Convolution

Chris Piech
CS109, Stanford University

What happens when you add random variables?

Sum of Independent Binomials

- Let X and Y be independent random variables
  - $X \sim \text{Bin}(n_1, p)$ and $Y \sim \text{Bin}(n_2, p)$
  - $X + Y \sim \text{Bin}(n_1 + n_2, p)$
- Intuition:
  - X has n_1 trials and Y has n_2 trials
    - Each trial has same “success” probability p
  - Define Z to be $n_1 + n_2$ trials, each with success prob. p
  - $Z \sim \text{Bin}(n_1 + n_2, p)$, and also $Z = X + Y$

If only it were always that simple

The Insight to Convolution Proofs

$P(X + Y = n)$?

What is the probability that X + Y = n?

X	Y	k	
0	n	0	$P(X = 0, Y = n)$
1	$n - 1$	1	$P(X = 1, Y = n-1)$
2	$n - 2$	2	$P(X = 2, Y = n-2)$
	$\cdots$		
n	0	n	$P(X = n, Y = 0)$

The Insight to Convolution Proofs

P(X + Y = n)?

What is the probability that X + Y = n?

Since this is the OR or mutually exclusive events

$$P(X+Y=n)=\sum_{k=0}^{n}P(X=k,Y=n-k)$$

If the random variables are independent

$$=\sum_{k=0}^{n}P(X=k)P(Y=n-k)$$

Sum of Independent Poissons

Recall the Binomial Theorem

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Sum of Independent Poissons

- Let X and Y be independent random variables
  - X ~ Poi(λ_1) and Y ~ Poi(λ_2)
  - X + Y ~ Poi(λ_1 + λ_2)
- Proof: (just for reference)
  - Rewrite (X + Y = *n*) as (X = *k*, Y = *n* – *k*) where 0 ≤ *k* ≤ *n*

$$P(X+Y=n)=\sum_{k=0}^{n}P(X=k,Y=n-k)=\sum_{k=0}^{n}P(X=k)P(Y=n-k)$$

$$=\sum_{k=0}^{n}e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}=e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}$$

  - Noting Binomial theorem: $(\lambda_1+\lambda_2)^n=\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}$
  - $P(X+Y=n)=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\left(\lambda_1+\lambda_2\right)^n$ so, X + Y = n ~ Poi(λ_1 + λ_2)

Reference: Sum of Independent RVs

- Let X and Y be independent Binomial RVs
  - $X \sim \text{Bin}(n_1, p)$ and $Y \sim \text{Bin}(n_2, p)$
  - $X + Y \sim \text{Bin}(n_1 + n_2, p)$
  - More generally, let $X_i \sim \text{Bin}(n_i, p)$ for $1 \leq i \leq N$, then

$$\left(\sum_{i=1}^{N} X_i\right) \sim \text{Bin}\left(\sum_{i=1}^{N} n_i,\ p\right)$$

- Let X and Y be independent Poisson RVs
  - $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$
  - $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$
  - More generally, let $X_i \sim \text{Poi}(\lambda_i)$ for $1 \leq i \leq N$, then

$$\left(\sum_{i=1}^{N} X_i\right) \sim \text{Poi}\left(\sum_{i=1}^{N} \lambda_i\right)$$

Convolution of Probability Distributions

We talked about sum of Binomial and Poisson…who's missing from this party?

Uniform.

Summation: not just for the 1%

Dance, Dance Convolution

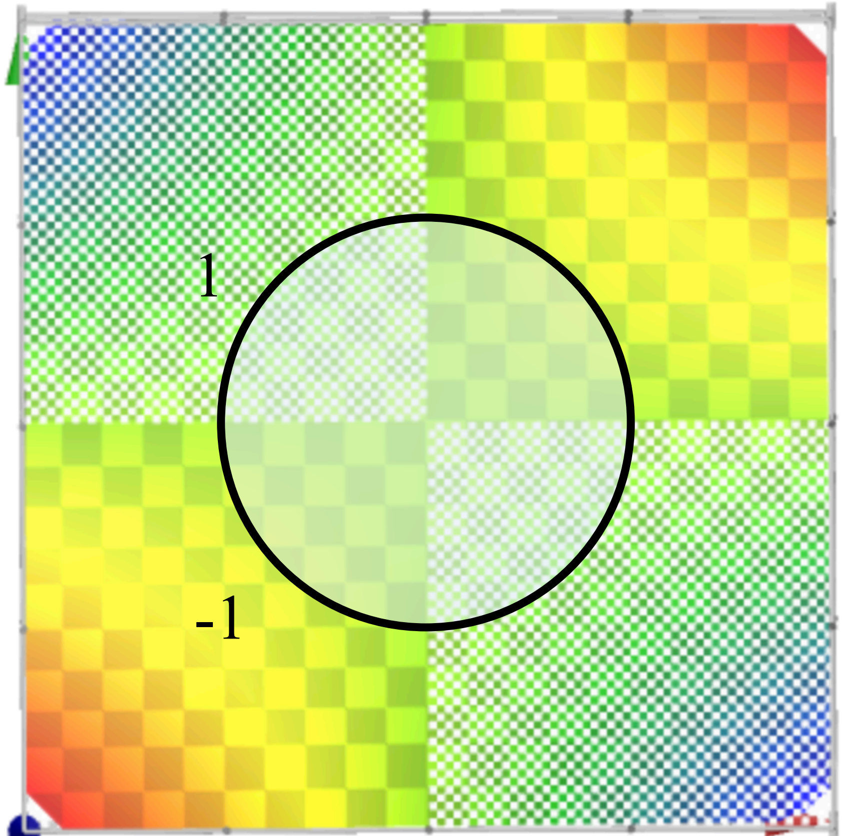
- Let X and Y be independent random variables
  - Probability Density Function (PDF) of X + Y:

$$f_{X+Y}(a) = \int_{y=-\infty}^{\infty} f_X(a-y)\, f_Y(y)\, dy$$

  - In discrete case, replace $\int_{y=-\infty}^{\infty}$ with $\sum_{y}$, and $f(y)$ with $p(y)$

Integration with Constraint

$$\iint\limits_{x^2+y^2<1} f_{x,y}\ dy\ dx = \int_{-1}^{1} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} f_{x,y}\ dy\ dx$$



Dance, Dance Convolution

- Let X and Y be independent random variables
  - Cumulative Distribution Function (CDF) of X + Y:

$$F_{X+Y}(a) = P(X+Y \le a)$$

$$= \iint_{x+y \le a} f_X(x) f_Y(y)\, dx\, dy = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{a-y} f_X(x)\, dx\; f_Y(y)\, dy$$

CDF of X

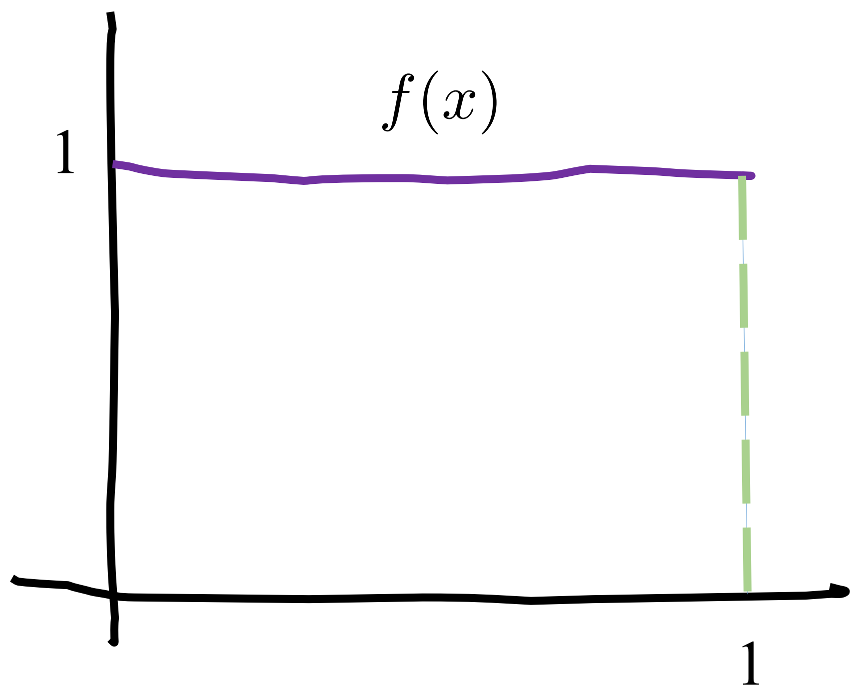
$$= \int_{y=-\infty}^{\infty} F_X(a-y)\, f_Y(y)\, dy$$

PDF of Y

  - In discrete case, replace $\int_{y=-\infty}^{\infty}$ with $\sum_y$, and $f(y)$ with $p(y)$

Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$



For both X and Y

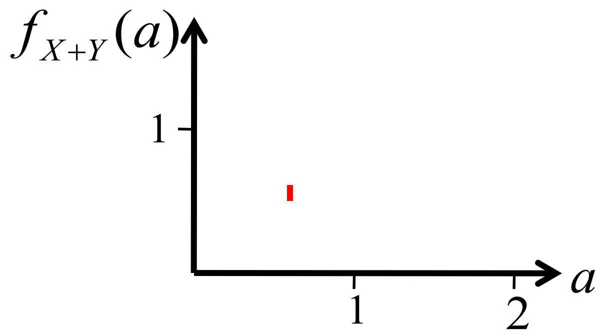
Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

When $a = 0.5$:

$$\begin{aligned} f_{X+Y}(0.5) &= \int_{y=?}^{y=?} f_X(0.5-y)dy \\ &= \int_{0}^{0.5} f_X(0.5-y)dy \\ &= \int_{0}^{0.5} 1dy \\ &= 0.5 \end{aligned}$$



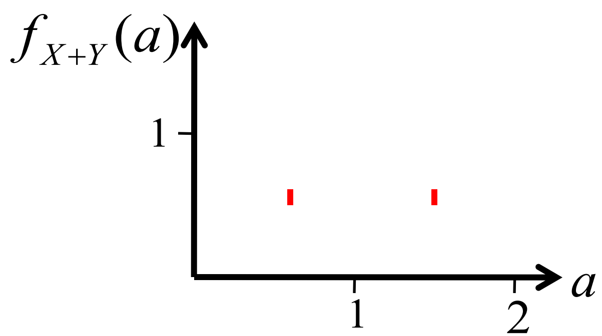
Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

When $a = 1.5$:

$$f_{X+Y}(1.5) = \int_{y=?}^{y=?} f_X(1.5-y)dy$$
$$= \int_{0.5}^{1} f_X(1.5-y)dy$$
$$= \int_{0.5}^{1} 1dy$$
$$= 0.5$$



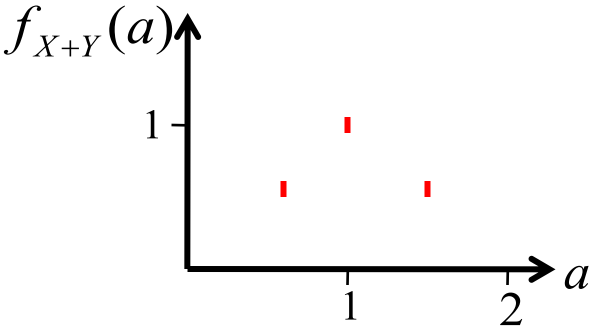
Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

When $a = 1$:

$$\begin{aligned} f_{X+Y}(1) &= \int_{y=?}^{y=?} f_X(1-y)dy \\ &= \int_0^1 f_X(1-y)dy \\ &= \int_0^1 1dy \\ &= 1 \end{aligned}$$



Sum of Independent Uniforms

- Let X and Y be independent random variables
  - X ~ Uni(0, 1) and Y ~ Uni(0, 1) → $f(x) = 1$ for $0 \le x \le 1$
  - What is PDF of X + Y?

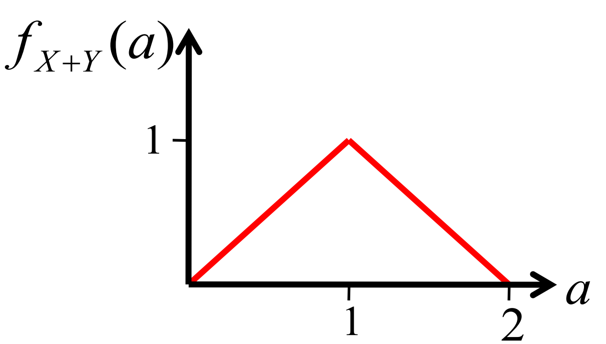
$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\, f_Y(y)\, dy = \int_{y=0}^{1} f_X(a-y)\, dy$$

  - When $0 \le a \le 1$ and $0 \le y \le a$, $0 \le a - y \le 1$ → $f_X(a-y) = 1$

$$f_{X+Y}(a) = \int_{y=0}^{a} dy = a$$

  - When $1 \le a \le 2$ and $a-1 \le y \le 1$, $0 \le a - y \le 1$ → $f_X(a-y) = 1$

$$f_{X+Y}(a) = \int_{y=a-1}^{1} dy = 2 - a$$



  - Combining: $f_{X+Y}(a) = \begin{cases} a & 0 \le a \le 1 \\ 2-a & 1 < a \le 2 \\ 0 & \text{otherwise} \end{cases}$

Sum of Independent Normals

- Let X and Y be independent random variables
  - $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$
  - $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

- Generally, have *n* independent random variables $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, ..., n$:

$$\left(\sum_{i=1}^{n} X_i\right) \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Virus Infections

- Say you are working with the WHO to plan a response to a the initial conditions of a virus:
  - Two exposed groups
  - P1: 50 people, each independently infected with $p = 0.1$
  - P2: 100 people, each independently infected with $p = 0.4$
  - Question: Probability of more than 40 infections?

Sanity check: Should we use the Binomial Sum-of-RVs shortcut?

A. YES!

B. NO!

C. Other/none/more

Virus Infections

- Say you are working with the WHO to plan a response to a the initial conditions of a virus:
  - Two exposed groups
  - P1: 50 people, each independently infected with p = 0.1
  - P2: 100 people, each independently infected with p = 0.4
  - A = # infected in P1 $\quad$ A ~ Bin(50, 0.1) $\approx$ X ~ N(5, 4.5)
  - B = # infected in P2 $\quad$ B ~ Bin(100, 0.4) $\approx$ Y ~ N(40, 24)
  - What is P(≥ 40 people infected)?
  - P(A + B ≥ 40) $\approx$ P(X + Y ≥ 39.5)
  - X + Y = W ~ N(5 + 40 = 45, 4.5 + 24 = 28.5)

$$P(W \geq 39.5) = P\left(\frac{W-45}{\sqrt{28.5}} > \frac{39.5-45}{\sqrt{28.5}}\right) = 1 - \Phi(-1.03) \approx 0.8485$$

Linear Transform

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

$$Y = X + X = 2 \cdot X$$

$$Y \sim N(2\mu, 4\sigma^2)$$

$$Y = X + X = 2 \cdot X$$

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

$$Y \sim N(2\mu, 2\sigma^2)$$

X is not independent of X

End sum of independent vars