CS70: Lecture 26.

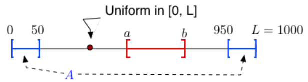
Continuous Probability

1. Examples
2. Events
3. Continuous Random Variables
4. Expectation
5. Normal Distribution
6. Central Limit Theorem

Continuous Probability - Pick a real number.

Choose a real number X, uniformly at random in $[0,1000]$.

What is the probability that X is exactly equal to $100\pi = 314.1592625\ldots$? Well, ..., 0.



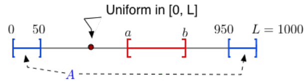
Let $[a,b]$ denote the **event** that the point X is in the interval $[a,b]$.

$$Pr[[a,b]] = \frac{\text{length of } [a,b]}{\text{length of } [0,L]} = \frac{b-a}{L} = \frac{b-a}{1000}.$$

Intervals like $[a,b] \subseteq \Omega = [0,L]$ are **events.**
More generally, events in this space are unions of intervals.
Example: the event A - "within 50 of 0 or 1000" is
$A = [0,50] \cup [950,1000]$. Thus,

$$Pr[A] = Pr[[0,50]] + Pr[[950,1000]] = \frac{1}{10}.$$

Continuous Probability - Pick a random real number.



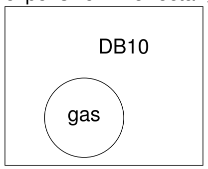
Note: A **radical** change in approach. For a finite probability space, $\Omega = \{1,2,\ldots,N\}$, we started with $Pr[\omega] = p_\omega$. We then defined $Pr[A] = \sum_{\omega \in A} p_\omega$ for $A \subset \Omega$. We can use the same approach for countable Ω.

For a continuous space, e.g., $\Omega = [0,L]$, we cannot start with $Pr[\omega]$, because this will typically be 0. Instead, we start with $Pr[A]$ for some events A. Here, we started with A = interval, or union of intervals.

Thus, the probability is a function from events to $[0,1]$. Can any function make sense? No! At least, it should be additive!. In our example, $Pr[[0,50] \cup [950,1000]] = Pr[[0,50]] + Pr[[950,1000]]$.

Shooting..

A James Bond example. In Spectre, Mr. Hinx is chasing Bond who is in a Aston Martin DB10 .
Hinx shoots at the DB10 and hits it at a random spot.
What is the chance Hinx hits the gas tank?
Assume the gas tank is a one foot circle and the DB10 is an expensive 4×5 rectangle.



$\Omega = \{(x,y) : x \in [0,4], y \in [0,5]\}$.

The size of the event is $\pi(1)^2 = \pi$.
The "size" of the sample space which is 4×5.

Since uniform, probability of event is $\frac{\pi}{20}$.

Continuous Random Variables: CDF

$Pr[a < X \le b]$ instead of $Pr[X = a]$.
For all a and b: specifies the behavior!
Simpler: $P[X \le x]$ for all x.

Cumulative probability Distribution Function of X is

$$F_X(x) = Pr[X \le x]^1$$

$$Pr[a < X \le b] = Pr[X \le b] - Pr[X \le a] = F_X(b) - F_X(a).$$

Idea: two events $X \le b$ and $X \le a$.
Difference is the event $a < X \le b$.
Indeed: $\{X \le b\} \setminus \{X \le a\} = \{X \le b\} \cap \{X > a\} = \{a < X \le b\}$.

[1] The subscript X reminds us that this corresponds to the RV X.

Example: CDF

Example: Value of X in $[0,L]$ with $L = 1000$.

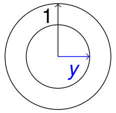
$$F_X(x) = Pr[X \le x] = \begin{cases} 0 & \text{for } x < 0 \\ \frac{x}{1000} & \text{for } 0 \le x \le 1000 \\ 1 & \text{for } x > 1000 \end{cases}$$

Probability that X is within 50 of center:

$$\begin{aligned} Pr[450 < X \le 550] &= Pr[X \le 550] - Pr[X \le 450] \\ &= \frac{550}{1000} - \frac{450}{1000} \\ &= \frac{100}{1000} = \frac{1}{10} \end{aligned}$$

Example: CDF

Example: hitting random location on dartboard.
Random location on circle.



Random Variable: Y distance from center.
Probability within y of center:

$$\begin{aligned} Pr[Y \leq y] &= \frac{\text{area of small circle}}{\text{area of dartboard}} \\ &= \frac{\pi y^2}{\pi} = y^2. \end{aligned}$$

Hence,

$$F_Y(y) = Pr[Y \leq y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}$$

Calculation of event with dartboard..

Probability between .5 and .6 of center?
Recall CDF.

$$F_Y(y) = Pr[Y \leq y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \leq y \leq 1 \\ 1 & \text{for } y > 1 \end{cases}$$

$$\begin{aligned} Pr[0.5 < Y \leq 0.6] &= Pr[Y \leq 0.6] - Pr[Y \leq 0.5] \\ &= F_Y(0.6) - F_Y(0.5) \\ &= .36 - .25 \\ &= .11 \end{aligned}$$

Poll

Consider the example of a dartboard of unit radius. RV Y is distance of the random spot from the center, and let F(y) be its CDF. Let p1 = F(0.3) and p2 = F(0.6). Then, p2/p1 is equal to

- 1/2
- 2
- 1/4
- 4

Density function.

Is the dart more likely to be near .5 or .1?
Probability of “Near x” is $Pr[x < X \leq x+\delta]$.
Goes to 0 as δ goes to zero.
Try

$$\frac{Pr[x < X \leq x+\delta]}{\delta}.$$

The limit as δ goes to zero.

$$\begin{aligned} \lim_{\delta \to 0} \frac{Pr[x < X \leq x+\delta]}{\delta} &= \lim_{\delta \to 0} \frac{Pr[X \leq x+\delta] - Pr[X \leq x]}{\delta} \\ &= \lim_{\delta \to 0} \frac{F_X(x+\delta) - F_X(x)}{\delta} \\ &= \frac{d(F_X(x))}{dx}. \end{aligned}$$

Density

Definition: (Density) A **probability density function** for a random variable X with cdf $F_X(x) = Pr[X \leq x]$ is the function $f_X(x)$ where

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du.$$

Thus,

$$Pr[X \in (x, x+\delta]] = F_X(x+\delta) - F_X(x) \approx f_X(x)\delta.$$

Examples: Density.

Example: uniform over interval $[0, 1000]$

$$f_X(x) = F'_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{1000} & \text{for } 0 \leq x \leq 1000 \\ 0 & \text{for } x > 1000 \end{cases}$$

Example: uniform over interval $[0, L]$

$$f_X(x) = F'_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{L} & \text{for } 0 \leq x \leq L \\ 0 & \text{for } x > L \end{cases}$$

Examples: Density.

Example: “Dart” board.
Recall that

$$F_Y(y) = Pr[Y \le y] = \begin{cases} 0 & \text{for } y < 0 \\ y^2 & \text{for } 0 \le y \le 1 \\ 1 & \text{for } y > 1 \end{cases}$$

$$f_Y(y) = F'_Y(y) = \begin{cases} 0 & \text{for } y < 0 \\ 2y & \text{for } 0 \le y \le 1 \\ 0 & \text{for } y > 1 \end{cases}$$

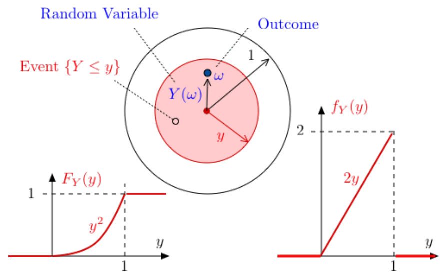
The cumulative distribution function (cdf) and probability distribution function (pdf) give full information.
Use whichever is convenient.

Poll

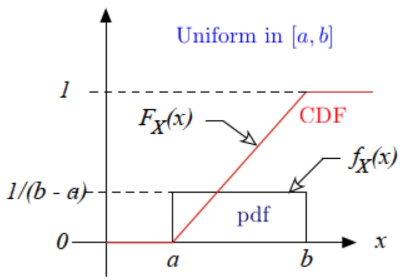
Let $F(x) = ax^2$ for $0 \le x \le 10$ be the CDF of a RV X that takes value in [0, 10]. Then, PDF f(x) must be

- 50x
- 100x
- x/50
- x/100

Target



$U[a,b]$

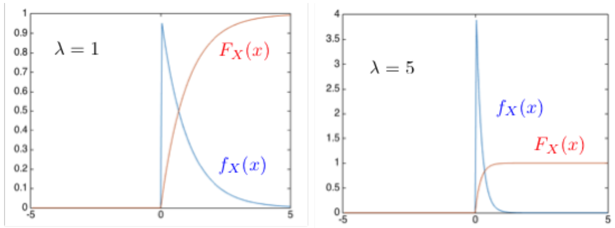


$Expo(\lambda)$

The exponential distribution with parameter $\lambda > 0$ is defined by

$$f_X(x) = \lambda e^{-\lambda x} 1\{x \ge 0\}$$

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \ge 0. \end{cases}$$



Note that $Pr[X > t] = e^{-\lambda t}$ for $t > 0$.

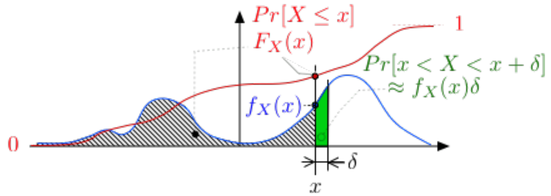
Random Variables

Continuous random variable X, specified by

1. $F_X(x) = Pr[X \le x]$ for all x.
   Cumulative Distribution Function (cdf).
   $Pr[a < X \le b] = F_X(b) - F_X(a)$
   1.1 $0 \le F_X(x) \le 1$ for all $x \in \Re$.
   1.2 $F_X(x) \le F_X(y)$ if $x \le y$.
2. Or $f_X(x)$, where $F_X(x) = \int_{-\infty}^{x} f_X(u)du$ or $f_X(x) = \frac{d(F_X(x))}{dx}$.
   Probability Density Function (pdf).
   $Pr[a < X \le b] = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$
   2.1 $f_X(x) \ge 0$ for all $x \in \Re$.
   2.2 $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

Recall that $Pr[X \in (x, x+\delta)] \approx f_X(x)\delta$. Think of X taking discrete values $n\delta$ for $n = \dots, -2, -1, 0, 1, 2, \dots$ with $Pr[X = n\delta] = f_X(n\delta)\delta$.

A Picture



The pdf $f_X(x)$ is a nonnegative function that integrates to 1.

The cdf $F_X(x)$ is the integral of f_X.

$$Pr[x < X < x+\delta] \approx f_X(x)\delta$$
$$Pr[X \le x] = F_x(x) = \int_{-\infty}^{x} f_X(u)du$$

Some Examples

1. ***Expo*** **is memoryless.** Let $X = Expo(\lambda)$. Then, for $s, t > 0$,

$$\begin{aligned} Pr[X > t+s \mid X > s] &= \frac{Pr[X > t+s]}{Pr[X > s]} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= Pr[X > t]. \end{aligned}$$

'Used is a good as new.'

2. Scaling ***Expo*****.** Let $X = Expo(\lambda)$ and $Y = aX$ for some $a > 0$. Then

$$\begin{aligned} Pr[Y > t] &= Pr[aX > t] = Pr[X > t/a] \\ &= e^{-\lambda(t/a)} = e^{-(\lambda/a)t} = Pr[Z > t] \text{ for } Z = Expo(\lambda/a). \end{aligned}$$

Thus, $a \times Expo(\lambda) = Expo(\lambda/a)$.

Some More Examples

3. Scaling Uniform. Let $X = U[0,1]$ and $Y = a + bX$ where $b > 0$. Then,

$$\begin{aligned} Pr[Y \in (y, y+\delta)] &= Pr[a+bX \in (y, y+\delta)] = Pr[X \in (\frac{y-a}{b}, \frac{y+\delta-a}{b})] \\ &= Pr[X \in (\frac{y-a}{b}, \frac{y-a}{b} + \frac{\delta}{b})] = \frac{1}{b}\delta, \text{ for } 0 < \frac{y-a}{b} < 1 \\ &= \frac{1}{b}\delta, \text{ for } a < y < a+b. \end{aligned}$$

Thus, $f_Y(y) = \frac{1}{b}$ for $a < y < a+b$. Hence, $Y = U[a, a+b]$.

4. Scaling pdf. Let $f_X(x)$ be the pdf of X and $Y = a + bX$ where $b > 0$. Then

$$\begin{aligned} Pr[Y \in (y, y+\delta)] &= Pr[a+bX \in (y, y+\delta)] = Pr[X \in (\frac{y-a}{b}, \frac{y+\delta-a}{b}] \\ &= Pr[Pr[X \in (\frac{y-a}{b}, \frac{y-a}{b} + \frac{\delta}{b}] = f_X(\frac{y-a}{b})\frac{\delta}{b}. \end{aligned}$$

Now, the left-hand side is $f_Y(y)\delta$. Hence,

$$f_Y(y) = \frac{1}{b} f_X(\frac{y-a}{b}).$$

Expectation

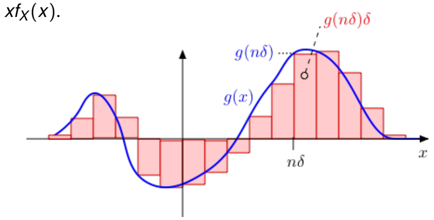
Definition The **expectation** of a random variable X with pdf $f(x)$ is defined as

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Justification: Say $X = n\delta$ w.p. $f_X(n\delta)\delta$. Then,

$$E[X] = \sum_n (n\delta) Pr[X = n\delta] = \sum_n (n\delta) f_X(n\delta)\delta = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Indeed, for any g, one has $\int g(x)dx \approx \sum_n g(n\delta)\delta$. Choose $g(x) = x f_X(x)$.



Expectation of function of RV

Definition The expectation of a function of a random variable is defined as

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx.$$

Justification: Say $X = n\delta$ w.p. $f_X(n\delta)\delta$. Then,

$$E[h(X)] = \sum_n h(n\delta) Pr[X = n\delta] = \sum_n h(n\delta) f_X(n\delta)\delta = \int_{-\infty}^{\infty} h(x) f_X(x) dx.$$

Indeed, for any g, one has $\int g(x)dx \approx \sum_n g(n\delta)\delta$. Choose $g(x) = h(x) f_X(x)$.

Fact Expectation is linear. **Proof:** As in the discrete case.

Variance

Definition: The **variance** of a continuous random variable X is defined as

$$\begin{aligned} var[X] &= E((X - E(X))^2) = E(X^2) - (E(X))^2 \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx - \left(\int_{-\infty}^{\infty} x f(x) dx\right)^2. \end{aligned}$$

Important Facts

- Concepts of independence developed for the discrete RVs apply to the continuous RVs: For independent RVs X, Y, $Pr[X \in A, Y \in B] = Pr[X \in A]Pr[Y \in B]$ and $E[XY] = E[X]E[Y]$.
- Concept of conditional probability for continuous RVs is very similar to that for discrete RVs: $h_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$, if $f_X(x) > 0$.
- Formulas/concepts for covariance, LLSE ($L[Y|X]$) and MMSE ($E[Y|X]$) are the same.
- For $X = U[a,b]$, $E[X] = \frac{a+b}{2}$, and $var[X] = \frac{(b-a)^2}{12}$.
- For $X = Expo(\lambda)$, $E[X] = 1/\lambda$, and $var[X] = 1/\lambda^2$.

Poll

Suppose life of a lightbulb has $Expo(\lambda)$ distribution with $1/\lambda = 500$ days. Given that the lightbulb has survived for 250 days, what's the expected remaining life?

- 250 days
- 500 days
- 750 days

Motivation for Gaussian Distribution

Key fact: The sum of many small independent RVs has a Gaussian distribution.

This is the Central Limit Theorem. (See later.)

Examples: Binomial and Poisson suitably scaled.

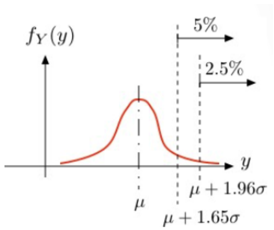
This explains why the Gaussian distribution (the bell curve) shows up everywhere.

Normal Distribution.

For any μ and σ, a **normal** (aka **Gaussian**) random variable Y, which we write as $Y = \mathcal{N}(\mu, \sigma^2)$, has pdf

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2}.$$

Standard normal has $\mu = 0$ and $\sigma = 1$.



Note: $Pr[|Y - \mu| > 1.65\sigma] = 10\%$; $Pr[|Y - \mu| > 2\sigma] = 5\%$.

Scaling and Shifting

Theorem Let $X = \mathcal{N}(0,1)$ and $Y = \mu + \sigma X$. Then

$$Y = \mathcal{N}(\mu, \sigma^2).$$

Proof: $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$. Now,

$$\begin{aligned} f_Y(y) &= \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}. \quad \square \end{aligned}$$

Expectation, Variance.

Theorem If $Y = \mathcal{N}(\mu, \sigma^2)$, then

$$E[Y] = \mu \text{ and } var[Y] = \sigma^2.$$

Proof: It suffices to show the result for $X = \mathcal{N}(0,1)$ since $Y = \mu + \sigma X$,....

Thus, $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$.

First note that $E[X] = 0$, by symmetry.

$$\begin{aligned} var[X] &= E[X^2] = \int x^2 \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\} dx \\ &= -\frac{1}{\sqrt{2\pi}} \int x d \exp\{-\frac{x^2}{2}\} = \frac{1}{\sqrt{2\pi}} \int \exp\{-\frac{x^2}{2}\} dx \text{ by IBP}^2 \\ &= \int f_X(x) dx = 1. \quad \square \end{aligned}$$

[2]Integration by Parts: $\int_a^b f dg = [fg]_a^b - \int_a^b g df$.

Central Limit Theorem.

Law of Large Numbers: For any set of independent identically distributed random variables, X_i, $A_n = \frac{1}{n}\sum X_i$ "tends to the mean."

Say X_i have expecation $\mu = E(X_i)$ and variance σ^2.

Mean of A_n is μ, and variance is σ^2/n.

Let
$$S_n := \frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then,

$$E(S_n) = \frac{1}{\sigma/\sqrt{n}}(E(A_n) - \mu) = 0$$
$$Var(S_n) = \frac{1}{\sigma^2/n}Var(A_n) = 1.$$

Central limit theorem: As n goes to infinity the distribution of S_n approaches the standard normal distribution.

Central Limit Theorem

Central Limit Theorem

Let $X_1, X_2, \ldots$ be i.i.d. with $E[X_1] = \mu$ and $var(X_1) = \sigma^2$. Define

$$S_n := \frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then,

$$S_n \to \mathcal{N}(0,1), \text{as } n \to \infty.$$

That is,

$$Pr[S_n \le \alpha] \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\alpha} e^{-x^2/2}dx.$$

Proof: See EE126.

CI for Mean

Let $X_1, X_2, \ldots$ be i.i.d. with mean μ and variance σ^2. Let

$$A_n = \frac{X_1 + \cdots + X_n}{n}.$$

The CLT states that

$$\frac{A_n - \mu}{\sigma/\sqrt{n}} = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1) \text{ as } n \to \infty.$$

Thus, for $n \gg 1$, one has

$$Pr[-2 \le \frac{A_n - \mu}{\sigma/\sqrt{n}} \le 2] \approx 95\%.$$

Equivalently,

$$Pr[\mu \in [A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}]] \approx 95\%.$$

That is,

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu.$$

CI for Mean

Let $X_1, X_2, \ldots$ be i.i.d. with mean μ and variance σ^2. Let

$$A_n = \frac{X_1 + \cdots + X_n}{n}.$$

The CLT states that

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1) \text{ as } n \to \infty.$$

Also,

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu.$$

Recall: Using Chebyshev, we found that (see Lec. 22, slide 6)

$$[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu.$$

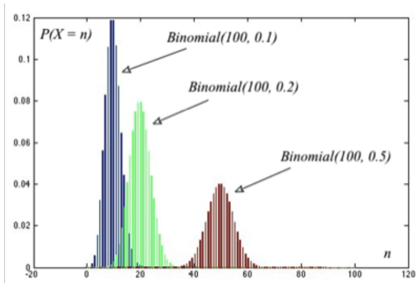
Thus, the CLT provides a smaller confidence interval.

Coins and normal.

Let $X_1, X_2, \ldots$ be i.i.d. $B(p)$. Thus, $X_1 + \cdots + X_n = B(n,p)$.

Here, $\mu = p$ and $\sigma = \sqrt{p(1-p)}$. CLT states that

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{p(1-p)n}} \to \mathcal{N}(0,1).$$



Coins and normal.

Let $X_1, X_2, \ldots$ be i.i.d. $B(p)$. Thus, $X_1 + \cdots + X_n = B(n,p)$.
Here, $\mu = p$ and $\sigma = \sqrt{p(1-p)}$. CLT states that

$$\frac{X_1 + \cdots + X_n - np}{\sqrt{p(1-p)n}} \to \mathcal{N}(0,1)$$

and

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } \mu$$

with $A_n = (X_1 + \cdots + X_n)/n$.
Hence,

$$[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } p.$$

Since $\sigma \le 0.5$,

$$[A_n - 2\frac{0.5}{\sqrt{n}}, A_n + 2\frac{0.5}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } p.$$

Thus,

$$[A_n - \frac{1}{\sqrt{n}}, A_n + \frac{1}{\sqrt{n}}] \text{ is a } 95\% - \text{CI for } p.$$

Poll

Consider repeated coin flipping for estimating the probability of heads. To have the CI width of 0.02, the number of flips should be at least

- 100
- 1000
- 10000
- 100000

Summary

Continuous Probability

1. pdf: $Pr[X \in (x, x+\delta]] = f_X(x)\delta$.
2. CDF: $Pr[X \leq x] = F_X(x) = \int_{-\infty}^{x} f_X(y)dy$.
3. $U[a,b]$, $Expo(\lambda)$, target.
4. Expectation: $E[X] = \int_{-\infty}^{\infty} x f_X(x)dx$.
5. Expectation of function: $E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x)dx$.
6. Variance: $var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$.
7. Gaussian: $\mathcal{N}(\mu, \sigma^2) : f_X(x) = ...$ "bell curve"
8. CLT: X_n i.i.d. $\implies \frac{A_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$
9. CI: $[A_n - 2\frac{\sigma}{\sqrt{n}}, A_n + 2\frac{\sigma}{\sqrt{n}}] =$ 95%-CI for μ.