Welcome Back the Multinomial!

- Multinomial distribution
  - *n* independent trials of experiment performed
  - Each trial results in one of *m* outcomes, with respective probabilities: $p_1, p_2, \ldots, p_m$ where $\sum_{i=1}^{m} p_i = 1$
  - X_i = number of trials with outcome *i*

$$P(X_1 = c_1, X_2 = c_2, \ldots, X_m = c_m) = \binom{n}{c_1, c_2, \ldots, c_m} p_1^{c_1} p_2^{c_2} \ldots p_m^{c_m}$$

where $\sum_{i=1}^{m} c_i = n$ and $\binom{n}{c_1, c_2, \ldots, c_m} = \dfrac{n!}{c_1! c_2! \cdots c_m!}$

Hello Die Rolls, My Old Friend…

- 6-sided die is rolled 7 times
  - Roll results: 1 one, 1 two, 0 three, 2 four, 0 five, 3 six

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 0, X_6 = 3)$$

$$= \frac{7!}{1!1!0!2!0!3!}\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^0\left(\frac{1}{6}\right)^3 = 420\left(\frac{1}{6}\right)^7$$

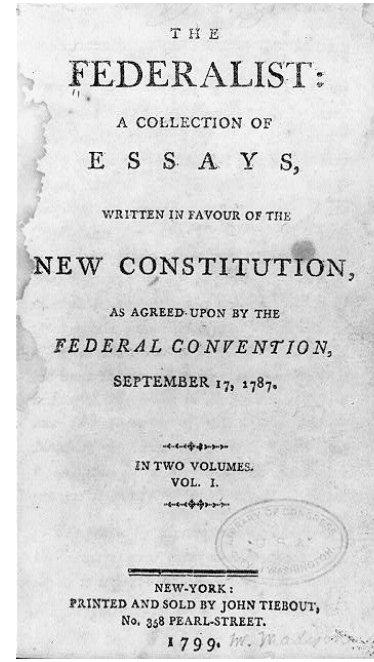
- This is generalization of Binomial distribution
  - Binomial: each trial had 2 possible outcomes
  - Multinomial: each trial has *m* possible outcomes

Probabilistic Text Analysis

- Ignoring order of words, what is probability of any given word you write in English?
  - P(word = “the”) > P(word = “transatlantic”)
  - P(word = “Stanford”) > P(word = “Cal”)
  - Probability of each word is just multinomial distribution
- What about probability of those same words in someone else’s writing?
  - P(word = “probability” | writer = you) >

    P(word = “probability” | writer = non-CS109 student)
  - After estimating P(word | writer) from known writings, use Bayes’ Theorem to determine P(writer | word) for new writings!

Old and New Analysis


THE
FEDERALIST:
A COLLECTION OF
ESSAYS,
WRITTEN IN FAVOUR OF THE
NEW CONSTITUTION,
AS AGREED UPON BY THE
FEDERAL CONVENTION,
SEPTEMBER 17, 1787.
IN TWO VOLUMES.
VOL. I.
NEW-YORK:
PRINTED AND SOLD BY JOHN TIEBOUT,
No. 358 PEARL-STREET.
1799.

- Authorship of “Federalist Papers”
  - 85 essays advocating ratification of US constitution
  - Written under pseudonym “Publius”
    - Really, Alexander Hamilton, James Madison and John Jay
  - Who wrote which essays?
    - Analyzed probability of words in each essay versus word distributions from known writings of three authors

- Filtering Spam
  - P(word = “Viagra” | writer = you)

    << P(word = “Viagra” | writer = spammer)

Independent Discrete Variables

- Two discrete random variables X and Y are called <u>**independent**</u> if:

$$p(x, y) = p_X(x)\, p_Y(y) \quad \text{for all } x, y$$

- Intuitively: knowing the value of X tells us nothing about the distribution of Y (and vice versa)
  - If two variables are <u>**not**</u> independent, they are called <u>***dependent***</u>
- Similar conceptually to independent *events*, but we are dealing with multiple <u>***variables***</u>
  - Keep your events and variables distinct (and clear)!

Coin Flips

- Flip coin with probability *p* of “heads”
  - Flip coin a total of *n* + *m* times
  - Let X = number of heads in first *n* flips
  - Let Y = number of heads in next *m* flips

$$P(X = x, Y = y) = \binom{n}{x} p^{x} (1-p)^{n-x} \binom{m}{y} p^{y} (1-p)^{m-y}$$

$$= P(X = x) P(Y = y)$$

  - X and Y are independent
  - Let Z = number of total heads in *n* + *m* flips
  - Are X and Z independent?
    - What if you are told Z = 0?

Web Server Requests

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a "bot" (probability = (1 – *p*)), independently
  - X = # requests from humans/day   (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day   (Y | N) ~ Bin(N, 1 - *p*)

$$P(X = i, Y = j) = P(X = i, Y = j \mid X + Y = i + j)P(X + Y = i + j)$$
$$+ P(X = i, Y = j \mid X + Y \neq i + j)P(X + Y \neq i + j)$$

  - Note: $P(X = i, Y = j \mid X + Y \neq i + j) = 0$

$$P(X = i, Y = j \mid X + Y = i + j) = \binom{i+j}{i} p^i (1-p)^j$$

$$P(X + Y = i + j) = e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}$$

$$P(X = i, Y = j) = \binom{i+j}{i} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}$$

Web Server Requests (cont.)

- Let N = # of requests to web server/day
  - Suppose N ~ Poi(λ)
  - Each request comes from a human (probability = *p*) or from a “bot” (probability = (1 – *p*)), independently
  - X = # requests from humans/day (X | N) ~ Bin(N, *p*)
  - Y = # requests from bots/day (Y | N) ~ Bin(N, 1 - *p*)

$$P(X=i,Y=j)=\frac{(i+j)!}{i!\,j!}p^i(1-p)^j e^{-\lambda}\frac{\lambda^{i+j}}{(i+j)!}=e^{-\lambda}\frac{(\lambda p)^i}{i!}\cdot\frac{(\lambda(1-p))^j}{j!}$$

$$=e^{-\lambda p}\frac{(\lambda p)^i}{i!}\cdot e^{-\lambda(1-p)}\frac{(\lambda(1-p))^j}{j!}=P(X=i)P(Y=j)$$

where X ~ Poi(λp) and Y ~ Poi(λ(1 – *p*))

  - X and Y are independent!

Independent Continuous Variables

- Two continuous random variables X and Y are called **independent** if:

  P(X ≤ *a*, Y ≤ *b*) = P(X ≤ *a*) P(Y ≤ *b*) for any *a*, *b*

- Equivalently:

$$F_{X,Y}(a,b) = F_X(a)F_Y(b) \quad \text{for all } a, b$$
$$f_{X,Y}(a,b) = f_X(a)\, f_Y(b) \quad \text{for all } a, b$$

- More generally, joint density factors separately:

$$f_{X,Y}(x,y) = h(x)g(y) \quad \text{where } -\infty < x, y < \infty$$

Pop Quiz (Just Kidding…)

- Consider joint density function of X and Y:

$$f_{X,Y}(x,y) = 6e^{-3x}e^{-2y} \quad \text{for } 0 < x, y < \infty$$

  - Are X and Y independent? Yes!

Let $h(x) = 3e^{-3x}$ and $g(y) = 2e^{-2y}$, so $f_{X,Y}(x,y) = h(x)g(y)$

- Consider joint density function of X and Y:

$$f_{X,Y}(x,y) = 4xy \quad \text{for } 0 < x, y < 1$$

  - Are X and Y independent? Yes!

Let $h(x) = 2x$ and $g(y) = 2y$, so $f_{X,Y}(x,y) = h(x)g(y)$

  - Now add constraint that: $0 < (x + y) < 1$
  - Are X and Y independent? No!
    - Cannot capture constraint on x + y in factorization!

The Joy of Meetings

- Two people set up a meeting for 12pm
  - Each arrives independently at time uniformly distributed between 12pm and 12:30pm
  - X = # min. past 12pm person 1 arrives   X ~ Uni(0, 30)
  - Y = # min. past 12pm person 2 arrives   Y ~ Uni(0, 30)
  - What is P(first to arrive waits > 10 min. for other)?

$P(X+10<Y)+P(Y+10<X)=2P(X+10<Y)$ by symmetry

$$2P(X+10<Y)=2\iint_{x+10<y} f(x,y)\,dx\,dy=2\iint_{x+10<y} f_X(x)\,f_Y(y)\,dx\,dy$$

$$=2\int_{y=10}^{30}\int_{x=0}^{y-10}\left(\frac{1}{30}\right)^2 dx\,dy=\frac{2}{30^2}\int_{y=10}^{30}\left(\int_{x=0}^{y-10}dx\right)dy=\frac{2}{30^2}\int_{y=10}^{30}\left(x\Big|_0^{y-10}\right)dy=\frac{2}{30^2}\int_{y=10}^{30}(y-10)\,dy$$

$$=\frac{2}{30^2}\left(\frac{y^2}{2}-10y\right)\Bigg|_{10}^{30}=\frac{2}{30^2}\left[\left(\frac{30^2}{2}-300\right)-\left(\frac{10^2}{2}-100\right)\right]=\frac{4}{9}$$

Dependent RVs: Imperfection on Disk

- Disk surface is a circle of radius R
  - A single point imperfection uniformly distributed on disk

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi R^2} & \text{if } x^2 + y^2 \le R^2 \\ 0 & \text{if } x^2 + y^2 > R^2 \end{cases} \quad \text{where} \quad -\infty < x, y < \infty$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \frac{1}{\pi R^2} \int_{x^2+y^2 \le R^2} dy = \frac{1}{\pi R^2} \int_{y=-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy = \frac{2\sqrt{R^2-x^2}}{\pi R^2}$$

$$f_Y(y) = \frac{2\sqrt{R^2-y^2}}{\pi R^2} \quad \text{where} - R \le y \le R, \text{ by symmetry}$$

  - Note: $f_{X,Y}(x,y) \ne f_X(x)\,f_Y(y)$
  - Distance to origin: $D = \sqrt{X^2 + Y^2}$, $P(D \le a) = \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2}$

$$E[D] = \int_0^R P(D > a)\,da = \int_0^R (1 - \frac{a^2}{R^2})\,da = \left(a - \frac{a^3}{3R^2} \right)\Bigg|_0^R = \frac{2R}{3}$$

Independence of Multiple Variables

- *n* random variables X_1, X_2, …, X_n are called <u>**independent**</u> if:

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

- Analogously, for continuous random variables:

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

Independence is Symmetric

- If random variables X and Y independent, then
  - X independent of Y, and Y independent of X
- Duh!? Duh, indeed...
  - Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (I.I.D.) continuous random vars
  - Say $X_n > X_i$ for all $i = 1, \ldots, n - 1$ (i.e. $X_n = \max(X_1, \ldots, X_n)$)
    - Call X_n a "record value"
  - Let event A_i indicate X_i is "record value"
    - Is A_{n+1} independent of A_n?
    - Is A_n independent of A_{n+1}?
    - Easier to answer: Yes!
    - By symmetry, $P(A_n) = 1/n$ and $P(A_{n+1}) = 1/(n+1)$
    - $P(A_n A_{n+1}) = (1/n)(1/(n+1)) = P(A_n)P(A_{n+1})$