Thm **Markov Inequality** - relate deviation to mean. (start here: 5 No pgs 1-4.)

Thur. Apr 17/25

If X is any non-negative r.v., then

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Proof:

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

$$= \int_{0}^{\infty} x\,p(x)\,dx \quad \because X \geq 0$$

$$\geq \int_{a}^{\infty} x\,p(x)\,dx. \quad \text{for any } a \geq 0$$

$$\geq \int_{a}^{\infty} a\,p(x)\,dx \quad \because x > a \text{ in the integrated region}$$

$$= a \cdot P(X \geq a).$$

$$\therefore \boxed{P(X \geq a) \leq \frac{E[X]}{a}} \quad \text{for } a > 0.$$

See 5.5 for post'y discussion

Thm **Chebyshev Inequality**: - relates deviation to variance.

If X is any rv $\bar{c}$ finite variance (and so mean), then for any $k > 0$,

$$P(|X - E[X]| \geq k\sigma) \leq \frac{1}{k^2}.$$

Proof:
- define $Y = (X - \mathbb{E}[X])^2$.
- Y is non-negative (≥ 0.)
- apply Markov

$$P(Y \geq b^2) \leq \frac{E[Y]}{b^2}$$

- but $\mathbb{E}[Y] = \sigma^2$

$$\rightarrow P(|X - \mathbb{E}[X]| \geq b) \leq \frac{\sigma^2}{b^2}$$

let $b = k\sigma$.

$$\boxed{P(|X - \mathbb{E}[X]| \geq k\sigma) \leq 1/k^2}$$

6

- In Markov inequality why does the r.v. have to be strictly positive?
- Look at the chain of inequalities for general $X \in \mathbb{R}$, which could be negative.

$$\Rightarrow \quad \mathbb{E}[X] = \int_{-\infty}^{\infty} x p(x)$$

$$= \underbrace{\int_{-\infty}^{0} x p(x)}_{\text{I}} + \underbrace{\int_{0}^{a} x p(x)}_{\text{II}} + \underbrace{\int_{a}^{\infty} x p(x)}_{\text{III}}.$$

→ replacing $x \to a$ in the integrand of III still gives the inequality:

$$\mathbb{E}[X] \geq \underbrace{\int_{-\infty}^{0} x p(x)}_{\text{I}} + \underbrace{\int_{0}^{a} x p(x)}_{\text{II}} + \underbrace{a \cdot \int_{a}^{\infty} p(x)}_{\text{III}}$$

$$\mathbb{E}[X] \geq \underbrace{\int_{-\infty}^{0} x \cdot p(x)}_{\text{I}} + \underbrace{a \int_{a}^{\infty} p(x)}_{\text{III}}.$$

remove II for next inequality

→ this is the step that gets us into trouble if $p(x) > 0$ for $x < 0$.

(*)

⇒ We want $\boxed{\mathbb{E}[X] \geq a \cdot \int_{a}^{\infty} p(x)}$ * for the inequality.

- but if X can be < 0 then I is negative and we can no longer claim (*).

e.g. If $e \geq a + b$ we can only claim $e \geq b$ if $a = 0$.

or in numbers: $5 > 4 + 0.9$ & so $5 > 4$ & $5 > 0.9$

But $5 > -20 + 17$ but $5 \not> 17$ ⇒ a negative term ruins it.

- note Markov used mean of dist^n (1'st order moment) $\frac{\mu}{a}$

Chebyshev " variance " " (2'nd order). $\frac{\sigma^2}{b^2}$

→ can we use higher orders.?

Yes → Chernoff Bound (tightest). → uses moment generating fcn.

e.g. $X \sim \text{Binomial}(n,p)$. → Go to 6.5

For $p = 1/4$: $P(X \geq \frac{3n}{4}) \leq \cancel{\frac{2}{3}}\ \frac{1}{3}$ Markov.

$P(X \geq \frac{3n}{4}) \leq \cancel{\frac{4}{n}}\ \frac{1}{3n}$ Chebyshev.

$P(X \geq \frac{3n}{4}) \leq \cancel{\left(\frac{16}{27}\right)^{\frac{n}{4}}}\ 3^{-n/2}$ Chernoff..

from 6.5

Law of Large Numbers: weak version.

Def^n: For iid r.v.'s $X_1, X_2 \cdots X_n$, the sample mean is

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

- note, $\bar{X}$ is itself a r.v. $\bar{c}$ a mean & variance,

mean: $\mathbb{E}[\bar{X}] = \frac{\mathbb{E}X_1 + \mathbb{E}X_2 + \cdots + \mathbb{E}X_n}{n}$

$= \frac{n \cdot \mathbb{E}[X]}{n}$

$= \mathbb{E}[X].$

variance: recall for iid $\text{var}(X+Y) = \text{var}X + \text{var}Y$

$\text{var}(aX) = a^2\,\text{var}X$

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X_1) + \cdots + \text{Var}(X_n)}{n^2} = \text{var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$= \frac{n \cdot \text{Var}(X)}{n^2}$

$= \frac{\text{var}(X)}{n}.$

7

→ 7.5

Chernoff Bound

6.5

- another (tighter!) bound on tail prob's of a dist^n.

→ we used first & second moments (mean & var.) to get tail prob's. → this uses all of them.

Def^n: the n'th moment of a r.v. X is:

$$\mathbb{E}[X^n]$$

The n'th central moment is:

$$\mathbb{E}\left[(X - \mathbb{E}X)^n\right]$$

Def^n: the moment generating function of a r.v. X is a function $M_X(s)$:

$$\boxed{M_X(s) \equiv \mathbb{E}\left[e^{sX}\right]}$$

Why would you do this → lets us read all the moments of X from the coeff's of the Taylor Exp
→ just take derivatives.

- recall $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

∴ X is real:

$$e^{sX} = \sum_{k=0}^{\infty} \frac{(sX)^k}{k!} = \sum_{k=0}^{\infty} \frac{s^k}{k!} X^k$$

k'th-moment!

$$M_X(s) = \mathbb{E}\left[e^{sX}\right] = \sum_{k=0}^{\infty} \frac{s^k}{k!} \mathbb{E}[X^k]$$

So given the function $M_X(s)$, we can read off the k'th moment.

$$\mathbb{E}[X^k] = \frac{d}{ds^k} M_X(s)\Big|_{s=0}.$$

~~Example:~~

Useful fact: MGF of sums of r.v.

Say $Y = X_1 + X_2 + \cdots + X_n$ where X_i are indep. but can have different distn's.

Then $M_Y(s) = \mathbb{E}[e^{sY}]$

$$= \mathbb{E}\, e^{s(X_1 + \cdots X_n)}$$

$$= \mathbb{E}\, e^{sX_1} \cdot e^{sX_2} \cdots e^{sX_n}$$

$$= \mathbb{E} e^{sX_1}\, \mathbb{E} e^{sX_2} \cdots \mathbb{E} e^{sX_n} \quad (\because \text{independent})$$

$$\boxed{M_Y(s) = M_{X_1}(s) \cdot M_{X_2}(s) \cdots M_{X_n}(s)}$$

· iff the X_i are iid:

$$\boxed{M_Y(s) = M_X(s)^n}$$

Example: Binomial(n,p) - n coin flips $\bar{c}$ with weight p $(0,1)$ out-comes.

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

Where each $X_i \sim$ Bernoulli(p) : 1 $\bar{c}$ prob. p

0 $\bar{c}$ prob. $q = 1-p$.

$$\rightarrow M_{X_i}(s) = p \cdot e^{s \cdot 1} + q e^{s \cdot 0}$$

$$= pe^s + q$$

$$\therefore \boxed{M_X(s) = (pe^s + q)^n}$$

· check $\left.\frac{d}{ds} M_X^{n=1}(s)\right|_{s=0} = \left.pe^s\right|_{s=0} = p$ = the mean, as expected.

Back to Chernoff Bnd

- we use the MGF to capture the moments & apply Markov's inequality to the expectation.

- if X is a r.v., then for any $a \in \mathbb{R}$ we have,

$$P(X \geq a) = P(e^{sX} \geq e^{a}), \quad \text{for } s>0$$
$$P(X \leq a) = P(e^{sX} \geq e^{a}) \quad \text{for } s<0.$$

- note e^{sX} is a positive r.v. for all $s \in \mathbb{R}$.
- Markov's inequality $\rightarrow$ $P(X \geq a) = P(e^{sX} \geq e^{sa})$

$$\leq \frac{\mathbb{E}\, e^{sX}}{e^{sa}}, \quad s>0$$

$$\Rightarrow \boxed{\begin{array}{l} P(X \geq a) \leq e^{-sa} M_X(s), \quad s>0 \\ P(X \leq a) \leq e^{-sa} M_X(s), \quad s<0 \end{array}}.$$

- note: this holds for all s values $\rightarrow$ we can choose an optimal one freely.

$$\boxed{P(X \geq a) \leq \min_{s>0} e^{-sa} M_X(s)}$$

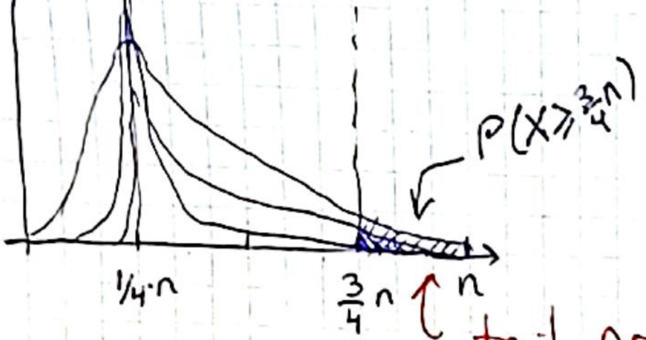
Example: Bernoulli (n,p) - n coin flips, 1 $\bar{c}$ prob p
$M_X(s) = (pe^{s}+q)^n$ 0 $\bar{c}$ " $q = 1-p$.

- $P(X \geq a) \leq e^{-sa} (pe^{s}+q)^n$
  $\rightarrow$ take deriv. of rhs & solve for optimal s:

$$e^{s_0} = \frac{aq}{p(n-a)}$$

- for $a = \frac{3}{4}n$, $p = 1/4$: $\boxed{P(X \geq \frac{3}{4}n) \leq 3^{-n/2}}$ - decays exponentially with n.

Bernoulli
$(n, 1/4)$
p''



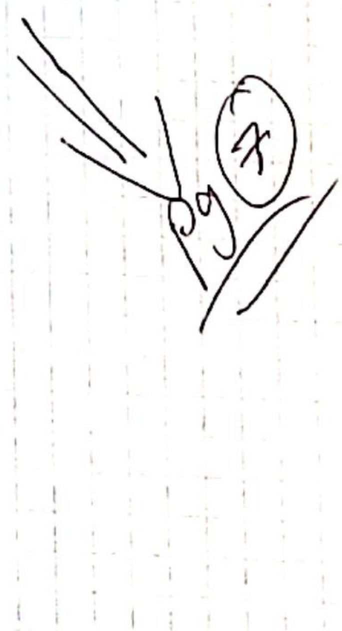
tail prob. ↓ with n.

Markov. $P(X \geq \frac{3}{4}n) \leq \frac{1}{3}$ - constant

Chebyshev: $P(X \geq \frac{3}{4}n) \leq \frac{1}{3n} = (3n)^{-1}$ - negative power

Chernoff: $P(X \geq \frac{3}{4}n) \leq 3^{-n/2}$ - negative exponential.

- this is just the empirical average of n flips tightening about the pop^n mean as we do more & more measurements.



Thm WLLN: - says the tail prob'y of the empirical mean $\xrightarrow{n\to\infty} 0$

Let $X_1, X_2, \ldots, X_n$ be iid r.v.'s with finite mean $\mathbb{E}X_i = \mu$.
Then $\forall\, \epsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0.$$

Proof: Apply Chebyshev's inequality:

CE: $P(|\bar{X} - \mathbb{E}[\bar{X}]| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon}$.

So, $P(|\bar{X} - \mu| \geq \epsilon) \leq \frac{var(\bar{X})}{\epsilon^2}$

$$\boxed{P(|\bar{X} - \mu| \geq \epsilon) = \frac{var(X)}{n\epsilon^2}}$$

$\rightarrow 0$ as $n \to \infty$ for any $\epsilon > 0$.

Why "weak"?

Jensen's Inequality

Ex., Recall the variance of a r.v. is positive.

$$\therefore Var(X) = \mathbb{E}(X - \mathbb{E}[X])^2$$
$$= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2]$$
$$= \mathbb{E}X^2 - (\mathbb{E}X)^2$$
$$> 0$$

$$\therefore \mathbb{E}[X^2] > (\mathbb{E}[X])^2$$

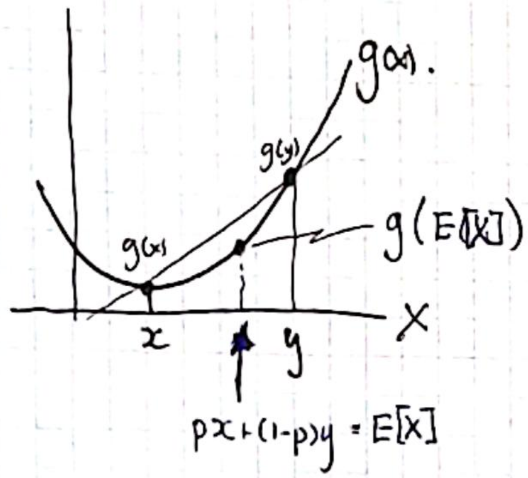
- defining $g(x) = x^2$ we have:

$$\mathbb{E}[g(x)] \geq g(\mathbb{E}[x])$$

→ this turns out to hold for general <u>convex</u> g.

Def^n: A fcn in one variable is ~~so~~ convex if the line segment connecting any two pts. on the graph lie completely above the graph.

Convex Function:



- not a proof but makes it easy to remember.
- say we have a pmf on the two values $X \sim x$ & y.
  - x c̄ prob. p
  - y c̄ prob. $q = 1-p$.

→ $\mathbb{E}[X] = px + qy$
- lies btwn x & y

$px+qy$ = convex comb^n of x & y. $= \mathbb{E}[X]$

- now look at $g(x)$ & $g(y)$.
- their convex sum $= pg(x) + qg(y)$ lies on the line connecting $(x, g(x))$ & $(y, g(y))$ in the graph.

$$\left.\begin{array}{l} (x, g(x)) \longrightarrow (px, pg(x)) \\ (y, g(y)) \longrightarrow (qy, qg(y)) \end{array}\right\} \text{convex comb}^n \; (\underbrace{px+qy}_{= \mathbb{E}[X]}, \underbrace{pg(x)+qg(y)}_{= \mathbb{E}[g(X)]})$$

~~note (px,~~

- by the def^n of a convex fcn the line connecting the two pts lies above the fcn.

$$\therefore \boxed{g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]} \text{ for convex } g(x).$$

→ generalizes to multiple variables, discrete or cont's in the obvious way.

→ called Jensen's Inequality.

Converse result: "impossibility", you can not do better than
Achievability: non-constructive, show a code exists. Existence.

ASYMPTOTIC EQUIPARTITION THEORY

1

Thur. Apr. 24/25

- follows from **Law of Large Numbers**:
  - Let $X_1, X_2, \dots$ be an iid sequence with a finite mean $\mathbb{E}_x[X] = \mu < \infty$
  - Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the running sample mean for $n = 1, 2, \dots$
  - The $\{\bar{X}_n\}_{n=1}^{\infty}$ form a sequence of r.v.'s themselves.
  - The sequence $\bar{X}_1, \bar{X}_2,$ converges to the distribution mean, μ, in the sense: For all $\epsilon > 0$,

$$\boxed{\lim_{n\to\infty} \mathbb{P}\left(|\bar{X}_n - \mu| > \epsilon\right) = 0} \quad \text{or} \quad \bar{X}_n \xrightarrow{\text{iid}} \mu \text{ in probability}$$

  - the probability the sample mean diverges from the pop^n goes to zero as $n \to \infty$. - the sample mean is unlikely to be far off from the true mean, μ.
  - this is precisely why "taking the average" is a good estimate of the true pop^n mean.

Example: Polling Problem

- you have a referendum & a fraction p of the pop^n is going to vote 'Yes'

Problem: conduct a poll & estimate $p \rightarrow$ how many to poll?
- we could do the whole pop^n but thats just conducting the referendum.

- randomly sample the pop^n uniformly
- i'th person returns $X_i = \begin{cases} 1 & \text{'Yes'} \\ 0 & \text{'No'} \end{cases}$
- note $\mathbb{E}[X_i] = p$ (unknown).
- $\bar{X}_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ = fraction of 'Yes' in the sample.

- You're boss wants you to predict p but the will always be some error if $n < N_{pop}$.
- so he asks for an estimate $\bar{c}$ a small error. Say, $\boxed{|\bar{X}_n - p| < 0.01 \quad (1\%)}$ → he want's this guaranteed!
  - let's say you sample $n = 10{,}000$
  - but you can't really meet the guarantee!
    - Say you accidently polled every who say 'No'
      ⇒ there is a $(1-p)^n$ chance of this happening
    → you can't give a guarantee in absolute terms of an error. that it will be small
  - instead you can offer this:

$\boxed{\text{I cannot guarantee } \bar{c} \text{ certainty the error will be small, but, I can guarantee that the error will be small } \bar{c} \text{ high probability.}}$

OR $\boxed{P(|\bar{X}_{10,000} - p| \geq 0.01) \leq \text{something "small"}}$

- to get a handle on the rhs recall Chebyscheus Inequality

$$P(|Y - \mathbb{E}[Y]| \geq \epsilon) \leq \frac{\sigma_Y^2}{\epsilon^2}.$$

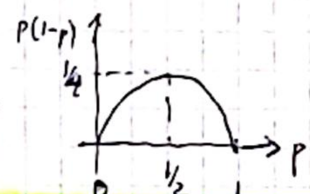
Here $Y = \frac{X_1 + X_2 + \cdots + X_n}{n} = \bar{X}_n$

$$\mathbb{E}[Y] = p$$

$$\sigma_Y^2 = var(Y) = \frac{\sigma_{X_i}^2}{n}$$

So, $P(|\bar{X}_{10,000} - p| \geq 0.01) \leq \frac{\sigma_{X_i}^2}{n\epsilon^2}$ ← recall $\sigma^2_{Bern.} = p(1-p)$.

$$P(|\bar{X}_{10,000} - p| \geq 0.01) \leq \frac{p(1-p)}{10 \cdot 10^{-4}} = p(1-p) \leq \frac{1}{4}.$$



* So, we can say if we sample 10,000 people the

* error being more than 1% occurs with less than 25% prob.

→ if we want the prob. of an error greater than 0.01 to be less than 5% ⟶ take $n = 50,000$.

→ so the prob. of a large error (>0.01) less than 5% requires $n = 50,000$ sample size.

→ you get a trade off in parameters:

$$P(|\bar{X}_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$$

or: $\boxed{P(|\bar{X}_n - p| \geq \epsilon) \leq \delta \quad \text{where } \delta = \frac{1}{4n\epsilon^2}}$

Shop signs:
- good
- cheap
- fast

⇒ Pick Two

Convergence of Random Variables.

- we all get concepts of convergence in high school & first year calculus.

$\Rightarrow$ how to apply these to sequences of r.v.'s ?

Recall WLLN's:

For any $\epsilon > 0$, $P(|\bar{X}_n - \mu|) \to 0$ as $n \to \infty$

- we'd like to say the "$\bar{X}_n$ converges to μ"
- need to define the word converges.
  - $\to$ use the WLLN's! to motivate the defn.

Defn: Convergence in Probability.

A sequence Y_n converges in probability to a scalar a if:

$$\boxed{\text{For any } \epsilon > 0, \ \lim_{n \to \infty} P(|Y_n - a| \geq \epsilon) = 0.}$$

- write $Y_n \xrightarrow{ip} a$.
- so the WLLN's states: $\bar{X}_n \xrightarrow[n \to \infty]{ip} \mu$

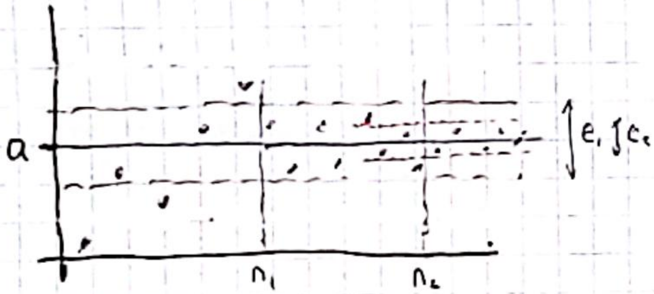
A Picture: compare to ordinary idea of convergence of a sequence of numbers.

- recall we say a sequence: $a_n \to a$

when a_n gets arbitrarily close and stays close to a eventually.

• **Ordinary Sequence:** $a_n \to a$ (a number):

"a_n gets arbitrarily close & stays close to a".



- for every $\epsilon > 0$, there exists N_0 such that for every $n \geq N_0$, we have $|a_n - a| \leq \epsilon$.

• **Convergence in Probability**

Sequence: $Y_n \to a$

"the prob. Y_n falls outside a ϵ-band around a falls to 0."

$$\forall \epsilon > 0, \; P(|Y_n - a| \geq \epsilon) \xrightarrow[n\to\infty]{} 0$$

- almost all the prob. of Y_n eventually gets concentrated arbitrarily close to a.

Some Properties: of cip.

• parellel properties of convergence in sequences.

Suppose $X_n \xrightarrow{ip} a$, $Y_n \xrightarrow{ip} b$ in probability: (no assumption of indep. within the X_n's & Y_n sequence or with each other).

① If $g(\cdot)$ is continuous, then $g(X_n) \to g(a)$
new r.v.'s

② $X_n + Y_n \longrightarrow a + b$

• **Note:** Even if $X_n \xrightarrow{ip} a$, it does <u>not</u> mean

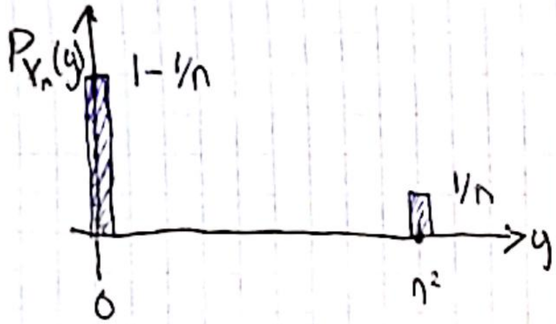
$$\mathbb{E}[X_n] \longrightarrow a$$

"convergence ip of rv's **does <u>not</u>** imply convergence of expectations" (will see in example).

Example 1. - discrete r.v.

$P(Y_n = 0) = 1 - 1/n$

$P(Y_n = n^2) = 1/n$



→ most prob. concentrated at 0 but still a small prob. at large value.

- intuitively you'd think $Y_n \xrightarrow[n\to\infty]{ip} 0$.
- check the def^n:
  - choose $\epsilon > 0$ & find $P(|Y_n - 0| \geq \epsilon)$.
  - see if it $\to 0$ as $n \to \infty$.
  - → clearly $P(|Y_n - 0| \geq \ ) = P(Y_n \geq \epsilon)$

    $= \frac{1}{n}$ for n large enough (

    $\to 0$ as $n \to \infty$

So, $Y_n \xrightarrow[n\to\infty]{ip} 0$

- So, what is $\mathbb{E}[Y_n]$ as $n \to \infty$?

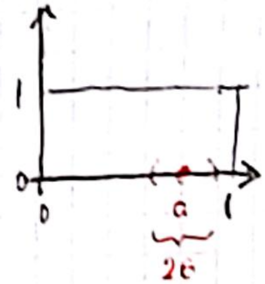
$$\mathbb{E}[Y_n] = 0 \cdot P(Y_n = 0) + n^2 \cdot P(Y_n = n^2)$$
$$= n^2 \cdot \frac{1}{n}$$
$$= n$$
$$\to \infty \text{ as } n \to \infty.$$

convergence i.p. $\not\to$ convergence in expectation

"fat tails"

- cip has to do c̄ the bulk of the prob. → only cares if the tail→0
- expectation is sensitive to outliers
  - → the tail prob. may be small but it's assigned to large values.

Example 2



i) Choose $X_i \sim$ iid uniform on $[0,1]$

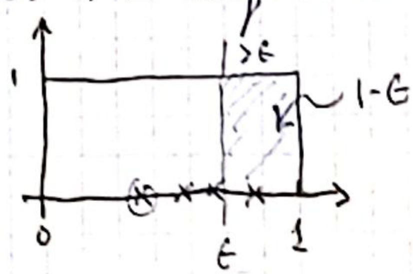
- are the X_i convergent to a value?

No. Choose any $a \in [0,1]$

No matter what ϵ interval you place around it there is alway a finite prob. of $1-2\epsilon$ that sequence values will fall outside it.

ii) Choose a related r.v.:

$Y_n = \min\{X_1, X_2 \cdots X_n\}$ the minimum of the first n values of the sequence.



- notice Y_n can only decrease

$Y_{n+1} \leq Y_n$

- since it is bounded below by 0.

it seems reasonable to think it converges to 0.

Check:

① guess $a=0$

② pick some $\epsilon > 0$.

③ evaluate $P(|Y_n - a| > \epsilon)$.

$\rightarrow P(|Y_n - 0| > \epsilon) = P(Y_n \geq \epsilon)$ ← need to show $\rightarrow 0$ as $n \rightarrow \infty$

Two cases: i) $\epsilon > 1$: $P(Y_n \geq 1^+) = 0$ ✓

$0 < \epsilon \leq 1$ ii) $P(Y_n \geq \epsilon) = P(X_1 \geq \epsilon) P(X_2 \geq \epsilon) \cdots P(X_n \geq \epsilon)$. ($\because$ iid)

$= (1-\epsilon)^n$ ← $1-\epsilon < 1$

$\rightarrow 0$ as $n \rightarrow \infty$

$\therefore Y_n = \min\{X_1 \cdots X_n\} \xrightarrow[n\to\infty]{ip.} 0$.