Example

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

X \ Y	y_1	y_2
x_1	0.04	0.36
x_2	0.30	0.30

$P(x_1)=P(x_1,y_1)+P(x_1,y_2)$
[i.e., sum across]

X \ Y	y_1	y_2	P(x)
x_1	0.04	0.36	0.4 ← $P(x_1)$
x_2	0.30	0.30	0.6 ← $P(x_2)$
	0.34	0.66	

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

X \ Y	y_1	$P(x \mid y_1)$
x_1	0.04	0.04 / 0.34
x_2	0.30	0.30 / 0.34
	P(0.34)	

(Recall that $P(x|y_1)$ is short hand for the probability that the random variable X takes the value x, given that the random variable Y has value y_1)

Probabilistic Queries

Bold face because these are vectors of variables

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

Generic Query: $P(\mathbf{Y}|\mathbf{E})$
- 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., $\underset{\mathbf{Y}}{\text{Argmax}}\ P(\mathbf{Y},\mathbf{E})$

MAP Query (most probably configuration of variables):

$$MAP(\mathbf{W} \mid \mathbf{E}) = \underset{\mathbf{W}}{\text{Argmax}}\ P(\mathbf{W},\mathbf{E}) \qquad (\mathbf{W}=\mathbf{Y} \cup \mathbf{X})$$

		Y: y_1	Y: y_2	
X	x_1	0.04	0.36	0.4
X	x_2	0.30	0.30	0.6
		0.34	0.66	

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** (Arg max P(x), Arg max P(y))

Review

Independence

This can cause confusion. If P(Y) is zero, the other case cannot be used (divide by zero). However, in this case, Y never happens, and so we (a priori) have a choice to declare whether X is independent from Y or not. However, under scrutiny, the choice does make sense, and allows consistency with the second definition. Note that the second formula works in this (weird) case because if P(Y)=0, then P(X,Y) is also 0.

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

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

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

Review

Conditional Independence

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

Equivalent, sometimes more convenient definition

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

Discrete Distributions (Bernoulli)

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

$p(x=1|\mu)=\mu$ and $p(x=0|\mu)=1-\mu$

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

Study this trick!

x is an indicator variable which is constrained to be “1” for exactly on value, and “0” for the rest.

Code for sampling a Bernoulli

```
a=rand()

if (a<u) return heads
else return tails
```

Discrete Distributions (Binomial)

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

$$Bin(m|N,\mu) = \underbrace{\binom{N}{m}}_{\substack{\text{Number of}\\ \text{ways to get}\\ m \text{ heads}\\ \text{in } N \text{ tosses.}}} \bullet \underbrace{\mu^{m}(1-\mu)^{N-m}}_{\substack{\text{Probility of each}\\ \text{way to get } m \text{ heads}\\ \text{in } N \text{ tosses}}}$$

where $\binom{N}{m} \equiv \dfrac{N!}{(N-m)!m!}$

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).

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

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 $\binom{N}{m_1 \;\; m_2 \;\; \ldots \;\; m_K} = \left(\dfrac{N!}{m_1! \;\; m_2! \;\; \ldots \;\; m_K!} \right)$

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.

Probability Density Functions

$p:\mathbb{R}\mapsto\mathbb{R}$ is a probability density function for X if $p(x)\geq 0$ and

$$\int_{Val(X)} p(x)dx = 1$$

$$P(a \leq X \leq b) = \int_a^b p(x)dx$$ (Probality of the event that $x \in [a,b]$)

$$P(X \in \Delta X) \cong p(x)|\Delta X|$$ (For small ΔX)

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

Example one

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

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

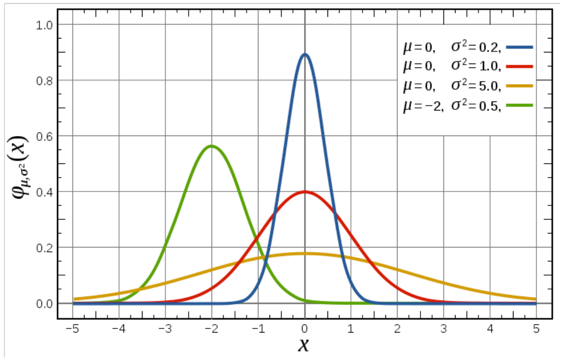
$$\int_{0.4}^{0.6} p(x)dx = \int_{0.4}^{0.6} \kappa\, dx = (0.2)\kappa = 1$$

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

Example two

The univariate Gaussian (or Normal) distribution

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



Example Three

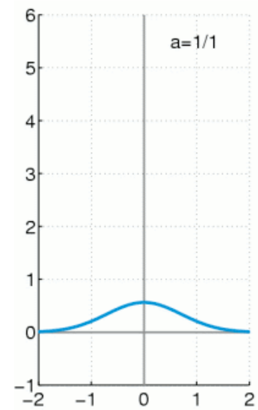
A continuous random variable can take on the exact values 0.3 and 0.6 with equal probability, and nothing else.

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 $\delta(x)$.

Delta "function" demo



Dirac delta function

The Dirac delta (generalized) function

$\delta(x) = 0$, 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)

$$\iint_{Val(X)\times Val(Y)} p(x,y)\,dx\,dy = 1$$

$$P(a_X \le X \le b_X,\ a_Y \le Y \le b_Y) = \int_{a_Y}^{b_Y}\int_{a_X}^{b_X} p(x,y)\,dx\,dy$$

Example--- multivariate Gaussian

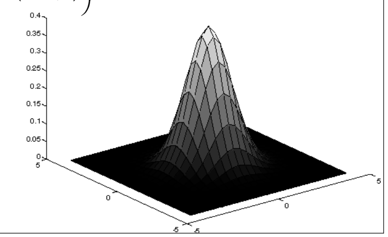
$$\mathbb{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{k}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left(\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\,\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

k is the number of variables (dimension)

If the variables are independent, then the covariance is diagonal

$$\mathbb{N}(\boldsymbol{\mu},\boldsymbol{\sigma}^2) = \frac{1}{(2\pi)^{\frac{k}{2}}\prod_{i=1}^{k}\sigma_i}\exp\left(\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\left(diag(\sigma^2)\right)^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

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



Marginalization

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

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 $p(\mathbf{x}_a, \mathbf{x}_b)$ with variables partitioned into $\mathbf{x}_a$ and $\mathbf{x}_b$ we have:

$p(\mathbf{x}_a)$ is also Gaussian

and

$p(\mathbf{x}_a \mid \mathbf{x}_b)$ is also Gaussian

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