B) $TTT = \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \frac{8}{27}$

1 $HTT = \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{27} \cdot 3 = \frac{12}{27}$

2 $HHT = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{27} \cdot 3 = \frac{6}{27}$

3 $HHH = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{27}$

$\underbrace{8+12}_{20} + \underbrace{6+1}_{7} = 27 \quad \frac{27}{27} = 1$

Midterm
Advanced Statistics.

Zhetessov
Nur

(P1). You have a coin whose prob of heads is $P(H) = \frac{1}{3}$. Toss it 3 times. Let X be the number of heads.

a) Find PMF and CDF of X.
b) Find $E(X)$ and $Var(X)$.

a) $n = 3$.
X - number of heads.
$X=0$ if $\Omega_1 = \{TTT\}$
$X=1$ if $\Omega_2 = \{HTT, THT, TTH\}$
$X=2$ if $\Omega_3 = \{HHT, HTH, THH\}$
$X=3$ if $\Omega_4 = \{HHH\}$

PMF - Probability Mass Function will be

X	0	1	2	3
$P(X)$	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$

$P(\Omega) = \frac{8}{27} + \frac{12}{27} + \frac{6}{27} + \frac{1}{27} = \frac{20+7}{27} = \frac{27}{27} = 1$

correct ✓

$\cdot 3 = \frac{12}{27}$

$3 = \frac{6}{27}$

$\frac{27}{27} = 1$ ✓

CDF – Cumulative distribution function will be

$$\cancel{F(x) = \begin{cases} \frac{8}{27} & \text{if } x<0 \\ \frac{20}{27} & \text{if } 0 \le x < 1 \\ \frac{26}{27} & \text{if } 1 \le x < \end{cases}}$$

$$F(x) = \begin{cases} 0 & \text{if } x<0 \\ \frac{8}{27} & \text{if } 0 \le x < 1 \\ \frac{20}{27} & \text{if } 1 \le x < 2 \\ \frac{26}{27} & \text{if } 2 \le x < 3 \\ \frac{27}{27} = 1 & \text{if } x \ge 3 \end{cases}$$

b) $E(X) = 0 \cdot \frac{8}{27} + 1 \cdot \frac{12}{27} + 2 \cdot \frac{6}{27} + 3 \cdot \frac{1}{27} =$

$= 0 + \frac{12}{27} + \frac{12}{27} + \frac{3}{27} = \frac{27}{27} = \underline{1}$

$E(X^2) = 0^2 \cdot \frac{8}{27} + 1^2 \cdot \frac{12}{27} + 2^2 \cdot \frac{6}{27} + 3^2 \cdot \frac{1}{27} =$

$= 0 + \frac{12}{27} + \frac{24}{27} + \frac{9}{27} = \frac{45}{27}$

$Var(X) = E(X^2) - [E(X)]^2 = \frac{45}{27} - (1)^2 =$

$= \frac{45}{27} - \frac{27}{27} = \frac{18}{27}$

$\sigma(X) = \sqrt{Var(X)} = \sqrt{\frac{18}{27}}$

$E(X) = 1, \quad Var(X) = \frac{18}{27} = \frac{2}{3}$

P2. Daily profit of a trader is $\sim N(300, 100)$

$P(X < 200)$ - ?

$P(X < 200) = P\left(\frac{X-\mu}{\sigma} < \frac{200-\mu}{\sigma}\right) =$

$= P\left(\frac{X-\mu}{\sigma} < \frac{200-300}{100}\right) = P(Z < -1) =$

$= \Phi(-1) = \int_{-\infty}^{-1} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx \approx 0,1587$

The probability that a trader makes less than 200 on particular day is 15.87%

P3 Students scores have $\mu = 65$ and ~~...~~ $\sigma = 25$
Class of 100 students taking the exam. Using CLT approximate $P(X > 70)$

$n = 100$ $\mu = 65$ $\sigma = 25$ $P(X \geq 70) - ?$	Our sample size is larger than 30 students, therefore the Central Limit Theorem holds true for our sample. Which means we assume that it is normally distributed.

$$P(\bar{X} \geq 70) = 1 - P(\bar{X} < 70) =$$

$$= 1 - P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{70 - 65}{25/\sqrt{100}}\right) = 1 - P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{5}{2{,}5}\right) =$$

$$= 1 - P(Z < 2) = 1 - \Phi(2) =$$

$$= 1 - \int_{-\infty}^{2} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}} dx \approx 1 - 0.9772 =$$

$$= 0{,}0228$$

The Probability that aver. score exceed 70 is 2.28%

(P4) Let $X_1, \dots, X_n$ be random sample from distribution whose mean μ is known, but standard deviation σ is unknown. Show that the variance estimator

$$Var(X) = \sum_i \frac{(X_i - \mu)^2}{(n-1)} \text{ is biased.}$$

Prop.

$$E[\overline{X}_n] = \mu$$

$$E[(S_n)^2] = \sigma^2$$

$$E[\overline{X}_n] = E\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n}\left[E[X_1] + \dots + E[X_n]\right] =$$

$$= \frac{1}{n} \cdot n \cdot \mu = \mu$$

$$E[(S_n)^2] = E\left[\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2\right]$$

$$(n-1)(S_n)^2 = \sum (X_i - \overline{X}_n)^2$$

$$(n-1)(S_n)^2 = \sum (X_i - \mu + \mu - \overline{X}_n)^2$$

$$(n-1)(S_n)^2 = \sum \left[(X_i - \mu)^2 - 2(\overline{X}_n - \mu) \sum (X_i - \mu) + n(\mu - \overline{X}_n)^2\right]$$

$$(n-1)(S_n)^2 = \sum (X_i - \mu)^2 - 2(\overline{X}_n - \mu) \cdot n \cdot (\overline{X}_n - \mu) + n(\mu - \overline{X}_n)^2$$

$$(n-1)(S_n)^2 = \sum (X_i - \mu)^2 - n(\mu - \overline{X}_n)^2$$

$$E[(n-1)(S_n)^2] = E\left[\sum (X_i - \mu)^2 - n(\mu - \overline{X}_n)^2\right]$$

$$E[(n-1)(S_n)^2] = E\left(\sum (X_i - \mu)^2 - nE[(\mu - \bar{X}_n)^2]\right)$$

$$(n-1)(S_n)^2 = n \cdot \sigma^2$$

$$(n-1)(S_n)^2 \quad n \cdot \sigma^2$$

$$(S_n)^2 = \frac{n \cdot \sigma^2}{n-1}$$

biased

(P5) p- fraction supporting new law
n is chosen to estimate p

1) Assume n = 1000 and 418 support law
Find 95% conf. interval for p.

$n = 1000$

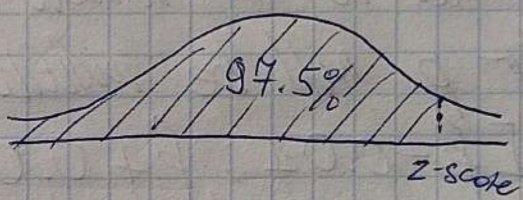
$\hat{p} = \frac{418}{1000} = 0.418$

$C = 95\%$

CI - ?

$\hat{p} \pm z * SE_{\hat{p}}$

$\hat{p} \pm z * \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$



z-score for C-confidence interval is 1.96.

$\hat{p} = 0.418$

$(1-\hat{p}) = 0.582$

$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.418 \cdot 0.582}{1000}} = 0.0156$

Upper bound for CI is:

$0.418 + 1.96 * 0.0156$

0.4486

Lower bound for CI is:

$0.418 - 1.96 * 0.0156$

$0.418 - 0.0306$

0.3874

Confidence interval is $[0.3874 ; 0.4486]$

b) Find n for which 95% conf. interval for p will be within 0.01 error.

$[\hat{p} - 0.01 ; \hat{p} + 0.01]$

Error is a standard error $SE_{\hat{p}}$ calculated by $SE_{\hat{p}} = Z * \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$

$SE_{\hat{p}}$ should be less than 0.01

$SE_{\hat{p}} \leq 0.01$

$Z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.01$

For 95% confidence interval z-score is 1.96

$\hat{p} = 0.418$

$(1-\hat{p}) = 0.582$

n - unknown

$1.96 * \sqrt{\frac{0.418 \cdot 0.582}{n}} \leq 0.01$

let's square all multipliers

$(1.96)^2 * \frac{0.418 \cdot 0.582}{n} \leq (0.01)^2$

Then I take out n, change sign.

$$n \geq \frac{(1.96)^2 \cdot 0.418 \cdot 0.582}{(0.01)^2}$$

$$n \geq \frac{7.523 \cdot 0.418 \cdot 0.582}{(0.01)^2} \geq \frac{1.8318}{10^{-4}}$$

$(0.01)^2 \Rightarrow 1 \cdot 10^{-2} \cdot 10^{-2} \Rightarrow 10^{-4}$

$n \geq 18317,5 \Rightarrow n = 18318$