AMAT 362 - Probability for Statistics

LECTURE 17 : INTRO TO CONTINUOUS RVs

For DISCRTE PROBABILITY we can talk about the probability of <u>individual outcomes</u>

<u>Ex</u> Suppose I can land a kickflip 1/5 times. What's the probability that in 10 tries I land it 2 times?

$$P(\text{land } 2x) = \binom{10}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^8 \approx 30\%$$

Binom(n, p, k) w/ $n = 10$ $p = 1/5$ $k = 2$

What if I try 100 times and want 20 successes?

$$b\left(n=100, k=\tfrac{1}{5}\right) = \binom{100}{20}\left(\frac{1}{5}\right)^{20}\left(\frac{4}{5}\right)^{80} \approx 9\%$$

$p = 20$

Or $n = 1000$ $k = 200$

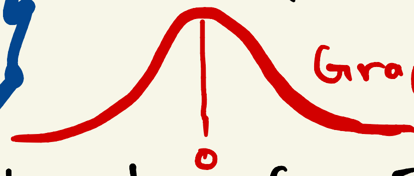
$$\binom{1000}{200}\left(\frac{1}{5}\right)^{200}\left(\frac{4}{5}\right)^{800} \approx 3\%$$

WHAT'S GOING ON HERE?!?

RECALL from

POINTWISE CLT ~ Lec 15

Let $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$



Graph of Unit Normal PDF

If I want to estimate for $S_n = X_1 + \dots + X_n$ ← IID RVs w/ μ & σ

$$\boxed{P(S_n = a) \approx \frac{\varphi\left(\frac{a - n\mu}{\sqrt{n\sigma^2}}\right)}{\sqrt{n\sigma^2}}}$$

"Pointwise CLT"
assuming X_i integer valued

N.B. If X_i are Bernoulli then $S_n \sim \text{Binom}$, $\mu = p$, $\sigma = \sqrt{pq}$

NOT SO GOOD APPROX

Ex $n = 10$, $p = 1/5$, $k = 2 = a$ in above

$$P(S_{10} = 2) \approx \frac{\varphi(0)}{\sqrt{10 \cdot 1/5 \cdot 4/5}} = \frac{1}{\sqrt{10}} \cdot \frac{\varphi(0)}{\sqrt{\frac{1}{5} \cdot \frac{4}{5}}}$$

$$\varphi(0) = \frac{1}{\sqrt{2\pi}} e^{-0^2/2}$$

$n=10$: $\boxed{31.5\%}$

$$P(S_{100} = 20) = \frac{1}{\sqrt{100}} \cdot \frac{\varphi(0)}{\sqrt{\frac{1}{5} \cdot \frac{4}{5}}} = \frac{1}{\sqrt{2\pi \cdot 100 \cdot \frac{1}{5} \cdot \frac{4}{5}}} \to \boxed{9.97\%}$$

$n=100$

$$P(S_{1000} = 200) = \frac{1}{\sqrt{1000}} \cdot \frac{\varphi(0)}{\sqrt{1/5 \cdot 4/5}} \to \boxed{3.15\%}$$

$n=1000$

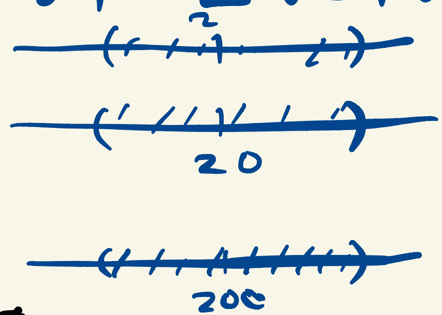
$$\boxed{\text{N.B. As } n \to \infty \quad P(S_n = a) \to 0 \;!}$$

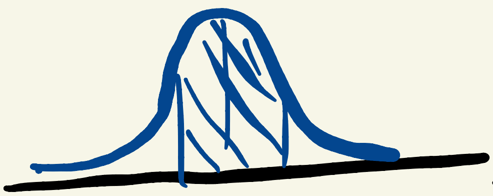
WHAT THIS SHOWS US IS THAT FOR LARGE DATA SETS TALKING ABOUT PROBABILITIES OF INDIVIDUAL OUTCOMES IS MEANINGLESS!

PHILOSOPHICAL TAKE-AWAY

INSTEAD, we should talk about

RANGES of EVENTS
(or intervals)





$\dfrac{\#\text{ Successes} - \text{Avg}}{\sqrt{npq}}$

For our Example: Better to talk about a range ($E(S_n)$ "$n\cdot p$")

of values $1 \to 3$ kickflips, i.e. $\pm \dfrac{3-2}{\sqrt{10\cdot\frac{1}{5}\cdot\frac{4}{5}}} = .79$ standard devs

For $n=100$ that amounts to

So when $n=10$ ± 1 KF

$$\frac{(20+x)-20}{\sqrt{100\cdot\frac{1}{5}\cdot\frac{4}{5}}} = \frac{x}{\sqrt{100\cdot\frac{1}{5}\cdot\frac{4}{5}}} \overset{\text{WANT}}{=} \frac{1}{\sqrt{10\cdot\frac{1}{5}\cdot\frac{4}{5}}} \;(= .79) \rightsquigarrow x = \sqrt{10} = 3.3$$

Becomes ± 3.3 KFs when $n=100$

For $n=1000$ $\quad x = \sqrt{\dfrac{1000}{10}} = \sqrt{100} = 10$ & Becomes ± 10 KFs

Here Prob of $\pm .79\sigma$ of the average will have **EQUAL PROBABILITY** $= 57\%$ rather than decreasing probability w/ n.

$\rightsquigarrow P(-.79\sigma, .79\sigma) = \Phi(.79) - \Phi(-.79) = 2\Phi(.79) - 1 = 2\cdot .7851 - 1$

Of course, as we will see, there are lots of situations where a continuous range of values more accurately models RANDOM EXPERIMENTS

DISCRETE ⟷ CONTINUOUS

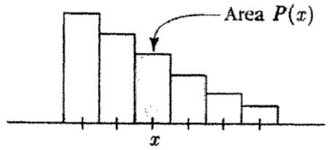
From Pitman

A DICTIONARY

Discrete Distributions

Point Probability:

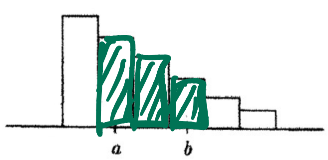
Draw PMF table as a histogram



$$P(X = x) = P(x)$$

So $P(x)$ is the probability that X has integer value x.

Interval Probability:



$$P(a \le X \le b) = \sum_{a \le x \le b} P(x)$$

the relative area under a histogram between $a - 1/2$ and $b + 1/2$.

Constraints: Non-negative with Total Sum 1

$P(x) \ge 0$ for all x and $\boxed{\sum_{\text{all } x} P(x) = 1}$

Normalization Axiom

Expectation of a Function g of X, e.g., X, X^2:

$$E(g(X)) = \sum_{\text{all } x} g(x)P(x)$$

provided the sum converges absolutely.

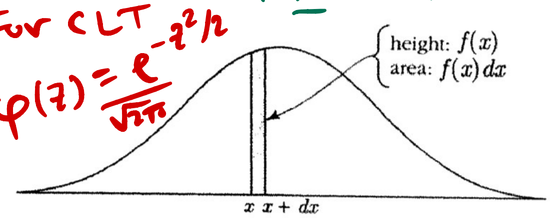
$\sum \longrightarrow \int$

$E[X] = \sum x P(X=x)$

Distributions Defined by a Density

Infinitesimal Probability: PDF ← Draw

For CLT $\varphi(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$

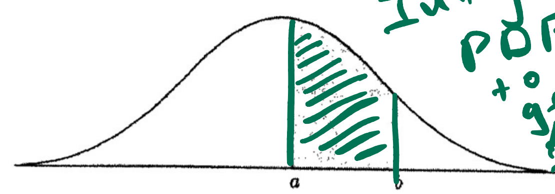


$$P(X \in dx) = f(x)dx$$

The *density* $f(x)$ gives the probability per unit length for values near x.

Interval Probability:

Integrate PDF to get Area = Probability



$$P(a \le X \le b) = \int_a^b f(x)dx$$

the area under the graph of $f(x)$ between a and b.

Constraints: Non-negative with Total Integral 1

$f(x) \ge 0$ for all x and $\boxed{\int_{-\infty}^{\infty} f(x)dx = 1}$

Expectation of a Function g of x, e.g., X, X^2:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

provided the integral converges absolutely.

$E[X] = \int x f_X(x)$

So far this dictionary of Discrete → Continuous was motivated by approximations

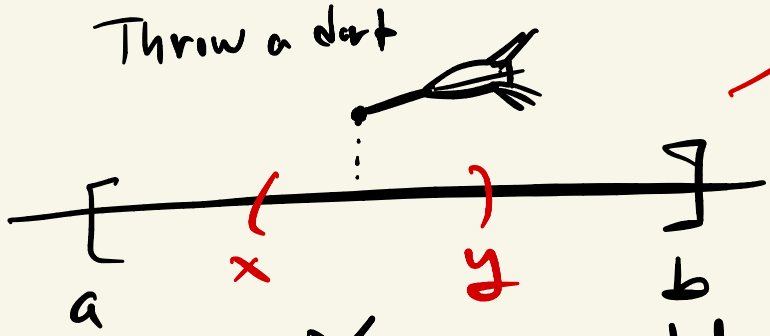
- Binomial (or any large Sum) $\xrightarrow{\text{Baby CLT}}$ (Grownup CLT) Normal Distribution
- Geometric (Poisson) $\xrightarrow{\text{(TODAY!)}}$ Exponential Distribution
- Uniform on {a, b} ⟶ Uniform [a, b] or (a, b) b/c endpoints don't matter

Roll a die ⟶ Throw a dart

UNIFORM (in 1-D)

Probability is proportional to relative length

Throw a dart



X = horizontal position of dart

$P(a<X<b) = \frac{b-a}{b-a} = 1$

" something happens"

$$\text{Prob}(x<X<y) = \frac{y-x}{b-a}$$

$$= \frac{\text{length}(x,y)}{\text{length}(a,b)}$$

= width of dart board

$$\boxed{P(X=x)=0}$$

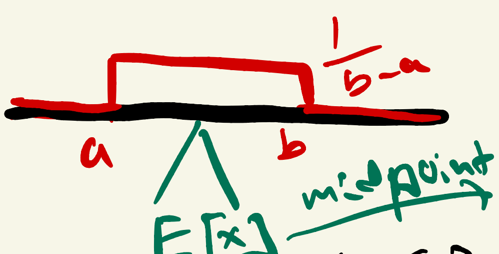
PDF for $X \sim \text{Unif}[a,b]$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{o.w.} \end{cases}$$

$$E[X] = \int_a^b x \cdot f_X(x)\,dx = \frac{1}{b-a} \cdot \int_a^b x\,dx = \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b$$

$$= \frac{b^2-a^2}{2(b-a)}$$

$$= \frac{(b+a)\cancel{(b-a)}}{2\cancel{(b-a)}}$$



$$\boxed{E[X] = \frac{b+a}{2}}$$

To compute VAR[X]

$$E[X^2] = \int_a^b x^2 f_X(x)\,dx = \frac{1}{b-a} \frac{x^3}{3}\Big|_a^b = \frac{b^3-a^3}{3(b-a)}$$

$$= \frac{\cancel{(b-a)}(b^2+ab+a^2)}{3\cancel{(b-a)}}$$

why? $\rightarrow$ $b^3 + \cancel{ab^2} + \cancel{ba^2} - \cancel{ab^2} - \cancel{a^2b} - a^3 = b^3 - a^3$

$\text{Var}[X] = E[X^2] - E[X]^2$

$E[X^2] = \frac{b^2 + ab + a^2}{3}$ For Unif $[a,b]$ The "Second Moment"

$V[X] = E[X^2] - E[X]^2$

$= \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4}$

$= \frac{\overset{b^2}{\cancel{4b^2}} + \cancel{4ab} + \overset{a^2}{\cancel{4a^2}} - \cancel{3b^2} - \overset{-2ab}{\cancel{6ab}} - \cancel{3a^2}}{12}$

$\boxed{V[X] = \frac{(b-a)^2}{12}}$

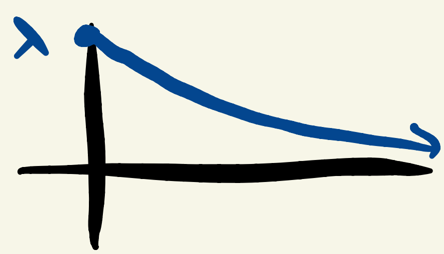
Compare w/ Unif $\{a, a+1, \dots b\}$ $V[X] = \frac{(b-a+1)^2 - 1}{12}$

EXPONENTIAL R.V.

Governs decay, lifetime, waiting times!

<u>Def</u> A continuous RV T is exponentially distributed w/ parameter λ if its PDF is

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & \text{o.w.} \end{cases}$$

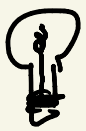


for $T \sim Exp(\lambda)$

$\hookrightarrow E[T] = \frac{1}{\lambda}$ AND $SD[T] = \frac{1}{\lambda}$

OFTEN YOU'RE GIVEN THIS

Ex



Suppose a piece of equipment has lifetime that is exponentially distributed.

Given $E[T]$ → The average lifetime is 100 working hours. What is the probability it lasts more than 200 hours?

Ans

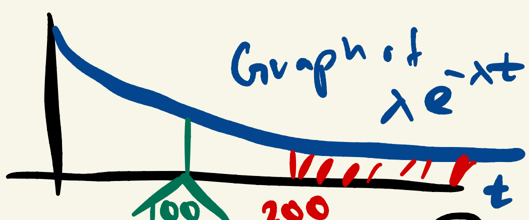
$$E[T] = \frac{1}{\lambda} = 100 \Rightarrow \boxed{\lambda = \frac{1}{100}}$$

$$P(T>200) = \int_{200}^{\infty} \lambda e^{-\lambda t}\, dt = -e^{-\lambda t}\Big|_{200}^{\infty}$$

$$= e^{-\lambda \cdot 200} - e^{-\lambda \cdot \infty}$$

$$\boxed{= e^{-2}}$$

ANS!



$\to \lambda e^{-\lambda t} \to 0$, ∞

Compare w/ Poisson/Geometric

Prob of failing in a given hour is $\frac{1}{100} = p$

What's the probability first failure comes after 200 hours

$$\text{Prob of no failures in 200 trial hours} = (1-p)^{200}$$

$$\approx (e^{-p})^{200} = e^{-\frac{200}{100}} = e^{-2}$$