Review

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.

Review

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.

Review

Example one

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

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

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

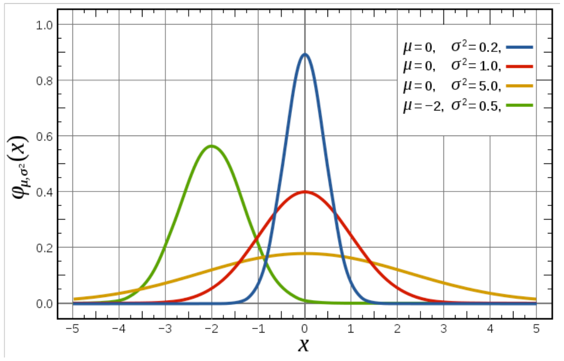
$$\kappa = \frac{1}{0.2} = 5 \quad \text{and thus} \quad p(x) = \begin{cases} 5 & x \in [0.4, 0.6] \\ 0 & \text{otherwise} \end{cases}$$

Review

Example two

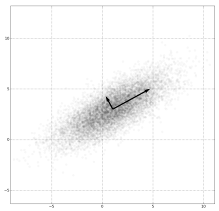
The univariate Gaussian (or Normal) distribution

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



Review

Multivariate Gaussian



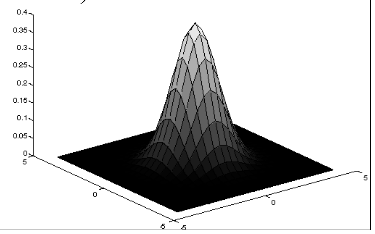
$$\mathbb{N}(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{k}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left(\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\,\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

k is the number of variables (dimension)

If the variables are independent, then the covariance is diagonal

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

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



IE, simply the product of univariate normals

Review

Conditional Distributions

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

Can get this by marginalizing

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

Review

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

$p(\mathbf{x}_a\,|\,\mathbf{x}_b)$ 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.

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

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

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

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: $p(x) = \begin{cases} 5 & x \in [0.4, 0.6] \\ 0 & \text{otherwise} \end{cases}$

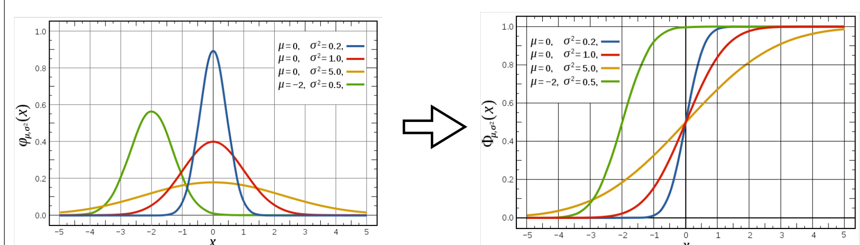
Sampling Continuous Distributions

- 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

$$F(x) = P(X \le x)$$

$$= \int_{-\infty}^{x} p(x)dx \quad \text{(continuous distributions)}$$



Sampling Continuous Distributions

We know how to sample y uniformly from [0,1]

We want to map $y \Rightarrow x \in [-\infty,\infty]$ where is x distributed as $p(x)$

For simplicity, map them monotonically (bigger $y \Rightarrow$ bigger x)

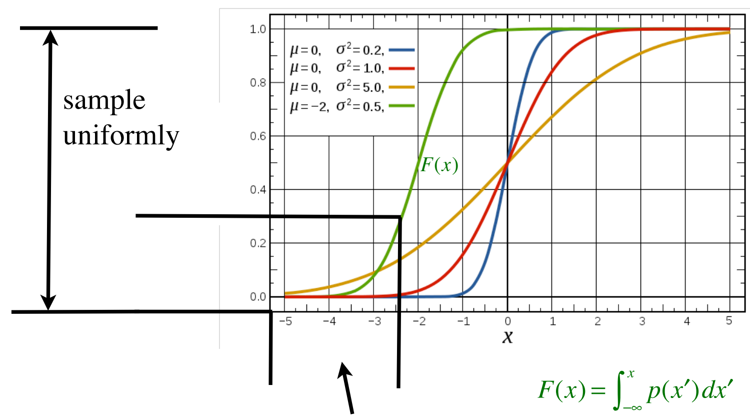
All samples in U=[0,y] should map to total probability y over $p(x)$.

We know how to sample y uniformly from [0,1]

We want to map $y \Rightarrow x \in [-\infty,\infty]$ where is x distributed as $p(x)$

For simplicity, map them monotonically (bigger $y \Rightarrow$ bigger x)

All samples within U=[0,y] should map to total probability y from $p(x)$.



Sampling Continuous Distributions

We know how to sample y uniformly from [0,1]

We want to map $y \Rightarrow x \in [-\infty,\infty]$ where is x distributed as $p(x)$

For simplicity, map them monotonically (bigger $y \Rightarrow$ bigger x)

All samples in U=[0,y] should map to total probability y in $p(x)$

So U=[0,y] maps into $P=[-\infty,x]$, where $y=\int_{-\infty}^{x} p(x')dx' = F(x)$

In other words, $x = F^{-1}(y)$

Sampling Continuous Distributions

- To sample a distribution p(x) (crude instructional algorithm)

Prepare an approximation of $F(x)$
in a vector $F=(x_1, x_2, x_3, \ldots, x_N)$

Loop
  sample $y \in [0,1]$
  find i so that $F(x_i) < y$ and $F(x_{i+1}) > y$
  report $(x_i + x_{i+1})/2$

Example (from Bishop, PRML)

Estimating the mean of a univariate Gaussian

Assume that the variance is known.
Given data points x_i, what is the "best" estimate for the mean?

Think for a moment about the joint distribution of the mean and the observations (both are random variables)
i.e., we are interested in $p(u,\{x_i\})$

The question is particularly about the conditional density $p(u|\{x_i\})$

Example (from Bishop, PRML)

Estimating the mean of a univariate Gaussian

The question is about the conditional density $p(u|\{x_i\})$

$$p(u|\{x_i\}) \propto p(\{x_i\}|u) \quad \text{(assuming uniform prior)}$$

$$p(\{x_i\}|u) = \prod_i p(x_i|u)$$

$$\propto \prod_i e^{-\frac{(x_i-u)^2}{2\sigma^2}} \quad \text{(remember, variance is a known constant)}$$

$$p(u|\{x_i\}) \propto p(\{x_i\}|u) \quad \text{(assuming uniform prior)}$$

$$p(\{x_i\}|u) = \prod_i p(x_i|u)$$

$$\propto \prod_i e^{-\frac{(x_i-u)^2}{2\sigma^2}}$$

To find a "best" answer, we can adjust u to make the above *likelihood* big

$$u_{ML} = \arg\max_u \left(p\left(u|\{x_i\}\right) \right)$$

We can maximize the likelihood by minimizing the negative log

We can maximize the likelihood by minimizing the negative log

$$-\log\left(p\left(u|\{x_i\}\right)\right) = -\log\left(\prod_i e^{-\frac{(x_i-u)^2}{2\sigma^2}}\right) \propto \sum_i (x_i-u)^2$$

$$u_{ML} = \arg\min_u \left(\sum_i (x_i-u)^2 \right)$$

Differentiating and setting to zero reveals that

$$u = \frac{1}{N}\sum x_i$$