Additional references for probability

Wasserman, “All of statistics”
(Vision lab has a few copies that we can lend out).

Forsyth and Ponce chapter
http://luthuli.cs.uiuc.edu/~daf/book/bookpages/pdf/probability.pdf

Your favorite intro to probability book (e.g., “Mathematical statistics and Data Analysis,” by John Rice.)

Google (and WikiPedia)

Bernoulli

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

$$\mathrm{p}(x=1|\mu)=\mu$$

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

Binomial

How likely it is that we get m "heads" in N tosses?

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

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

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).

$p(\mathbf{x} \mid \mathbf{u})=\prod_{k=1}^{K} u_k^{x_k}$ (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 $\binom{N}{m_1 \; m_2 \; \ldots \; m_K} = \left(\frac{N!}{m_1! \; m_2! \; \ldots \; m_K!} \right)$

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

Related distributions (*)

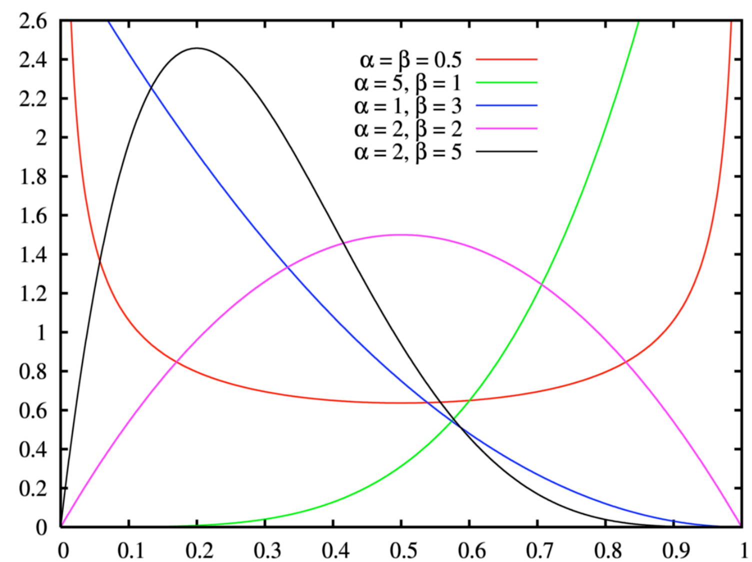
Beta (binary case)

Dirichlet (multi-outcome case)

These are “conjugate priors” for the two pairs of distributions just introduced.

(*) In class, we motivated these distributions as being conjugate to the discrete ones just covered, but this was confusing because these are continuous distributions, which we covered **next**.

$$Beta(u \mid a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1} (1-u)^{b-1}$$



Conjugacy

Informal definition: Given a likelihood function

$l(\theta \mid x) = p(x \mid \theta)$ (we reverse θ and x when we call it a likelihood function)

a (prior) distribution is natural distribution where the posterior,

$p(\theta \mid x) \propto p(x \mid \theta) p(\theta)$, has the same form as $p(\theta)$.

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 (infinite number of possibilities).

To deal with this tecnicality, 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

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

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

$$P(X \in \Delta X) \cong p(x)|\Delta X| \qquad \text{(For small } \Delta X\text{)}$$

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.

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

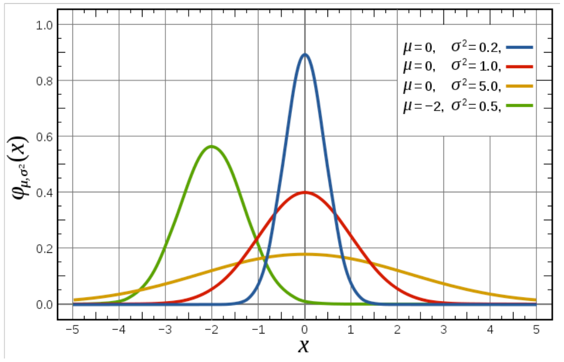
$$\int_{0.4}^{0.6} p(x)\,dx = \int_{0.4}^{0.6} \kappa\,dx = (0.2)\kappa = 1$$

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

Example two

The univariate Gaussian (or Normal) distribution

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



Joint Density Functions

Analogous to univariate case (illustrated with two variables)

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

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

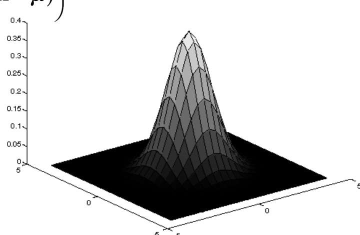
Example--- multivariate Gaussian

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

If the variables are independent, then the covariance is diagonal

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

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



Marginalization

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

Conditional Distributions

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

Expectation

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

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

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

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