Example

\[
P(x_1, y_2) = P(X=x_1 \text{ AND } Y=y_2)
\]

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>y_1</td>
</tr>
<tr>
<td>x_1</td>
<td>0.04</td>
</tr>
<tr>
<td>x_2</td>
<td>0.30</td>
</tr>
</tbody>
</table>

\[
P(x_1) = P(X=x_1) + P(X=x_2)
\]

[Recall that \(P(x)\) is short hand for the probability that the random variable \(X\) takes the value \(x\), similarly for \(P(y)\)].

Probabilistic Queries

Organize variables into:
- Evidence (observed), \(E\)
- Query (what you want to know), \(Y\)
- Hidden (leftover), \(X\) (for completeness)

Generic Query: \(P(Y|E)\)

This leads to a distribution over \(Y\) given the evidence

Note that \(X\) is marginalized out

We can use this to make a decision

Simplest is most probable, i.e., \(\arg\max_Y P(Y,E)\)

MAP Query (most probably configuration of variables):

\[MAP(W|E) = \arg\max_W P(W,E) \quad (W=Y \cup X)\]
Arg max $P(x, y)$ is $(x_1, y_2)$
Arg max $P(x)$ is $(x_2)$
Arg max $P(y)$ is $(y_2)$

Arg max $P(x,y)$ is not necessarily $(\text{Arg max } P(x), \text{Arg max } P(y))$

Independence

$X \perp Y \iff P(X|Y) = P(X)$ or $P(Y)=0 \quad *$

$X \perp Y \iff P(X,Y) = P(X)P(Y) \quad *$

Note that Bishop uses $\parallel$ instead of $\perp$

Conditional Independence

$X \perp Y | Z \iff P(X|Y,Z) = P(X|Z) \quad \text{or } P(Y,Z)=0 \quad *$

Equivalent, sometimes more convenient definition

$X \perp Y | Z \iff P(X,Y|Z) = P(X|Z)P(Y|Z) \quad *$

Discrete Distributions (Bernoulli)

$x \in \{0,1\} \quad (\text{e.g., } 1 \text{ is "heads" and } 0 \text{ is "tails"})$

$p(x = 1|\mu) = \mu \quad \text{and } p(x = 0|\mu) = 1 - \mu$

$Bern(x|\mu) = \mu^x(1-\mu)^{(1-x)}$

Study this trick!

$x$ is an indicator variable which is constrained to be “1” for exactly one value, and “0” for the rest.
Code for sampling a Bernoulli

```python
a = rand()
if (a < u) return heads
else return tails
```

Discrete Distributions (Binomial)

Probability distribution for getting $m$ "heads" in $N$ tosses.

$$Bin(m|N,\mu) = \binom{N}{m} \cdot \mu^m (1-\mu)^{N-m}$$

where

$$\binom{N}{m} = \frac{N!}{(N-m)!m!}$$

Example

$N=3$, $m=2$

HHT

HTH

THH

Multi-outcome Bernoulli

Simple extensions to Bernoulli to multiple outcomes (e.g., a six sided die).

Let $K$ be the number of outcomes.

Now we use vectors for $u$ and $x$, i.e., $\mathbf{u}$ and $\mathbf{x}$.

$\mathbf{x}$ is a vector of 0's and exactly one 1 for observed outcome (e.g., rolling 3 with a 6 sided die is $(0,0,1,0,0,0)$).

$$p(\mathbf{x} | \mathbf{u}) = \prod_{k=1}^{K} u_k^{x_k} \quad \text{(note that } \sum_{k=1}^{K} x_k = 1)$$

Multinomial

Extension of binomial to multiple outcomes.

Let $K$ be the number of outcomes.

$$Mult(m_1, m_2, \ldots, m_K) = \binom{N}{m_1 m_2 \ldots m_K} \prod_{k=1}^{K} \mu_k^{m_k}$$

where

$$\binom{N}{m_1 m_2 \ldots m_K} = \frac{N!}{m_1! m_2! \ldots m_K!}$$

and $\sum_{k=1}^{K} m_k = N$
Continuous Spaces

Outcome space is observation of real values (e.g., height, mass)

Example, a random variable, X, can take any value in [0,1] with equal probability.

We say that X is uniformly distributed over [0,1].

Here, \( P(X=x) = 0 \) (uncountable number of possibilities).

To deal with this, we use Probability Density Functions.

Example one

A random variable is uniformly distributed between 0.4 and 0.6, and never occurs outside of that range.

\[
p(x) = \begin{cases} \kappa & x \in [0.4,0.6] \\ 0 & \text{otherwise} \end{cases}
\]

\[
\int_{0.4}^{0.6} p(x) \, dx = \int_{0.4}^{0.6} \kappa \, dx = 0.2\kappa = 1
\]

\[
\kappa = \frac{1}{0.2} = 5 \quad \text{and thus} \quad p(x) = \begin{cases} 5 & x \in [0.4,0.6] \\ 0 & \text{otherwise} \end{cases}
\]

Example two

The univariate Gaussian (or Normal) distribution

\[
N(\mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

Note that \( P \in [0,1] \) but \( p(x) \) can be larger than 1.
Example Three

A continuous random variable can take on the exact values 0.3 and 0.6 with equal probability, and nothing else.

This is really a discrete distribution in disguise.

This PDF is not a function, let alone a continuous function.

If we want to use a PDF to represent it, we can use the "generalized" function \( \delta(x) \).

Dirac delta function

The Dirac delta (generalized) function

\[ \delta(x) = 0, \text{ where } x \neq 0 \]

\[ \int \delta(x) \, dx = 1 \]

\[ \int \delta(x-a) f(x) \, dx = f(a) \]

Example three (continued)

Recall our "function" which was the PDF of a continuous random variable that took the exact values 0.3 and 0.6 with equal probability, and nothing else.

\[ p(x) = \frac{1}{2} \delta(x-0.3) + \frac{1}{2} \delta(x-0.6) \]
Joint Density Functions

Analogous to univariate case (illustrated with two variables)

\[ \int_{Val(X) \times Val(Y)} \int p(x,y) \, dx \, dy = 1 \]

\[ P(a_x \leq X \leq b_x, \ a_y \leq Y \leq b_y) = \int_{a_y}^{b_y} \int_{a_x}^{b_x} p(x,y) \, dx \, dy \]

Marginalization

\[ p(x) = \int_{-\infty}^{\infty} p(x,y) \, dy \]

Example--- multivariate Gaussian

\[ N(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^\frac{1}{2}} \exp\left(\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right) \]

If the variables are independent, then the covariance is diagonal

\[ N(\mu, \Sigma^2) = \frac{1}{(2\pi)^\frac{k}{2} \prod \sigma_i} \exp\left(\frac{1}{2} (x - \mu)^T \left(\text{diag}(\sigma^2)^{-1}\right) (x - \mu)\right) \]

\[ = \prod_{i=1}^{k} N(\mu_i, \sigma_i^2) \]

Conditional Distributions

\[ p(y \mid x) = \frac{p(x,y)}{p(x)} \quad \text{where } p(x) \neq 0 \]

Can get this by marginalizing

\[ p(x) = \int_{-\infty}^{\infty} p(x,y) \, dy \]
Gaussian Facts

For a multivariate Gaussian $p(x_a, x_b)$ with variables partitioned into $x_a$ and $x_b$ we have:

$p(x_a)$ is also Gaussian

and

$p(x_a \mid x_b)$ is also Gaussian

Chapter 2.3 of Bishop has a very thorough treatment of the Gaussian distribution.