Continuous RVs Continued: Independence, Conditioning, Gaussians, CLT

CS 70, Summer 2019

Lecture 25, 8/6/19

Not Too Different From Discrete...

Discrete RV:
X and Y are independent iff for all a, b:

$$\mathbb{P}[X = a, Y = b] = \mathbb{P}[X = a] \cdot \mathbb{P}[Y = b]$$

Continuous RV:
X and Y are independent iff for all $a \le b$, $c \le d$:

$$\mathbb{P}[a \le X \le b, c \le Y \le d] = \mathbb{P}[a \le X \le b] \times \mathbb{P}[c \le Y \le d]$$

A Note on Independence

For continuous RVs, what is weird about the following?

$$\underbrace{\mathbb{P}[X = a, Y = b]}_{=0} = \underbrace{\mathbb{P}[X = a]}_{=0} \cdot \underbrace{\mathbb{P}[Y = b]}_{=0}$$

What we **can** do: consider a interval of length dx around a and b!

$$\mathbb{P}[X=a, Y=b] \approx \mathbb{P}[X \in [a, a+dx], Y \in [b, b+dy]]$$
$$= \mathbb{P}[X \in [a, a+dx]]\,\mathbb{P}[Y \in [b, b+dy]]$$
$$\approx (f_X(a)\,dx)(f_Y(b)\,dy)$$

Independence, Continued

If X, Y are independent, their joint density is the product of their individual densities:

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$$

Example: If X, Y are independent exponential RVs with parameter λ:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$
$$= (\lambda e^{-\lambda x})(\lambda e^{-\lambda y})$$
$$= \lambda^2 e^{-\lambda(x+y)}$$

Example: Max of Two Exponentials

Let $X \sim \text{Expo}(\lambda)$ and $Y \sim \text{Expo}(\mu)$.
X and Y are **independent**.
Compute $\mathbb{P}[\max(X, Y) \ge t]$.

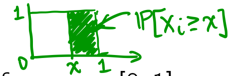
$$= 1 - \mathbb{P}[\max(X,Y) \le t]$$
$$= 1 - \mathbb{P}[X \le t, Y \le t]$$
independence → $= 1 - \mathbb{P}[X \le t]\,\mathbb{P}[Y \le t]$
use CDF → $= 1 - (1 - e^{-\lambda t})(1 - e^{-\mu t})$

$$e^{-\lambda t} + e^{-\mu t} - e^{-(\lambda+\mu)t}$$

Use this to compute $\mathbb{E}[\max(X, Y)]$.

Tail Sum: $\mathbb{E}[\max] = \int_0^\infty \mathbb{P}[\max \ge t]\,dt$
$$= \int_0^\infty (e^{-\lambda t} + e^{-\mu t} - e^{-(\lambda+\mu)t})\,dt$$
post-integration → $= \frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\lambda+\mu}$

Min of n Uniforms



Let $X_1, \dots, X_n$ be **i.i.d.** and uniform over $[0, 1]$.

↳ $\mathbb{P}[a \le X \le b] = b - a$ for $0 \le a \le b \le 1$

$f_X(x) = \begin{cases} 1 & \text{over } [0,1] \\ 0 & \text{o.w.} \end{cases}$

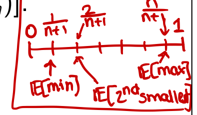
What is $\mathbb{P}[\min(X_1, \dots, X_n) \le x]$?

↳ $1 - \mathbb{P}[\min \ge x]$
$= 1 - \mathbb{P}[X_1 \ge x, \dots, X_n \ge x]$
ind → $= 1 - \mathbb{P}[X_1 \ge x] \cdot \ldots \cdot \mathbb{P}[X_n \ge x]$
$= 1 - (1-x)^n$

Use this to compute $\mathbb{E}[\min(X_1, \dots, X_n)]$.

Tail Sum: $\int_0^\infty \underbrace{\mathbb{P}[\min \ge x]}_{\text{prev.}}\,dx$
$$= \int_0^1 (1-x)^n\,dx$$
$$= \left(-\frac{(1-x)^{n+1}}{n+1}\right)\Big|_0^1 = 0 - \left(-\frac{1}{n+1}\right) = \frac{1}{n+1}$$

n iid unif([0,1]) in expectation:



Min of n Uniforms

from prev. slide

What is the CDF of $\min(X_1, \ldots, X_n)$?

$$F_{min}(x) = \mathbb{P}[\min \le x] = 1-(1-x)^n$$

What is the PDF of $\min(X_1, \ldots, X_n)$?

$$\frac{d}{dx}\left(1-(1-x)^n\right)$$

$$0 - n(1-x)^{n-1}(-1)$$

$$f_{min}(x) = n(1-x)^{n-1}$$

Memorylessness of Exponential

We can't talk about independence without talking about **conditional probability**!

Let $X \sim \text{Expo}(\lambda)$. X is **memoryless**, i.e.

$$\mathbb{P}[X \geq s + t | X > t] = \mathbb{P}[X \geq s]$$

LHS: $\dfrac{\mathbb{P}[\{X \ge s+t\} \cap \{X > t\}]}{\mathbb{P}[X > t]}$ redundant event

$$= \frac{\mathbb{P}[X \ge s+t]}{\mathbb{P}[X>t]} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} \qquad = \int_{s+t}^{\infty} \lambda e^{-\lambda x}\,dx$$

$$= e^{-\lambda s} = \mathbb{P}[X \ge s]$$

Conditional Density

What happens if we condition on events like $X = a$? These have 0 probability!

The same story as discrete, except we now need to define a conditional **density**:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

convention: set this to 0 when $f_X(x) = 0$.

Think of $f_{Y|X}(y|x)$ as

$$\mathbb{P}[Y \in [y, y+dy] | X \in [x, x+dx]]$$

Conditional Density, Continued

Given a conditional density $f_{Y|X}$, compute

$$\mathbb{P}[Y \le y|X = x] = \int_{-\infty}^{y} f_{Y|X}(z|x)\,dz$$

$\approx \mathbb{P}[Y = y | X = x]$ ($[y, y+dy]$, $[x, x+dx]$)

If we know $\mathbb{P}[Y \le y|X = x]$, compute

$$\mathbb{P}[Y \le y] = \int_{-\infty}^{\infty} \mathbb{P}[Y \le y | X = x] f_X(x)\,dx$$

Total prob. Rule.

If discrete: case on x! $\sum_x \mathbb{P}[Y \le y | X = x] \cdot \mathbb{P}[X = x]$

Go with your gut! What worked for discrete also works for continuous.

Example: Sum of Two Exponentials

Let X_1, X_2 be **i.i.d** Expo(λ) RVs.
Let $Y = X_1 + X_2$.

What is $\mathbb{P}[Y < y|X_1 = x]$?

↳ $\mathbb{P}[X_1 + X_2 < y | X_1 = x]$

$= \mathbb{P}[X_2 < y - x]$

CDF of expo: $1 - e^{-\lambda(y-x)}$

What is $\mathbb{P}[Y < y]$?

case on values for X_1!

$$\int_0^y \mathbb{P}[Y < y | X_1 = x] \cdot \mathbb{P}[X_1 = x]$$

$$= \int_0^y \left(1 - e^{-\lambda(y-x)}\right) f_{X_1}(x)\,dx$$

$$= \int_0^y \left(1 - e^{-\lambda(y-x)}\right) \lambda e^{-\lambda x}\,dx.$$ → Exercise.

Example: Total Probability Rule

What is the CDF of Y? Exercise

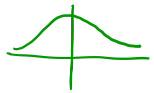
What is the PDF of Y? "

Break

If you could immediately gain one new skill, what would it be?

The Normal (Gaussian) Distribution

X is a **normal** or **Gaussian** RV if:



$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(x-\mu)^2/2\sigma^2}$$

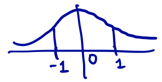
Symm. about its mean

Parameters: μ, σ

Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$

$\mathbb{E}[X] = \mu$ $\quad$ $\text{Var}(X) = \sigma^2$

Standard Normal: $\mu = 0, \sigma^2 = 1$



Gaussian Tail Bound

Let $X \sim \mathcal{N}(0,1)$.

Easy upper bound on $\mathbb{P}[|X| \geq \alpha]$, for $\alpha \geq 1$?

(Something we've seen before...)

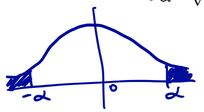
Chebyshev:

$$\mathbb{P}[|X - 0| \geq \alpha] \leq \frac{\text{Var}(X)}{\alpha^2} \leq \frac{1}{\alpha^2}$$

Gaussian Tail Bound, Continued

Turns out we can do better than Chebyshev.

Idea: Use $\int_\alpha^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \leq \int_\alpha^\infty \frac{x}{\sqrt{2\pi}} e^{-x^2/2} dx$



Shaded: $|X| \geq \alpha$

$$\mathbb{P}[|X| \geq \alpha] = 2\,\mathbb{P}[X \geq \alpha] = 2\int_\alpha^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$\leq 2\int_\alpha^\infty \frac{x}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$= 2\left[\left(\frac{1}{\sqrt{2\pi}}\right)\left(e^{-x^2/2}\right)(-1)\right]\Big|_\alpha^\infty$$

$$= \frac{2}{\sqrt{2\pi}} e^{-\alpha^2/2}$$

Shifting and Scaling Gaussians

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = \frac{X-\mu}{\sigma}$. Then:

$$Y \sim \mathcal{N}(0, 1)$$

Proof: Compute $\mathbb{P}[a \leq Y \leq b]$.

Notes: out of scope.

Change of variables: $x = \sigma y + \mu$.

Shifting and Scaling Gaussians

Can also go the other direction:

If $X \sim \mathcal{N}(0,1)$, and $Y = \mu + \sigma X$:

Y **is still Gaussian!**

$\mathbb{E}[Y] = \mathbb{E}[\mu + \sigma X] = \mu + \sigma \underbrace{\mathbb{E}[X]}_{0} = \mu$

$\text{Var}(Y) = \text{Var}(\mu + \sigma X) = \text{Var}(\sigma X) = \sigma^2 \underbrace{\text{Var}(X)}_{=1} = \sigma^2$

Sum of Independent Gaussians

Let X, Y be **independent** standard Gaussians.

Let $Z = [aX + c] + [bY + d]$.
Then, Z is **also Gaussian!** (Proof optional.)

$\mathbb{E}[Z] = \mathbb{E}[aX + c + bY + d] = a\cancel{\mathbb{E}[X]}^{\to 0} + b\cancel{\mathbb{E}[Y]}^{\to 0} + c + d$

$= c + d$

$\text{Var}(Z) = \text{Var}(aX + bY + c + d)$
$= \text{Var}(aX + bY)$ shift
$= \text{Var}(aX) + \text{Var}(bY)$
$= a^2 \text{Var}(X) + b^2 \text{Var}(Y) = a^2 + b^2$

Example: Height Exercise

Consider a family of a two parents and twins with the same height. The parents' heights are independently drawn from a $\mathcal{N}(65, 5)$ distribution. The twins' height are independent of the parents', and from a $\mathcal{N}(40, 10)$ distribution.

Let H be the sum of the heights in the family. Define relevant RVs:

Example: Height

$\mathbb{E}[H] =$

$\text{Var}[H] =$

Sample Mean

We sample a RV X **independently** n times.
X has mean μ, variance σ^2.

Denote the **sample mean** by $A_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$

$\mathbb{E}[\cancel{X} A_n] = \mathbb{E}\left[\frac{1}{n}(X_1 + \ldots + X_n)\right] = \frac{1}{n}\left(\mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]\right)$

$= \frac{1}{n}(\cancel{n}\mu) = \mu$

$\text{Var}(\cancel{X} A_n) = \text{Var}\left(\frac{1}{n}(X_1 + \ldots + X_n)\right)$
$= \frac{1}{n^2}[\text{Var}(X_1) + \ldots + \text{Var}(X_n)]$
$= \frac{1}{n^{\cancel{2}}}(\cancel{n}\sigma^2) = \frac{\sigma^2}{n}$

The Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots, X_n$ be **i.i.d.** RVs with mean μ, variance σ^2. (Assume mean, variance, are finite.)

Sample mean, as before: $A_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$
Recall: $\mathbb{E}[A_n] = \mu$
$\text{Var}(A_n) = \frac{\sigma^2}{n}$

Normalize the sample mean:

normalized sample mean $A'_n = \frac{A_n - \mu}{\sigma / \sqrt{n}}$ standard deviation

Then, as $n \to \infty$, $\cancel{\mathbb{P}[A'_n] \to}$ easier way: A'_n as $n \to \infty$ follows a $\mathcal{N}(0,1)$ dist.

Example: Chebyshev vs. CLT

Let $X_1, X_2, \ldots$ be **i.i.d** RVs with $\mathbb{E}[X_i] = 1$ and $\text{Var}(X_i) = \frac{1}{2}$. Let $A_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$.

$\mathbb{E}[A_n] = 1$ ← expectation of single sample.

$\text{Var}(A_n) = \frac{\sigma^2}{n} = \frac{1}{2n}$ σ^2 ← variance of a single sample

Normalize to get A'_n: $\frac{A_n - \mathbb{E}[A_n]}{\sqrt{\text{Var}(A_n)}}$

$\mathbb{E}[A'_n] = \mathbb{E}\left[\frac{A_n - 1}{\sigma / \sqrt{n}}\right] = 0$

$\text{Var}(A'_n) = \text{Var}\left(\frac{A_n - 1}{\sigma / \sqrt{n}}\right) = 1$ Try to check this!

Example: Chebyshev vs. CLT

Upper bound $\mathbb{P}[A'_n \geq 2]$ for **any** n.
(We don't know if A'_n is **non-neg** or **symmetric**.)

$$\mathbb{P}[A_n' \geq 2] \leq \mathbb{P}[|A_n' - \underset{\uparrow \atop \mathbb{E}[A_n']}{0}| \geq 2] \leq \frac{\mathrm{Var}(A_n')}{2^2}$$

$$\leq \frac{1}{4}$$

If we take $n \to \infty$, upper bound on $\mathbb{P}[A'_n \geq 2]$?

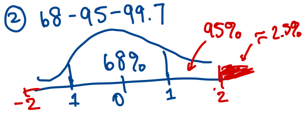
① Gaussian Tail

$$\lim_{n \to \infty} \mathbb{P}[A_n' \geq 2] \leq \text{const} \times e^{-2^2/2} = \text{const} \times \frac{1}{e^2}$$

↓

approaches $\mathcal{N}(0,1)$ dist.

② 68-95-99.7



Summary

- Independence and conditioning also generalize from the **discrete** RV case.
- The Gaussian is a very important continuous RV. It has several nice properties, including the fact that adding independent Gaussians gets you another Gaussian
- The CLT tells us that if we take a **sample average** of a RV, the distribution of this average will approach a **standard normal**.