Joint, marginal, etc. (context: multivariate R.V.)

\[ Z = (X, Y) \quad \mathcal{Z} = \mathcal{X} \times \mathcal{Y} \]

(joint) pdf of \( Z \)

"Joint pmf" on \( X \times Y \)

\[ p(x, y) = P[X = x, Y = y] \]

joint dist.

Can represent elementary events of \( \mathcal{Z}(x, y) \) as tables

<table>
<thead>
<tr>
<th>( x )</th>
<th>( x = 0 )</th>
<th>( x = 1/2 )</th>
<th>( x = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>0</td>
<td>1/2</td>
<td>0</td>
</tr>
<tr>
<td>( p(x, y) )</td>
<td>1/4</td>
<td>1/4</td>
<td>1/2</td>
</tr>
</tbody>
</table>

cts. case: if \( X \) and \( Y \) are cts

\[ P[ \text{box region} ] = \int_{\text{box}} p(x, y) \, dx \, dy \]

Marginal distribution (in the context of a joint dist.)

dist. of a component of a random vector

\[ P[X = x] = \sum_{y \in \mathcal{Y}} P[X = x, Y = y] \quad \text{"sum rule"} \]

"Marginalizing out \( Y \)"

Independence:

\[ X \text{ is independent of } Y \iff p(x, y) = p(x)p(y) \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y} \]

notation: \( X \perp Y \)
- R.V. \( X_1, \ldots, X_n \) are "mutually independent" \( \iff p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i) \quad \forall x_1, \ldots, x_n \in \mathbb{R} \).

\[ Z = (X, Y) \quad \text{disjoint} \quad P \{ Z \in \text{box}, Y = y \} = \int_{x \in \text{box}} p(x, y) \, dx \]

\[ \text{cond. i.e.} \]

- for events \( A \cap B \), suppose \( P(B) \neq 0 \)

\[ \text{then } P(A|B) = \frac{P(A \cap B)}{P(B)} \]

Aside:

\( A \cap B \) we \( A \), we \( B \)

\( x = x, \ y = y \)

\[ p(x, y) = \frac{p(x \mid y) \cdot p(y)}{p(y)} \]

\[ P(B) = \sum_{A \in \text{partition of } B} P(A \cap B) \]

For discrete R.V.:

"conditional pmf"

\[ p(x \mid y) \triangleq P(X = x \mid Y = y) \triangleq \frac{P(X = x, Y = y)}{P(Y = y)} \]

\[ P(x \mid y) \propto p(x, y) \]

"conditional pdf"

\[ p(x \mid y) \triangleq \frac{p(x, y)}{p(y)} \]

- For continuous R.V.:
\[ p(y) \]

subtle part: \( p(x|y) \) is undefined when \( p(y) = 0 \)

\[
\begin{align*}
\text{[note independent \( \times - y \) \]} & \quad p(x, y) = p(x|y)p(y) = p(y|x)p(x) = p(x) \\
\text{example of conditioning:} & \quad \mathbb{P}(\text{having cancer | tumor measurement} = 2) \\
\text{Bayes rule:} & \quad \text{invert conditioning} \\
& \quad \text{by def.:} \quad p(y|x) = \frac{p(x,y)}{p(x)} \\
& \quad \Rightarrow p(x,y) = p(y|x)p(x) = p(x|y)p(y) \\
& \quad \text{"product rule"}
\end{align*}
\]

\[
\begin{align*}
\text{Chain rule:} & \quad \text{successive application of product rule} \\
p(x_1, \ldots, x_n) &= p(x_n|x_{1:n-1})p(x_{1:n-1}) \\
& \quad \cdots \\
p(x_1, \ldots, x_n) &= p(x_n|x_{1:n-1})p(x_{1:n-2})p(x_{1:n-2})p(x_{1:n-2}) \\
& \quad \text{Convention:} \quad l: 0 = 0, \quad p(x_1 | x_0) = p(x_1) \quad \text{[can be simplified when making conditional independence assumptions from directed graphical model]} \\
p(x_i | x_{1:i-1}) & \quad \text{points in graph} \\
\text{[can also choose any \( n \) permutation]} \\
\end{align*}
\]

\[ X \text{ is \text{ cond. indep. of } Y \ given } Z \quad \text{notation: } X \perp Y | Z \]

\[ \iff p(x, y | z) = p(x|z)p(y|z) \quad \forall x \in X \quad \forall y \in Y \quad \forall z \in Z \text{ s.t. } p(z) \neq 0 \]

exercise: prove that if \( X \perp Y | Z \)

Note: the paradox about conditional probabilities that I was looking for was indeed (as Jacob Louis Hoover rightfully pointed out) the Borel-Kolmogorov paradox - have a look at it if you're curious!
then \( p(x|y,z) = p(x|z) \)

\[ \text{[conditional analog of } \times \Delta Y \Rightarrow p(x|Y) = p(x)\]}

**Example:**
- \( \mathbf{Z} = \) indicator whether another comes a genetic disease
- \( \mathbf{X} = \) aml has usage
- \( \mathbf{Y} = \) "2" 

\[ \mathbf{X} \perp 
\mathbf{Y} \mid \mathbf{Z} \]

Here, \( \mathbf{X} \) is set "marginal" of \( \mathbf{Y} \)
\( \text{i.e. } \mathbf{X} \notin \mathbf{Y} \) \( \Rightarrow \) \( p(x,y) = p(x)p(y) \)

\[ p(x,y,z) = p(z|x)p(y|z)p(x) \]

**Example of pairwise indep. R.V.s that are not mutually indep.**
- \( \mathbf{X} = \) can step as \( \mathbb{I} \)
- \( \mathbf{Y} = \) indep can step

\[ \mathbf{X} \perp 
\mathbf{Y} \]

**Other notation:**

- **Expectation/mean** of a R.V.

\[ \mathbb{E}[X] \triangleq \mathbb{E}(X|\mathbf{Z}) \quad \text{for discrete R.V.} \]

\[ \sum_{x} x p(x) \] \( \text{[1 ch. R.V.]} \)

\[ \mathbb{E}[X] \text{ is a linear operator } \Rightarrow \mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] \]

**Variance:**

\[ \text{Var}[X] \triangleq \mathbb{E}[(X-\mathbb{E}[X])^2] \]

\[ = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \]

\[ \text{"measures dispersion"} \]

\[ \sigma = \sqrt{\text{Var}[X]} \]
parametric models:

Let a family of distributions

\[ \mathcal{D} = \{ p(x; \theta) \mid \theta \in \Theta \} \]

possible pdf's/pmf's depending on parameter \( \theta \)

abuse of notation \( \mathcal{D} \leftrightarrow p(x; \theta) \mid \theta \in \Theta \)

better notation \( \mathcal{D} = \{ p(x; \theta) \mid \theta \in \Theta \} \)

Relation:\n\[ X \sim \text{Bern}(\theta) \]

\( \theta \) \text{ is the parameter of the Bernoulli distribution with parameter } \theta.

Probability mass function of Bernoulli distribution:

\[ p(x; \theta) = \text{Bern}(x; \theta) = \theta^x (1-\theta)^{1-x} \]

Another example:
\[ X \sim N(x; \mu, \sigma^2) \]

Bernoulli: result of a coin flip \( X = \{0, 1\} \)

\[ \Pr\{X = 1\mid \theta\} = \theta \Rightarrow \Theta = [0, 1] \]

\[ p(x; \theta) = \theta^x (1-\theta)^{1-x} = \text{Bern}(x; \theta) \]

Another example:
\[ p(x; (\mu, \sigma^2)) = N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \]

\[ N(x; (\mu, \sigma^2)) \]