15-381: Artificial Intelligence

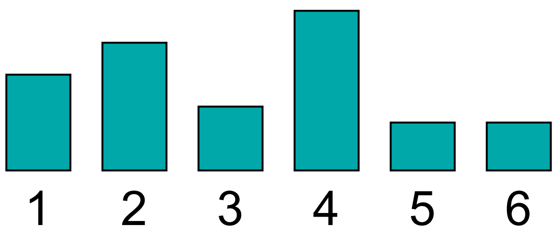
Probabilistic Reasoning and Inference:
Statistics and distributions

Outline

- Continuous distributions
  - Probability density functions, Cumulative density functions
  - Recap on the probability rules
- Gaussian distribution, multivariate Gaussian
- Density estimation example
  - Joint density estimation
  - Naïve density estimation
- Preview of Bayesian networks

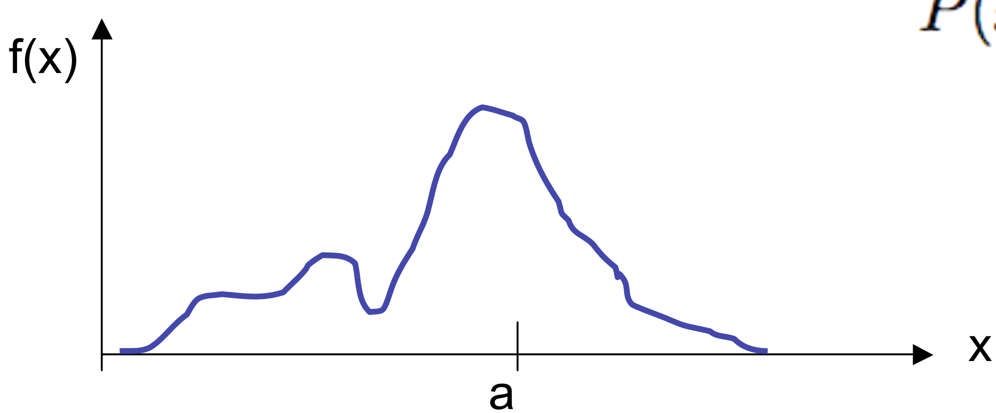
Probability Density Function

- Discrete distributions



$$\sum_i P(X = x_i) = 1$$

- Continuous: Cumulative Density Function (CDF): *F(a)*



$$P(x \leq a) = \int_{-\infty}^{a} f(\tau) d\tau$$

Cumulative Density Functions

- Total probability

$$P(\Omega) = \int_{-\infty}^{\infty} f(x)dx = 1$$

- Probability Density Function (PDF)

$$\frac{d}{dx}F(x) = f(x)$$

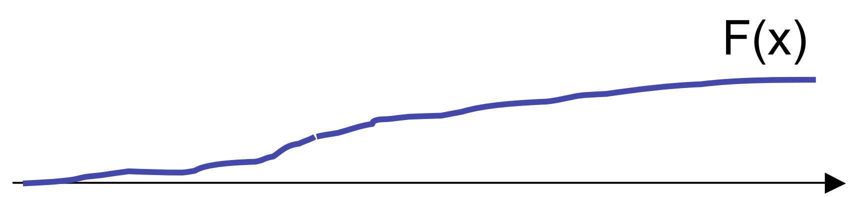
- Properties:

$$P(a \leq x \leq b) = \int_{b}^{a} f(x)dx = F(b) - F(a)$$

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to \infty} F(x) = 1$$

$$F(a) \geq F(b) \; \forall a \geq b$$



Expectations

- Mean/Expected Value:

$$E[x] = \bar{x} = \int x f(x) dx$$

- Variance:
  - Note:

$$Var(x) = E[(x - \bar{x})^2] = E[x^2] - (\bar{x})^2$$

- In general:

$$E[x^2] = \int x^2 f(x) dx$$

$$E[g(x)] = \int g(x) f(x) dx$$

Multivariate

- Joint for (x,y)

$$P((x,y) \in A) = \int \int_A f(x,y) dx dy$$

- Marginal:

$$f(x) = \int f(x,y) dy$$

- Conditionals:

$$f(x|y) = \frac{f(x,y)}{f(y)}$$

- Chain rule:

$$f(x,y) = f(x|y)f(y) = f(y|x)f(x)$$

Bayes Rule

- Standard form:

$$f(x|y) = \frac{f(y|x)f(x)}{f(y)}$$

- Replacing the bottom:

$$f(x|y) = \frac{f(y|x)f(x)}{\int f(y|x)f(x)dx}$$

Binomial

- Distribution:

- Mean/Var: $x \sim Binomial(p, n)$

$$P(x = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E[x] = np$$

$$Var(x) = np(1-p)$$

Uniform

- Anything is equally likely in the region [a,b]
- Distribution: $x \sim U(a,b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases}$$

- Mean/Var

$$E[x] = \frac{a+b}{2}$$

$$Var(x) = \frac{a^2+ab+b^2}{3}$$

Gaussian (Normal)

- If I look at the height of women in country xx, it will look approximately Gaussian
- Small random noise errors, look Gaussian/Normal

- Distribution:

$$x \sim N(\mu, \sigma^2)$$

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

- Mean/var

$$E[x] = \mu$$
$$Var(x) = \sigma^2$$