Last Class

- (Finished defining CDF)
- $\mathbb{E}[g(X)] = \begin{cases} \int g(x) f(x)\,dx & \text{cont} \\ \sum g(x) p(x) & \text{disc} \end{cases}$

  $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
- $\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$

  $= \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
- $\text{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$

  $= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
- $X \perp\!\!\!\perp Y$ iff $\mathbb{P}[X \leq x, Y \leq y] = F_X(x) F_Y(y)$ FOR ALL x, y
- If $X \perp\!\!\!\perp Y \Rightarrow \text{Cov}(X,Y) = 0$

Today

- Law of Large Numbers
- Gaussian random variables
- Central Limit Thm

class poll

$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

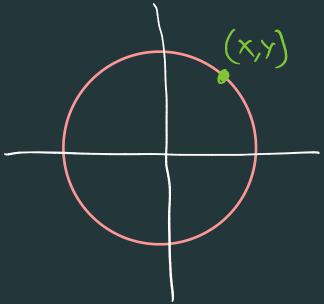
(X, Y) are uniformly distributed around unit circle $\{(x, y) : x^2 + y^2 = 1\}$

(a) $X \perp\!\!\!\perp Y$ and $Cov(X, Y) \neq 0$

(b) $X \perp\!\!\!\perp Y$ and $Cov(X, Y) = 0$

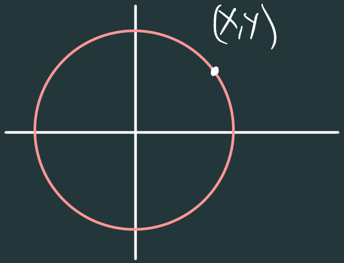
(c) $X \not\perp\!\!\!\perp Y$ and $Cov(X, Y) \neq 0$

✓ (d) $X \not\perp\!\!\!\perp Y$ and $Cov(X, Y) = 0$



class poll: solution

(X, Y) are uniformly distributed around unit circle $\{(x, y) : x^2 + y^2 = 1\}$



- Is $X \perp\!\!\!\perp Y$?

Suppose $X = x$ (eg, $X=0$)

$\Rightarrow Y \in \{\sqrt{1-x^2}, -\sqrt{1-x^2}\}$ (ie, $Y \in \{-1, 1\}$)

However Y is a cont dist^n on $[-1, 1]$

$\Rightarrow$ Not independent

- $\mathrm{Cov}(X,Y) = \underbrace{\mathbb{E}[XY]}_{} - \underset{=0}{\mathbb{E}[X]}\ \underset{=0}{\mathbb{E}[Y]}$ (by symmetry)

For $X = x$, $\mathbb{E}[Y \mid X = x] = 0$

$= 0$

linearity of expectation

for any rvs X and Y, and any constants $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

note 1: no assumptions! (in particular, does not need independence)

holds because $\mathbb{E} \equiv \int$ or $\sum$

linearity of expectation

Note - $Var(X+a) = Var(X)$ for any const a

for any rvs X and Y, and any constants $a, b \in \mathbb{R}$

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

note 1: no assumptions! (in particular, does not need independence)

note 2: does not hold for variance in general

- for general X, Y (assume $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ for convenience, see next pg for general pf)

$$Var(aX + bY) = \mathbb{E}\left[(aX+bY)^2\right] = a^2\,\mathbb{E}[X^2] + b^2\,\mathbb{E}[Y^2] + 2ab\,\mathbb{E}[XY]$$

$$= \boxed{a^2\,Var(X) + b^2\,Var(Y) + 2ab\,Cov(X,Y)}$$

- when X and Y are independent

$$Var(aX + bY) = a^2\,Var(X) + b^2\,Var(Y)$$

variance of linear combinations

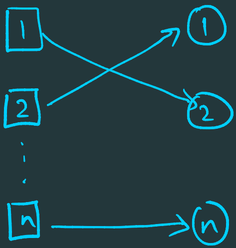
Let $\mu_X = \mathbb{E}[X]$, $\mu_Y = \mathbb{E}[Y]$ $\Rightarrow$ $\mathbb{E}[aX+bY] = a\mu_X + b\mu_Y$

$$\begin{aligned}
\mathrm{Var}(aX+bY) &= \mathbb{E}\left[\left(aX+bY-\mathbb{E}[aX+bY]\right)^2\right] \\
&= \mathbb{E}\left[\left(a(X-\mu_X)+b(Y-\mu_Y)\right)^2\right] \\
&= \mathbb{E}\left[a^2(X-\mu_X)^2 + b^2(Y-\mu_Y)^2 + 2ab(X-\mu_X)(Y-\mu_Y)\right] \\
&\overset{\text{linearity of expectation}}{=} a^2\underbrace{\mathbb{E}\left[(X-\mu_X)^2\right]}_{\mathrm{Var}(X)} + b^2\underbrace{\mathbb{E}\left[(Y-\mu_Y)^2\right]}_{\mathrm{Var}(Y)} + 2ab\underbrace{\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]}_{\mathrm{Cov}(X,Y)}
\end{aligned}$$

$$\Rightarrow \boxed{\mathrm{Var}(aX+bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X,Y)}$$

using linearity of expectation (# of derangements of a permutations)

the TAs get lazy and distribute graded assignments among n students u.a.r.
on average, how many students get their own hw?



Computing these kinds of probabilities is tricky because of correlations

Eg - What is $\mathbb{P}$[No one gets their own hw]?

(Ans: $1 - 1/1! + 1/2! - \ldots + (-1)^n/n!$, which is not easy to prove...)

- However it is easy to compute the expected # via linearity of expectations!

using linearity of expectation

the TAs get lazy and distribute graded assignments among n students u.a.r.
on average, how many students get their own hw?

Let $X_i = \mathbb{1}\left[\text{student i gets her hw}\right]$ (indicator rv) $= \begin{cases} 1 & \text{if student } i \text{ gets hw}_i \\ 0 & \text{ow} \end{cases}$

$N =$ number of students who get their own hw $= \sum_{i=1}^{10} X_i$

then we have:

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}[X_i]$$

$X_i \sim \text{Ber}(1/n)$ given no other info

Note - The X_is are highly correlated

$$= \sum_{i=1}^{n} \mathbb{P}[X_i = 1] = \sum_{i=1}^{n} \frac{1}{n} = 1$$

normal distribution

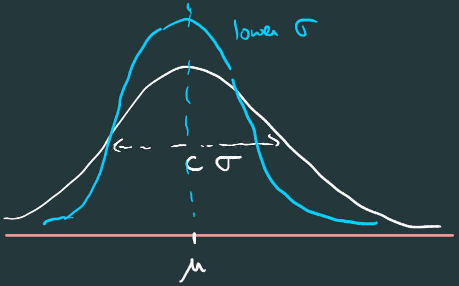
normal distribution

rv X is said to be normally distributed with mean μ and variance σ^2 (in notation, $X \sim \mathcal{N}(\mu, \sigma^2)$) if its pdf $f(\cdot)$ is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty.$$

No 'close form' for CDF

the pdf looks like



properties of the normal distribution

1. the pdf is symmetric around the mean μ: if $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\mathbb{P}[X \leq \mu - a] = \mathbb{P}[X \geq \mu + a]$$

2. (Linear transformation) If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Y = aX + b : \left.\begin{array}{l} \mathbb{E}[Y] = a\mu + b \\ \mathrm{Var}[Y] = a^2\sigma^2 \end{array}\right] \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

standard Normal

$$Z = \frac{X - \mu}{\sigma} : \left.\begin{array}{l} \mathbb{E}[Z] = 0 \\ \mathrm{Var}[Z] = \sigma^2/\sigma^2 = 1 \end{array}\right] \sim \mathcal{N}(0, 1)$$

3. If $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, and $X \perp\!\!\!\perp Y$, then

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

'Stable distribution' under addition

$$\mathbb{E}[X+Y] = \mu_1 + \mu_2, \quad \mathrm{Var}(X+Y) = \sigma_1^2 + \sigma_2^2$$

cdf of normal distribution

if $X \sim \mathcal{N}(\mu, \sigma^2)$, then its cdf is given by

$$\mathbb{P}[X \le x] = \mathbb{P}\left[\frac{X-\mu}{\sigma} \le \underbrace{\frac{x-\mu}{\sigma}}_{y}\right] = \underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-x^2/2}\,dx}_{\Phi(y)}$$

- knowing cdf for $\mathcal{N}(0,1)$ is enough to find cdf for any normally distributed rv
- for $X \sim \mathcal{N}(0,1)$: cdf denoted $\Phi(x)$, available in most computing packages

  closely related to the error function $erf(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt$; in particular,

$$\Phi(x) = \frac{1}{2}\left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right)$$

What you may think is the CLT -

If $X_1, X_2, X_3 \ldots X_n$ are iid, then

$\sum X_i$ is 'like a Gaussian'

class poll: where is that Normal rv?

$\mathbb{E}[X_i] = 1/2$, $\mathrm{Var}(X_i) = 1/12$

$f(x) = 1$ for $x \in [0,1]$

given independent and identically distributed (iid) $X_i \sim U[0,1]$, which of the following is well approximated as a $\mathcal{N}(0,1)$?

(a) their sum $S_n = \sum_{i=1}^n X_i$ NO, $\mathbb{E}[S_n] = \frac{n}{2}$, $S_n \in [0, n]$

(b) their average $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ NO, $\mathbb{E}[\bar{X}_n] = \frac{1}{2}$, $S_n \in [0,1]$

(c) their centered sum $\hat{S}_n = \sum_{i=1}^n (X_i - 0.5)$ NO, $\mathbb{E}[\hat{S}_n] = 0$, $S_n \in [\frac{-n}{2}, \frac{n}{2}]$

(d) their centered average $\hat{Y}_n = \frac{1}{n}\sum_{i=1}^n (X_i - 0.5)$ NO - $\mathbb{E}[\hat{Y}_n] = 0$

$\mathrm{Var}(\hat{Y}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i)$ because X_i's are ⫫

$= \mathrm{Var}(X_i)/n$

(e) none of these ✓

$\hat{Z}_n = \sqrt{\frac{12}{n}} \sum_{i=1}^n (X_i - 0.5)$

$\sqrt{n}$ normalization!

sums and averages of independent rv

See notebook

- $X_1, X_2, \ldots$ are independent rv that are uniformly distributed over the interval $[0, 1]$; $\mathbb{E}[X_1] = 1/2$, $Var(X_1) = 1/12$.
- the pdf of X_1 looks like
- what about the pdf of $X_1 + X_2$ and $(X_1 + X_2)/2$?

sums and averages of independent rv

$\mu_x = \mathbb{E}[X_i]$
$\sigma_x^2 = Var(X_i)$

$$S_n = X_1 + \ldots + X_n \qquad \bar{X}_n = \frac{1}{n}[X_1 + \ldots + X_n]$$

- $\mathbb{E}[S_n] = n\mu_x$

  $\mathbb{E}[\bar{X}_n] = \mu_x$

  $Var(S_n) = n\sigma_x^2$

  $Var(\bar{X}_n) = \frac{\sigma_x^2}{n}$

- (roughly) sum of n i.i.d. random variables is $\sqrt{n}$ times as variable as any one of the random variables
- average of n i.i.d. random variables is $1/\sqrt{n}$ times as variable as any one of the random variables

law of large numbers

let $X_1, X_2, \ldots$ be a sequence of independent rvs with $\mathbb{E}[X_i] = \mu$ for all i

then, "almost" always

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow \mu \quad , \quad \text{as } n \to \infty$$

note: for any finite n, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ is still a random variable

'Ergodic Thm'

Philosophical Basis for simulation

- averaging iid experiments gives a 'good' estimate of the mean

?

Avg across space = Avg across time

central limit theorem ('scientific basis' for simulation)

'universality thms'

let $X_1, X_2, \ldots$ be a sequence of independent rvs with
$$\mathbb{E}[X_i] = \mu,\ Var(X_i) = \sigma^2 < \infty \text{ for all } i$$
then,
$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} \sigma\mathcal{N}(0,1) = \mathcal{N}(0,\sigma^2) \quad , \quad \text{as } n \to \infty$$

this suggests the following approximations for large n,

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{D}{\approx} \mathcal{N}(\mu, \sigma^2/n)$$

$$S_n = \sum_{i=1}^{n} X_i \overset{D}{\approx} \mathcal{N}(n\mu, n\sigma^2)$$

These are 'approximations' based on CLT-like thinking
In reality, if $n \to \infty$, these are not meaningful