

# Recap

- (Ch 5) Useful probability distributions
  - Bernoulli
  - Geometric
  - Binomial and multinomial

# Today

- (Ch 5) Useful probability distributions
  - Poisson
  - Continuous uniform
  - Exponential
  - Normal distribution and binomial approximation

# Binomial distribution

- Examples
  - If we roll a six-sided die  $N$  times, how many sixes will we see?
  - If I attempt  $N$  free throws, how many points will I score?
  - What is the sum of  $N$  independent and identically distributed Bernoulli trials?
- A discrete random variable  $X$  is **binomial** if

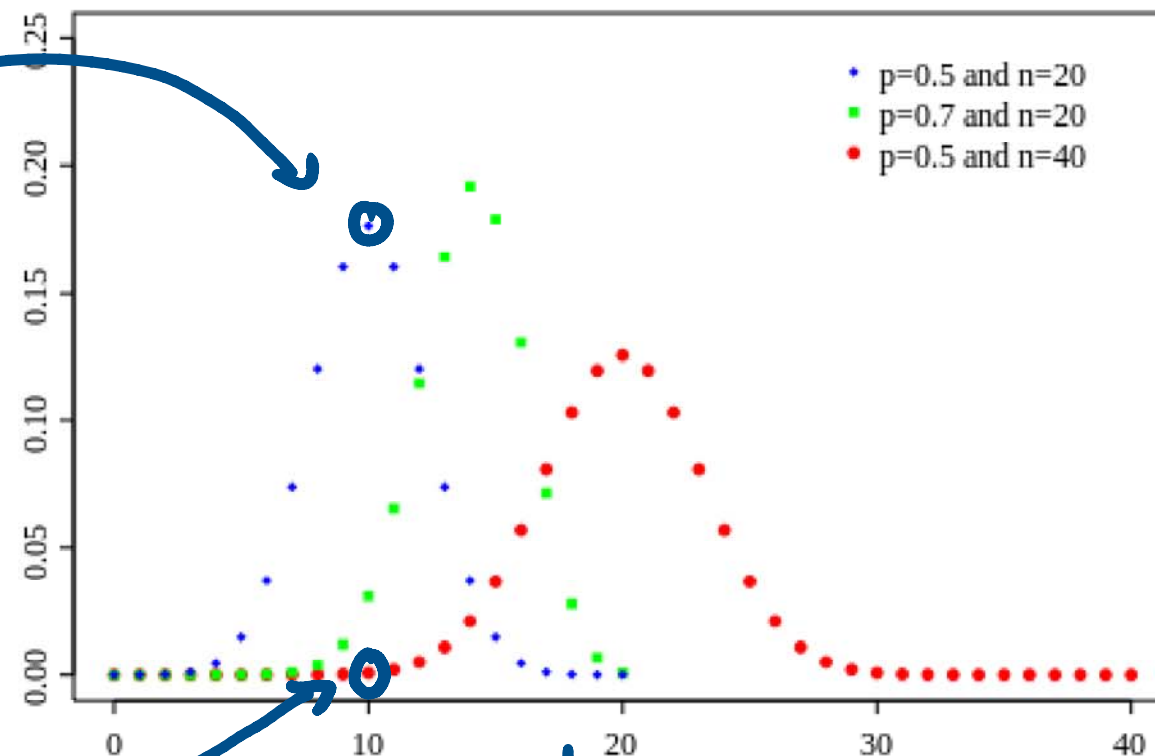
$$P(X = k) = \binom{N}{k} p^k (1 - p)^{N-k} \quad \text{for integer } 0 \leq k \leq N$$

$$\text{with } E[X] = Np \quad \text{and} \quad \text{var}[X] = Np(1 - p)$$

# Binomial distribution: coin example

- In  $N=20$  fair coin tosses, what is the probability of  $k=10$  heads?  
*p=0.5*

- In  $N=40$  fair coin tosses, what is the probability of  $k=10$  heads?



Source: Wikipedia

# Poisson distribution

- Examples

- How many calls does a call center receive in an hour?
- How many mutations occur per 100,000 nucleotides in a strand of DNA?
- How many **independent** incidents occur in an interval?

*average rate of  
↓ incidents*

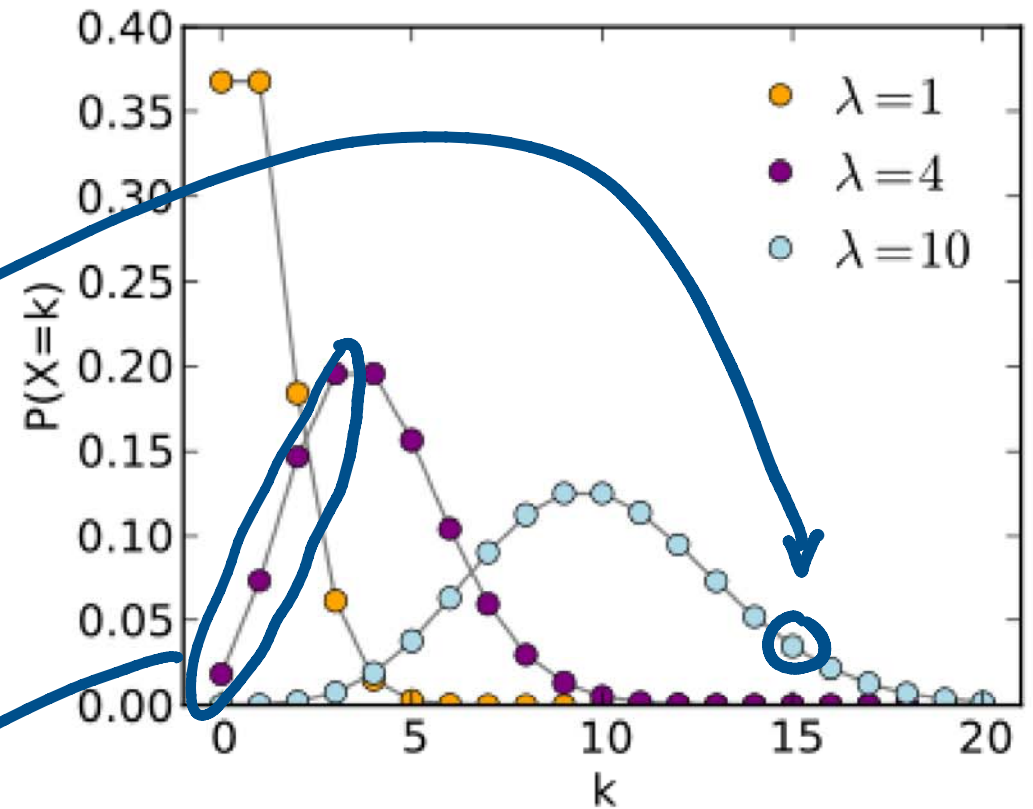
- A discrete random variable  $X$  is **Poisson** with intensity  $\lambda$  if

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{for integer } k \geq 0$$

with  $E[X] = \lambda$  and  $\text{var}[X] = \lambda$

# Poisson distribution: call center example

- If a call center receives  $\lambda = 10$  calls per hour on average, what is the probability that it receives  $k = 15$  calls in a given hour?
- If a call center receives  $\lambda = 4$  calls per hour on average, what is the probability that it receives less than 4 calls in a given hour?



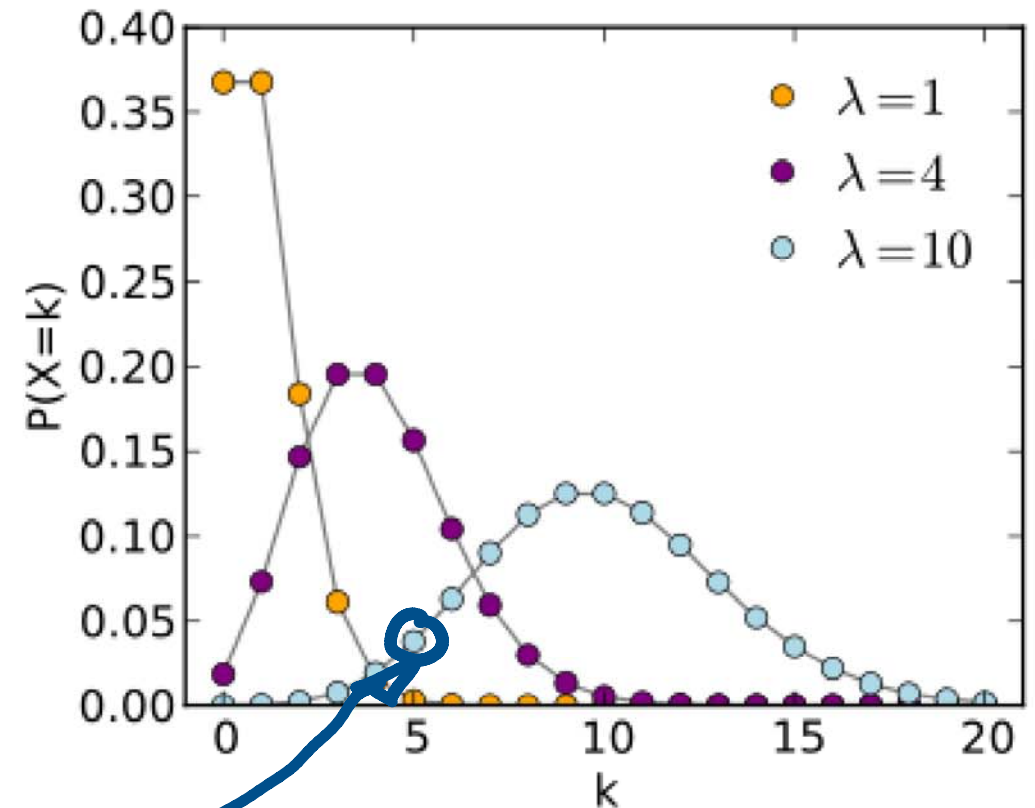
Source: Wikipedia

sum

# Poisson distribution: DNA mutation example

- If we scale the interval, then the intensity of the distribution scales by the same amount
- Suppose DNA mutations occur at an average rate of 1 per 100,000 nucleotides. What is the probability of 5 mutations in 1,000,000 nucleotides?

$$\lambda = 10$$



Source: Wikipedia

# Review: continuous random variables

- A continuous random variable is described by its probability density function  $p(x)$ , so that

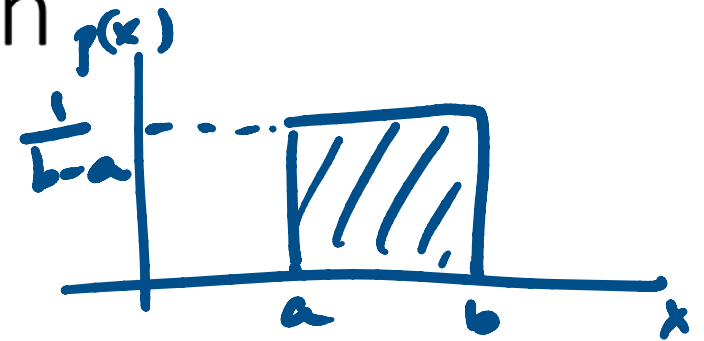
$$\int_a^b p(x) dx = P(X \in [a, b])$$

- Expected value of a continuous random variable  $X$

$$E[X] = \int_{-\infty}^{\infty} xp(x) dx$$

# Continuous uniform distribution

- A continuous random variable  $X$  is **uniform** if



$$p(x) = \frac{1}{b-a} \text{ for } x \text{ between constants } a \text{ and } b$$

$$\text{with } E[X] = \frac{a+b}{2} \text{ and } \text{var}[X] = \frac{(b-a)^2}{12}$$

- Examples

- Napoleon's height given that it is 62.5 inches rounded up to nearest 0.5 inch
- Orientation of a fidget spinner

$$h \in (62, 62.5]$$

$$\theta \in [0^\circ, 120^\circ)$$





# Exponential distribution

- A continuous random variable  $X$  is **exponential** if it represents the “time” until the next incident in a Poisson process with intensity  $\lambda$

$$p(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

$$\text{with } E[X] = \frac{1}{\lambda} \quad \text{and} \quad \text{var}[X] = \frac{1}{\lambda^2}$$

- Examples
  - How long until the next call to be received by a call center?
  - How many nucleotides in a strand of DNA until the next mutation?

# Normal (or Gaussian) distribution

- A continuous random variable  $X$  is **normal** if

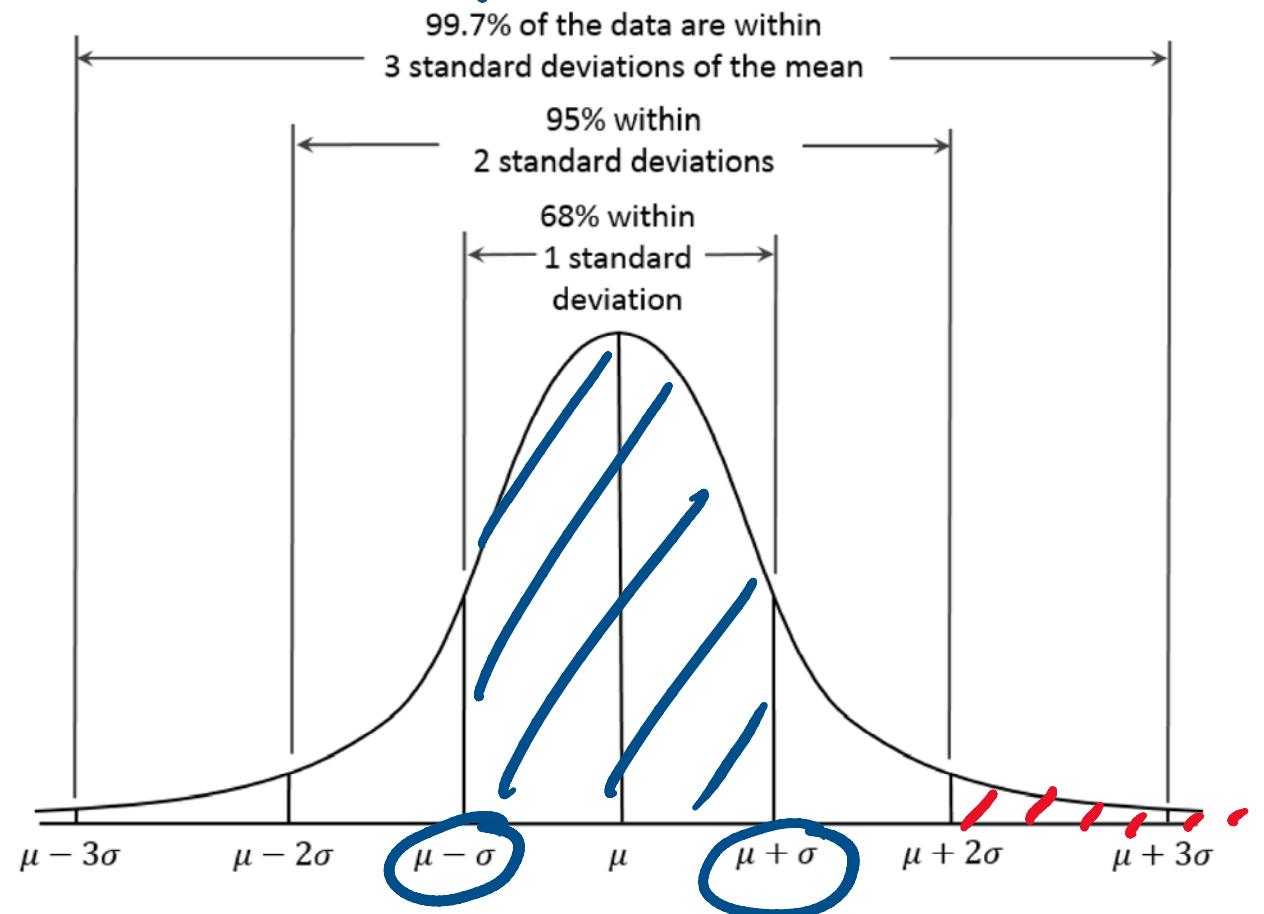
$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

with  $E[X] = \mu$  and  $\text{var}[X] = \sigma^2$  *std[x] =  $\sigma$*

- Lots of data in nature are *approx.* normally distributed
  - Height of adults
  - IQ and other test scores

# Spread of normally distributed data

*approx 99% in textbook*



Source: Wikipedia

# Standard normal distribution

- If we standardize a normally distributed random variable (by subtracting  $\mu$  and dividing by  $\sigma$ ), we get a random variable with standard normal distribution.
- A continuous random variable  $X$  is **standard normal** if

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

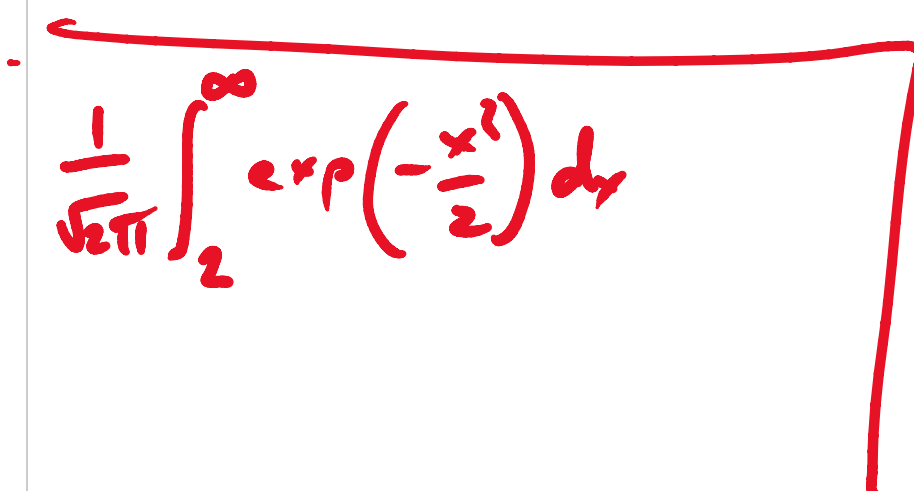
with  $E[X] = 0$  and  $\text{var}[X] = 1$

# Another way to look at spread of normal data

- Fraction of **normal** data within 1 standard deviation of the mean

$$\frac{1}{\sqrt{2\pi}} \int_{-1}^1 \exp\left(-\frac{x^2}{2}\right) dx \cong 0.68$$

- Fraction of **normal** data within  $b$  standard deviations of the mean


$$\frac{1}{\sqrt{2\pi}} \int_{-b}^b \exp\left(-\frac{x^2}{2}\right) dx$$

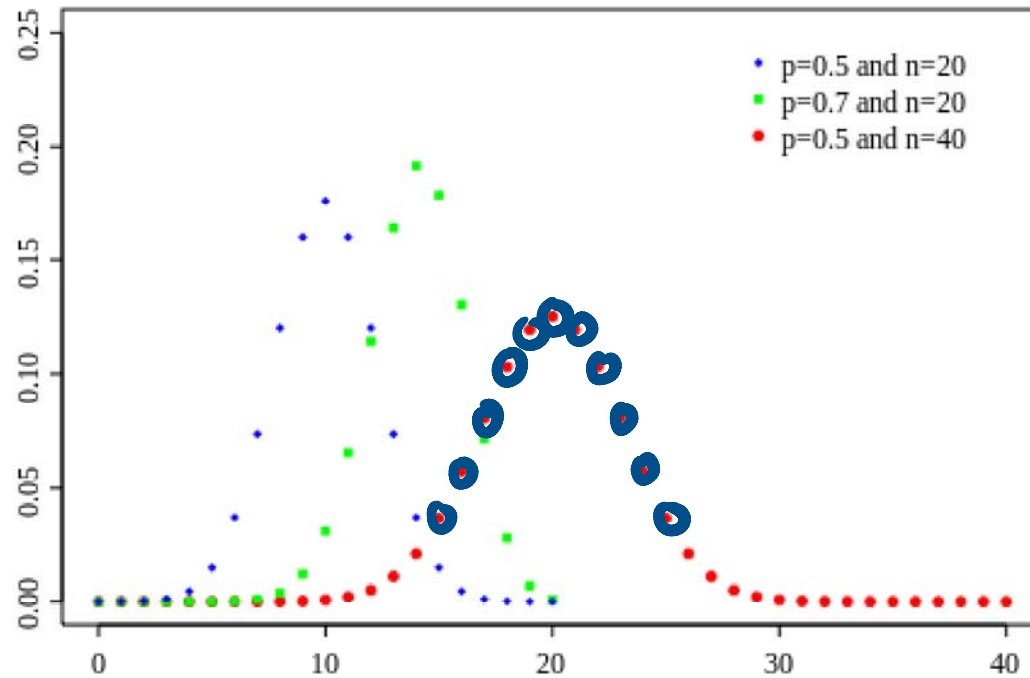
$$\frac{1}{\sqrt{2\pi}} \int_{-b}^b \exp\left(-\frac{x^2}{2}\right) dx$$

# Central limit theorem (CLT)

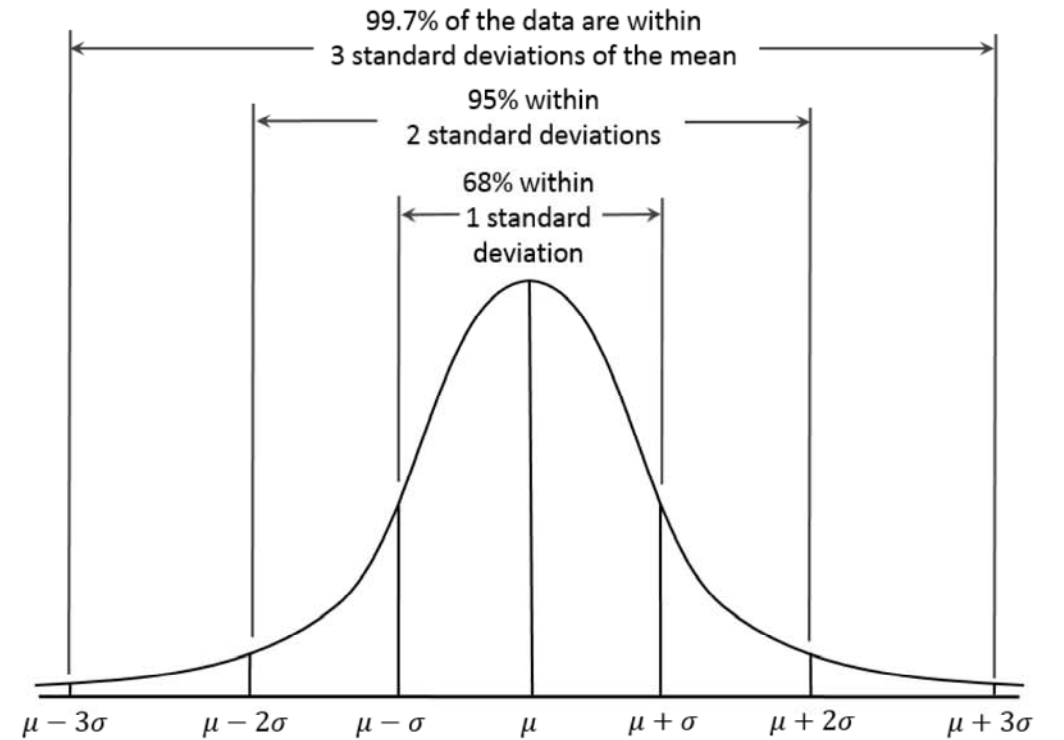
- The distribution of the **sum** of  $N$  independent and identically distributed (IID) random variables tends towards a normal distribution as  $N \rightarrow \infty$
- Even when the component random variables are not exactly IID, the result is approximately true and very useful in practice
- CLT helps explain the prevalence of normal distributions in nature
- A binomial random variable is the sum of IID Bernoulli random variables!

# Binomial approximation

## Binomial distributions



## Normal distribution



Exact:  $\binom{40}{15} 0.5^{15} 0.5^{25} + \binom{40}{16} 0.5^{16} 0.5^{24} + \dots + \binom{40}{25} 0.5^{25} 0.5^{15} \approx 0.92$

## Binomial approximation: coin example

- Let  $H$  be the number of heads observed in 40 tosses of a fair coin  $N$
- Estimate  $P(15 \leq H \leq 25)$  using a normal approximation  $p=0.5$

$$E[H] = Np = (40)(0.5) = 20 \quad \text{std}[H] = \sqrt{Np(1-p)} = \sqrt{10} \approx 3.16$$

$$P(15 \leq H \leq 25) = \frac{1}{\sqrt{2\pi}\sigma} \int_{15}^{25} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

*unnecessary*

$$= P\left(H \text{ is within } \boxed{\frac{5}{3.16}} \text{ std. deviations of the mean}\right)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\frac{5}{3.16}}^{\frac{5}{3.16}} \exp\left(-\frac{x^2}{2}\right) dx \approx 0.89$$