If you model \((X, \Theta)\) as \(Bernoulli(\Theta)\) and \(\Theta \sim \text{Uniform}(0, 1)\), the joint PDF of \((X, \Theta)\) would be

\[
    f_{X, \Theta}(x, \theta) = \begin{cases} 
    \theta & \text{for } (x = 0, 1) \\
    1 - \theta & \text{for } 0 < \theta < 1 \\
    0 & \text{else}
    \end{cases}
\]

Then (e.g.) \(P(X = 1) = P(X = 1 \text{ and } \Theta \text{ is anything between } 0 \text{ and } 1) = \int_0^1 \theta (1 - \theta)^{1 - 1} \, d\theta = \int_0^1 \theta \, d\theta = \frac{1}{2} \).

(2 May 19)

**Bivariate CDFs**

**Def.** The joint CDF of two rvs \(X\) and \(Y\) is the function \(F_{X,Y}(x, y)\) satisfying

\[
    F_{X,Y}(x, y) = P(X \leq x \text{ and } Y \leq y)
\]

for all \(-\infty < x < \infty\) and \(-\infty < y < \infty\).
Consequences of this definition:

1. If $(X,Y)$ has the joint CDF $F_{X,Y}(x,y)$, you can obtain the marginal CDF $F_X(x)$ from the joint CDF as $F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y)$, and similarly the marginal CDF $F_Y(y)$ is just $F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x,y)$.

2. The joint pdf and joint CDF are related in a manner similar to their relationship in the univariate (one RV at a time) case:
If \((X, Z)\) have a joint pdf \(f_{XZ}(x, z)\),

then
\[
F_{XZ}(x, z) = \int_{-\infty}^{x} \int_{-\infty}^{z} f_{XZ}(u, v) \, du \, dv
\]

and
\[
f_{XZ}(x, z) = \frac{d^2}{\partial x \partial y} F_{XZ}(x, z) = \frac{d^2}{\partial y \partial x} F_{XZ}(x, z)
\]

(at every \((x, z)\) where the partial derivatives exist).

**Consequence:**

- If \((X, Z)\) have a discrete joint distribution with joint pmf \(f_{XZ}(x, z)\), then the marginal pmf \(f_X(x)\) of \(X\) is
\[
f_X(x) = \sum_y f_{XZ}(x, y)
\]
and similarly for \(f_Z(y)\).
The idea behind marginal distributions is that it's harder to visualize a joint (2-dimensional) distribution than it is to visualize each of its 1-dimensional marginal distributions.

If $(X, Y)$ have a continuous joint distribution with joint pdf $f_{X,Y}(x, y)$, the marginal pdf $f_X(x)$ of $X$ is (marginalizing out $Y$)

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy \quad (\text{for all } -\infty < x < \infty)$$

and the marginal pdf $f_Y(y)$ of $Y$ is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \quad (\text{for all } -\infty < y < \infty).$$
This is a very long text that is not clear to read. It appears to be mathematical equations and diagrams, but the handwriting is not legible. It is difficult to extract meaningful content from this page.
Similarly, the support of \( f \) is \((0, 1)\), and its pdf is

\[
f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx = \int_{-\infty}^{\infty} \frac{21}{4} x^2 y \, dx
\]

\[
= \begin{cases} \frac{7}{2} y^{\frac{5}{2}} & \text{for } 0 < y < 1 \\ 0 & \text{else} \end{cases}
\]

\[(\text{W.A.5.11})\]

**Consequences:**

If you have the joint distribution \( f_{X,Y}(x, y) \), you can reconstruct the marginals \( f_X(x) \) and \( f_Y(y) \), but not the other way around: if all you have is the marginals, they do not uniquely determine the joint.
Example

(pop. 134)

Case 1: \[ X = \# \text{heads in } n \text{ tosses of fair coin 1} \]
and independently
\[ Y = \# \text{heads in } n \text{ tosses of fair coin 2} \]

Case 1:
\[ Y = X \]

\[ X \sim \text{Binomial} \left( n, \frac{1}{2} \right) \]
so
\[ f_{X}(x) = \binom{n}{x} \left( \frac{1}{2} \right)^x \left( 1 - \frac{1}{2} \right)^{n-x}, \quad x = 0, 1, \ldots, n \]

is also

\[ Y \sim \text{Binomial} \left( n, \frac{1}{2} \right) \]
so
\[ f_{Y}(y) = \binom{n}{y} \left( \frac{1}{2} \right)^y \left( 1 - \frac{1}{2} \right)^{n-y}, \quad y = 0, 1, \ldots, n \]

Since \( X \) and \( Y \) are independent in

Case 1,
\[ f_{X,Y}(x,y) = f_{X}(x) \cdot f_{Y}(y) \]
(as we'll see in a minute),
So in case 1, 
\[ f_{X,Y}(x, y) = \begin{cases} \left( x \right)^{\frac{1}{2}} \left( y \right)^{\frac{1}{2}} & \text{for } x = 0, 1, \ldots, n \text{ and } y = 0, 1, \ldots, n \\ 0 & \text{else} \end{cases} \]

However: In case 2, 
\[ X \text{ is Binomial } (n, \frac{1}{2}) \text{ and so is } Y \]
(same as in case 1), but their joint distribution (since \( Y = X \)) is

\[ f_{X,Y}(x, y) = \begin{cases} \left( x \right)^{\frac{1}{2}} \left( y \right)^{\frac{1}{2}} & \text{for } x = y = 0, \ldots, n \\ 0 & \text{else} \end{cases} \]

There is one situation in which the marginals uniquely determine the joint: when \( X \) and \( Y \) are independent.
Def. rvs $X$ and $Y$ are independent (non-weird) if for every sets $A$ and $B$ of real numbers $P(X \in A \text{ and } Y \in B) = P(X \in A) \cdot P(Y \in B)$.

Consequence

1. Immediately you get that if $X$ and $Y$ are independent

$$F_{X Y}(x, y) = P(X \leq x \text{ and } Y \leq y)$$

$$= P(X \leq x) \cdot P(Y \leq y)$$

$$= F_X(x) \cdot F_Y(y).$$

This is an iff: The converse also true.

2. Differentiating this equation once with respect to $x$ and once with respect to $y$ to get the result that $X, Y$ independent $\iff f_{X Y}(x, y) = f_X(x) \cdot f_Y(y)$.
Example

Suppose that continuous rvs $X$ and $Y$ have joint pdf

$$f_{X,Y}(x,y) = \begin{cases} 4xy & \text{for } 0 \leq x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The support $\mathcal{S}_{X,Y}$ of $f_{X,Y}$ is the region inside the unit circle. You can evaluate the normalization constant by computing

$$\int \int_{\mathcal{S}_{X,Y}} 4xy \, dx \, dy$$

and setting it equal to 1:

$$1 = \int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 4xy \, dy \, dx = \frac{\pi}{24}$$

So $k = \frac{24}{\pi}$.

Q: Are $X$ and $Y$ independent?
A: No, they can't be: since the only points with positive density satisfy $x^2 + y^2 \leq 1$, for any given value $y$ of $X$, the possible values of $X$ depend on $y$, and vice versa. Example:

Continuous rv $X$ and $Y$ have joint pdf $f_{X,Y}(x,y) = \begin{cases} k e^{-(x+y)} & \text{for } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{else} \end{cases}$

A: Are $X$ and $Y$ independent?

A: Yes, because (a) $e^{-(x+y)}$ factors into $(e^{-x})(e^{-y})$ and (b) the support $S_{X,Y}$ also has factors: $(x \geq 0) \& (y \geq 0)$.
Just choose \((k, k_x, k_y)\) such that
\[\int_{xy} k e^{-(x+y)} \, dxdy = 1, \int_{x} k_x e^{-x} \, dx = 1, \int_{y} k_y e^{-2y} \, dy = 1, \text{ and } k = k_x \cdot k_y.\]

You get \(k_x = 1, k_y = 2, k = 2\). \(\checkmark\)

Conditional probability distributions (or log \(\psi(A) > 0\)), we should be able to extend this idea to random variables. Start with \(I\) and \(X\) both discrete, so that we can talk about \(p(I=y | X=x)\).
Def. If $X$ and $Y$ have a discrete joint distribution with joint pmf $f_{XY}(x, y)$ and $X$ has marginal pmf $f_X(x)$, then for each $x$ such that $f_X(x) > 0$ define

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

to be the conditional pmf of $Y$ given $X$.

Example: gender & marijuana legalization preference at UCLA

(see doc. comm. notes)

$14$ out of $15$ students surveyed favored legalization
Now let's do the analogous thing for continuous \( X \) vs \( Y \).

If \( X \) and \( Y \) have a continuous joint distribution with joint pdf \( f_{X,Y}(x,y) \) and \( X \) (continuous) has marginal pdf \( f_X(x) \), then for each \( x \) such that \( f_X(x) > 0 \), define

\[
f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)} & \text{to be the conditional pdf of } Y \text{ given } X. \\
0 & \text{otherwise}
\end{cases}
\]

Continuity or earlier example:

\( X, Y \) have joint pdf

\[
f_{X,Y}(x,y) = \begin{cases} \frac{21}{4}x^2y & \text{for } 0 < x < y < 1 \\
0 & \text{else}
\end{cases}
\]
Let's work out $f_{Z|X}(y|x)$ and

$\int f_{Z|X}(x|y) \, dx$.

Earlier we saw that

\[
f_X(x) = \begin{cases} \frac{21}{8} x^2 (1-x^4) & \text{for } -1 \leq x \leq 1 \\ 0 & \text{else} \end{cases}
\]

and

\[
f_Y(y) = \begin{cases} \frac{7}{2} y^{5/2} & \text{for } 0 \leq y \leq 1 \\ 0 & \text{else} \end{cases}
\]

Immediately, then, for all $x$ for which $f_X(x) > 0$, namely $-1 < x < 1$

\[
f_{Z|X}(y|x) = \frac{f_{Z,X}(x,y)}{f_X(x)}
\]

\[
= \begin{cases} \frac{21}{4} x^2 y & \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ \frac{21}{8} x^2 (1-x^4) & \text{else} \end{cases}
\]
and this simplifies to

\[ f_{X|Y}(x|y) = \begin{cases} \frac{2y}{1-x^2} & 0 \leq x^2 \leq y \leq 1 \\ 0 & \text{else} \end{cases} \]

A few "slices" of this:

\[ \mathcal{X} = 0 \]

And in the other direction for \( 0 \leq x \leq 1 \):

\[ f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \]

\[ = \begin{cases} \frac{2y}{4 \times y} = \frac{3}{2} \frac{x^2}{y} & \text{for } 0 \leq y \leq x^2 \\ \frac{7}{2} \frac{y}{5} = \frac{3}{2} \frac{x^2}{y} & \text{for } 0 \leq x^2 \leq y \leq 1 \\ 0 & \text{else} \end{cases} \]
A few "sizes" of this

Note: when $X$ and $Y$ are continuous, $f_{X|Y}(y|x)$ may seem to involve conditioning on the event $X = x$, which (as we saw earlier) has probability 0. But that's not what's actually going on; strictly speaking $f_{X|Y}(y|x)$ is a limit:

$$f_{X|Y}(y|x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \mathbb{P}\left( \frac{x - \epsilon}{2} < X < x + \epsilon \right)$$
of a value of width \( \varepsilon \) around \( x^* \). 

Compute \( P(Y \in y | X \text{ is in the strip}) \), differentiate the result with respect to \( y \), and let \( \varepsilon \) go to 0. Thus you can think of \( f_{X|Y}(y|x) \) as the conditional pdf of \( Y \) given that \( X \) is close to \( x \).

Construct a joint pdf from marginals and conditionals. We know that (as long as no division by 0 happen)

\[
f_{X|Y}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}
\]

and

\[
f_{Y|X}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.
\]
Multiply equation 1 by $\frac{f_x(x)}{\gamma}$ and equation 2 by $\frac{f_\gamma(\gamma)}{\gamma}$ to get

$$f_{x\gamma}(x, \gamma) = f_x(x) f_{\gamma|\gamma}(\gamma \mid x)$$

$$= f_\gamma(\gamma) f_{X\gamma}(x \mid \gamma).$$

So there are two ways to construct a joint pdf from marginal pdfs and a conditional pdf.

A machine produces nuts and bolts, and the nut paired with a particular bolt in the manufacturing process is...
To learn about $\Theta$, we could take a vanishing sample of b(i)s (b(i)s). Let $\Theta = \text{proportion of defective bolts if the machine is allowed to run indefinitely.}$

Since we can only observe the machine for a finite (short) period and the bolt size can be small or not, we need to find a (not, bolt) pair that fits. For example, bolt diameter either too big or too small or too big. Let's call a (not, bolt) pair defective.
Count the # of defectives in the sample

\( N \) [call this \( N \)]

\( \frac{1}{\text{stationarity}} \)

\( \text{i.i.d.} \)

\( D_i \sim \text{Bernoulli}(\theta) \)

\( \theta \)

\( \Pr(i \text{ is defective}) \)

\( d_i \)

\( (i = 1, \ldots, n) \)

\[ N = \sum_{i=1}^{n} D_i \]

so the sum of \( N \) is fixed & known

\[ f_N(n | m, \theta) \]

\( \text{(sampled dist.)} \)

Suppose that

\[ m = 114, \quad N = 3 \]

A reasonable estimate of \( \theta \) would be

\[ \hat{\theta} = \frac{N}{m} = \frac{3}{114} = 2.6\% \]

but how much uncertainty do we have about \( \theta \) on the basis of this dataset?
Bayesian story

$\theta$ unknown $\in (0, 1)$

Dataset $D = (D_1, ..., D_m)$

$p(D | \theta)$ easy

$p(\theta | N)$ hard

Bayes's Theorem

Because

Bernoulli dataset $D = (D_1, ..., D_m)$ and the rv $\theta$ carry the same info about $\theta$

Total info about $\theta$ external to dataset + info about $\theta$ internal to dataset + normalization (arrows)

(7 May 18)