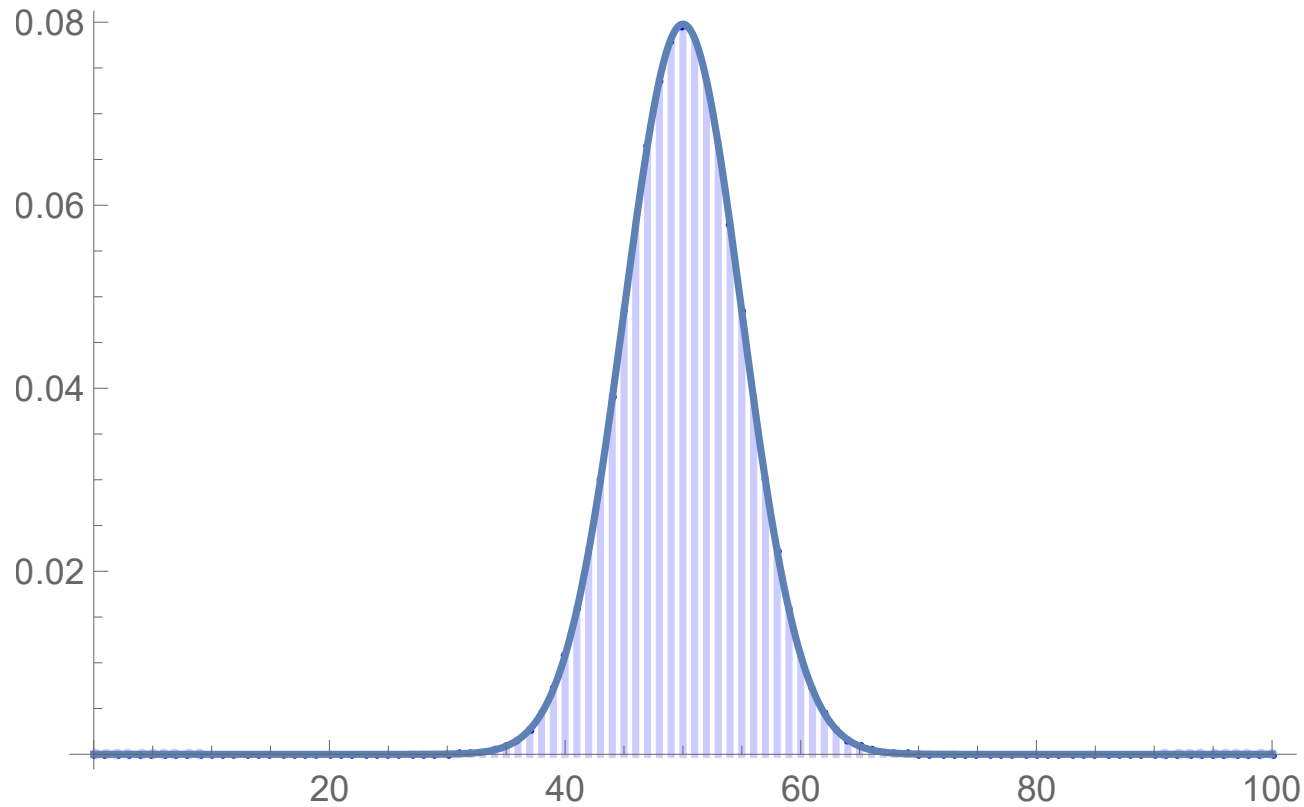
$$X \sim \mathsf{Bin}(n,p) \xrightarrow{\text{normal approximation}} \tilde{X} \sim N(np, npq)$$

Normal approximation of binomials



$$X \sim \mathsf{Bin}(n,p) \xrightarrow{\text{normal approximation}} \tilde{X} \sim N(np, npq)$$

Normal approximation of binomials



$$X \sim \text{Bin}(n, p) \xrightarrow{\text{normal approximation}} \tilde{X} \sim N(np, npq)$$

Contrast with Poisson approximation: approximate $\text{Bin}(n, \frac{\lambda}{n})$ by $\text{Pois}(\lambda)$

Revisit: deviation inequality (Lec 12)

Question

Toss a fair coin 100 times. How unlikely is it to get at least 75 heads?

Let $X \sim \text{Bin}(100, 1/2)$. Then $EX = 50$ and $\text{Var}(X) = 25$

- Markov inequality: $P(X \geq 75) \leq \frac{50}{75} = \frac{2}{3}$

Revisit: deviation inequality (Lec 12)

Question

Toss a fair coin 100 times. How unlikely is it to get at least 75 heads?

Let $X \sim \text{Bin}(100, 1/2)$. Then $EX = 50$ and $\text{Var}(X) = 25$

- Markov inequality: $P(X \geq 75) \leq \frac{50}{75} = \frac{2}{3}$
- Chebyshev inequality:

$$P(X \geq 75) \leq P(|X - 50| \geq 25) \leq \frac{25}{25^2} = 4\%$$

Revisit: deviation inequality (Lec 12)

Question

Toss a fair coin 100 times. How unlikely is it to get at least 75 heads?

Let $X \sim \text{Bin}(100, 1/2)$. Then $EX = 50$ and $\text{Var}(X) = 25$

- Markov inequality: $P(X \geq 75) \leq \frac{50}{75} = \frac{2}{3}$
- Chebyshev inequality:

$$P(X \geq 75) \leq P(|X - 50| \geq 25) \leq \frac{25}{25^2} = 4\%$$

- Normal approximation: approximate X by $\tilde{X} \sim N(50, 25)$

$$P(X \geq 75) \approx P(\tilde{X} \geq 75) = P\left(\frac{\tilde{X} - 50}{5} \geq \frac{75 - 50}{5}\right)$$
$$= 1 - \Phi(5) = 2.9 \times 10^{-7}$$

- Actual value: 2.8×10^{-7}

Revisit: confidence interval (Lec 12)

Question

Toss a coin n times. Construct confidence interval for the bias p.

- Number of heads: $X \sim \text{Bin}(n, p)$, where the bias p is unknown.
- Empirical frequency of heads $\hat{p} = \frac{X}{n}$ a reasonable estimate of p.
- Confidence interval of level 99%:

$$P(p \in [\hat{p} \pm 0.1]) \geq 99\%, \quad \text{for any } p$$

- How many times do we need to flip the coin?
- Same story for polling.

Sample size

- Chebyshev inequality ($q = 1 - p$):

$$P(p \notin [\hat{p} \pm 0.1]) = P(|X - np| > 0.1n)$$
$$\leq \frac{npq}{(0.1n)^2} \overset{pq \leq 1/4}{\leq} \frac{1}{4n \times 0.1^2} \leq 1\% \Rightarrow \boxed{n \geq 2500}$$

Sample size

- Chebyshev inequality ($q = 1 - p$):

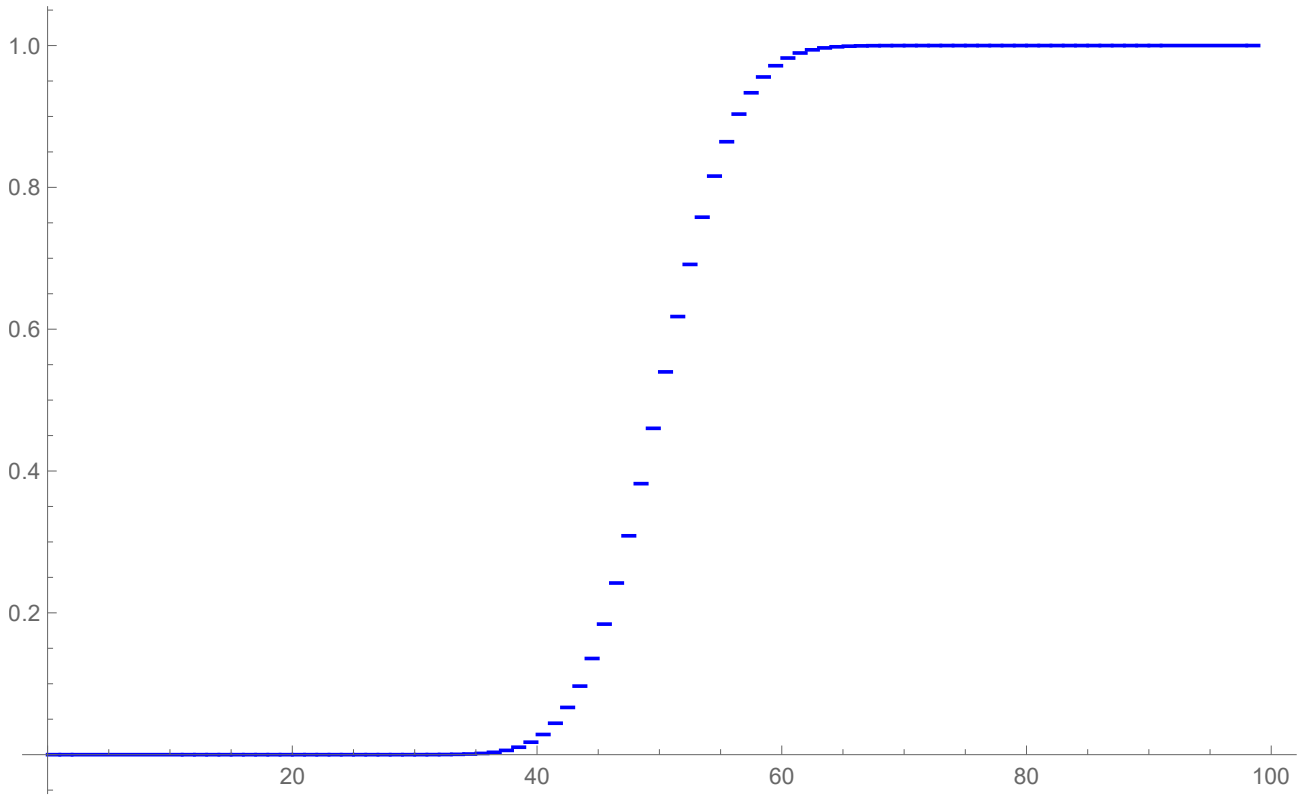
$$P(p \notin [\hat{p} \pm 0.1]) = P(|X - np| > 0.1n)$$
$$\leq \frac{npq}{(0.1n)^2} \overset{pq \leq 1/4}{\leq} \frac{1}{4n \times 0.1^2} \leq 1\% \Rightarrow \boxed{n \geq 2500}$$

- Normal approximation: replace $X \sim \text{Bin}(n, p)$ by $\tilde{X} \sim N(np, npq)$

$$P(p \notin [\hat{p} \pm 0.1]) = P(|X - np| > 0.1n)$$
$$\approx P(|\tilde{X} - np| > 0.1n)$$
$$= 2\Phi^c\left(\frac{0.1n}{\sqrt{npq}}\right) \overset{pq \leq 1/4}{\leq} 2\Phi^c(0.2\sqrt{n}) \leq 1\%$$
$$\Rightarrow \boxed{n \geq 166}$$

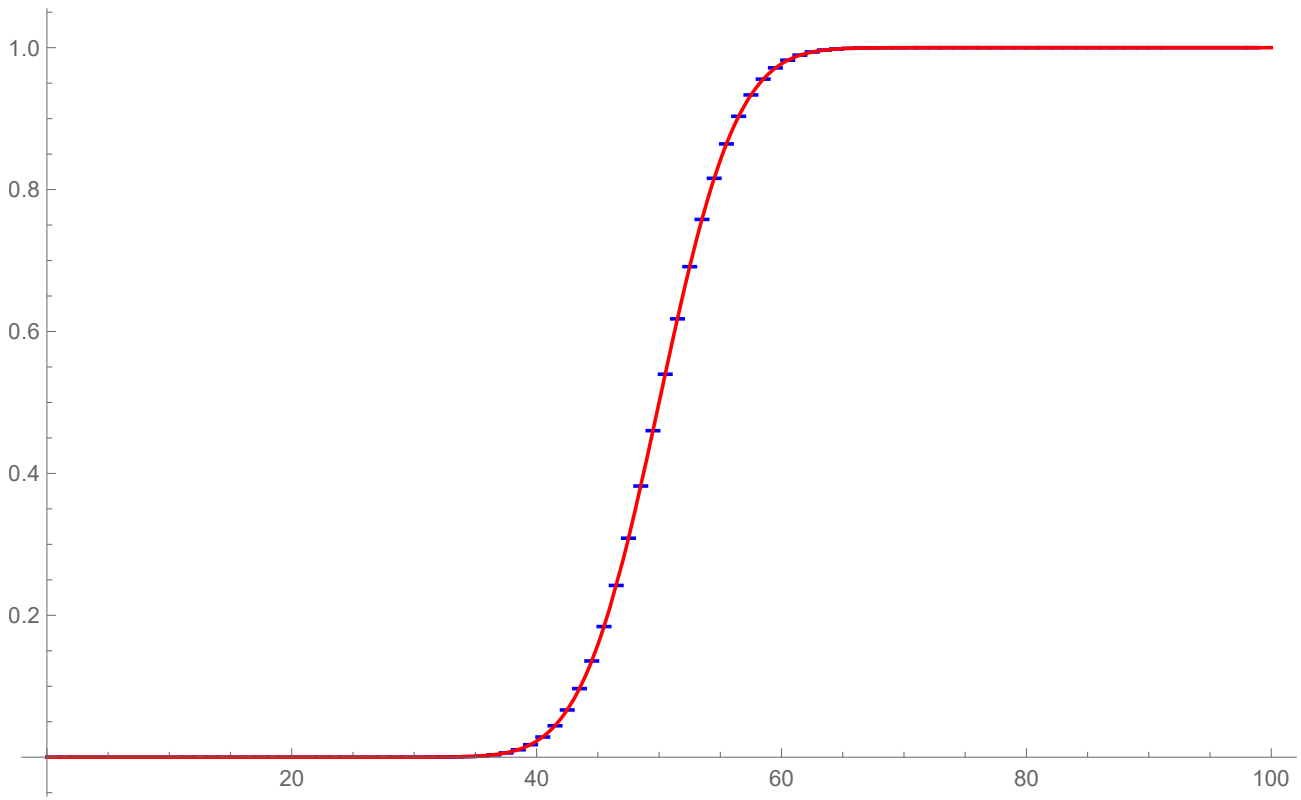
Normal approximation of binomial

CDF of Bin$(100, 1/2)$ and $N(50, 25)$:

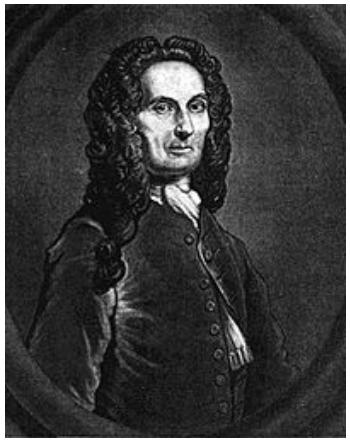
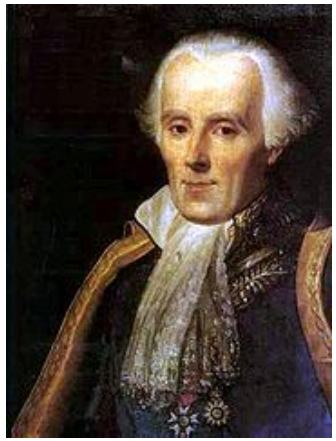


Normal approximation of binomial

CDF of Bin$(100, 1/2)$ and $N(50, 25)$:



de Moivre-Laplace CLT

- Let $X \sim \text{Bin}(n, p)$ and let $q = 1 - p$. Then $E(X) = np$ and $\text{SD}(X) = \sqrt{npq}$.
- CLT is formulated in the convergence of the CDF of the standardized random variable:

Theorem (CLT for binomials)

For any b,

$$P\left(\frac{X - np}{\sqrt{npq}} \leq b\right) \xrightarrow{n\to\infty} \Phi(b) = \int_{-\infty}^{b} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-x^2/2}}_{\varphi(x)}\, dx$$

Normal approximation of binomial: unstandardized

CDF of X vs $N(np, npq)$:

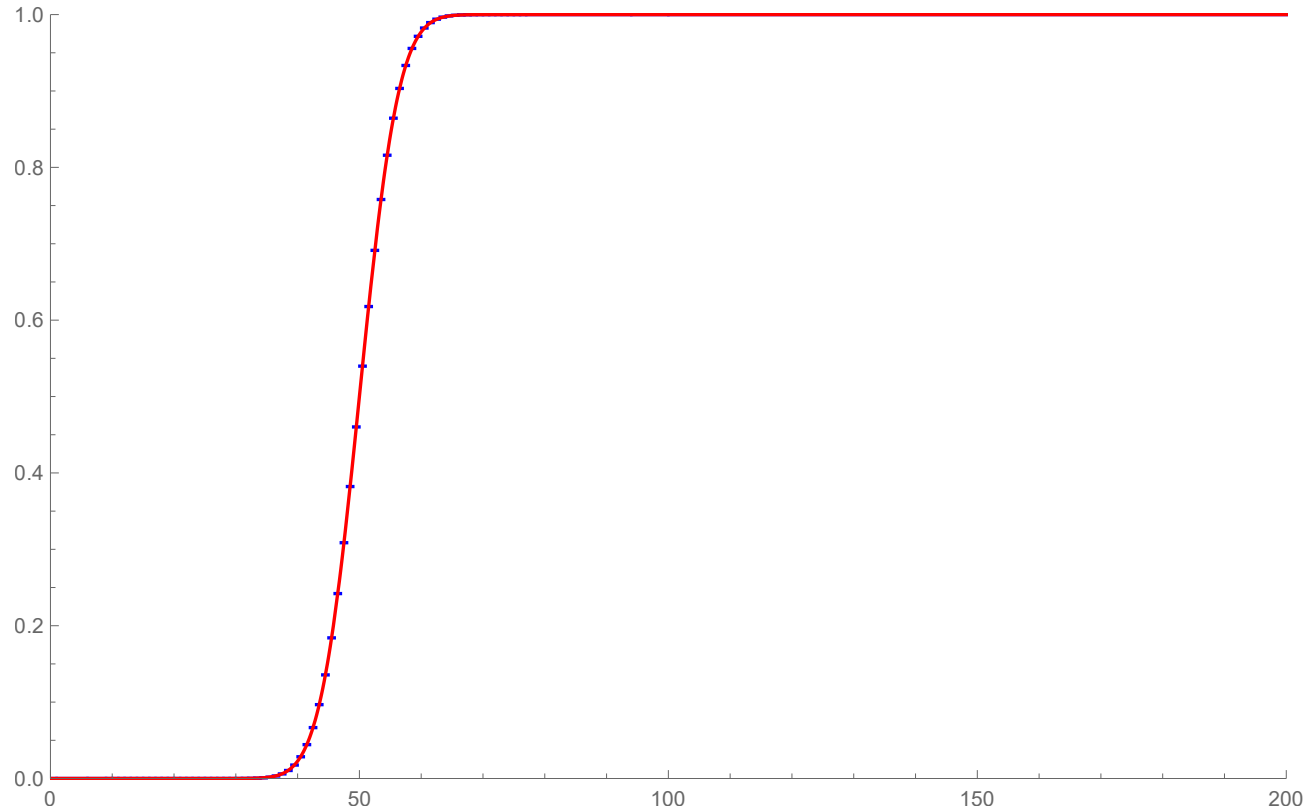


Figure: $n = 100, p = \frac{1}{2}$

Normal approximation of binomial: unstandardized

CDF of X vs $N(np, npq)$:

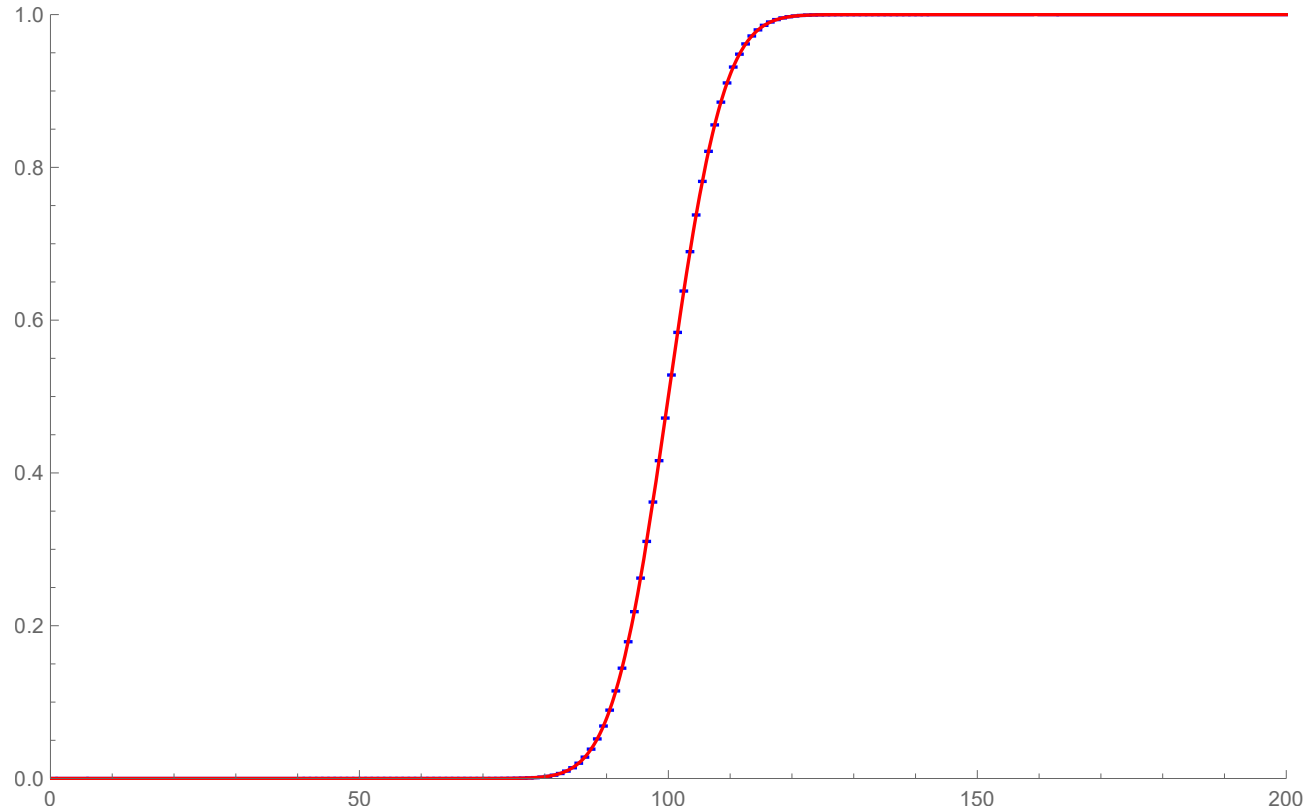


Figure: $n = 200, p = \frac{1}{2}$

Normal approximation of binomial: unstandardized

CDF of X vs $N(np, npq)$:

Figure: $n = 400, p = \frac{1}{2}$

Normal approximation of binomial: standardized

CDF of $\frac{X-np}{\sqrt{npq}}$ vs $N(0,1)$:

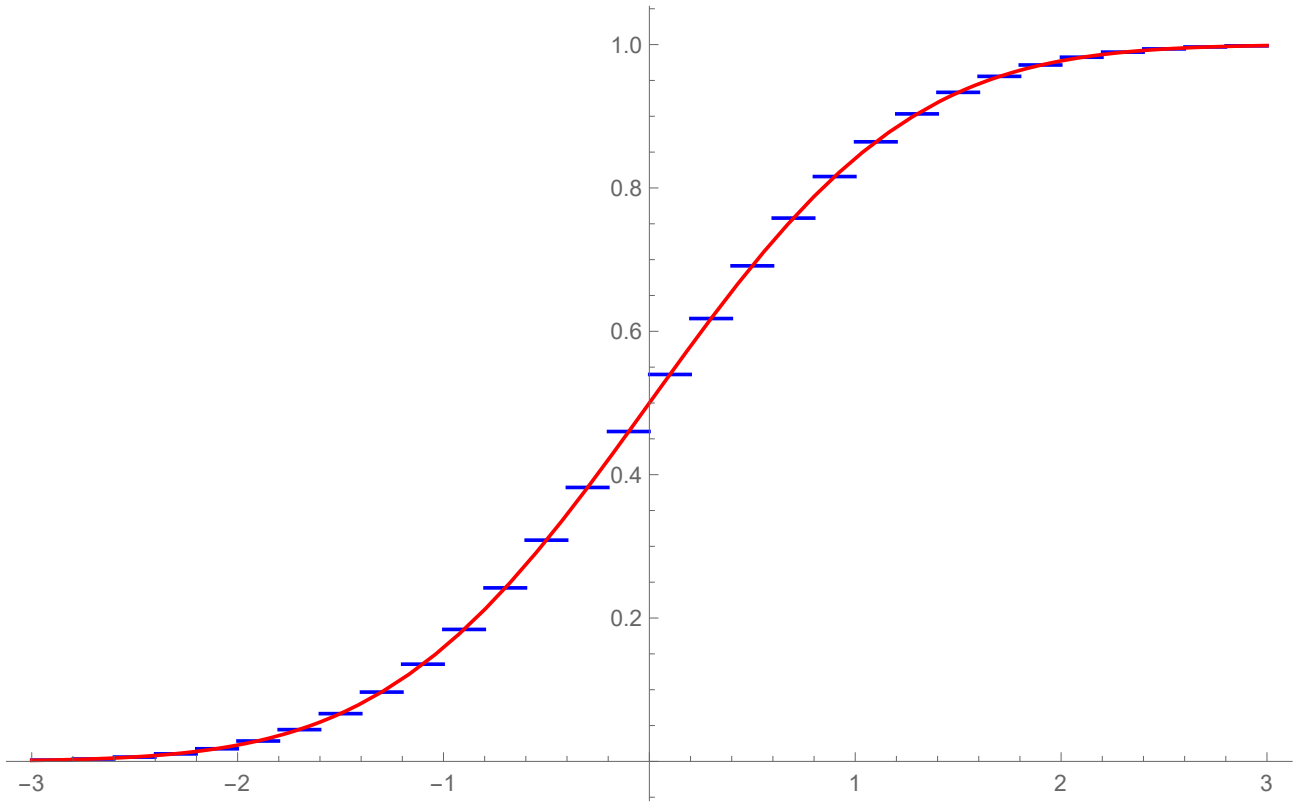


Figure: $n = 100, p = \frac{1}{2}$

Normal approximation of binomial: standardized

CDF of $\frac{X-np}{\sqrt{npq}}$ vs $N(0,1)$:

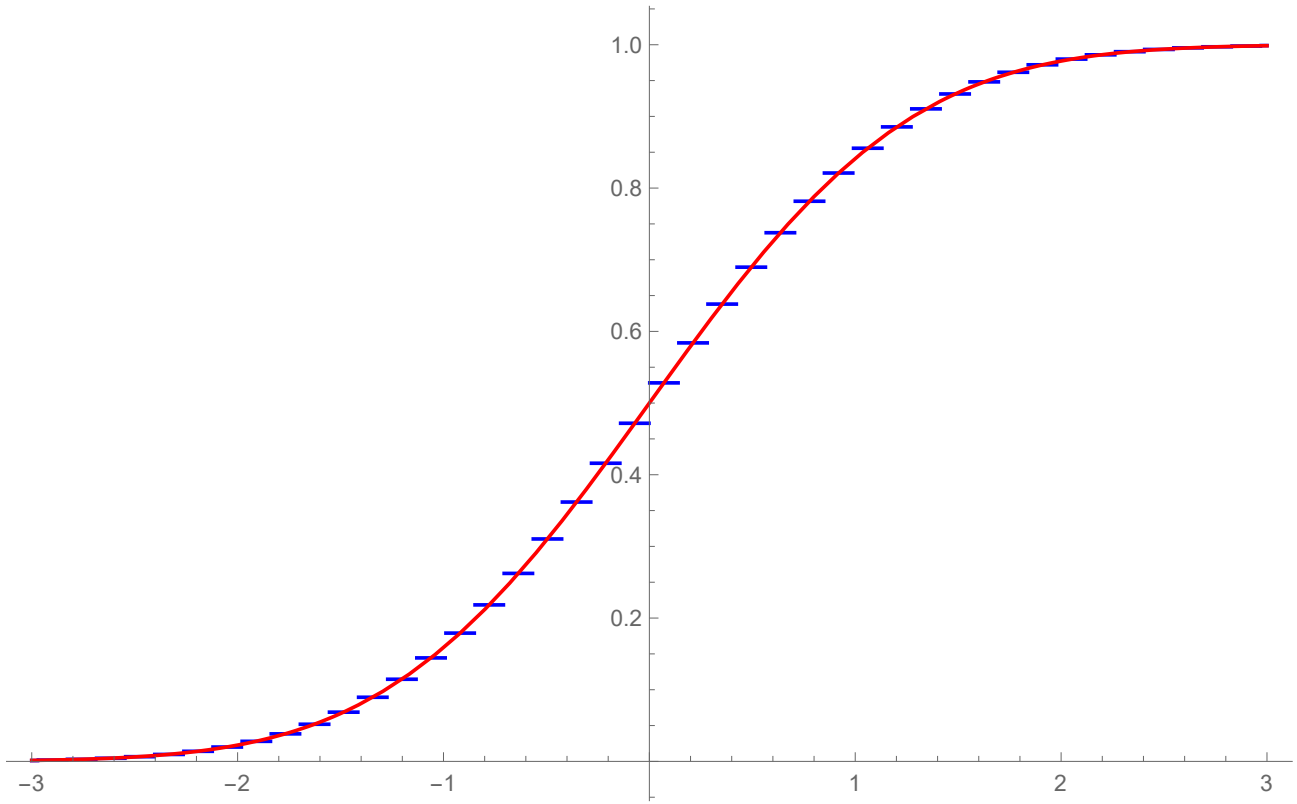


Figure: $n = 200, p = \frac{1}{2}$

Normal approximation of binomial: standardized

CDF of $\frac{X-np}{\sqrt{npq}}$ vs $N(0,1)$:

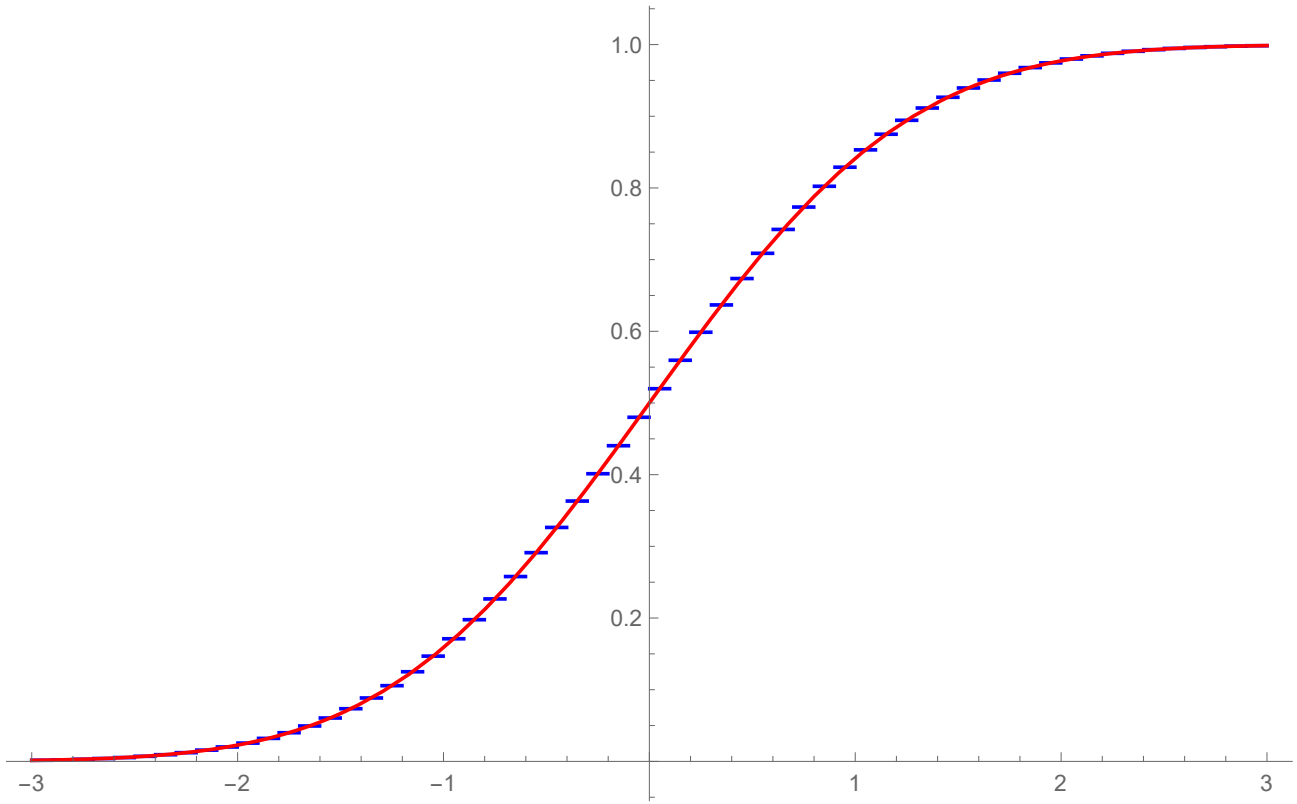


Figure: $n = 400, p = \frac{1}{2}$

Normal approximation of binomial: standardized

CDF of $\frac{X-np}{\sqrt{npq}}$ vs $N(0,1)$:

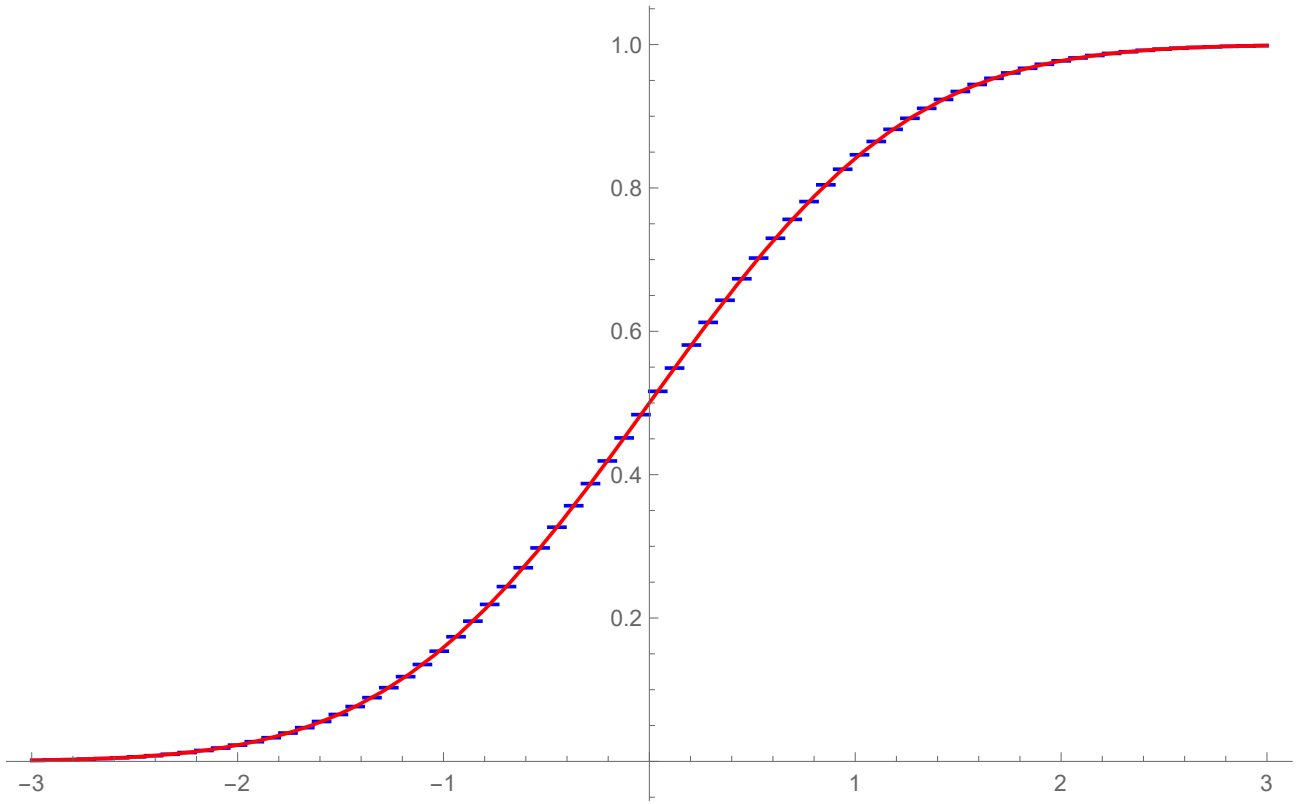


Figure: $n = 600, p = \frac{1}{2}$

Remarks on CLT

- CLT states that

$$\frac{X - np}{\sqrt{npq}} \text{ is approximately distributed as } N(0,1)$$

$$\implies X \text{ is approximately distributed as } \tilde{X} \sim N(np, npq)$$

Remarks on CLT

- CLT states that

$$\frac{X - np}{\sqrt{npq}} \text{ is approximately distributed as } N(0,1)$$

$$\implies X \text{ is approximately distributed as } \tilde{X} \sim N(np, npq)$$

In practice,

$$P(x \leq X \leq y) \approx P(x \leq \tilde{X} \leq y) = \Phi\left(\frac{y - np}{\sqrt{npq}}\right) - \Phi\left(\frac{x - np}{\sqrt{npq}}\right)$$

Remarks on CLT

- CLT states that

$$\frac{X - np}{\sqrt{npq}} \text{ is approximately distributed as } N(0,1)$$

$$\implies X \text{ is approximately distributed as } \tilde{X} \sim N(np, npq)$$

  In practice,

$$P(x \leq X \leq y) \approx P(x \leq \tilde{X} \leq y) = \Phi\left(\frac{y - np}{\sqrt{npq}}\right) - \Phi\left(\frac{x - np}{\sqrt{npq}}\right)$$

- After all, binomial (discrete) and normal (continuous) are very different distributions, so normal approximation has its limitations.
  - For example, $P(X \text{ is integer}) = 1$ but $P(\tilde{X} \text{ is integer}) = 0$
  - CLT says events like $P(x \leq X \leq y)$ can be approximated well.

General CLT (Lec 24)

Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed (i.i.d.) with mean μ and variance σ^2. Then

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \text{ is approximately distributed as } N(0,1)$$

General CLT (Lec 24)

Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed (i.i.d.) with mean μ and variance σ^2. Then

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \text{ is approximately distributed as } N(0,1)$$

- de Moivre-Laplace CLT is a special case for Bernoullis
- We can apply this to discrete random variables (e.g. dice) or continuous random variables (later)

General CLT (Lec 24)

Let $X_1, X_2, X_3, \ldots$ be independent and identically distributed (i.i.d.) with mean μ and variance σ^2. Then

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} \text{ is approximately distributed as } N(0,1)$$

- de Moivre-Laplace CLT is a special case for Bernoullis
- We can apply this to discrete random variables (e.g. dice) or continuous random variables (later)
- Universality of normal distribution:
  - Universal laws that do not depend too much on the underlying microscopic mechanism of the system

Justification of CLT

- Later in Lec 24 (B-H 10.3) we will justify CLT in its full generality
- Next let's justify CLT for binomial (cf. Grinstead-Snell Sec 9.1):

Goal

Let $X \sim \text{Bin}(n, p)$, where p is fixed. For any $a < b$,

$$P\left(a \leq \frac{X - np}{\sqrt{npq}} \leq b\right) \xrightarrow{n\to\infty} \Phi(b) - \Phi(a) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Sketch proof of de Moivre-Laplace CLT (optional)

Step 1 Approximate the binomial PMF near the mean:

$$P(X = np + k) \sim \frac{1}{\sqrt{2\pi pqn}} e^{-\frac{k^2}{2pqn}} = \frac{1}{\sqrt{pqn}} \varphi\left(\frac{k}{\sqrt{pqn}}\right),$$

provided that $|k|/\sqrt{n}$ is bounded from above. (We write $a_n \sim b_n$ if $\lim \frac{a_n}{b_n} = 1$.)

Sketch proof of de Moivre-Laplace CLT (optional)

Step 1 Approximate the binomial PMF near the mean:

$$P(X = np + k) \sim \frac{1}{\sqrt{2\pi pqn}} e^{-\frac{k^2}{2pqn}} = \frac{1}{\sqrt{pqn}} \varphi\left(\frac{k}{\sqrt{pqn}}\right),$$

provided that $|k|/\sqrt{n}$ is bounded from above. (We write $a_n \sim b_n$ if $\lim \frac{a_n}{b_n} = 1$.)

Step 2 Approximate summation by integral: fix $a < b$,

$$\begin{aligned} P\left(a \le \frac{X - np}{\sqrt{npq}} \le b\right) &= \sum_{k=a\sqrt{npq}}^{b\sqrt{npq}} P\left(X = np + k\right) \\ &\approx \sum_{k=a\sqrt{npq}}^{b\sqrt{npq}} \frac{1}{\sqrt{pqn}} \varphi\left(\frac{k}{\sqrt{pqn}}\right) \\ &\to \int_a^b \varphi(x)dx = \Phi(b) - \Phi(a). \end{aligned}$$

Justification of Step 1

- Recall Stirling approximation (Lec 11 or B-H Chap 3 Problem 36)

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Justification of Step 1

- Recall Stirling approximation (Lec 11 or B-H Chap 3 Problem 36)

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

- Define $h(x) \triangleq -x\log x - (1-x)\log(1-x)$.

$$\begin{aligned} & P(X = np + k) \\ =& \binom{n}{np+k} p^{np+k} q^{nq-k} = \frac{n!}{(np+k)!(nq-k)!} p^{np+k} q^{nq-k} \end{aligned}$$

Justification of Step 1

- Recall Stirling approximation (Lec 11 or B-H Chap 3 Problem 36)

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

- Define $h(x) \triangleq -x\log x - (1-x)\log(1-x)$.

$$\begin{aligned}
&P(X = np+k) \\
=&\binom{n}{np+k} p^{np+k} q^{nq-k} = \frac{n!}{(np+k)!(nq-k)!} p^{np+k} q^{nq-k} \\
\overset{\text{Stirling}}{\sim}& \frac{1}{\sqrt{2\pi pqn}} \exp\left(n \underbrace{h\left(p+\frac{k}{n}\right)}_{\text{Taylor expansion near } p} - nh(p) + k\log\frac{p}{q} \right) \\
\sim& \frac{1}{\sqrt{2\pi pqn}} \exp\left(-\frac{k^2}{2pqn}\right)
\end{aligned}$$