Sum of Independent Binomial RVs

- 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 $Z = X + Y$
- More generally: $X_i \sim \text{Bin}(n_i, p)$ for $1 \le i \le N$

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

Sum of Independent Poisson RVs

- Let X and Y be independent random variables
  - $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$
  - $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$
- Proof: (just for reference)
  - Rewrite $(X + Y = n)$ as $(X = k, Y = n - k)$ where $0 \le k \le n$

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

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

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

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 $1 \le i \le N$, 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 $1 \le i \le N$, then

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

Expected Values of Sums

- Let $g(X, Y) = X + Y$.
  - Compute $E[g(X, Y)] = E[X + Y]$
  - $E[X + Y] = E[X] + E[Y]$
- Generalized: $E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$
  - Holds regardless of dependency between X_i's
  - We'll prove this next time

Dance, Dance, Convolution

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

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

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

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

  - F_{X+Y} is called **convolution** of F_X and F_Y
  - Probability Density Function (PDF) of X + Y, analogous:

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

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

Sum of Independent Uniform RVs

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

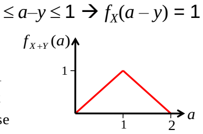
$$f_{X+Y}(a) = \int_{y=0}^{1} f_X(a-y)\ f_Y(y)\,dy = \int_{y=0}^{1} f_X(a-y)\,dy$$

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

Sum of Independent Normal RVs

- 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 your RCC checks dorm machines for viruses
  - 50 Macs, each independently infected with $p = 0.1$
  - 100 PCs, each independently infected with $p = 0.4$
  - A = # infected Macs $A \sim \text{Bin}(50, 0.1) \approx X \sim N(5, 4.5)$
  - B = # infected PCs $B \sim \text{Bin}(100, 0.4) \approx Y \sim N(40, 24)$
  - What is P(≥ 40 machine infected)?
  - $P(A + B \geq 40) \approx P(X + Y \geq 39.5)$
  - $X + Y = W \sim N(5 + 40 = 45, 4.5 + 24 = 28.5)$

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

- Be glad it's not swine flu!

Discrete Conditional Distributions

- Recall that for *events* E and F:

$$P(E \mid F) = \frac{P(EF)}{P(F)} \quad \text{where } P(F) > 0$$

- Now, have X and Y as discrete random variables
  - Conditional PMF of X given Y (where $p_Y(y) > 0$):

$$P_{X|Y}(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

  - Conditional CDF of X given Y (where $p_Y(y) > 0$):

$$F_{X|Y}(a \mid y) = P(X \leq a \mid Y = y) = \frac{P(X \leq a, Y = y)}{P(Y = y)}$$
$$= \frac{\sum_{x \leq a} p_{X,Y}(x, y)}{p_Y(y)} = \sum_{x \leq a} p_{X|Y}(x \mid y)$$

Operating System Loyalty

- Consider person buying 2 computers (over time)
  - X = 1st computer bought is a PC (1 if it is, 0 if it is not)
  - Y = 2nd computer bought is a PC (1 if it is, 0 if it is not)
  - Joint probability mass function (PMF):

Y \ X	0	1	$p_Y(y)$
0	0.2	0.3	0.5
1	0.1	0.4	0.5
$p_X(x)$	0.3	0.7	1.0

  - What is P(Y = 0 | X = 0)?

$$P(Y = 0 \mid X = 0) = \frac{p_{X,Y}(0,0)}{p_X(0)} = \frac{0.2}{0.3} = \frac{2}{3}$$

  - What is P(Y = 1 | X = 0)?

$$P(Y = 1 \mid X = 0) = \frac{p_{X,Y}(0,1)}{p_X(0)} = \frac{0.1}{0.3} = \frac{1}{3}$$

  - What is P(X = 0 | Y = 1)?

$$P(X = 0 \mid Y = 1) = \frac{p_{X,Y}(0,1)}{p_Y(1)} = \frac{0.1}{0.5} = \frac{1}{5}$$

And It Applies to Books Too…

P(Buy Book Y | Bought Book X)

Web Server Requests Redux

- Requests received at web server in a day
  - X = # requests from humans/day $X \sim \text{Poi}(\lambda_1)$
  - Y = # requests from bots/day $Y \sim \text{Poi}(\lambda_2)$
  - X and Y are independent → $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$
  - What is P(X = *k* | X + Y = *n*)?

$$P(X = k \mid X + Y = n) = \frac{P(X = k, Y = n-k)}{P(X + Y = n)} = \frac{P(X = k)P(Y = n-k)}{P(X + Y = n)}$$
$$= \frac{e^{-\lambda_1}\lambda_1^k}{k!} \cdot \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \cdot \frac{n!}{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n} = \frac{n!}{k!(n-k)!} \cdot \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n}$$
$$= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

  - $X \mid X + Y \sim \text{Bin}\left(X + Y, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$

Continuous Conditional Distributions

- Let X and Y be continuous random variables
  - Conditional PDF of X given Y (where $f_Y(y) > 0$):

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

$$f_{X|Y}(x \mid y)\,dx = \frac{f_{X,Y}(x,y)\,dx\,dy}{f_Y(y)\,dy}$$

$$\approx \frac{P(x \le X \le x+dx, y \le Y \le y+dy)}{P(y \le Y \le y+dy)} = P(x \le X \le x+dx \mid y \le Y \le y+dy)$$

  - Conditional CDF of X given Y (where $f_Y(y) > 0$):

$$F_{X|Y}(a \mid y) = P(X \le a \mid Y = y) = \int_{-\infty}^{a} f_{X|Y}(x \mid y)\,dx$$

  - Note: Even though P(Y = *a*) = 0, can condition on Y = *a*
    - Really considering: $P(a - \frac{\varepsilon}{2} \le Y \le a + \frac{\varepsilon}{2}) = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f_Y(y)dy \approx \varepsilon f(a)$

Let's Do an Example

- X and Y are continuous RVs with PDF:

$$f(x,y) = \begin{cases} \frac{12}{5} x(2-x-y) & \text{where } 0 < x, y < 1 \\ 0 & \text{otherwise} \end{cases}$$

  - Compute conditional density: $f_{X|Y}(x \mid y)$

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{\int_0^1 f_{X,Y}(x,y)\,dx}$$

$$= \frac{\frac{12}{5}x(2-x-y)}{\int_0^1 \frac{12}{5}x(2-x-y)\,dx} = \frac{x(2-x-y)}{\int_0^1 x(2-x-y)\,dx} = \frac{x(2-x-y)}{\left[x^2 - \frac{x^3}{3} - \frac{x^2 y}{2}\right]_0^1}$$

$$= \frac{x(2-x-y)}{\frac{2}{3} - \frac{y}{2}} = \frac{6x(2-x-y)}{4-3y}$$

Independence and Conditioning

- If X and Y are independent discrete RVs:

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P(X = x)P(Y = y)}{P(Y = y)} = P(X = x)$$

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x)$$

- Analogously, for independent continuous RVs:

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

Conditional Independence Revisited

- *n* discrete random variables $X_1, X_2, \ldots, X_n$ are called **conditionally independent** given Y if:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid Y = y) = \prod_{i=1}^{n} P(X_i = x_i \mid Y = y) \quad \text{for all } x_1, x_2, \ldots, x_n, y$$

- Analogously, for continuous random variables:

$$P(X_1 \le a_1, X_2 \le a_2, \ldots, X_n \le a_n \mid Y = y) = \prod_{i=1}^{n} P(X_i \le a_i \mid Y = y) \quad \text{for all } a_1, a_2, \ldots, a_n, y$$

- Note: can turn products into sums using logs:

$$\ln \prod_{i=1}^{n} P(X_i = x_i \mid Y = y) = \sum_{i=1}^{n} \ln P(X_i = x_i \mid Y = y) = K$$

$$\prod_{i=1}^{n} P(X_i = x_i \mid Y = y) = e^K$$

Mixing Discrete and Continuous

- Let X be a <u>continuous</u> random variable
- Let N be a <u>discrete</u> random variable
  - Conditional PDF of X given N:

$$f_{X|N}(x \mid n) = \frac{p_{N|X}(n \mid x) f_X(x)}{p_N(n)}$$

  - Conditional PMF of N given X:

$$p_{N|X}(n \mid x) = \frac{f_{X|N}(x \mid n) p_N(n)}{f_X(x)}$$

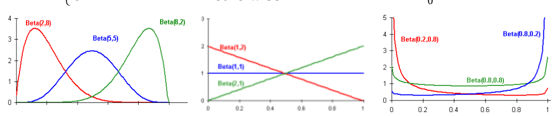
  - If X and N are independent, then:

$$f_{X|N}(x \mid n) = f_X(x) \qquad p_{N|X}(n \mid x) = p_N(n)$$

Beta Random Variable

- X is a **Beta Random Variable**: X ~ Beta(*a*, *b*)
  - Probability Density Function (PDF):

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{where} \quad B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$$



  - Symmetric when *a* = *b*
  - $E[X] = \frac{a}{a+b}$ $\qquad Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$

Flipping Coin With Unknown Probability

- Flip a coin (n + m) times, comes up with n heads
  - We don't know probability X that coin comes up heads
  - All we know is that: X ~ Uni(0, 1)
  - What is density of X given n heads in n + m flips?
  - Let N = number of heads
  - Given X = x, coin flips independent: N | X ~ Bin(n + m, x)
  - Compute conditional density of X given N = n

$$f_{X|N}(x|n) = \frac{P(N=n|X=x)\boxed{f_X(x)}^{\,1}}{P(N=n)} = \frac{\binom{n+m}{n} x^n (1-x)^m}{P(N=n)}$$

$$= \frac{1}{c} \cdot x^n (1-x)^m \quad \text{where} \quad c = \int_0^1 x^n (1-x)^m \, dx$$

Dude, Where's My Beta?!

- Flip a coin (n + m) times, comes up with n heads
  - Conditional density of X given N = n

$$f_{X|N}(x|n) = \frac{1}{c} \cdot x^n (1-x)^m \quad \text{where} \quad c = \int_0^1 x^n (1-x)^m \, dx$$

  - Note: $0 < x < 1$, so $f_{X|N}(x|n) = 0$ otherwise
  - Recall Beta distribution:

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \qquad B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$$

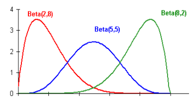
  - Hey, that looks more familiar now...
  - X | (N = n, n + m trials) ~ Beta(n + 1, m + 1)

Understanding Beta

- X | (N = n, m + n trials) ~ Beta(n + 1, m + 1)
  - X ~ Uni(0, 1)
  - Check this out, boss: $f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} = \frac{1}{B(a,b)} x^0 (1-x)^0$
    - Beta(1, 1) = Uni(0, 1) $= \frac{1}{\int_0^1 1\,dx} 1 = 1 \quad \text{where} \quad 0 < x < 1$
  - So, X ~ Beta(1, 1)
  - "Prior" distribution of X (before seeing any flips) is Beta
  - "Posterior" distribution of X (after seeing flips) is Beta
- Beta is a **<u>conjugate</u>** distribution for Beta
  - Prior and posterior parametric forms are the same!
  - Beta is also conjugate for Bernoulli and Binomial
  - Practically, conjugate means easy update:
    - Add number of "heads" and "tails" seen to Beta parameters

Further Understanding Beta

- Can set X ~ Beta(*a*, *b*) as prior to reflect how biased you think coin is apriori
  - This is a subjective probability!
  - Then observe n + m trials, where n of trials are heads



- Update to get posterior probability
  - X | (n heads in n + m trials) ~ Beta(a + n, b + m)
  - Sometimes call *a* and *b* the "equivalent sample size"
  - Prior probability for X based on seeing (*a* + *b* – 2) "imaginary" trials, where (*a* – 1) of them were heads.
  - Beta(1, 1) ~ Uni(0, 1) → we haven't seen any "imaginary trials", so apriori know nothing about coin