Review of Probability

ECE57000: Artificial Intelligence

David I. Inouye

Sep. 18, 2019

Announcements

- HW2 due on Friday (Sep. 20) at 12noon
- Quiz 3 on Monday (Sep. 23)

<u>Probability distributions</u> attach probabilities to all possible values of a random variable

- **<u>Probability mass functio</u>**n (PMF) is used for *discrete* random variables
- A PMF P for random variable X that satisfies the following:
  1. Domain of P must include all possible states of X
  2. Unit domain: $\forall\, x \in X, 0 \leq P(x) \leq 1$
  3. Sum to 1: $\sum_{x \in X} P(x) = 1$
  - (Write on board)

Probability distributions attach probabilities to all possible values of a random variable

- *Probability density function* (PDF) is used for *continuous* random variables
- A PDF p for random variable X that satisfies the following:
  1. Domain of p must include all possible states of X
  2. Non-negative: $\forall\, x \in X, p(x) \geq 0$ ** $p(x)$ **could be greater than 1**
  3. Integrate to 1: $\int_{x \in X} p(x) = 1$
- $p(x)$ is NOT a probability, rather ***integrating*** the PDF gives probabilities over **sets**

Integrate PDF to get probabilities that random variable lies within set (usually a range)

- The probability that X is less than q

$$\Pr(X \le q) = \int_{-\infty}^{q} p(x)dx$$

- The probability that X lies between a and b

$$\Pr(a \le X \le b) = \int_{a}^{b} p(x)dx$$

- The probability that X lies between (a and b) or between (c and d)

$$\Pr(a \le X \le b \text{ OR } c \le X \le d)$$
$$= \int_{a}^{b} p(x)dx + \int_{c}^{d} p(x)dx$$

Cumulative distribution function (CDF) is the integral of the PDF from the left up to query point q

- The CDF is the probability that X is less than q

$$F(q) \equiv \Pr(X \leq q) = \int_{-\infty}^{q} p(x)dx$$

- What does $F(\infty)$ equal?

- The probability between a and b can be written as:

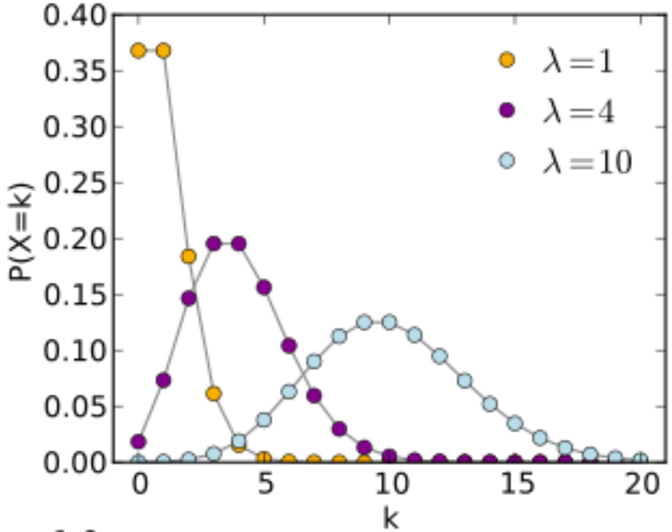
$$\Pr(a < X \leq b) = F(b) - F(a)$$

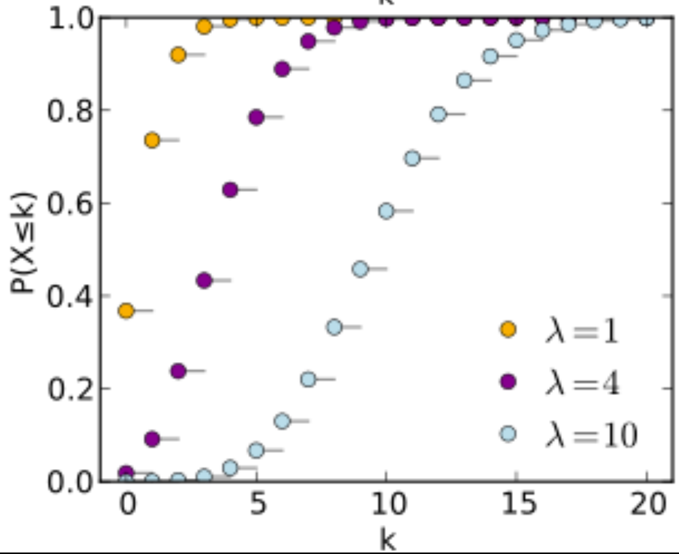
- The PDF is the derivative of CDF:

$$p(x) = \frac{dF(x)}{dx}$$

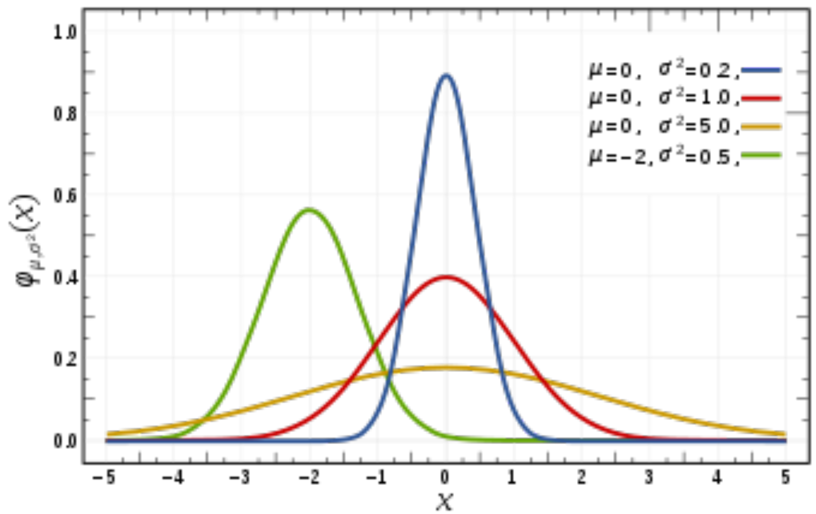
Examples of PMF/PDF and corresponding CDF

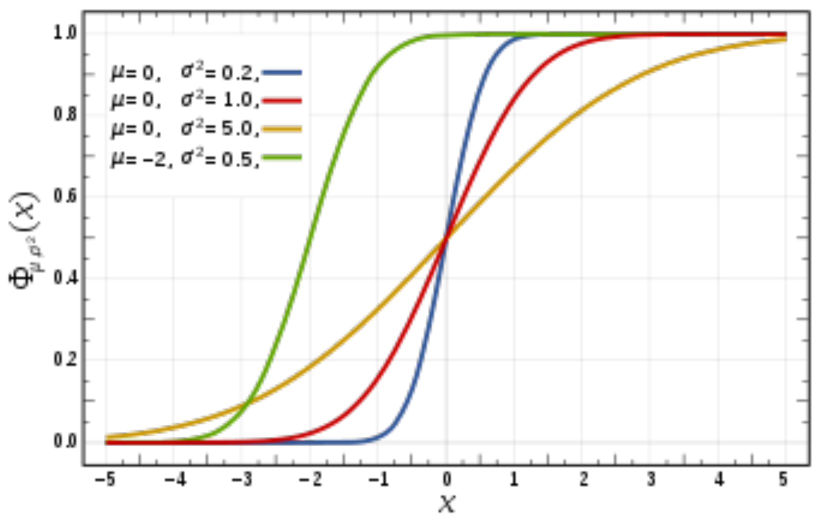
Discrete PMF/CDF





Continuous PDF/CDF





Notation: Tilde used to specify distribution of random variable ($\sim$ in LaTeX)

- $X \sim \mathcal{N}(\mu = 0, \sigma = 1)$
  - "Random variable X is **distributed** as a normal distribution with mean of zero and standard deviation of 1."
- $X \sim \mathrm{Uniform}(\alpha, \beta)$
  - "Random variable X is **distributed** as a uniform distribution with parameters α and β (parameters may be unknown)."
- $X \sim P(x)$ or $X \sim \mathbb{P}(x)$
  - "Random variable X is **distributed** as the distribution represented by PMF/PDF $P(x)$ or $\mathbb{P}(x)$."

Notation: Semicolon ";" (or sometimes bar "|") often used to specify parameters

- $p(x\,;\alpha,\beta)=\frac{1}{\beta-\alpha}$
  - "The PDF of X is parameterized by α and β."
  - This is the uniform distribution between α and β.
- $P(x\,;\lambda)$
  - "The PMF of X is parameterized by λ."
- $p(x \mid \mu,\sigma)$
  - "The PDF of X is parameterized by μ and σ."

Joint distribution of multiple variables

- Joint PDF/PMF is a function of two or more random variables (or a random vector)

- Joint PDF/PMF can be written as:

$$p(x, y), \qquad p(x_1, x_2), \qquad p(\boldsymbol{x})$$

- If $X \in [-1, 1]$ and $Y \in [-1,1]$ is the following a valid PDF?

$$p(x, y) = xy$$

- If $X \in [0, 1]$ and $Y \in [0,1]$ is the following a valid PDF?

$$p(x, y) = 4xy$$

Marginal distribution is sum/integral over other variables

- Example: Height and weight, "What is the distribution of height regardless of weight?"
- Given joint distribution $P(x, y)$ the marginal is:

$$P(x) = \sum_{y \in \mathcal{Y}} P(x, y) \text{ and } P(y) = \sum_{x \in \mathcal{X}} P(x, y)$$

- Given joint distribution $P(x, y)$ the marginal is:

$$p(x) = \int_{y \in \mathcal{Y}} p(x, y) dy \text{ and } p(y) = \int_{x \in \mathcal{X}} p(x, y) dx$$

- Example: $P(x, y) = \big[[0.1, 0.4], [0.3, 0.2]\big]$
- Example: $p(x, y) = 4xy$

Conditional distribution is the distribution *given* some other event

- What is the distribution of weight *given* that a person is x inches tall?
- Conditional density is the joint PDF/PMF *renormalized* by marginal density of event:

$$p(y \mid x) \equiv \frac{p(x, y)}{p(x)}$$

- Example: $P(x, y) = \big[[0.1, 0.4], [0.3, 0.2]\big]$
- Example: $p(x, y) = 4xy$

Chain rule (or product rule) of probability

- The joint distribution can be written as product of conditional PDFs/PMFs:

$$p(x_1, x_2) = p(x_1)p(x_2|x_1)$$
$$p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_1, x_2)$$

- This can be written as:

$$p(x_1, x_2, \dots, x_d) = \prod_{i=1}^{d} p(x_i|x_1, \dots, x_{i-1})$$

- Consequence (order doesn’t matter):

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

Bayes rule: Enables conversion between one conditional and the other (they are ***different***)

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

(derive on board, show reverse)

When are $p(x|y)$ and $p(y|x)$ equal?

<u>Independence</u> means that one variable is not affected by the other variable

- Example: Flip two coins, X and Y are 0 or 1.
- Counterexample: Roll dice for number X; then flip that number of coins and count the number of heads Y.
- Formally, PDF/PMF can be written as product of functions that only involve x or y (but not both)

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

- Usually, these are the marginal densities:

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

- Equivalent definition:

$$p(x|y) = p(x) \text{ and } p(y|x) = p(y)$$

Next time: Expectations, variance, covariance, functions of RV, empirical distribution and expectation, entropy, KL divergence