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 $a \leq b$, $c \leq d$:

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

$$\mathbb{P}[a \leq X \leq b] \times \mathbb{P}[c \leq Y \leq d]$$

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

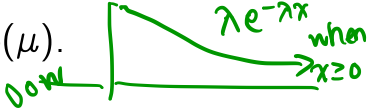
If X, Y are independent, their joint density is the product of their individual densities:

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

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

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

Example: Max of Two Exponentials



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

$$= 1 - \mathbb{P}[\max(X,Y) \leq t]$$

$$= 1 - \mathbb{P}[X \leq t, Y \leq t]$$

independence $\rightarrow$ $$= 1 - \mathbb{P}[X \leq t]\,\mathbb{P}[Y \leq t]$$

use CDF $\rightarrow$ $$= 1 - (1 - e^{-\lambda t})(1 - e^{-\mu t})$$

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

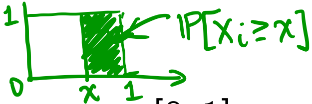
Use this to compute $\mathbb{E}[\max(X, Y)]$.

Tail Sum: $$\mathbb{E}[\max] = \int_0^\infty \mathbb{P}[\max \geq t]\, dt$$

$$= \int_0^\infty \left(e^{-\lambda t} + e^{-\mu t} - e^{-(\lambda+\mu)t}\right) dt$$

post-integration $\longrightarrow$ $$= \frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\lambda+\mu}$$

Min of n Uniforms



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

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

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

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

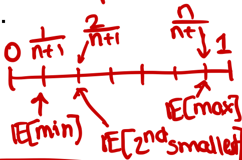
$\hookrightarrow 1 - \mathbb{P}[\min \geq x]$

$= 1 - \mathbb{P}[X_1 \geq x, \ldots, X_n \geq x]$

ind $\rightarrow$ $= 1 - \mathbb{P}[X_1 \geq x] \cdot \ldots \cdot \mathbb{P}[X_n \geq x]$

$= 1 - (1-x)^n$

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



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

Tail Sum: $\displaystyle\int_0^\infty \underbrace{\mathbb{P}[\min \geq x]}_{\text{prev.}}\, dx$

$= \displaystyle\int_0^1 (1-x)^n \, dx$

$= \left(-\dfrac{(1-x)^{n+1}}{n+1}\right)\Bigg|_0^1 = 0 - \left(-\dfrac{1}{n+1}\right) = \dfrac{1}{n+1}$

Min of n Uniforms

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

from prev. slide

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

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: $\dfrac{\mathbb{P}[\{X \geq s+t\} \cap \{X > t\}]}{\mathbb{P}[X > t]}$

redundant event

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

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

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 $\underset{Y|X}{f}(y|x)$ as

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

Conditional Density, Continued

Given a conditional density $f_{Y|X}$, compute $\approx \mathbb{P}[Y=y \mid X=x]$

$[y, y+dy]$ $[x, x+dx]$

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

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

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

Total Prob. Rule.

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

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 $Y = X_1 + X_2$.

CDF of expo:

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

$\hookrightarrow \mathbb{P}[X_1 + X_2 < y \mid X_1 = x]$

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

$1 - e^{-\lambda(y-x)}$

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

case on values for X_1!

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

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

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

Example: Total Probability Rule

What is the CDF of Y? Exercise

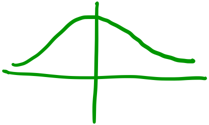
What is the PDF of Y? "

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:



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

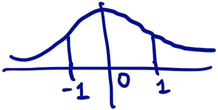
Symm. about its mean

Parameters: μ, σ

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

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

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:

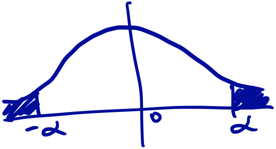
$$\mathbb{P}[|X-0| \geq \alpha] \leq \frac{Var(X)}{\alpha^2}$$

$$\leq \frac{1}{\alpha^2}$$

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!**

$$\mathbb{E}[Y] = \mathbb{E}[\mu + \sigma X] = \mu + \sigma \cancel{\mathbb{E}[X]}^{\,0}$$
$$= \mu$$

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

Sum of Independent Gaussians

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

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

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

$= c + d$

$\mathrm{Var}(Z) =$ $\mathrm{Var}(aX + bY + c + d)$

$= \mathrm{Var}(aX + bY)$ ✓ shift

$= \mathrm{Var}(aX) + \mathrm{Var}(bY)$

$= a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y) = a^2 + b^2$

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

$$\mathbb{E}[H] =$$

$$\mathrm{Var}[H] =$$

Sample Mean

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

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

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

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

$$\mathrm{Var}(\cancel{X}\,A_n) = \mathrm{Var}\left(\frac{1}{n}(X_1+\ldots+X_n)\right)$$

$$= \frac{1}{n^2}\left[\mathrm{Var}(X_1)+\ldots+\mathrm{Var}(X_n)\right]$$

$$= \frac{1}{\cancel{n^2}}(\cancel{n}\sigma^2) = \frac{\sigma^2}{n}$$

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. $\rightarrow A_n' = \dfrac{A_n - \mu}{\underbrace{\sigma/\sqrt{n}}_{\text{standard deviation}}}$

Then, as $n \to \infty$, ~~$\mathbb{P}[A_n'] \to$~~ 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 $\mathrm{Var}(X_i) = \frac{1}{2}$. Let $A_n = \frac{X_1+X_2+\ldots+X_n}{n}$.

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

$\mathrm{Var}(A_n) =$ $\frac{\sigma^2}{n} = \frac{1}{2n}$

σ^2 ← variance of a single sample

Normalize to get A_n':

$\frac{A_n - \mathbb{E}[A_n]}{\sqrt{\mathrm{Var}(A_n)}}$

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

$\mathrm{Var}(A_n') = \mathrm{Var}\left(\frac{A_n - 1}{\sigma/\sqrt{n}}\right) = 1$

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**.)

$$\mathbb{P}[A_n' \geq 2] \leq \mathbb{P}[|A_n' - \underbrace{0}_{\mathbb{E}[A_n']}| \geq 2] \leq \frac{\mathrm{var}(A_n')}{2^2}$$

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

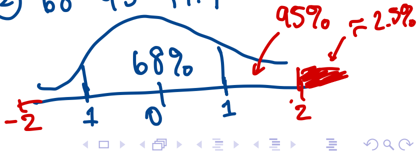
If we take $n \to \infty$, upper bound on $\mathbb{P}[A'_n \geq 2]$?

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**.