Continuous Conditional Distributions (Review)

- Let $X$ and $Y$ be continuous random variables
  - Recall, conditional PDF of $X$ given $Y$ (where $f_Y(y) > 0$):
    \[
    f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}
    \]
Let’s Do an Example (Review)

- 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: \( f_{x|y}(x \mid y) \)

\[
f_{x|y}(x \mid y) = \frac{f_{x,y}(x, y)}{f_y(y)} = \frac{f_{x,y}(x, y)}{\int_0^1 f_{x,y}(x, y) \, dx}
\]

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

\[
= \frac{x(2 - x - y)}{\frac{2}{3} - \frac{y}{2}} = \frac{6x(2 - x - y)}{4 - 3y}
\]
Independence and Conditioning

- If $X$ and $Y$ are independent discrete RVs:

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

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

- Analogously, for independent continuous RVs:

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

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

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

- Analogously, for continuous random variables:

  \[
P(X_1 \leq a_1, X_2 \leq a_2, \ldots, X_n \leq a_n \mid Y = y) = \prod_{i=1}^{n} P(X_i \leq a_i \mid Y = y)
  \]
  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 continuous random variable
- Let $N$ be a discrete random variable
  - Conditional PDF of $X$ given $N$:
    $$f_{X|N}(x | n) = \frac{p_{N|X}(n | x)f_X(x)}{p_N(n)}$$
  - Conditional PMF of $N$ given $X$:
    $$p_{N|X}(n | x) = \frac{f_{X|N}(x | n)p_N(n)}{f_X(x)}$$
  - If $X$ and $N$ are independent, then:
    $$f_{X|N}(x | n) = f_X(x) \quad p_{N|X}(n | x) = p_N(n)$$
Beta Random Variable

- X is a **Beta Random Variable**: $X \sim \text{Beta}(a, b)$
  - Probability Density Function (PDF): $f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$ where $\int_0^1 x^{a-1} (1-x)^{b-1} \, dx = B(a,b)$

- Symmetric when $a = b$

- $E[X] = \frac{a}{a + b}$
- $\text{Var}(X) = \frac{ab}{(a + b)^2(a + b + 1)}$
**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: \(X \sim \text{Uni}(0, 1)\)
  - What is density of \(X\) given \(n\) heads in \(n + m\) flips?
  - Let \(N =\) number of heads
  - Given \(X = x\), coin flips independent: \((N \mid X) \sim \text{Bin}(n + m, x)\)
  - Compute conditional density of \(X\) given \(N = n\)

\[
f_{X|N}(x \mid n) = \frac{P(N = n \mid X = x) f_X(x)}{P(N = n)} = \left( \frac{n + m}{n} \right) x^n (1 - x)^m 
\]

\[
= \frac{1}{c} \cdot x^n (1 - x)^m \quad \text{where} \quad c = \int_0^1 x^n (1 - x)^m \, dx
\]
Dude, Where’s My Beta?!

- Flip a coin \((n + m)\) times, comes up with \(n\) heads
  - Conditional density of \(X\) given \(N = n\)
    \[
f_{X|N}(x \mid n) = \frac{1}{c} \cdot x^n (1 - x)^m \quad \text{where} \quad c = \int_0^1 x^n (1 - x)^m \, dx
    \]
  - Note: \(0 < x < 1\), so \(f_{X|N}(x \mid n) = 0\) otherwise
  - Recall Beta distribution:
    \[
f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \quad B(a,b) = \int_0^1 x^{a-1} (1-x)^{b-1} \, dx
    \]
    - Hey, that looks more familiar now...
  - \(X \mid (N = n, n + m \text{ trials}) \sim \text{Beta}(n + 1, m + 1)\)
Understanding Beta

- $X \mid (N = n, m + n \text{ trials}) \sim \text{Beta}(n + 1, m + 1)$
  - $X \sim \text{Uni}(0, 1)$
  - Check this out, boss:
    - $\text{Beta}(1, 1) = \text{Uni}(0, 1)$
  - So, $X \sim \text{Beta}(1, 1)$
  - “Prior” distribution of $X$ (before seeing any flips) is Beta
  - “Posterior” distribution of $X$ (after seeing flips) is Beta

- Beta is a **conjugate** 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
Can set $X \sim \text{Beta}(a, b)$ 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
- $X \mid (n \text{ heads in } n + m \text{ trials}) \sim \text{Beta}(a + n, b + m)$
- 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.
- $\text{Beta}(1, 1) \sim \text{Uni}(0, 1) \Rightarrow$ we haven’t seen any “imaginary trials”, so apriori know nothing about coin
Welcome Back Our Friend: Expectation

• Recall expectation for discrete random variable:

\[ E[X] = \sum_{x} x P(x) \]

• Analogously for a continuous random variable:

\[ E[X] = \int_{-\infty}^{\infty} x f(x) \, dx \]

• Note: If X always between \( a \) and \( b \) then so is \( E[X] \)
  
  • More formally:

\[ \text{if } P(a \leq X \leq b) = 1 \text{ then } a \leq E[X] \leq b \]
Generalizing Expectation

- Let $g(X, Y)$ be real-valued function of two variables

- Let $X$ and $Y$ be discrete jointly distributed RVs:
  \[ E[g(X, Y)] = \sum_y \sum_x g(x, y) p_{X,Y}(x, y) \]

- Analogously for continuous random variables:
  \[ E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) \, dx \, dy \]
Expected Values of Sums

- Let $g(X, Y) = X + Y$. Compute $E[g(X, Y)] = E[X + Y]$

$$E[X + Y] = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} (x + y) f_{X,Y}(x, y) \, dx \, dy$$

$$= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} x f_{X,Y}(x, y) \, dy \, dx + \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} y f_{X,Y}(x, y) \, dx \, dy$$

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

$$= E[X] + E[Y]$$

- 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
Tie Me Up! : Bounding Expectation

• If random variable $X \geq a$ then $E[X] \geq a$
  
  if $P(a \leq X \leq \infty) = 1$ then $a \leq E[X] \leq \infty$
  
  ▪ Often useful in cases where $a = 0$
  
  ▪ But, $E[X] \geq a$ does not imply $X \geq a$ for all $X = x$
    
    ○ E.g., $X$ is equally likely to take on values -1 or 3. $E[X] = 1$.

• If random variables $X \geq Y$ then $E[X] \geq E[Y]$
  
  ▪ $X \geq Y \Rightarrow X - Y \geq 0 \Rightarrow E[X - Y] \geq 0$
  
  
  ▪ Substituting: $E[X] - E[Y] \geq 0 \Rightarrow E[X] \geq E[Y]$
  
  ▪ But, $E[X] \geq E[Y]$ does not imply $X \geq Y$ for all $X = x, Y = y$
Sample Mean

• Consider $n$ random variables $X_1, X_2, \ldots, X_n$
  ▪ $X_i$ are all independently and identically distributed (I.I.D.)
  ▪ Have same distribution function $F$ and $\text{E}[X_i] = \mu$
  ▪ We call sequence of $X_i$ a **sample** from distribution $F$

• Sample mean: $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$