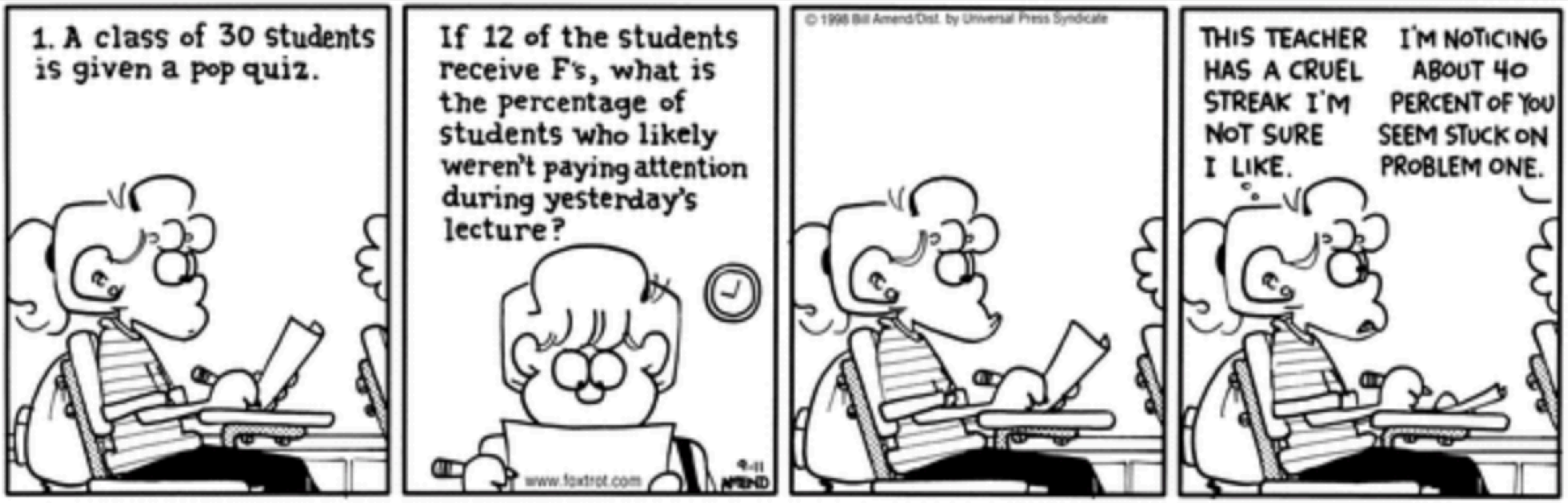
Stat 88: Probability & Mathematical Statistics in Data Science



Lecture 21: 3/10/2021

Sections 6.3, 6.4

Markov and Chebyshev's Inequalities

Bounding the tail probabilities

- **Markov's Inequality**: For a nonnegative rv X, and constant $c > 0$

$$P(X \geq c) \leq \frac{E(X)}{c}$$

- **Chebyshev's inequality**: For a random variable X, with mean μ and standard deviation σ, for any positive constant $c > 0$, we have:

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

- Ex: Is it possible that half of US flights have delay times at least 3 times the national average?

Chebyshev's inequality interpreted as distances

- Say that $E(X)$ is the origin, and we are measuring distances in terms of $SD(X)$.
- We want to know the chance that the rv X is at least $k\ SD'$s away from its mean:

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

- What if we are only interested in one tail? A certain type of light bulb has an average lifetime of 10,000 hours. The SD of bulb lifetimes is 550 hours. What decimal fraction of bulbs could last more than 11,980 hours?

Chebyshev or Markov?

- Suppose X is a non-negative random variable with expectation 60 and SD 5.

(a) What can we say about $P(X \geq 70)$?

(b) What is the chance that X is outside the interval $(50, 70)$?

(c) What about $P\left(X \in (50, 70)\right)$?

Exercise 6.5.6

Ages in a population have a mean of 40 years. Let X be the age of a person picked at random from the population.

a) If possible, find $P(X \geq 80)$. If it's not possible, explain why, and find the best upper bound you can based on the information given.

b) Suppose you are told in addition that the SD of the ages is 15 years. What can you say about $P(10 < X < 70)$?

c) With the information as in Part b, what can you say about $P(10 \leq X \leq 70)$?

Examples

Suppose that each year, Berkeley admits 15,000 students on average, with an SD of 5,000 students. Assuming that the application pool is roughly the same across years, find the smallest upper bound you can on the probability that Berkeley will admit at least 22,500 students in 2021.

Example

Suppose a list of numbers $x = \{x_1, \dots, x_n\}$ has mean μ and standard deviation σ . Let k be the smallest number of standard deviations away from μ we must go to ensure the range $(\mu - k\sigma, \mu + k\sigma)$ contains at least 50% of the data in x. What is k?

Example

A list of non negative numbers has an average of 1 and an SD of 2. Let p be the proportion of numbers over 4. To get an upper bound for p, you should:

a) Assume a binomial distribution

b) Use Markov's inequality.

c) Use Chebyshev's inequality

d) None of the above.

Example

Let X be a non-negative random variable such that $E(X) = 100 = Var(X)$.

a) Can you find $E(X^2)$ exactly? If not, what can you say?

b) Can you find $P\ (70 < X < 130)$ exactly? If not, what can you say?

7.1: Sums of Independent Random Variables

- Recall that expectation is additive, which we used many times.
$$(E(X+Y)=E(X)+E(Y))$$

- What about $Var(X+Y)$? Well, it depends.

- Consider tossing a fair coin 10 times. Let H be the number of heads and T be the number of tails in 10 tosses. Then $H+T=10$. Note that $Var(H), Var(T)\neq 0$, but $Var(H+T)=Var(10)=0$!

- But now let H_1 be the number of heads in the first 5 tosses, and H_2 the number of heads in the last 5 tosses. Will we have that $Var(H_1+H_2)=0$?

- It turns out that if X and Y are ***independent***, then we have that
$$Var(X+Y)=Var(X)+Var(Y)$$

Sums of iid random variables

- Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random variables with mean μ and variance σ^2. Define S_n to be their sum:
$$S_n = X_1 + X_2 + \cdots + X_n.$$

- We already know that $E(S_n) = \sum E(X_k) = n\mu$.
- Now we can further say that:

$$Var(S_n) = Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n) = n\sigma^2$$
$$SD(S_n) = \sqrt{n}\,\sigma$$

- Notice that the expected value grows as n, but the sd grows as $\sqrt{n}$.)

Variance of the Binomial distribution

- Recall that a binomial random variable $X \sim Bin(n, p)$ is the sum of n iid Bernoulli(p) random variables $I_1, I_2, \ldots, I_n$ where I_k is the indicator of success on the kth trial.
- What are the mean and variance of I_k? And therefore, what are the mean and variance of X? For what p will this variance be maximum?

Variance of Poisson (μ) and geometric(p)

- Recall that one way to get the Poisson rv is by approximating the Binomial(n, p) distribution when n is large and p is small. ($\mu = np$)

- SD of the binomial distribution is $\sqrt{np(1-p)}$.
- Note that if p is small, $(1-p) \approx 1$, and we can say that $np(1-p) \approx np$.
- This gives us that the SD of the Poisson(μ) distribution is $\sqrt{\mu}$

- Geometric(1/p) distribution: Fact: the variance of the geometric distribution is $\frac{1-p}{p^2}$

- Ex: (Waiting till the 10th success) Suppose you roll a die until the 10th success. Let R be the number of rolls required. Find $SD(R)$.