AMAT 362 - PROBABILITY for STATISTICS

LECTURE 13: TAIL PROBABILITIES

TOPICS FOR TODAY
- Normal Distributions: 68-95-99.7 Rule
- Z-values
- Chebyshev's Inequality & Law of Large Numbers

Motivating Question

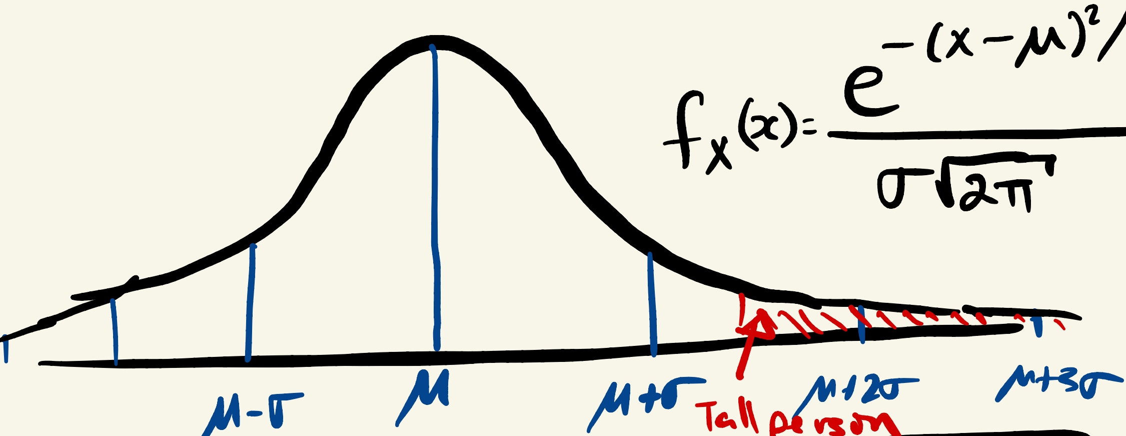
"Why are we impressed by tall people?"

THEY'RE RARE

↳ Height like many things follows a

NORMAL DISTRIBUTION

aka Gaussian or a "Bell Curve"



$$f_X(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} = \varphi_{\mu,\sigma}(x)$$

μ = mean of our Normally distributed RV

σ = std. dev $= \sqrt{\sigma^2}$

Rule for computing probabilities when X is a continuous R.V.

$$P(a<X<b) = \int_a^b \underbrace{f_X(x)}_{\text{Probability Density Function}}\,dx$$

Unfortunately for Normally distributed RVs, like height, this integral is impossible to compute exactly!

68 - 95 - 99.7 RULE

For Normally distributed X we have

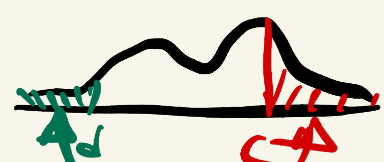
$\mu = E(X)$
$\sigma = SD(X)$

- $P(\mu - \sigma < X < \mu + \sigma) \approx 68\%$
- $P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 95\%$
- $P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 99.7\%$

"68% of the time X will land w/in 1 std dev of μ"

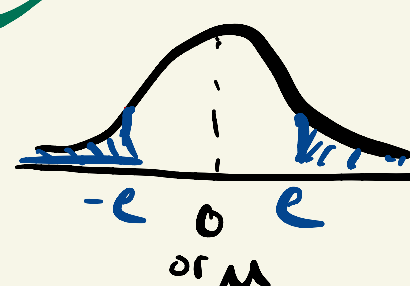
TAIL PROBABILITIES (For any RV X)

$P(X > c)$ = upper tail probability



$P(X < d)$ = lower tail probability

$P(|X| > e)$ = Two-tail probability



N.B. Usually we talk about

$P(|X - \mu| > e)$

68-95-99.7 Rule Rephrased

For a normally distributed R.V. X

- $P(|X-\mu|>\sigma) \approx 1-\underbrace{P(|X-\mu|\le\sigma)}_{68\%} \approx 32\%$
- $P(|X-\mu|>2\sigma) \approx 5\%$
- $P(|X-\mu|>3\sigma) \approx .3\%$

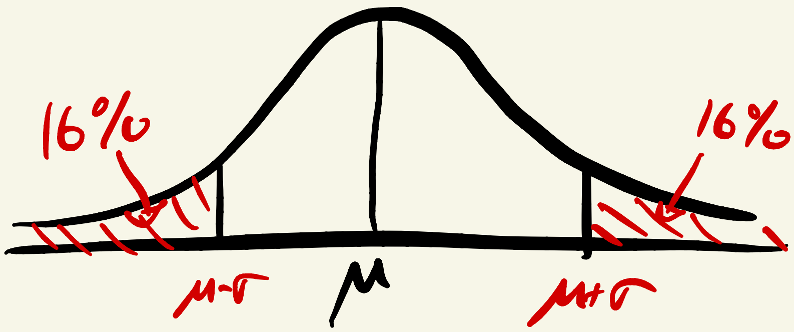
$\Rightarrow$

Upper Tail version

$P(X-\mu>\sigma)\approx 16\%$

$P(X-\mu>2\sigma)\approx 2.5\%$

$P(X-\mu>3\sigma)\approx .15\%$



The MOTIVATING QUESTION

For U.S. Men (>20 yo) $\mu = 69.1''$ $\sigma = 2.9''$

68-95-99.7 Rule $\sim 5'9''$ $\sim 3''$

$\Longrightarrow$ A 6' guy in the U.S. is in upper 16% of all men in the U.S.

$\rightarrow$ A 6'3" $\rightsquigarrow$ event 2.5% = prob

A 6'6" $\longrightarrow$ event w/ prob .15%

$\Longrightarrow$ Any NBA player is in upper tenth of a percent

Z-Values / Standardization

Given an observation x produced from an experiment / RV X

Compute $z = \frac{x-\mu}{\sigma}$ = measures # of std deviations above/below mean (+) (−)

$Z = \frac{X - \mu_X}{\sigma_X}$ is the standardization of X

<u>Ex</u> A 5' tall woman in the U.S.

$\mu = 63.7'' \to 5'3.7''$

$\sigma = 2.5''$

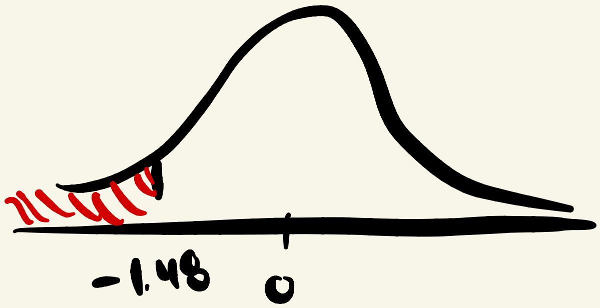
Corresponding z-value $\frac{60 - 63.7}{2.5}$

$z = -1.48$

<u>Q</u>: P(Woman ≤ 5' tall)

$P(Z \leq -1.48) \rightsquigarrow .0694$

Z-table lookup — probability ~7%



Standard Normal Distribution $\mu = 0$ $\sigma = 1$
Table of integrals

Limitation here is that "rareness" has only been quantified for events governed by a normal distribution.

FOR GENERAL (not just normal) RVs

WE CAN QUANTIFY "RARENESS" as follows

Markov's Inequality: For a RV X whose values are non-neg.

"Upper tail is bounded by expectation"

$$\boxed{P(X \geq c) \leq \frac{E(X)}{c}}$$

Ex: Store gets, on average, 1000 customers a week.

Q: What's the probability of getting 1400 or more customers a week?

$$P(X \geq 1400) \leq \frac{1000}{1400} \approx 71.4\%$$

CHEBYSHEV'S INEQUALITY

For ANY RV X

$$\boxed{P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}}$$

Ex: Same store as above $\mu = 1000$ customers $\sigma^2 = 200$ customers

$\sigma \approx 14$

$$P(X \geq 1400) \leq P(|X - 1000| \geq 400) \leq \frac{200}{400^2} = \frac{1}{800} !$$

Another Q: # customers in a week

$$P(950 < \overset{\#\text{customers in a week}}{\overset{\shortparallel}{X}} < 1050) = P(|X - 1000| < 50)$$

$$= 1 - P(|X - 1000| \geq 50)$$

$$P(|X - 1000| \geq 50) \leq \frac{200}{50^2}$$

$$.08$$

$$\Longrightarrow \quad P(950 < X < 1050) \geq 1 - .08 = 92\%$$

LAW OF AVERAGES / LAW OF LARGE NUMBERS

"For large samples or # of tries the empirical mean $(\bar{x})$ will approximate the true mean (μ) almost certainly"

Thm If $X_1, X_2, \ldots, X_n$ are IID RVs w/ $E(X_i) = \mu$, $SD(X_i) = \sigma$ then the empirical mean/average RV $A_n = \frac{X_1 + \cdots + X_n}{n}$ will satisfy for any $\varepsilon > 0$

$$P(|A_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Proof Apply Chebyshev $\quad Var(A_n) = Var\left(\frac{X_1 + \cdots + X_n}{n}\right)$

$$= n\sigma^2 / n^2$$

Ex

Q: How many times do I need to roll a fair 6-sided die to be w/in $1/2$ of $\mu = 7/2$ $\geq 95\%$ of the time?

A: $$P(|A_n - \mu| < \tfrac{1}{2}) \geq 95\%$$

equiv $$\underbrace{1 - P(|A_n - \mu| < 1/2)}_{P(|A_n - \mu| \geq 1/2)} \leq \underbrace{1 - .95}_{.05}$$

Apply Law of Large Numbers on $\sigma^2 = \frac{35}{12}$

$$P(|A_n - \mu| \geq 1/2) \leq \frac{\sigma^2}{n(1/2)^2} \leq .05$$

$$\Rightarrow \quad n \geq 234$$