🧀🧀🧀 03 Mystery distributions: Identify from data

A researcher collects data from three different experiments and computes summary statistics:

Dataset A: $n = 500$ observations, all values are either 0 or 1. Sample mean ≈ 0.23, sample variance ≈ 0.177.

Dataset B: $n = 1000$ observations, values range from 0 to 47. Sample mean ≈ 12.1, sample variance ≈ 11.8.

Dataset C: $n = 800$ observations, values are positive reals ranging from 0.001 to 14.2. Sample mean ≈ 2.5, sample variance ≈ 6.3.

For each dataset:

- a. Identify the most likely distribution family.
- b. Estimate the parameter(s) of that distribution from the summary statistics.
- c. For Dataset B, the researcher notices that these are counts of customer complaints per day at a call center. Does this context support your answer? What if instead they were counts of “successes” in 50

Bern, Bin, geometric,

poisson, expon,

uniform, ~~normal~~

$\boxed{p}$ $\boxed{p(1-p)}$

λ λ

🧀 🧀 01 Convergence modes: The vanishing spike

Define X_n as follows: with probability $\frac{1}{n}$, $X_n = n$; otherwise $X_n = 0$.

- a. Compute $E[X_n]$ and $\mathrm{Var}[X_n]$. What happens as $n \to \infty$?
- b. Show that $X_n \xrightarrow{P} 0$ (converges in probability to 0) by computing $P[|X_n| > \epsilon]$ for any $\epsilon > 0$.
- c. Does $X_n \to 0$ almost surely? Hint: Consider $\sum_{n=1}^{\infty} P[X_n \neq 0]$. Use the Borel-Cantelli lemma intuition: if events happen "infinitely often" then convergence a.s. fails.
- d. Explain in 2-3 sentences: why can a sequence have vanishing probability of being far from 0, yet still "spike" infinitely often with probability 1?

Discrete Distributions

🧀 04 The “obvious” Bernoulli that isn’t

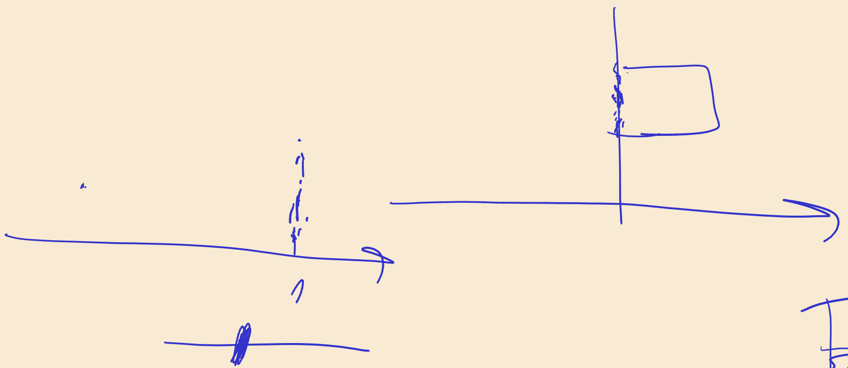
A weighted die shows 6 with probability $\frac{1}{3}$ and each of 1–5 with probability $\frac{2}{15}$.

- a. Define a Bernoulli random variable X for “rolling a 6.” State p and compute $E[X]$ and $\mathrm{Var}[X]$.
- b. Define a different Bernoulli random variable Y for “rolling an even number.” Compute $E[Y]$ and $\mathrm{Var}[Y]$.
- c. For which event is the variance larger? Explain intuitively why maximum Bernoulli variance occurs at $p = 0.5$.

1 0

+ − 1

$p = 1$



$E[(X - E(X))^2] \to 0$

$\boxed{p(1-p)}$

$p - p^2$ $p = 0.5$

303. Ավանում կա 2800 բնակիչ: Նրանցից յուրաքանչյուրը ամսական մոտ 6 անգամ գնացքով մեկնում է քաղաք, ուղևորության օրերը մեկը մյուսից անկախ ընտրելով պատահականորեն: Գտնել գնացքի այն փոքրագույն տարողությունը, որի դեպքում այն ամբողջությամբ կլցվի միջին հաշվով 100 օրում ոչ ավելի, քան մեկ անգամ (գնացքը գնում է օրական մեկ անգամ):

100 99

2800

1 – 6

6

2800-ից $x \geq$ 100

30 – 6

$P = \frac{6}{30} = 0.2$

0.2
0.8

2800

$\boxed{2800 \cdot 0.2 = 560}$

$a = n \cdot p = 560$

$x = 0.2$

$n \cdot x = 0.8$

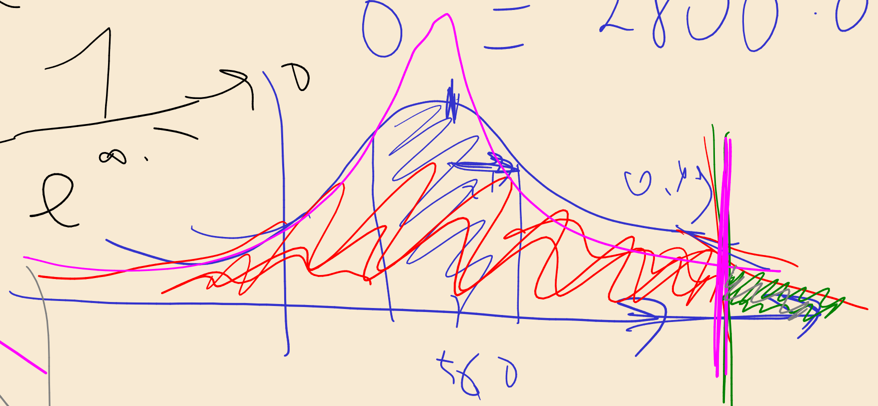
$n \cdot x = 0.2 \cdot 0.8$

$$\boxed{\int_{-\infty}^{c} f(x)\,dx} \qquad \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\frac{1}{e^{\infty}} \to 0$$

$$\sigma^2 = 2800 \cdot 0.2 \cdot 0.8 = 448$$

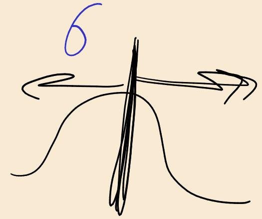
$$\sigma = 21$$



560

0.99

560-21 $\qquad$ 560+21



σ

f

0.01

$= 0.99$

0.99

560 $\quad$ 600 $\quad$ c $\quad$ 0.01

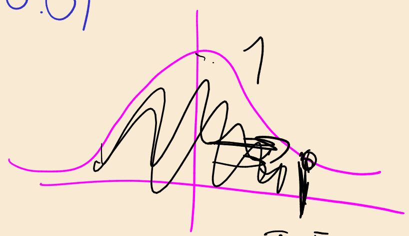
$$\boxed{P(Z < c) = 0.99}$$

$Z_{0.?} = 0.495$

0.99 $\quad \boxed{609}$ $\quad$ 610

560 $\quad$ 100

$560 + 2.33 \cdot \sqrt{448}$



2.5

0, 1 $\qquad$ 99.7

95 $\qquad$ 3

2

547

$$\boxed{560 + 2{,}5 \cdot 21}$$

7, ~~6~~, 3, 4, 2, 2, 4, 3.

P

52

2^n 4 4 4

13 in

80 20

$P(X=K) = \dfrac{C_4^2 \, C_{N-K}^{2-2}}{C_{52}^2}$

$n \cdot \dfrac{K}{N}$

$n \cdot \dfrac{K}{N} \dfrac{N-K}{N} \dfrac{N-N}{N-1}$

1 2 3 4 5