$\mu = \mathbb{E}X$ first moment

Last time:

6.1–6.2 Variance $\left\langle \begin{array}{l} \mathbb{E}(X-\mu)^2 \\ \underbrace{\mathbb{E}X^2}_{\text{second moment}} - \mu^2 \end{array}\right.$

6.3 Markov's inequality

$$\mathbb{P}(X \geq c) \leq \frac{\mu}{c} \quad \text{for } x \geq 0$$

"one-sided tail"

Today:

Sec 6.4 Chebyshev's inequality.

$\sigma = SD(X)$

"Can we do better if we also know $\sigma^2 = Var(X)$?"
$= \mathbb{E}(X-\mu)^2$.

$\to \mathbb{P}(|X-\mu| \geq c)$ (for $c>0$)

$= \mathbb{P}(\boxed{(X-\mu)^2} \geq c^2)$

$\leq \dfrac{\sigma^2}{c^2}$

(Treat $(X-\mu)^2$ as the non-negative R.V. in Markov's)

"Chebyshev's inequality."

Reason why we do not work on $|X-\mu|$: $\mathbb{E}|X-\mu|$

The chance that the absolute value of the deviation is large is small

Restatement:

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

"Normalize the data"

① Remove mean / Treat mean as origin

② Measure distances in units of $SD(X) = \sigma$

i.e. say something is +/− k SD's away.

$$\mathbb{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Example.

$X \quad \mu = 60 \quad \sigma = 5.$

X is outside of $(50, 70)$

$$\mathbb{P}(X \notin (50,70)) = \mathbb{P}(|X-60| \geq 10) \leq \frac{5^2}{10^2} = \frac{1}{4}$$

$(-2SD, +2SD)$

$$\mathbb{P}(|X-\mu| \geq 2\sigma) \leq \frac{1}{4}$$

Chebyshev's inequality tells us that no matter how the distr. looks like, the chance that X is $\underset{\text{many}}{k}$ SD's away from its mean (two-sided tail) is $\underset{\text{small}}{\text{no}}$ larger than $\frac{1}{k^2}$

Recall: Markov's inequality tells us that the chance a non-negative R.V. is larger than $\underset{\text{many}}{k}$ times its mean (one-sided tail) is no larger than $\underset{\text{small}}{\frac{1}{k}}$

Question: Can we do one-sided tail using Chebyshev's inequality?

Yes, just discard the other half

Example $X \quad \mu = 60 \quad \sigma = 5$ non-negative try $\mu = 15 \quad \sigma = 10$

$\mathbb{P}(X \geq 70)$

① Markov's. $\mathbb{P}(X \geq 70) \leq \dfrac{60}{70} = \dfrac{6}{7}$

② Chebyshev's: $\mathbb{P}(X \geq 70) \leq \mathbb{P}(X \leq 50 \text{ or } X \geq 70)$
$\leq \frac{1}{4}$ (better / tighter)

Not always

Usually, even if the R.V. is nonnegative and we are considering a single-sided tail, we prefer to use Chebyshev's inequality as long as we know $\sigma = SD$.

Sec 7.1 (Var/SD of) sums of indpt. R.V.s

- Binom(n,p)
- Additivity / linearity of Expectation

$$\mathbb{E}\sum_i X_i = \sum_i \mathbb{E}X_i \quad \text{indpt or not}$$

Not true for Variance!

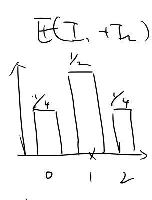
What is the variance of an indpt. / dependent sum?

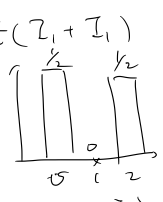
Example. Consider a fair coin toss it twice

$$I_j = \begin{cases} 1 & \text{if the } j^{th} \text{ toss is H} \\ 0 & \text{o/w.} \end{cases} \quad j = 1,2$$

$I_1 + I_2$ twice vs. $I_1 + I_1$ double.

$\mathbb{E}(I_1 + I_2) = \mathbb{E}(I_1 + I_1) = 1$





$Var(I_1 + I_2) \leq Var(I_1 + I_1)$

$= \frac{1}{4}\cdot 1^2 + \frac{1}{2}\cdot 0^2 + \frac{1}{4}\cdot 1^2 \qquad = \frac{1}{2}\cdot 1^2 + \frac{1}{2}\cdot 1^2$

$= \frac{1}{2} \qquad = 1.$

Conclusion: Variance of the sums is dependent on whether they are indpt or not.

It can be shown that $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$ for X_1, X_2 indpt.

Note: $\begin{cases} \mathbb{E}(-X) = -\mathbb{E}X \\ \mathbb{E}(kX) = k\mathbb{E}X \end{cases}$ vs. $\begin{cases} Var(-X) = Var(X) \\ Var(kX) = k^2 Var(X) \end{cases}$

① $Var(X_1 - X_2) = Var(X_1) + Var(X_2)$

② $Var(X_1 + X_2) = Var(X_1) + Var(X_2) \;\not\Rightarrow\; X_1 \,\&\, X_2$ indpt.
$\Leftarrow$

Generally, for indpt R.V's $X_1, X_2, X_3, \ldots X_n$

$$Var(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^n Var(X_i)$$

As an instance, for i.i.d. R.V.'s $X_1, X_2, \ldots, X_n$

$$Var(X_1 + X_2 + \cdots + X_n) = n\,Var(X_1)$$
$$SD(\ldots) = \sqrt{n}\,SD(X_1)$$

Example.

1. Binomial.

$X \sim Binom(n,p)$. $\mu = \mathbb{E}X = np$

$X = \underbrace{I_1 + I_2 + \cdots + I_n}_{n \text{ i.i.d. Bern}(p)}$

$Var(I_1) = \mathbb{E}I_1^2 - (\mathbb{E}I_1)^2$
$= \mathbb{E}I_1 - (\mathbb{E}I_1)^2$
$= p - p^2$

$I_1^2 = I_1$ since I_1 can only be 0 & 1.

$Var(X) = n\,Var(I_1)$
$= n(p - p^2)$
$= np(1-p) = npq \quad (q = 1-p)$

When varying p, npq reaches its maximum at $p = \frac{1}{2}$.
$npq = \frac{1}{4}n$

This means the coin is most variable if it is fair.

2 Poisson

$X \sim Poisson(\mu)$

$Y \sim Binom(n,p)$, $np = \mu$, n large, p small $(q \approx 1)$

By law of small numbers. $X \overset{d}{\approx} Y$

$Var(X) \simeq Var(Y) = npq = np - np^2$ (higher order term.)
$= \mu \cdot q$
$\approx \mu$

$\mu = \mathbb{E}X = Var(X).$

3. Geometric / Negative Binomial

$\begin{cases} T_1 \sim Geom(p) \text{ on } 1,2,3,\ldots & \mathbb{E}T_1 = \frac{1}{p} \\ T_r \sim NegBinom(r,p) & \mathbb{E}T_r = \frac{r}{p}. \end{cases}$

T_r is the sum of r i.i.d. Geom(p)'s.

$Var(T_1) = \dfrac{q}{p^2} \left(= \dfrac{1-p}{p^2}\right)$

$Var(T_r) = \dfrac{rq}{p^2}.$

Sec. 7.2 Var in SRS

$X \sim HyperGeom(N, G, n)$

$X = I_1 + I_2 + \cdots + I_n \qquad \mathbb{E}I_k = \frac{G}{N} (=p) \quad \mathbb{E}X = np$

where I_k is the indicator for the k^{th} sample

$Var(X) = \underbrace{\mathbb{E}X^2}_{\mathbb{E}(X^2)} - \underbrace{(\mathbb{E}X)^2}_{✓}$

"Method of indicators for variance" $\mathbb{E}X^2$

① Expand.

$$\mathbb{E}X^2 = \mathbb{E}\left(\sum_{j=1}^n I_j\right)^2 = \mathbb{E}\sum_{j=1}^n\sum_{k=1}^n I_j I_k = \sum_{j=1}^n\sum_{k=1}^n \mathbb{E}I_j I_k$$

1° $\mathbb{E}I_j^2 = \mathbb{E}I_j = \mathbb{E}I_1 = \frac{G}{N}$

2° $\mathbb{E}I_j I_k \; (j \neq k)$
$= \mathbb{E}I_1 I_2 = \mathbb{P}(I_1 = I_2 = 1)$ by symmetry
$= \frac{G}{N}\cdot\frac{G-1}{N-1}$

② Use symmetry $= \sum_{j=1}^n \mathbb{E}I_j^2 + \sum_{j\neq k}\mathbb{E}I_j I_k$
$= n\,\mathbb{E}I_1 + (n^2 - n)\,\mathbb{E}I_1 I_2$

③ Calculate prob. $= n\cdot\dfrac{G}{N} + (n^2-n)\cdot\dfrac{G(G-1)}{N(N-1)}$

$Var(X) = \quad - (\mathbb{E}X)^2$

$= \quad - \left(n\frac{G}{N}\right)^2.$

$= \ldots$ (read textbook for detail)

$= n\cdot\dfrac{GB}{N^2}\cdot\boxed{\dfrac{N-n}{N-1}} \qquad (B = N-G)$

"Finite population Correction" f.p.c

It is <1 as long as $N < \infty$ and $n > 1$

When N is large and n is small comparing to N. f.p.c ≈ 1

For sampling w/. repl. $Y \sim Binom(n, p = \frac{G}{N})$

$Var(Y) = npq = n\cdot\dfrac{GB}{N^2}.$

Sec. 7.3 Law of large number.
(Law of Average)

We say that we will get about 50 heads in 100 coin tosses

What does this mean?

$X_n \sim Binom(n, \frac{1}{2})$

$\mu_n = \mathbb{E}X_n = \frac{1}{2}n$ This looks fine.

$\sigma_n = SD(X_n) = \sqrt{n}\,SD(X_1) = \sqrt{npq} = \frac{1}{2}\sqrt{n} \nearrow \infty$ as $n \to \infty$

"the sum is more spread out if we toss more".

Consider $A_n = \dfrac{X_n}{n}$:= proportion of H's we get.

$SD(A_n) = \frac{1}{n}SD(X_n) = \dfrac{1}{2\sqrt{n}} \searrow 0$ as $n \to \infty$

"the mean is less spread out if we toss more".

Generally, suppose $X_1, X_2, \ldots, X_n$ are i.i.d R.V.'s w/. mean μ & sd σ. $S_n := X_1 + X_2 + \cdots + X_n$

$A_n = \dfrac{S_n}{n}$ is the average

Then $\mathbb{E}S_n = n\mu \quad Var(S_n) = n\sigma^2,\; SD(S_n) = \sqrt{n}\,\sigma$

$\mathbb{E}A_n = \underbrace{\mu}_{\text{is constant}},\quad Var(A_n) = \dfrac{\sigma^2}{n} \downarrow 0,\quad SD(A_n) = \dfrac{\sigma}{\sqrt{n}} \downarrow 0$

"the square root law"
if we double our sample size / # of trials we may only get $\sqrt{2}$ more precise.

Recall: A R.V. is not likely to be several SD's away from its mean.

$\forall c > 0$

$$\mathbb{P}(|A_n - \mu| \geq c) \longrightarrow 0 \quad \text{as } n \to \infty$$

"The (weak) law of large numbers"

proof: $\leq \dfrac{Var(A_n)}{c^2} \to 0$

$$A_n \xrightarrow{d} \mu. \quad n \to \infty$$