

What's a conf interval?



$$p = \frac{300}{1000}$$

$$31 \quad 100$$

What's p ?

How sure are you abt that?

$$\hat{p} = 0.32$$

$$\frac{32}{100} = \frac{320}{1000}$$

$$p \notin [0.30, 0.32]$$

$$\underline{80\%}$$

$$E(X) \quad \text{var}(X)$$

$$X, Y$$

$$Z = X|Y$$

$y \backslash x$	0	1	2	3
0	0	0	$\frac{1}{3}$	0
1	0	$\frac{1}{3}$	$\frac{1}{6}$	0
2	$\frac{1}{2}$	0	0	$\frac{1}{2}$

$$Z_0 = X|(Y=0)$$

$$Z_1 = X|(Y=1)$$

$$Z_2$$

	0	1	2	3
Z_0	0	0	1	0
Z_1	0	$\frac{2}{3}$	$\frac{1}{3}$	0
Z_2	$\frac{1}{2}$	0	0	$\frac{1}{2}$

$$E[X|Y=y] = E[Z_y]$$

$$E[X|Y=0] = E[Z_0] = 2$$

$$E[X|Y=1] = E[Z_1] = \frac{4}{3}$$

Amogh Gupta, Sylvia Jin, Aekus Bhathal, Abinav Routhu, Debayan Bandyopadhyay

Roast us here: <https://tinyurl.com/csm70-feedback20>

1 Confidence Intervals

1. Define i.i.d. variables $A_k \sim \text{Bern}(p)$ where $k \in \{1, \dots, n\}$. Assume we can declare that $\Pr \left[\left| \frac{1}{n} \sum_{k=1}^n A_k - p \right| \geq 0.25 \right] = 0.01$.

(a) Please give a 99% confidence interval for p given A_k .

point estimate \pm margin of error
mean standard dev = $\sqrt{\text{var}}$

(b) We know that the variables X_1, \dots, X_n , are i.i.d. random variables and have variance σ^2 . We also have the observation that $A_n = \frac{X_1 + \dots + X_n}{n}$. We want to estimate the mean, μ , of each X_j .

Prove that we have 95% confidence that μ lies in the interval $\left[A_n - 4.5 \frac{\sigma}{\sqrt{n}}, A_n + 4.5 \frac{\sigma}{\sqrt{n}} \right]$

That is, $\Pr \left[A_n - 4.5 \frac{\sigma}{\sqrt{n}} \leq \mu \leq A_n + 4.5 \frac{\sigma}{\sqrt{n}} \right] \geq 95\%$

(c) Give the 99% confidence interval for μ .

2. We have a die with 6 faces labeled 1, 2, 3, 4, 5, 6.

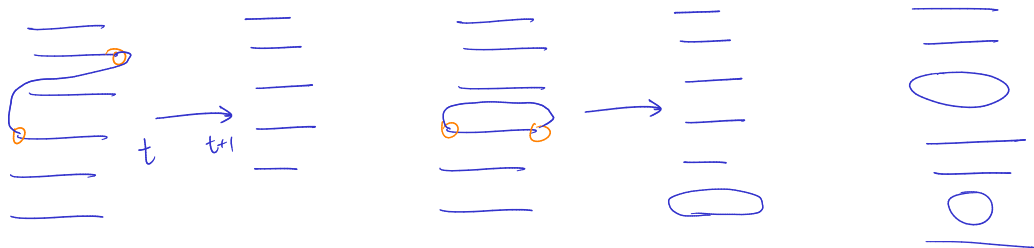
(a) Develop a 99% confidence interval for the sum of n samples.

(b) Now, suppose the die's face values are just 6 consecutive integers $k + 1, k + 2, \dots, k + 6$, but we do not know k . For example, if $k = 6$, the die faces would take on the values 7, 8, 9, 10, 11, 12. If we observe that the average of the n samples is 15.5, develop a 99% confidence interval for the value of k .

2 Conditional Expectation

1. Looping Ropes

Frobenius has n ropes in his backyard, which he likes a lot. But when he goes to work everyday, they grow a mind of their own and begin behaving weirdly. At every timestep t , two ends of a rope(s) are uniformly chosen at random and knotted together. If the two ends are from the same rope, they form a loop. If the two ends are from different ropes, they join together to form a new rope. By the time Frobenius comes home, this process has completed (meaning no more loose ends are left). **How many loops can Frobenius expect to see?** **Bonus:** Does this converge as $n \rightarrow \infty$?



X_n is # of loops when n ropes are left.

I is indicator of loop formation.

$$\begin{aligned} \textcircled{1} E[X_n] &= E[X_n | I=1]Pr[I=1] + E[X_n | I=0]Pr[I=0] \\ &= (E[X_{n-1}] + 1)Pr[I=1] + (E[X_{n-1}])Pr[I=0] \\ &= E[X_{n-1}] + \underbrace{Pr[I=1]}_{\frac{1}{2n-1}} \\ &= E[X_{n-1}] + \frac{1}{2n-1} \end{aligned}$$

$$\textcircled{2} E[X_1] = 1$$

$$\begin{aligned} \textcircled{3} E[X_2] &= E[X_2] + \frac{1}{5} \\ &= E[X_1] + \frac{1}{3} + \frac{1}{5} \\ &= 1 + \frac{1}{3} + \frac{1}{5} \end{aligned}$$

$$E[X_n] = \sum_{i=1}^n \frac{1}{2i-1}$$

2. If there are no distractions, it takes me 30 minutes to walk to the store. However, if I pass someone with a cat, I stop and pet the cat and chat to the owner. The number Y of cats I pass is a Poisson random variable with mean 2. Each time I stop, the number of minutes I spend petting the cat and chatting is an exponential random variable with PDF:

$$f_X(x) = 0.5e^{-0.5x} \mathbb{1}_{\{x \geq 0\}}$$

- (a) If I see a single cat, what is the expectation and variance of the time spent petting the cat and chatting to its owner?

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \int_0^{\infty} x \cdot (0.5e^{-0.5x}) dx = \boxed{2}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x) dx = \int_0^{\infty} x^2 \cdot (0.5e^{-0.5x}) dx = 8$$

$$\text{Var}(X) = E[X^2] - E[X]^2 = 8 - 2^2 = \boxed{4}$$

- (b) What is the conditional expectation $E[X | Y]$ of the total time spent petting cats and chatting to their owners, as a function of Y ?

- (c) Using the smoothing law (or law of iterated expectation, or law of total expectation), calculate $E[X]$.

3. Rolling Chopsticks

The content mentors were trying to eat noodles in a new way. Rather than eating noodles by chopsticks directly, they tried eating noodles by rolling one noodle on the chopstick and eat it. This is seemingly a hard way to eat noodles so the probability they successfully eat a noodle on each attempt is p .

- (a) Suppose they start attempting to eat a noodle, and eat the noodle on the attempt X . What is the distribution of X ? What is the distribution of unsuccessful attempts to eat that noodle, X' in terms of X ?

- (b) Let Y be the unsuccessful attempts that they will make trying to eat 2 noodles. What is the distribution of Y ?

- (c) Content with their distribution Y and eating 2 noodles, the content mentors want to find the distribution Z for the total unsuccessful attempts of eating the whole bowl of R noodles. They were planning to proceed as part b) but then Aekus, a random variable distribution enthusiast, suggested to use $P(Z = k) = \binom{r+k-1}{k} (1-p)^k p^r$ where $r = R$.

The distribution Z is defined by 2 parameters: 1) the number of successful attempts r and 2) the probability of a successful attempt p so we will write Z as $Z(r, p)$.

Show that Z is the sum of independent X' random variables by using induction on r where $r = 1$ is the base case, and the content mentors can use Z as their distribution.

(Hint: Remember the "Hockey stick" identity $\sum_{i=0}^{k-1} \binom{n+i}{i} = \binom{n+k}{k-1}$)

- (d) What is the expected value of total unsuccessful attempts of eating the whole bowl of R noodles, the random variable Z ?

3 Linear Least-Squares Estimation

1. Linear Least Squares Estimate

Linear Least Squares Estimate (LLSE) The LLSE of Y given X , denoted by $L[Y|X]$, is the linear function $a + bX$ that minimizes

$$C(g) = E(|Y - a - bX|^2).$$

Let's try to derive a formula for $L[Y|X]$ in the form of properties of distribution of X and Y

- (a) Write $C(g)$ as linear function of $E(Y^2)$, $E(X^2)$, $E(Y)$, $E(X)$ and $E(XY)$
- (b) Find the values of a and b that minimize the expression in part a. To simplify the calculation use

$$\text{Cov}(X, Y) = E(YX) - E(Y)E(X) \text{ and } \text{Var}(X) = E(X^2) - E(X)^2.$$

- (c) Let's put everything together and find the formula for $L[Y|X]$