Announcements

Slides for last lecture with some improvements posted.

Somewhat legible scan of chapter 2 of K&F available on website.

Office hour tomorrow is troublesome (but could do it from 10:00-11:00).

Discrete Distributions (Binomial)

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

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

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

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).

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

Multinomial

Extension of binomial to multiple outcomes.
Let K be the number of outcomes.

$$Mult(m_1, m_2, \ldots, m_K) = \begin{pmatrix} & N & & \\ m_1 & m_2 & \ldots & m_K \end{pmatrix} \prod_{k-1}^{K} \mu_k^{m_k}$$

where $\begin{pmatrix} & N & & \\ m_1 & m_2 & \ldots & m_K \end{pmatrix} = \left(\frac{N!}{m_1! \; m_2! \; \ldots \; m_K!} \right)$

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.

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

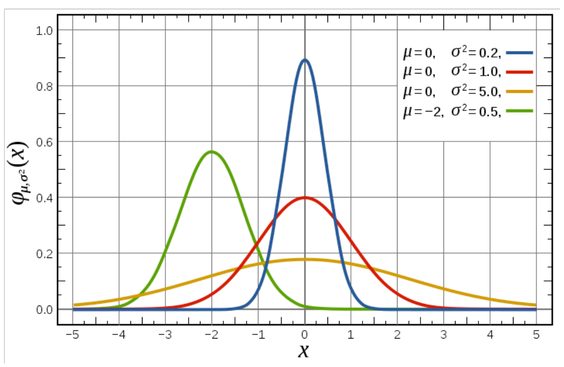
$$\int_{0.4}^{0.6} p(x)\,dx = \int_{0.4}^{0.6} \kappa\,dx = (0.2)\kappa = 1$$

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

Example two

The univariate Gaussian (or Normal) distribution

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



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

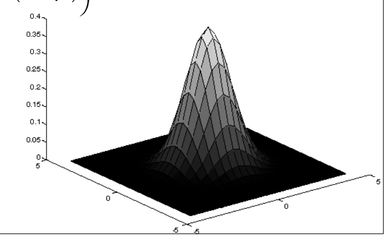
$$\mathbb{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{k}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left(\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\,\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

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

$$p(y\,|\,x) = \frac{p(x,y)}{p(x)} \quad \text{where } p(x)\neq 0$$

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.

Expectation

$$E_p[X] = \sum_x x \cdot P(x) \qquad \text{(discrete)}$$

$$E_p[X] = \int x \cdot p(x)\, dx \qquad \text{(continuous)}$$

$$E_p[X+Y] = E_p[X] + E_p[Y]$$

Implicit definition of a new random variable

Variance

Recall that this is our symbol for independent.

$$Var(X) = E_p\left[\left(X - E_p[X]\right)^2\right]$$

$$Var(X+Y) = Var(X) + Var(Y) \qquad \text{(when X } \perp \text{ Y)}$$

$$Var(aX) = a^2 \cdot Var(X)$$

Standard deviation, $\sigma_X = \sqrt{Var(X)}$

Sampling Continuous Distributions

- Suppose you want to generate samples from (i.e., simulate a probability distribution).
- The typical tool at your disposal is a pseudo random number generator returning approximately uniformly distributed rational numbers in [0,1]
- Sampling Bernoulli processes is straightforward
- Variants of uniform distributions are also easy
- Example: $p(x) = \begin{cases} 5 & x \in [0.4, 0.6] \\ 0 & \text{otherwise} \end{cases}$

Sampling Continuous Distributions

- N(0,1) is less obvious (there are standard fast methods)
- A general approach for sampling a continuous distribution (sometimes call inverse transformation sampling) is based on the cumulative distribution function, CDF, denoted by F(x)

Cumulative Distribution Function

$$F(x) = P(X \le x)$$
$$= \int_{-\infty}^{x} p(x)\,dx \quad \text{(continuous distributions)}$$