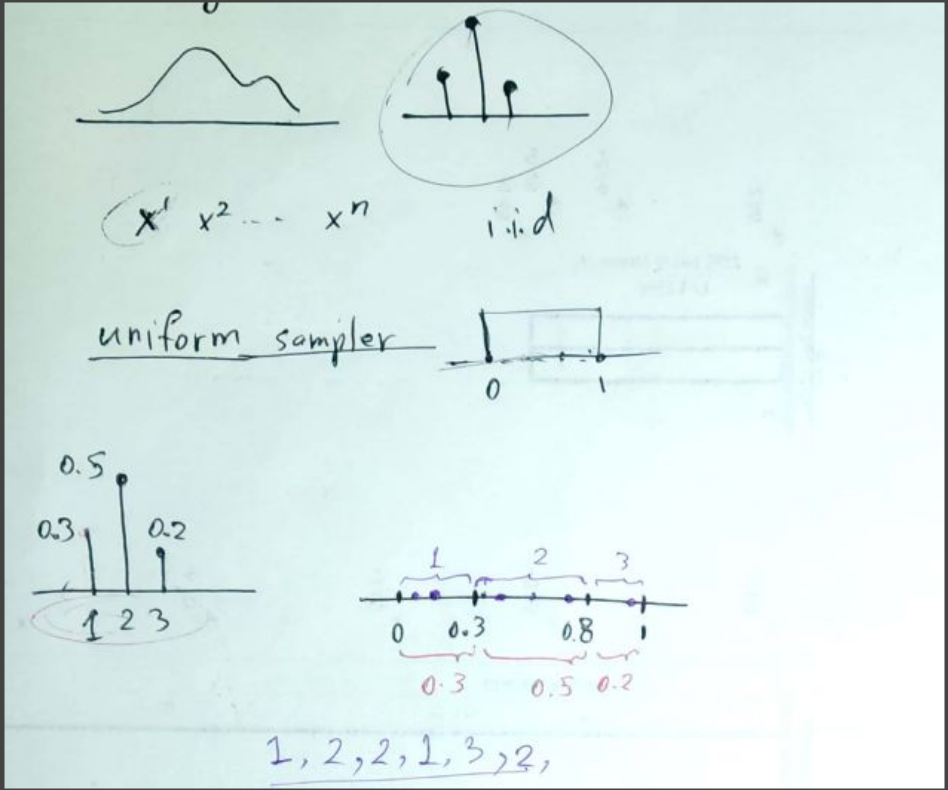
Mathematics for AI

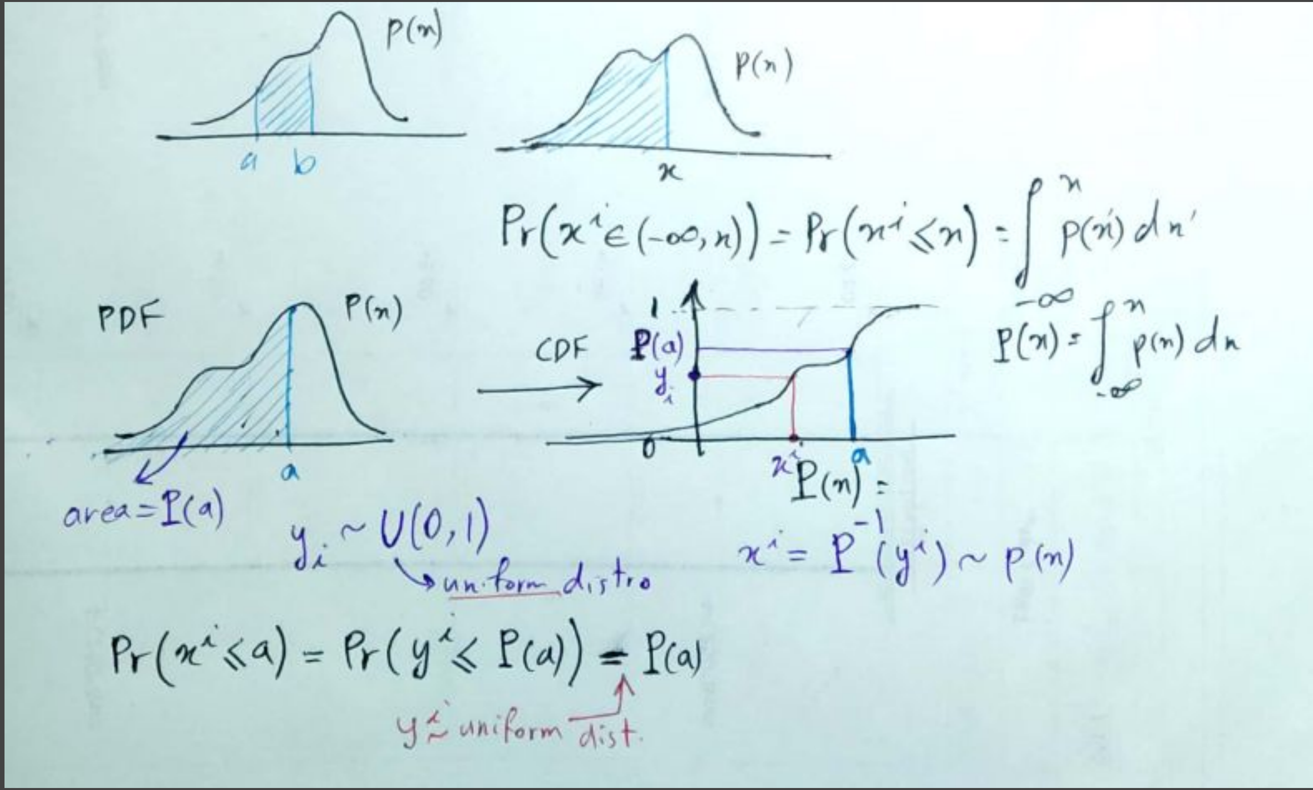
Lecture 26

Sampling (Cont.)

Sampling (Review)



Sampling Continuous Distributions



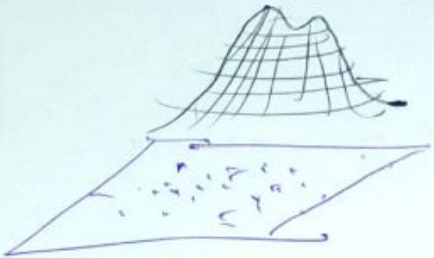
Sampling Multivariable Distributions

What About 2D & ND distributions?

$$P(x,y) \xrightarrow{\text{sample}} (x^1,y^1), (x^2,y^2) \cdots, (x^m,y^m)$$

$$P\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$$

$$P(X) = P\left(\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}\right) = P(X_1, X_2, \dots, X_n)$$



sample $X^1, X^2, \dots, X^m$

$$\begin{bmatrix} X_1^1 \\ X_2^1 \\ \vdots \\ X_n^1 \end{bmatrix}, \begin{bmatrix} X_1^2 \\ X_2^2 \\ \vdots \\ X_n^2 \end{bmatrix} \cdots, \begin{bmatrix} X_1^m \\ X_2^m \\ X_n^m \end{bmatrix}$$

Sampling Multivariable Distributions - Independent Case

$$P(X) = P(x_1, \text{——}, x_n) = P_1(x_1)\, P_2(x_2) \cdots P_n(x_n)$$

$$\left.\begin{array}{l} x_1^i \sim P_1 \\ x_2^i \sim P_2 \\ \vdots \\ x_n^i \sim P_n \end{array}\right\} \Rightarrow (x_1^i, x_2^i, \dots, x_n^i) \sim P$$

Sampling Multivariable Distributions - Using Chain Rule

what if variables are not independent?

$p(x,y)$, x, y are not independent.

$$p(x,y) = p(y|x)\, p(x)$$

$$p(x) = \sum_y p(x,y)$$

$$p(y|x) = \frac{p(x,y)}{p(x)} = \frac{p(x,y)}{\sum_{y'} p(x,y')}$$

x^i = take a sample from $p(x)$

y^i = take a sample from $p(y|x^i)$

$\Rightarrow$ (x^i, y^i) is a sample from $p(x,y)$

$$\Pr(X = x^i, Y = y^i) = \Pr(X = x^i)\Pr(Y = y^i | X = x^i)$$

$$= p(x^i)\, p(y^i|x^i) = p(x^i, y^i)$$

Sampling Multivariable Distributions - Using Chain Rule

MA24 III

$$P(X_1, X_2, \dots, X_n) = P(X_n \mid X_1 - X_{n-1})\, P(X_1, X_2, \dots, X_{n-1})$$

$$= P(X_n \mid X_1 - X_{n-1})\, P(X_{n-1} \mid X_1 - X_{n-2})\, P(X_1 - X_{n-2})$$

$$\vdots$$

$$P(X_1, X_2, \dots, X_n) = P(\underline{X_n} \mid X_1 - X_{n-1})\, P(\underline{X_{n-1}} \mid X_1 - X_{n-2})$$

$$P(\underline{X_{n-2}} \mid X_1 - X_{n-3}) \cdots P(\underline{X_3} \mid X_1, X_2)\, P(X_2 \mid X_1)\, \underline{P(X_1)}$$

$$\left.\begin{array}{l}
X_1^i \sim P(X_1) \\
X_2^i \sim P(X_2 \mid X_1^i) \\
X_3^i \sim P(X_3 \mid X_2^i, X_1^i) \\
\quad \vdots \\
X_{n-1}^i \sim P(X_{n-1} \mid X_1^i, X_2^i, \dots, X_{n-2}^i) \\
X_n^i \sim P(X_n \mid X_1^i, X_2^i, \dots, X_{n-1}^i)
\end{array}\right\} \text{might be hard / impossible to compute.}$$

$$\Longrightarrow (X_1^i, X_2^i, \dots, X_n^i) \sim P(X_1, X_2, \dots, X_n)$$

Example: Sampling from a Markov Chain

$X_1, X_2, \ldots, X_n, \ldots$

$$X_1 \to X_2 \to X_3 \to \cdots \to X_{n-1} \to X_n$$

$\forall t \quad P(X_t \mid X_{t-1}, X_{t-2}, \ldots, X_2, X_1) = P(X_t \mid X_{t-1})$

Markov Property (for Markov Chains)

X_t is conditionally independent of $X_{t-2}, X_{t-3}, \ldots, X_2, X_1$ given X_{t-1}.

$$P(X_1, X_2, \ldots, X_n) = P(X_n \mid X_1 - X_{n-1})\, P(X_{n-1} \mid X_1 - X_{n-2}) \cdots P(X_3 \mid X_2, X_1)\, P(X_2 \mid X_1)\, P(X_1)$$

$$= P(X_n \mid X_{n-1})\, P(X_{n-1} \mid X_{n-2}) \cdots P(X_3 \mid X_2)\, P(X_2 \mid X_1)\, P(X_1)$$

Example: Sampling from a Markov Chain

$$X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \cdots$$

$$X_1^i \sim P(X_1)$$

$$X_2^i \sim P(X_2 \mid X_1^i)$$

$$X_3^i \sim P(X_3 \mid X_2^i)$$

$$\vdots$$

$$X_n^i \sim P(X_n \mid X_{n-1}^i)$$

Sampling Using Change of Variables

K. N. Toosi
University of Technology

$P_X(X) \quad X = (X_1, X_2, \ldots, X_n)$

$$X = f(Y)$$

We can write the random variable X as $X = f(Y)$ for some function f. where $Y \sim P_Y(Y)$ & $P_Y(Y)$ is easy to sample from. ~~E.g. $P_Y(Y)$ is~~

$$Y^i \sim P_Y(Y)$$

$$X^i = f(Y^i)$$

$P_Y(Y)$ is easy to sample from

→ Example $P_Y(Y) = P_Y(Y_1, Y_2, \ldots, Y_n)$ is a Markov Chain

→ Example $P_Y(Y) = P_Y(Y_1, Y_2, \ldots, Y_n) = P(Y_1)\, P(Y_2) \ldots P(Y_n)$

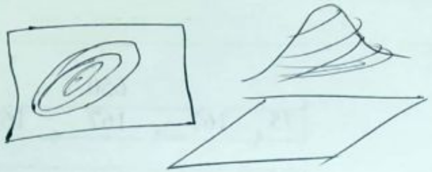
$Y_1, Y_2, \ldots, Y_n$ are independent.

Example: Multivariable Gaussian Distribution

V

Example:

Draw samples from a multivariable Gaussian distribution $N(\overset{0}{\cancel{\mu}}, \Sigma)$



$$Y \sim N\left(0, \begin{bmatrix} 1 & & \emptyset \\ & 1 & \\ \emptyset & & \ddots \\ & & & 1 \end{bmatrix}\right) = N(0, I)$$

$$X \sim N(0, \Sigma)$$

$Y \sim N(0, I) \rightarrow$ stardard normal distribution

$N(0, I)$ is easy to sample from because $Y_1, Y_2, \dots, Y_n$ are independent.

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

Example: Multivariable Gaussian Distribution

K. N. Toosi
University of Technology

are independent

Let Y^i be a sample from $N(0, I)$

$$X^i = A Y^i$$

A: $n \times n$ matrix

AY is a Gaussian distribution

$$X = AY \qquad \mu_X = E\{AY\} = AE\{Y\} = \vec{0}$$

$$\Sigma_X = Cov\{AY\} = A\,Cov\{Y\}\,A^T$$

$$= A I A^T = AA^T$$

Choose A such that $AA^T = \Sigma$.

Σ covariance matrix $\Rightarrow$ Positive (semi) definite.

Decompose Σ into AA^T

there are many choices for A

for any orthogonal H: $AA^T = AHH^{-1}A = AHH^TA^T = AH(AH)^T$

Example: Multivariable Gaussian Distribution

MA26 (VI)

Draw samples from $N(0, \Sigma)$

$X^1, X^2, \dots, X^m$

1- Decompose Σ as $\Sigma = AA^T$

for $i = 1 \dots m$:

$Y^i \sim N(\vec{0}, I)$

$X^i = AY^i$

$\Rightarrow X^i$ are samples from $N(0, \Sigma)$

Using Cholesky Decomposition

$$\underset{\mathbb{R}^n}{X^i} = \underset{\mathbb{R}^{n\times n}}{A}\,\underset{\mathbb{R}^n}{Y^i}$$

we need N^2 multiplications

$$\Sigma = AA^T$$

Cholesky Decomposition

$$\underset{\text{positive semi-definite}}{\Sigma} = L\underset{\text{lower-triangular}}{L^T}$$

$$X^i = LY^i$$

$\frac{N(N+1)}{2}$ multipilications

Example: Multivariable Gaussian Distribution

K. N. Toosi
University of Technology

- Draw Samples from $\mathcal{N}(\mu, \Sigma)$:

$$\Sigma = AA^T$$

$$X^i = AY^i + \mu$$

- ~~MA~~ Markov Chain Monte Carlo (MCMC)

```
L = np.linalg.choolskey(Σ)
Ys = [Y¹ — Yᵐ] = np.random.randn(n,m)
Xs = L @ Ys + mu.reshape((-1,1))
```

Markov Chain Monte Carlo methods

K. N. Toosi
University of Technology