MTH 224, Spring 2024

Instructor: Bahman Angoshtari

Lecture 18

Section 5.6: Normal approximation of the binomial distribution.

18.1. The normal distribution (continued)

- So far, we have defined standard normal distribution. Next, we define general normal distribution. We need the concept of the z-score of a random variable.

Definition 18.1. Let X be a random variable with mean $\mu = \mathbb{E}[X]$ and standard variation $\sigma = \sqrt{\mathrm{Var}(X)}$. Then, the random variable Z defined by $Z = \frac{X-\mu}{\sigma}$ is called the z-score of X.

- For any r.v. X (not necessarily normal), $\mathbb{E}[Z] = \frac{\mathbb{E}[X-\mu]}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$ and $\mathrm{Var}(Z) = \frac{1}{\sigma^2}\mathrm{Var}(X) = 1$.
- We say that a random variable is normally distributed with mean μ and variance σ^2 (notation $X \sim N(\mu, \sigma^2)$) if its z-score $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
- In other words, $X \sim N(\mu, \sigma^2)$ if $X = \mu + \sigma Z$ for some $Z \sim N(0,1)$. Let us calculate the pdf of X by first calculating its cdf and then differentiating it.
- The cdf of X is $F(x) = \mathbb{P}(X \le x) = P(\mu + \sigma Z \le x) = \mathbb{P}\left(Z \le \frac{x-\mu}{\sigma}\right) = \phi\left(\frac{x-\mu}{\sigma}\right)$.
- By the chain rule, the pdf of X is:

$$f(x) = F'(x) = \phi'\left(\frac{x-\mu}{\sigma}\right) \cdot \frac{1}{\sigma} = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

- Note also that the expected value and variance of X are:

$$\mathbb{E}[X] = \mathbb{E}(\mu + \sigma Z) = \mu + \sigma \underset{=0}{\underset{\downarrow}{\mathbb{E}[Z]}} = \mu$$

$$\mathrm{Var}(X) = \mathrm{Var}(\mu + \sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2.$$

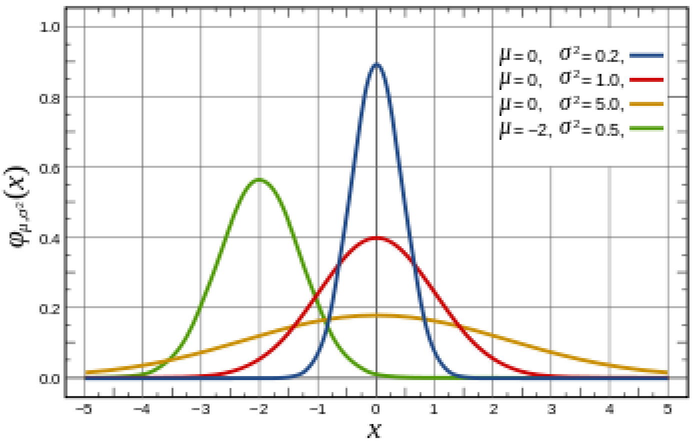
- This leads us to the following definition:

Definition 18.2. A r.v. X has normal distribution with mean μ and variance σ^2 (denoted by $X \sim \mathrm{N}(\mu, \sigma^2)$) if X has the pdf:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}.$$

We have $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$, and $F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \phi\left(\frac{x-\mu}{\sigma}\right)$, in which $\phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \mathrm{d}z$ is the standard normal cdf.

- Different values of μ correspond to horizontal shifts of the density function, and different values of σ make the density's graph narrower/broader (see figure below).



- The following result is an important property of the normal distribution. In short, sum of independent normally distributed random variables is also a normally distributed random variable.

THEOREM 18.3. *Assume that* $X_1, \ldots, X_n$ *are independent and* $X_i \sim \mathrm{N}\left(\mu_i, \sigma_i^2\right)$ *for* $i = 1, \ldots, n$. *Define* $S = X_1 + \cdots + X_n$. *Then,* $S \sim \mathrm{N}\left(\mu_1 + \cdots + \mu_n, \sigma_1^2 + \cdots + \sigma_n^2\right)$.

PROOF. Outside the scope of the class. $\square$

EXAMPLE 18.4. The height (in inches) of a random student is: male $\sim \mathrm{N}(70, 6.76)$, female $\sim \mathrm{N}(64, 8.41)$. In a class of 32 with 15 M and 17 F, what is the probability that the average height exceeds 68 inches?

SOLUTION. We assume all heights are independent. For $j = 1, \ldots, 17$, $X_j \sim \mathrm{N}(64, 8.41)$, and for $i = 1, \ldots, 15$, $Y_i \sim \mathrm{N}(70, 6.76)$. Then

$$\begin{aligned} H &= \frac{1}{32}\left(X_1 + \cdots + X_{17} + Y_1 + \cdots + Y_{15}\right) \sim N\left(\frac{64 \cdot 17 + 70 \cdot 15}{32}, \frac{8.41 \cdot 17 + 6.76 \cdot 15}{32^2}\right) \\ &= N(66.8, 0.239). \end{aligned}$$

Therefore, $\mathbb{P}(H \geq 68) = \mathbb{P}\left(\frac{H - 66.8}{\sqrt{0.239}} \geq \frac{68 - 66.8}{\sqrt{0.239}}\right) = \mathbb{P}(Z \geq 2.45) \approx 0.007$.

18.2. Normal approximation of the binomial distribution

The normal approximation to the binomial distribution states that for large n and a fixed p, we have that $\operatorname{Bin}(n, p) \approx N\left(\mu = np, \sigma^2 = np(1-p)\right)$. In other words, if $X \sim \operatorname{Bin}(n, p)$ then its z-score, $Z = \frac{X - np}{\sqrt{np(1-p)}}$, is approximately a $\mathrm{N}(0, 1)$ random variable.

THEOREM. *($p = \frac{1}{2}$; De Moivre, 1733, general p; Laplace, 1812) Let* $X \sim \operatorname{Bin}(n, p)$ *and define* $Z = \frac{X - np}{\sqrt{np(1-p)}}$.
Then for every t, $\lim_{n \to +\infty} \mathbb{P}(Z \leq t) = \phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$. *In particular,*

$$\lim_{n \to +\infty} \mathbb{P}\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) = \phi(b) - \phi(a) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-x^2/2} dx.$$

EXERCISE 18.5. We flip a fair coin 100 times. Estimate the probability that the # of heads is between 40 and 60.

SOLUTION. $X \sim \text{Bin}\left(100, \frac{1}{2}\right)$, $np = 50$, $\sqrt{np(1-p)} = 5$. Therefore

$$\mathbb{P}(40 \leq X \leq 60) = \mathbb{P}\left(\frac{40-50}{5} \leq \frac{X-50}{5} \leq \frac{60-50}{5}\right) \approx \phi(2) - \phi(-2) = 2\phi(2) - 1 \approx 0.954.$$

Exact value for $X \sim \text{Bin}\left(100, \frac{1}{2}\right)$ is 96.5%.

EXERCISE 18.6. 51% of the newborn children are boys. In a certain community, more girls than boys were born in 2011. The total number of children born was 1000. How likely was this event?

SOLUTION. We have $X \sim \#\text{ boys} \sim \text{Bin}(1000, 0.51)$,

$$\begin{aligned}
&\mathbb{P}(\#\text{ of girls} > \#\text{ of boys}) = \mathbb{P}(X < 500) \\
&= \mathbb{P}\left(\frac{X-510}{\sqrt{510 \cdot 0.49}} \leq \frac{500-510}{\sqrt{510 \cdot 0.49}}\right) \\
&\approx \mathbb{P}(Z < -0.63) = \phi(-0.63) = 1 - \phi(0.63) \approx 0.264.
\end{aligned}$$

So, there is about 26.4% chance (the exact number is 25.3%) that more girls are born.