Lecture 23

Standard normal distribution

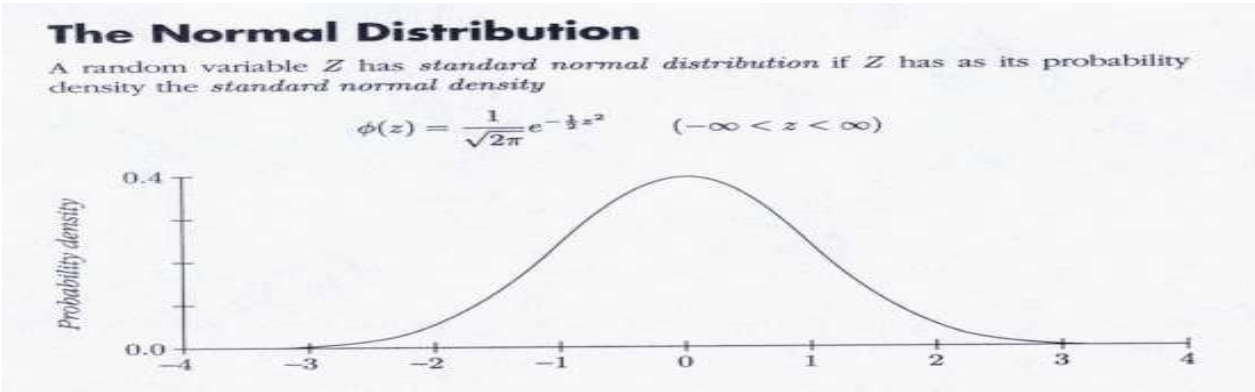
$N(\mu, \sigma^2)$

Normal tables

Central limit theorem

Binomial approximation

Standard normal distribution: $\phi(z) \;=\; \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$



- Standard normal integrals

$$\int_{-\infty}^{\infty} \phi(z)\, dz = 1, \quad \int_{-\infty}^{\infty} z\phi(z)\, dz = 0, \quad \int_{-\infty}^{\infty} z^2\phi(z)\, dz = 1,$$

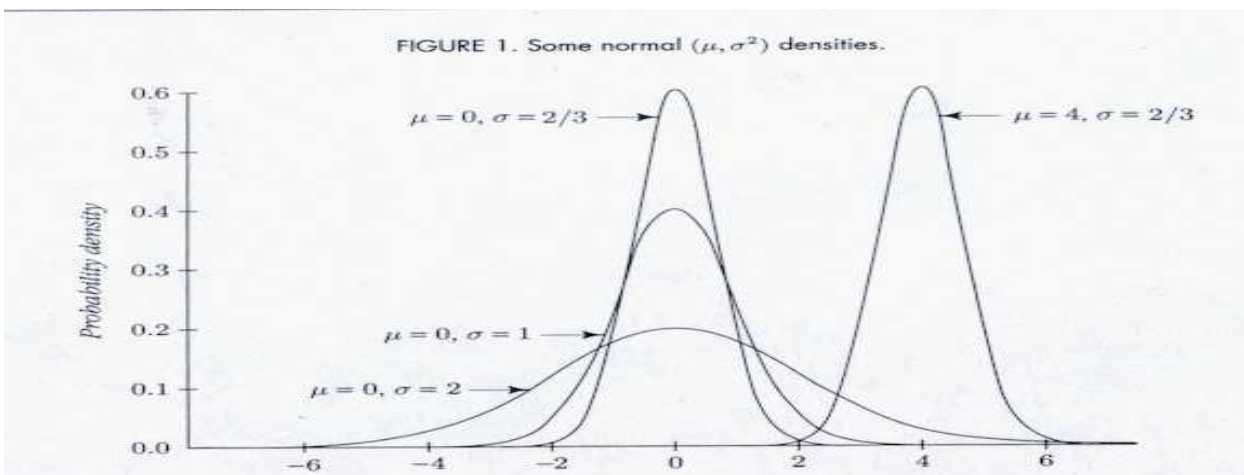
- Normal <u>distribution</u> function

$$\Phi(z) \;=\; P\{Z \le z\} \;=\; \int_{-\infty}^{z} \phi(z)\, dz$$

$$\Phi(a,b) \;=\; P\{a \le Z \le b\} \;=\; \Phi(b) - \Phi(a)$$

General normal distribution

- $X \sim N(\mu, \sigma^2) \iff f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$
- $E\{X\} = \int_{-\infty}^{\infty} x f_X(x)\, dx = \mu$
- $Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\, dx = \sigma^2$



$$X \sim N(\mu, \sigma^2) \iff Z = X^* = \frac{X-\mu}{\sigma} \sim N(0,1)$$

Standard Normal: $Z \sim N(0,1)$

Event	$Z < 1$	$0 < Z < 1$	$Z < 3$	$Z > 3$
Prob	.8413	.3413	.9987	.0013

Intelligence: $IQ \sim N(100, 225)$: $(IQ - 100)/15 \sim N(0,1)$

Event	$IQ < 115$	$100 < IQ < 115$	$IQ < 145$	$IQ > 145$
Prob	.8413	.3413	.9987	.0013

Heights: $H \sim N(178, 36)$: $(H - 178)/6 \sim N(0,1)$

Event	$H < 184$	$178 < H < 184$	$H < 196$	$H > 196$
Prob	.8413	.3413	.9987	.0013

Facts to check

$\int_{-\infty}^{\infty} \phi(z)\, dz = 1$

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\, dz\right)^2 &= \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\, dx \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)}\, dxdy \\
&= \int_{0}^{2\pi}\int_{0}^{\infty} re^{-\frac{1}{2}r^2}\, drd\theta \\
&= 2\pi\left[-e^{-\frac{1}{2}r^2}\right]_0^{\infty} \quad = \quad 2\pi
\end{aligned}$$

$\int_{-\infty}^{\infty} z\phi(z)\, dz = 0$

By symmetry, since $\int_{-\infty}^{\infty} |z|\phi(z)\, dz < \infty$.

$$\int_{-\infty}^{\infty} e^{-z^2}\, dz \;=\; \sqrt{\pi}$$

"A mathematician is one to whom *that*
is as obvious as twice two is four is to you.

Liouville was a mathematician."

– Lord Kelvin

$\int_{-\infty}^{\infty} z^2 \phi(z)\, dz = 1$

$$
\begin{aligned}
\int_{-\infty}^{\infty} z^2 \phi(z)\, dz &= \int_{-\infty}^{\infty} z^2 \, \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} \, dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-z) \frac{d}{dz} \left(e^{-\frac{1}{2}z^2} \right) \, dz \\
&= \frac{1}{\sqrt{2\pi}} \left[(-z) e^{-\frac{1}{2}z^2} \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \, dz \\
&= 1
\end{aligned}
$$

$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad X^* = (X - \mu)/\sigma \sim N(0, 1).$

- $E\{X^*\} = 0$ and $Var(X^*) = 1$ is simply standardisation.
- Normality: Next lecture.

Central limit theorem

- $X_1, X_2, \ldots, X_n$ independent random variables with same distribution. (Discrete or continuous.)
- $E\{X_k\} = \mu, \qquad Var(X_k) = \sigma^2 < \infty$
- $S_n = X_1 + \cdots + X_n$
- $\Rightarrow \quad E\{S_n\} = n\mu, \quad Var(S_n) = n\sigma^2.$
- $\Rightarrow \quad (S_n - n\mu)/(\sqrt{n}\sigma)$ is standardised.

$$\boxed{P\left\{a < \frac{S_n - n\mu}{\sqrt{n}\sigma} \leq b\right\} \approx \Phi(b) - \Phi(a)}$$

PROOF: Not today!

Figure 3. Distribution of the sum of n die rolls for $n = 1, 2, 4, 8, 16, 32$.

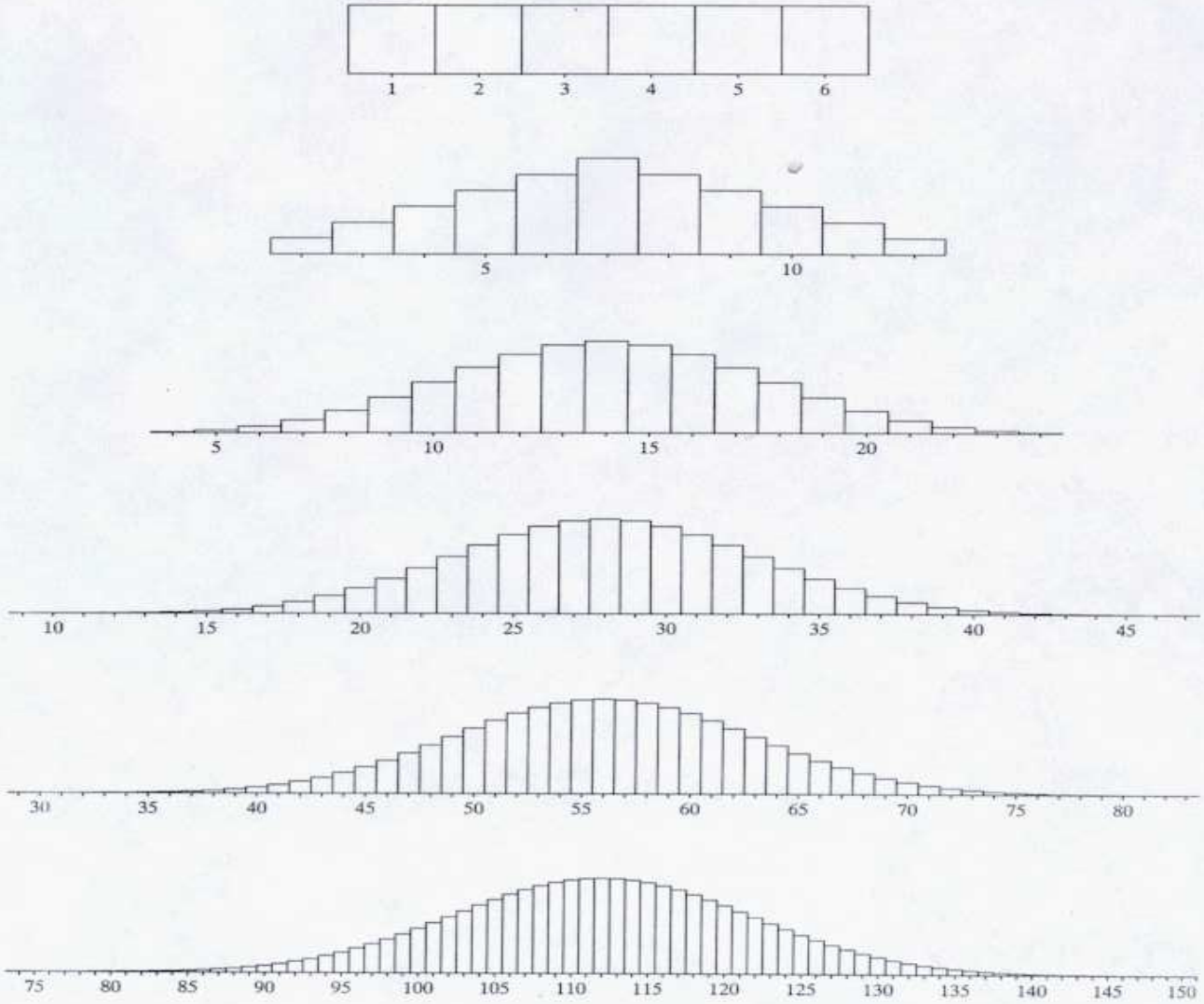


Figure 4. Distribution of S_n for $n = 1, 2, 4, 8, 16, 32$.

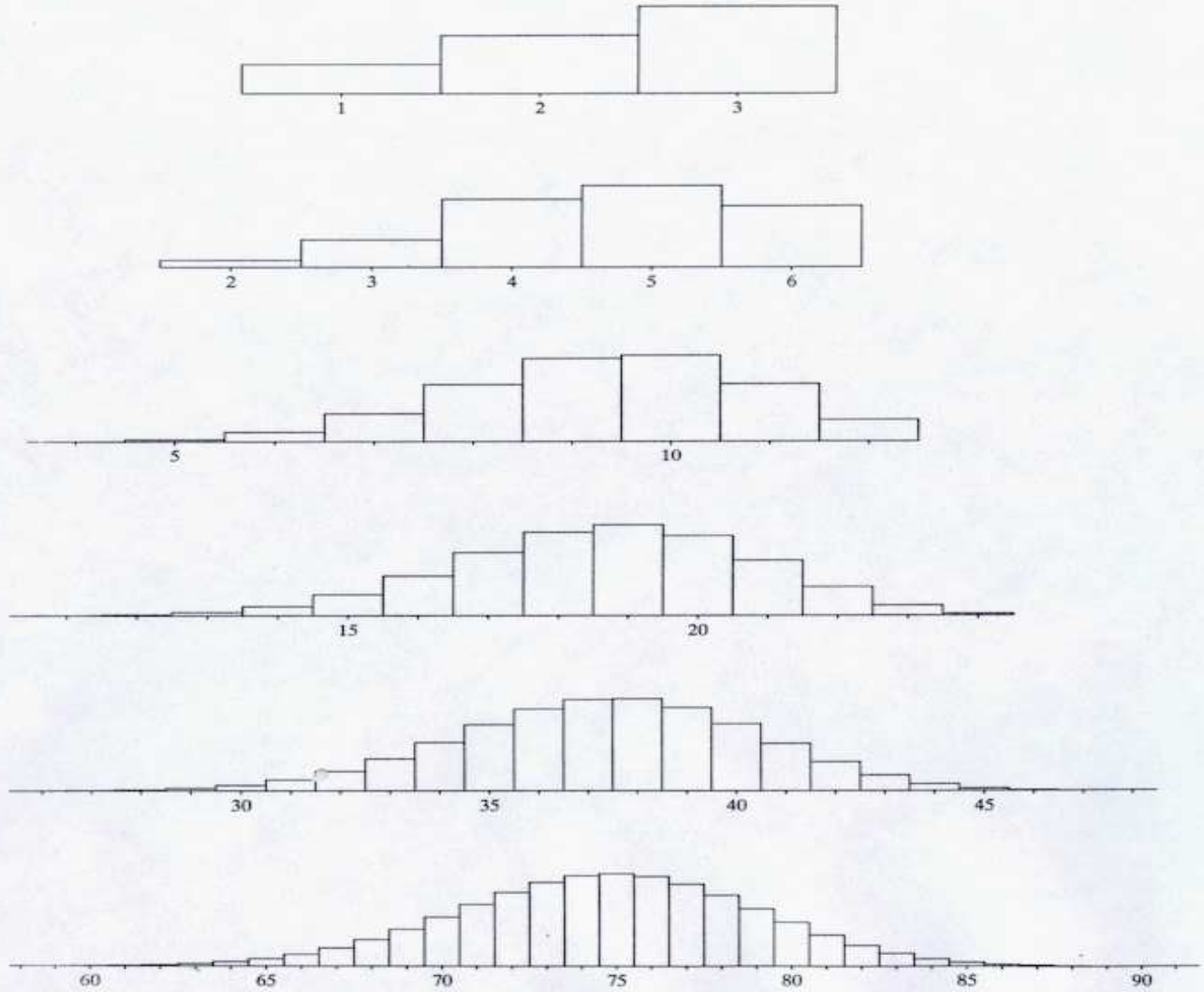


Figure 5. Distribution of S_n for $n = 1, 2, 4, 8, 16, 32$.

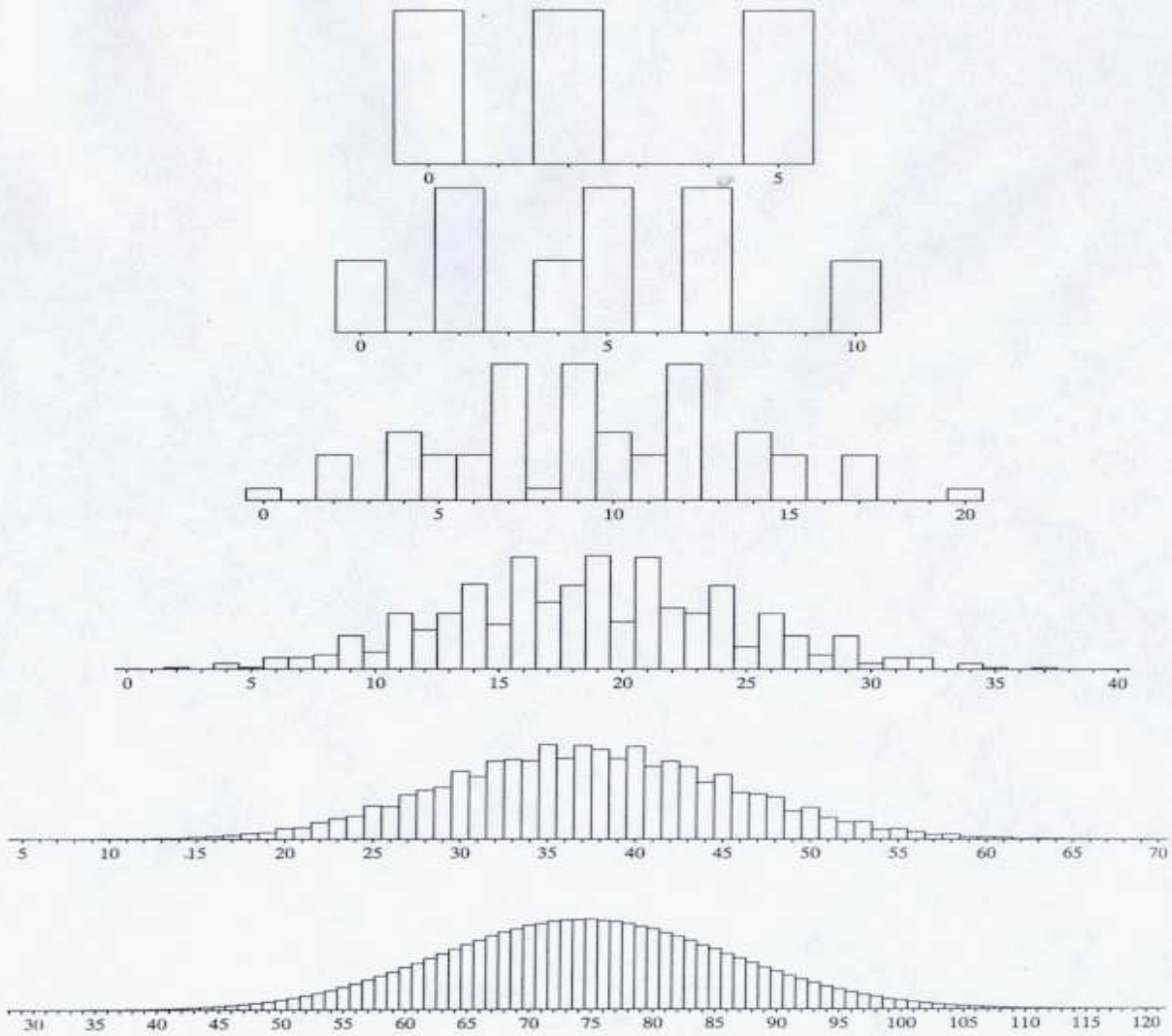
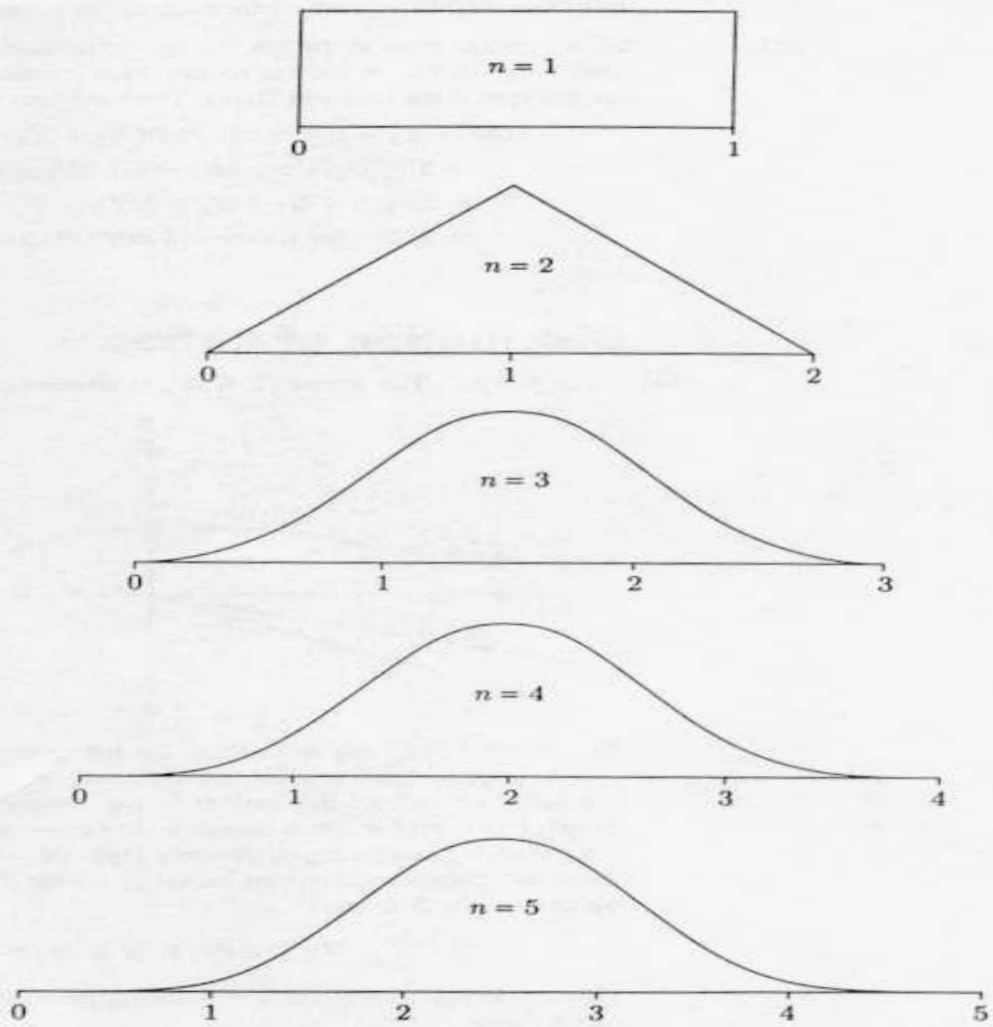


FIGURE 4. Density of the sum of n independent uniform $(0, 1)$ variables. The graphs are all centered at the mean with a constant horizontal distance on the page representing one standard unit in each graph. This shows how rapidly the shape of the distribution becomes normal as n increases.



Binomial limit theorem

- $X_1, X_2, \ldots, X_n$ independent Bernoulli random variables.
- $E\{X_k\} = p, \qquad Var(X_k) = pq$
- $S_n = X_1 + \cdots + X_n$
- $\Rightarrow \quad S_n \sim \text{Bin}(n, p)$.
- $\Rightarrow \quad (S_n - n\mu)/(\sqrt{n}\sigma)$ is standardised.

$$\boxed{P\left\{a \;\leq\; \frac{S_n - np}{\sqrt{npq}} \;\leq b\right\} \;\approx\; \Phi(b) - \Phi(a)}$$

PROOF: Corollary of the general theorem or see Pitman.

A better Binomial approximation

- The "standard one":

$$P\left\{a \;\leq\; \frac{S_n - np}{\sqrt{npq}} \;\leq b\right\} \;\approx\; \Phi(b) - \Phi(a)$$

$$\Longleftrightarrow \quad P\left\{a \;\leq\; S_n \;\leq b\right\} \;\approx\; \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right)$$

- The better approximation

$$P\left\{a \;\leq\; S_n \;\leq b\right\} \;\approx\; \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{npq}}\right)$$

Example 1: An unbiased coin is tossed 100 times. N is the number of heads seen.

$$
\begin{aligned}
P\{N = 50\} &\approx \Phi\left((50 + \frac{1}{2} - 50)/5\right) - \Phi\left((50 - \frac{1}{2} - 50)/5\right) \\
&= \Phi(0.1) - \Phi(-0.1) \\
&= 2\Phi(0.1) - 1 \\
&= 2 \times 0.5398 - 1 \\
&= 0.0796 \qquad \text{Exact} = 0.0795892
\end{aligned}
$$

$$
\begin{aligned}
P\{45 \le N \le 55\} &\approx \Phi\left((55 + \frac{1}{2} - 50)/5\right) - \Phi\left((45 - \frac{1}{2} - 50)/5\right) \\
&= \Phi(1.1) - \Phi(-1.1) \\
&= 2\Phi(1.1) - 1 \\
&= 2 \times 0.8643 - 1 \\
&= 0.7268 \qquad \text{Exact} = 0.728747
\end{aligned}
$$

Example 2:

A sample of 100 students is taken. If the population average of heights is 176cm, and population standard deviation of heights is 6cm, what is the probability that the average height (in the sample) will be over 178cm?

- $\mu = 176, \quad \sigma = 6$
- $\bar{x}_n = S_n/n$

$$
\begin{aligned}
P\{\bar{x}_{100} > 178\} &= P\left\{\frac{1}{100}S_{100} > 178\right\} \\
&= P\left\{\frac{S_{100} - 100 \times 176}{\sqrt{1006}} > \frac{100 \times 178 - 100 \times 176}{\sqrt{100} \times 6}\right\} \\
&\approx P\{Z > 3.3333\} \\
&= 1 - \Phi(3.3333) \;=\; 1 \;-\; 0.9996 \;=\; .0004
\end{aligned}
$$

Example 3:

As for Example 2. i.e. A sample of 100 students is taken from a population and population standard deviation of heights is 6cm. BUT

- $\bar{x}_{100}$ turned out to be 178.
- How likely is it that the population mean really is 176?
- 0.0004

The CLT and the LLN

- Law of large numbers (LLN) says that $\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ converges to the common mean μ.

- The central limit theorem (CLT) says that $(S_n - n\mu)/(\sigma\sqrt{n})$ is approximately standard normal.

- BUT

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\frac{1}{n}(S_n - n\mu)}{\frac{1}{n}(\sigma\sqrt{n})} = \frac{\bar{X}_n - \mu}{\sigma\sqrt{\frac{1}{n}}}$$

- $\Rightarrow \ (\bar{X}_n - \mu) \approx N(0, \sigma^2/n)$

- $\Rightarrow \ |\bar{X}_n - \mu|$ is of order $1/\sqrt{n}$