Dis 14A: Continuous Probability & CLT

- See Calculus & Probability Review on my website.

Review

Distributions:

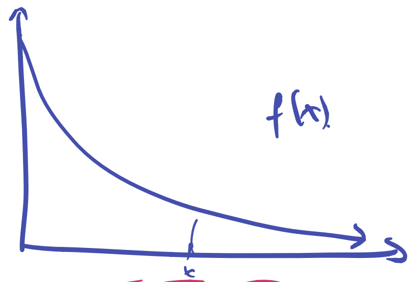
Exponential: $X \sim \text{Expo}(\lambda)$. ← parameter

"continuous time geometric".

pdf: $f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{o/w} \end{cases}$

$\mathbb{E}[X] = 1/\lambda$

$\text{Var}(X) = 1/\lambda^2$.



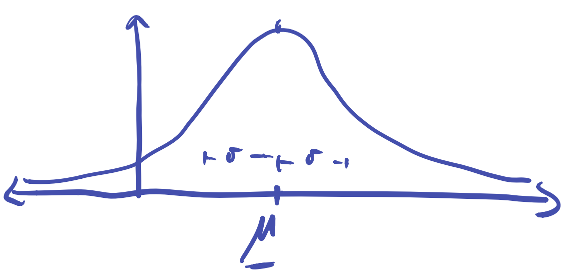
cdf: $\mathbb{P}[X \leq x] = 1 - e^{-\lambda x}$.

Normal: $X \sim N(\mu, \sigma^2)$

"bell-shaped curve"

pdf: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$\left.\begin{array}{l} \mathbb{E}[X] = \mu \\ \text{Var}(X) = \sigma^2 \end{array}\right\}$ parameters of a normal r.v. !!!



be familiar w/ the idea of a "z-score" (see impt facts)

impt facts:

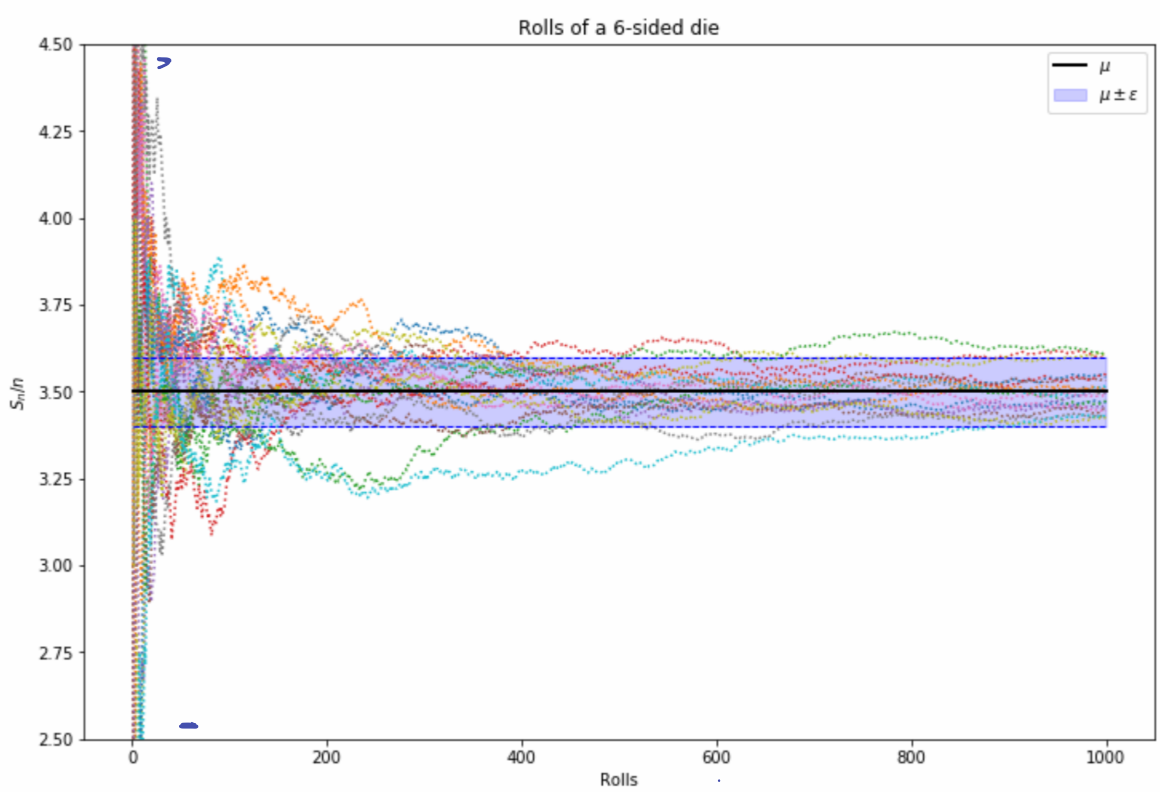
- shifting & scaling a normal gives another normal r.v. useful for "normalizing," i.e. converting $X \sim N(\mu, \sigma^2)$ into the standard normal $Z \sim N(0,1)$ by $Z = \frac{X-\mu}{\sigma}$.
- the sum of independent normals is again normal. (special! doesn't happen w/ geometrics or binomials).

i.e., $\begin{array}{l} X \sim N(0,1) \\ Y \sim N(0,1) \end{array} \Rightarrow Z = aX + bY \sim N(0, a^2 + b^2)$.

Central Limit Theorem

★ Intuition: the weak LLN tells us that if $X_1, \ldots, X_n$ are i.i.d. RV's w/ mean μ, their average will eventually tend to μ (the blue bounding box below).

Question: Exactly how close for a particular n? How confident can we be?



Central Limit Theorem (essentially): Let $X_1, \ldots, X_n$ be i.i.d. random variables. Then

$$S_n = X_1 + \cdots + X_n$$

is approximately Gaussian for large enough n.

How to use this in practice? Standardize the distribution. e.g., say $\mathbb{E}[X_i] = \mu$, $Var(X_i) = \sigma^2$. Then $\mathbb{E}[S_n] = n\mu$, $Var(S_n) = n\sigma^2$, so standardize w/ $Z = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \overset{d}{\sim} N(0,1)$.

★ instead of looking at S_n, use z-score's and look at Z.

★ work through my CLT handout, I guarantee it will make everything crystal clear.

$X_1, \dots, X_n$ are i.i.d., $X_i \sim \text{Expo}(\lambda)$

$S_n = X_1 + \dots + X_n \sim \boxed{N(n/\lambda, n/\lambda^2)}$

$\mathbb{E}[X_i] = 1/\lambda \quad \Rightarrow \quad \mathbb{E}[S_n] = n/\lambda$

$Var(X_i) = 1/\lambda^2 \qquad Var(S_n) = n/\lambda^2$

$$Z = \boxed{\frac{S_n - n/\lambda}{\sqrt{n/\lambda^2}}} \sim \boxed{N(0,1)}$$

$n = 100$
$\lambda = 10$

$S_n = (10, 1)$

$\mathbb{P}(|S_n - 10| \le 2) = \mathbb{P}(|Z| \le 2)$

standard normal cdf

$= \Phi(2) - \Phi(-2)$

$=$

S_n

S_n.

$= 95\% = 0.95$.

1 First Exponential to Die

Let X and Y be Exponential(λ_1) and Exponential(λ_2) respectively, independent. What is

$$\mathbb{P}(\min(X,Y)=X),$$

the probability that the first of the two to die is X?

hint: CDF of $X \sim \text{Expo}(\lambda)$ is $\mathbb{P}(X \le x) = 1 - e^{-\lambda x}$. $= \mathbb{P}(X \ge x)$.

$$\mathbb{P}(X<Y) = \text{``}\mathbb{P}(X<Y \mid Y=y)\ \mathbb{P}(Y=y)\text{''}$$

$$\left(\mathbb{P}(X<Y) = \sum_{y=0}^{\infty} \mathbb{P}(X<Y \mid Y=y)\cdot\mathbb{P}(Y=y)\right) \quad \text{discrete.}$$

$$= \int_0^{\infty} \mathbb{P}(X<y \mid Y=y)\cdot f_Y(y)\,dy.$$

$$= \int_0^{\infty} (1-e^{-\lambda_1 y})\lambda_2 e^{-\lambda_2 y}\,dy$$

$$= \lambda_2 \int_0^{\infty} e^{-\lambda_2 y} - e^{-(\lambda_1+\lambda_2)y}\,dy$$

$$= \lambda_2\left(\frac{-1}{\lambda_2}e^{-\lambda_2 y}\Big|_0^{\infty} - \frac{-1}{\lambda_1+\lambda_2}e^{-(\lambda_1+\lambda_2)y}\Big|_0^{\infty}\right)$$

$$= \lambda_2\left(\frac{1}{\lambda_2} - \frac{1}{\lambda_1+\lambda_2}\right) = 1 - \frac{\lambda_2}{\lambda_1+\lambda_2}$$

$$= \boxed{\frac{\lambda_1}{\lambda_1+\lambda_2}}$$

2 Chebyshev's Inequality vs. Central Limit Theorem

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the following distribution:

$$\mathbb{P}[X_i = -1] = \frac{1}{12}; \qquad \mathbb{P}[X_i = 1] = \frac{9}{12}; \qquad \mathbb{P}[X_i = 2] = \frac{2}{12}.$$

(a) Calculate the expectations and variances of X_1, $\boxed{\sum_{i=1}^n X_i}$, $\sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, and

$S_n = X_1 + \cdots + X_n$.

$$Z_n = \cancel{\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}}.$$

$\mathbb{E}[X_1] = 1$, $Var(X_1) = 1/2$.

$\mathbb{E}[S_n] = n$. $Var(S_n) = n/2$.

$Var(aX) = a^2 Var(X)$.

$$Z_n = \frac{S_n - n}{\sqrt{n/2}}$$

$\mathbb{E}[Z_n] = 0$

$Var(Z_n) = 1$.

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n| \geq 2]$.

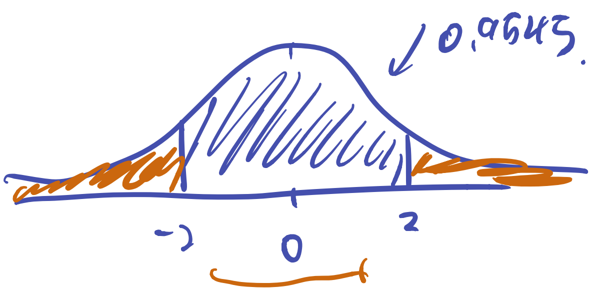
$$\mathbb{P}(|Z_n| \ge 2) \le \frac{Var(Z_n)}{2^2} = \frac{1}{4} = b.$$

(c) Can you use b to bound $\underbrace{\mathbb{P}[Z_n \geq 2]}_{\le \frac14}$ and $\underbrace{\mathbb{P}[Z_n \leq -2]}_{\le \frac14}$? $= 0.25$

Φ - standard normal cdf.

(d) As $n \to \infty$, what is the distribution of Z_n? $\sim N(0,1)$

(e) We know that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{P}[|Z| \leq 2] = \Phi(2) - \Phi(-2) \approx 0.9545$. As $n \to \infty$, can you provide approximations for $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$?



$$\frac{1 - 0.9545}{2} = 0.0275$$

$\mathbb{P}(Z_n \ge 2) \approx$

$\mathbb{P}(Z_n \le -2) \approx$.

3 Why Is It Gaussian?

Let X be a normally distributed random variable with mean μ and variance σ^2. Let $Y = aX + b$, where $a > 0$ and b are non-zero real numbers. Show explicitly that Y is normally distributed with mean $a\mu + b$ and variance $a^2\sigma^2$. The PDF for the Gaussian Distribution is $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. One approach is to start with the cumulative distribution function of Y and use it to derive the probability density function of Y.

[1.You can use without proof that the pdf for any gaussian with mean and sd is given by the formula $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ where μ is the mean value for X and σ^2 is the variance. 2. The drivative of CDF gives PDF.]