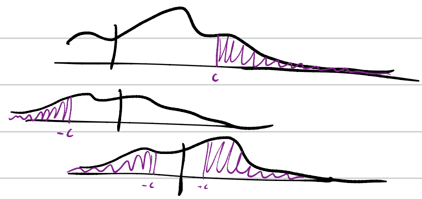
Tail Estimation

$c > 0$ throughout

Tail: $P[X \geq c]$

$P[X \leq -c]$

$P[\,|X-\mu| \geq c\,]$



Markov Inequality

Suppose we know $X \geq 0$. Take any $c > 0$.

$$P[X \geq c] \leq \frac{\mu_X}{c}$$

Why?

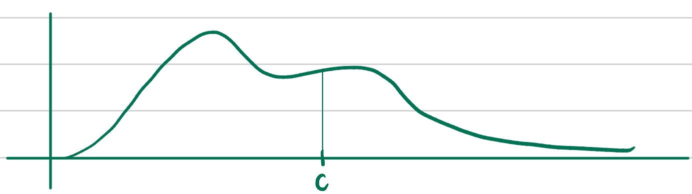
$$\mu_X = \int_0^\infty x f_X \, dx = \int_0^c x f_X \, dx + \int_c^\infty x f_X \, dx$$

$$\geq \int_0^c x f_X \, dx + \int_c^\infty c f_X \, dx$$

$$\geq \int_c^\infty c f_X \, dx$$

$$= c \int_c^\infty f_X \, dx \qquad = c \, P[X \geq c]$$

$X > 0$



Chebyshev Inequality

Any X, any $c > 0$. $\leadsto$ $c^2 > 0$ too.

$(X - \mu_x)^2 \geq 0$ apply Markov to $(X-\mu_x)^2$:

$$P[(X-\mu_x)^2 \geq c^2] \leq \frac{E[(X-\mu_x)^2]}{c^2}$$

Observe: ∘ $(X-\mu_x)^2 \geq c^2 \iff |X-\mu_x| \geq c$

∘ $E[(X-\mu_x)^2] = \sigma_X^2$

So Chebyshev:

$$\boxed{P[\,|X-\mu_x| \geq c\,] \leq \frac{\sigma_x^2}{c^2}}$$

Note of course: $P[X \geq c + \mu_x] \leq P[\,|X-\mu_x| \geq c\,]$

Example: A tire shop has avg. 500 customers/day.

(a) Odds more than 700 customers arrive today.

(b) " " " " " " "
w/ assumption $\sigma^2 = 100$

Solution: (a) $P[X \geq 700] \leq \frac{500}{700} \approx 71\%$

(b) $c = 200$, $P[|X - 500| \geq 200] \leq \frac{100}{200^2} \approx 0.25\%$

"$X \geq 700$" $\subset$ "$|X-500| \geq 200$"
$P[X \geq 700] \leq P[|X-500| \geq 200] \leq 0.25\%$

Sample Mean

$$M_n(X) = \frac{1}{n}(X_1 + \dots + X_n)$$

X_i IID, $X_i \sim X$

"Sample mean"

$\mu_M = \mu_X = E[M_n(X)]$
$\sigma_M = \frac{1}{\sqrt{n}}\sigma_X$

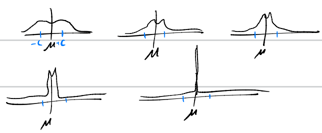
Law of Large Numbers

$$P[|M_n(X) - \mu_M| \geq c] \leq \frac{\sigma_X^2}{n c^2} \xrightarrow{n \to \infty} 0$$

($\frac{\sigma_X^2}{n} = \sigma_M^2$)

$$\Rightarrow \boxed{P[|M_n - \mu_X| \leq c] \xrightarrow[n \to \infty]{} 1}$$

LLN



13-B-02 2500 customers, \$1000 expected paid per customer, $\sigma = \$900$

S = total claims = $\sum X_i$

$\mu_S = E[S] = (2500)(1000) = 2.5 \times 10^6$ Let r = premium

$\sigma_S = 900\sqrt{2500} = 45{,}000$

↙ tot. premiums revenue

Want $P[S \leq 2500r] \geq .999$

(a) $S \sim (45000) Z + 2{,}500{,}000$

$P[(45000) Z + 2500000 \leq 2500r]$

$= P\left[\frac{2500r - 2{,}500{,}000}{45{,}000}\right] \geq .999$

$\rightsquigarrow \Phi^{-1}(.999) = 3.09$

$\rightsquigarrow r = \$1{,}055$

(b) $P[|S - \mu_S| \geq 2500r - \mu_S] \leq \frac{\sigma^2}{(2500r - \mu_S)^2} = .001$

i.e. $S \geq 2500r$

$\iff S - \mu_S \geq 2500r - \mu_S$

$\Rightarrow |S - \mu_S| \geq 2500r - \mu_S$

$r = \$1{,}569$

13-A-02

$$E[X] = 80 = \mu_X$$

$$10 = \sigma_X^2 \rightsquigarrow \frac{10}{n} = \sigma_{M_n(X)}^2$$

$$P\left[|M_n(X) - 80| \leq 5\right] \leq 10\% \quad , \quad 5\%$$

Chebyshev: $\dfrac{\sigma_X^2}{n c^2}$ $\quad c = 5 \ , \ \sigma_X^2 = 10$

$$\frac{10}{n 5^2} = .1 \rightsquigarrow n = 4$$

$$\frac{10}{n 5^2} = .05 \rightsquigarrow n = 8$$

13-A-05 10 Q's .55 chance each Q correct

(a)

X = score of a student = # Q's correct

$\begin{cases} 1 & .55 \\ 0 & .45 \end{cases}$ ⇒ $Var[X] = (.55)(.45)$

$\sim Bin(10, .55)$

$Var[X] = npq \to 10(.55)(.45) \to 2.475$

$Var[M_{15}(X)] = \frac{Var[X]}{15} \to \frac{2.475}{15} \to 0.165$

(b) $E[X] = np = 10(.55) = 5.5 \to E[M_{15}(X)]$

$P[\,|M_{15} - 5.5| > .5\,] \leq \frac{.165}{(.5)^2} = .66$

Meaning of Probability

- Epistemological
  - Classical
  - Evidential

- Subjective
  - Bayesianism

- Physical
  - Frequentism
  - Propensity

See notes online.