7 Probability Review

PVK 2019: MAD

bacdavid@student.ethz.ch

gitlab.ethz.ch/bacdavid

Schedule

1. Theory
    1. Setup
    2. Random variable
    3. Marginalization, Conditional Probability, and Bayes Rule
    4. Expectation Value and Variance
    5. Sampling
2. Exercises
3. Homework

Theory

Setup

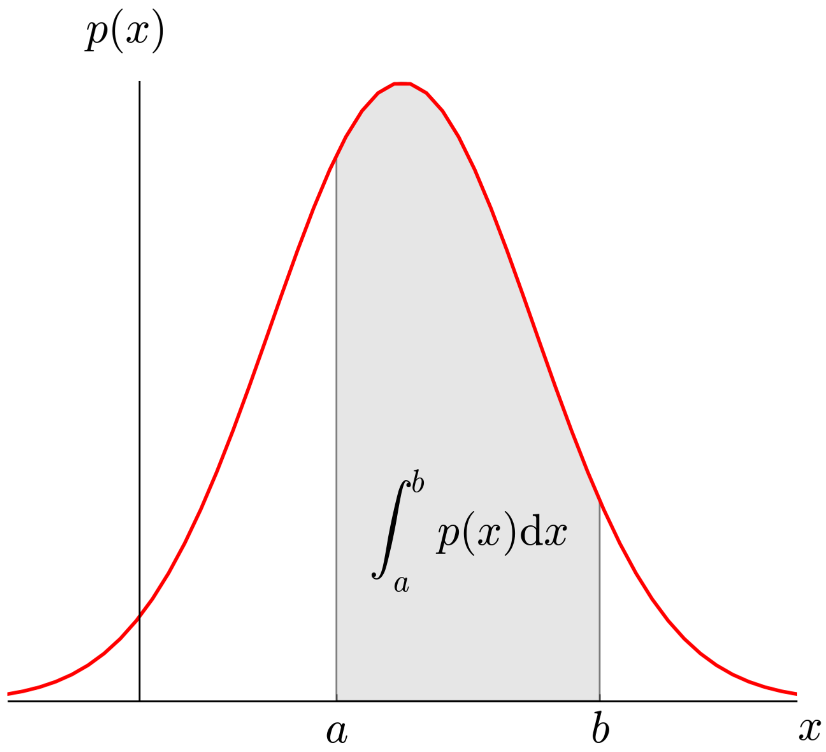
- Random variables $x, y, \ldots$
- Probability assignment $p(x), x \in \{x_1, \ldots\}$
- Probability density $p(x), x \in [x^-, x^+]$
- Event Probability $P(E)$
- Joint Probability $p(x, y, \ldots)$
- Conditional Probability $p(x \mid y)$
- Expecation value of x over $p(x)$ $\mathbb{E}_{p(x)}[x]$
- Variance of x over $p(x)$ $Var(x) = \sigma^2$

Random Variable

- Is not like a classical variable:
  - Classical variable: Defined by value
  - Random variable: Defined by distribution over values

- Discrete Random Variables
  - $p: \Omega \to \mathbb{R}$, where eg. Ω={0, 1, 2, …}
  - $\sum_{\Omega} p = 1, p \geq 0$

- Continuous Random Variables
  - $p: \Omega \to \mathbb{R}$, where eg. Ω=[0, 1]
  - $\int_{\Omega} p = 1, p \geq 0$



Marginalization, Conditional Probability, and Bayes Rule

- Marginalization:

$$p(x) = \int_{\mathcal{Y}} p(x, y) dy$$

- Conditional Probability

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

- Bayes Rule:

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

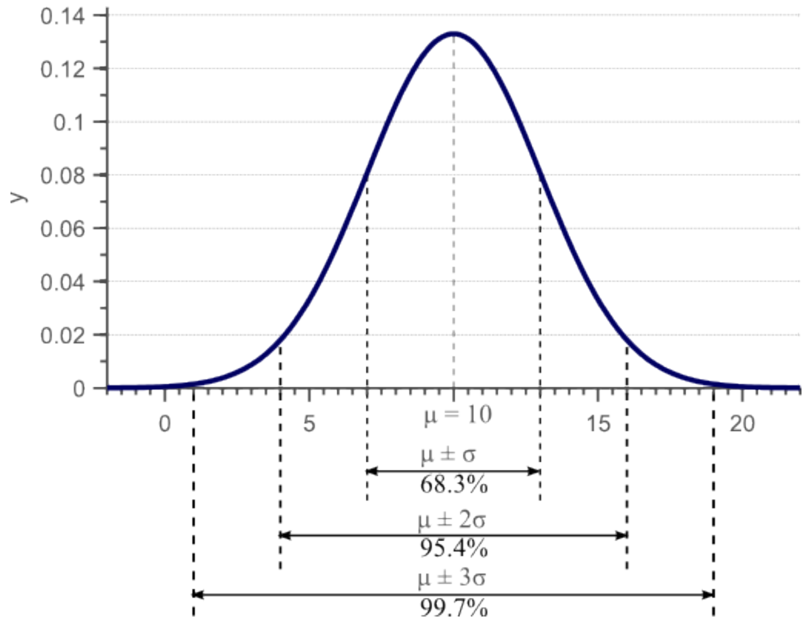
Expectation Value and Variance

- Expectation value:

$$\mathbb{E}[f(x)] = \int_{-\infty}^{\infty} f(x)p(x)dx$$

- Variance:

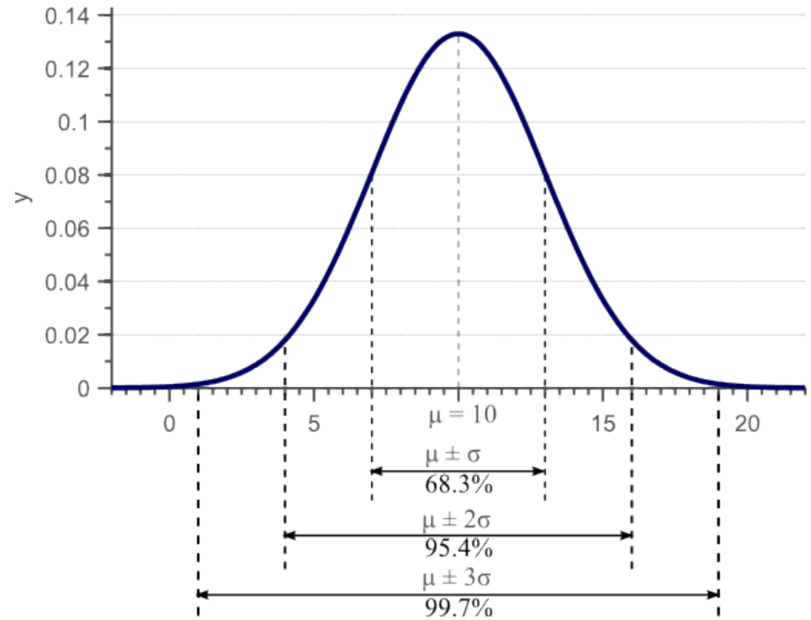
$$Var\big(f(x)\big) = \sigma^2$$
$$= \mathbb{E}[f^2(x)] - \mathbb{E}^2[f(x)]$$



Sampling

- Generate samples from a random distribution

- Three approaches (covered):
  - Inverse CDF
  - Rejection Sampling
  - Markov Chain Monte Carlo



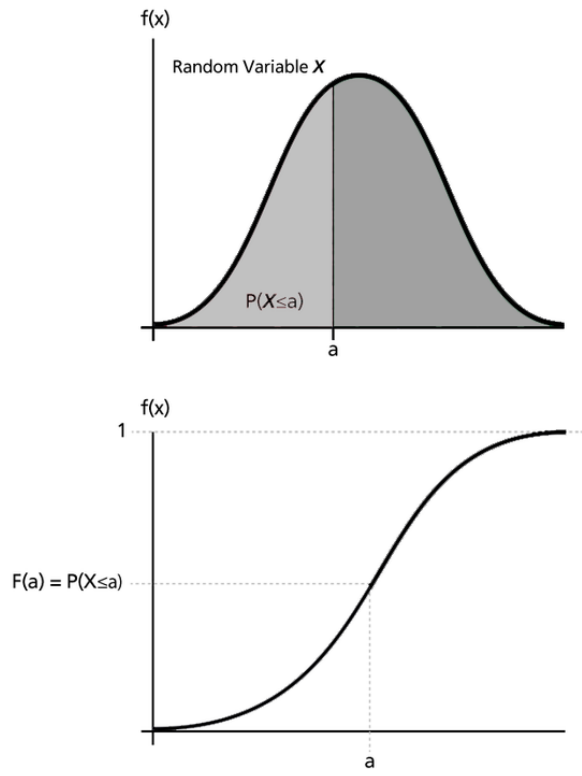
Sampling cont. *Inverse CDF sampling*

- Cumulative Distribution Function:

$$cdf(x) = \int_{-\infty}^{x} p(\bar{x})d\bar{x}$$

- Inverse CDF sampling:
  1. Sample $u \sim unif(0,1)$
  2. Generate sample $x = cdf^{-1}(u)$

- Difficult: Have to compute integral & inverse!

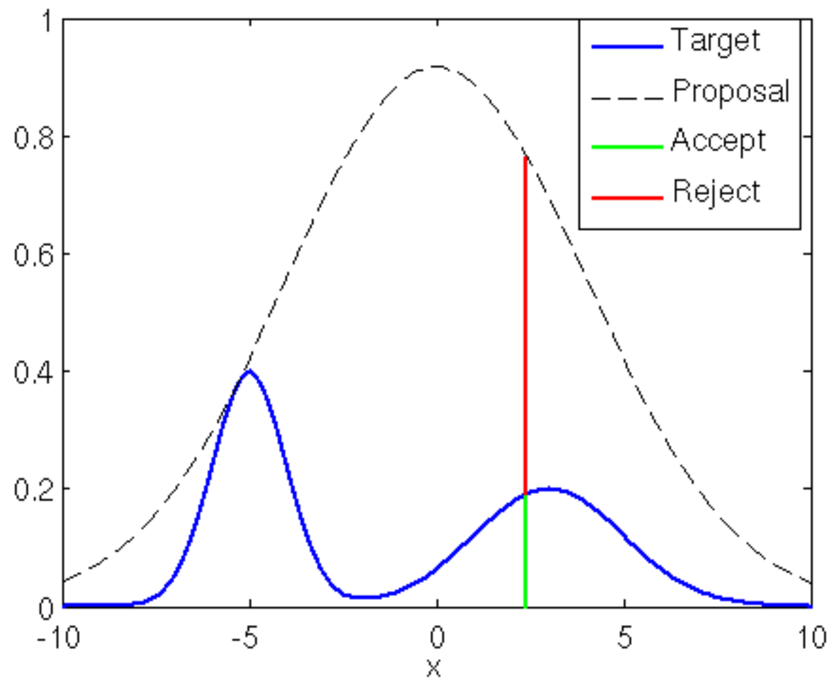


Sampling cont. *Rejection Sampling*

- Define envelope "proposal":
$$p(x) \leq L \cdot q(x)$$

- Rejection Sampling:
  1. Draw $y \sim q$ and $u \sim unif(0,1)$
  2. If $u \cdot L \cdot q(y) \leq p(y)$ keep the sample y, else discharge

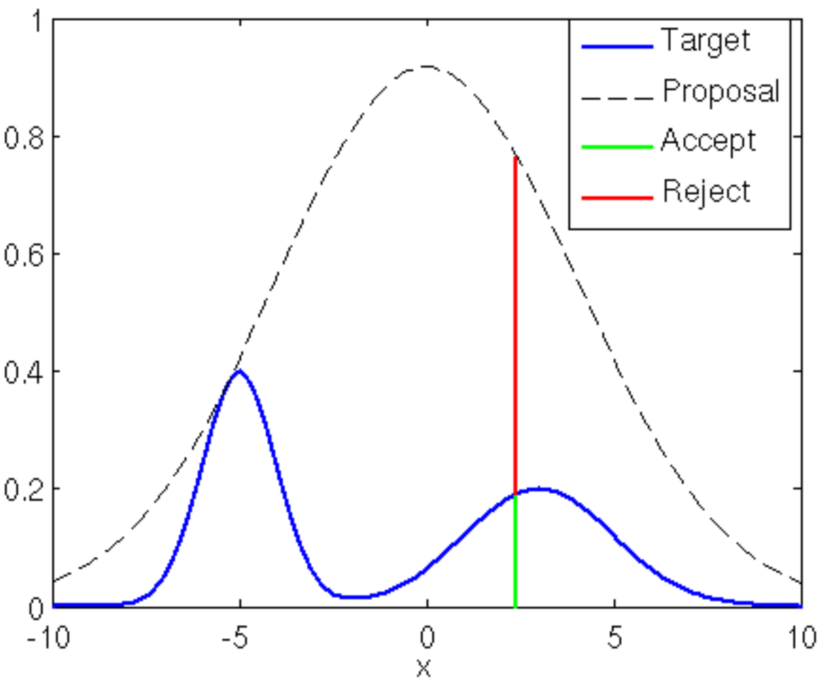
- Note: We are not sampling from p directly, only evaluating!



Sampling cont. *Rejection Sampling* Example

- Given:
  - We sample $y = -5$ and $u = 0.5$

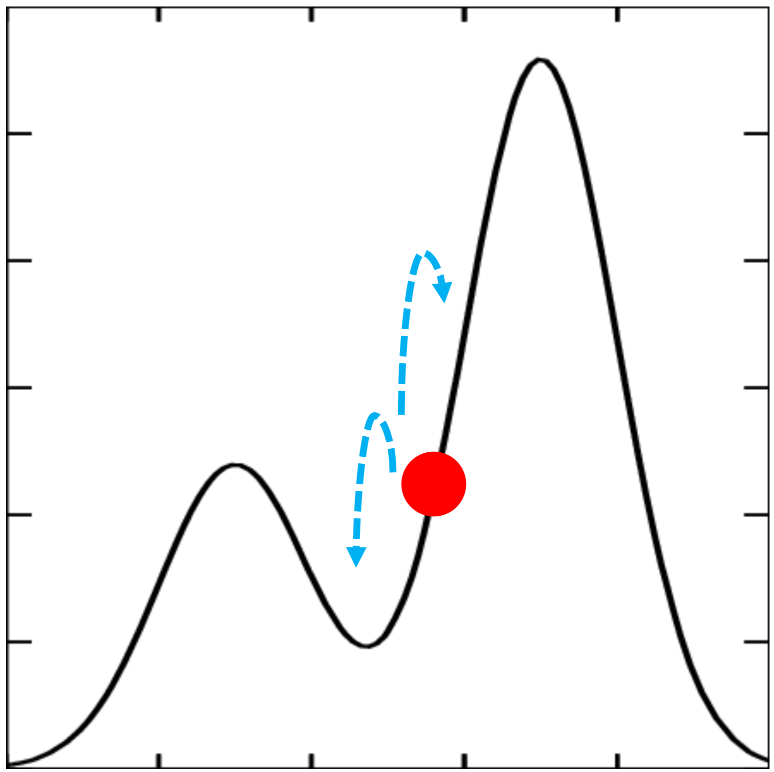
- Task:
  - Do we keep the sample?



Sampling cont. *MCMC*

- MCMC:
  1. Sample $y \sim \mathcal{N}(x, \sigma^2)$ and $u \sim unif(0,1)$
  2. If $u \cdot p(x) \leq p(y)$ set $x \leftarrow y$ else keep $x \leftarrow x$

- Note: If $p(y) \geq p(x)$ we always move to y



Sampling cont. *MCMC* Example

- Given:
  - Target distribution $p(x) = 2x, x \in [0, 1]$
  - $x_1 = 0.5$

- Task (independent):
  - If we sample $y = 0.6$ and $u = 0.001$ do we move? How does the buffer look like?
  - If we sample $y = 0.4$ and $u = 0.9$ do we move? How does the buffer look like?

Exercises

Exercise 1

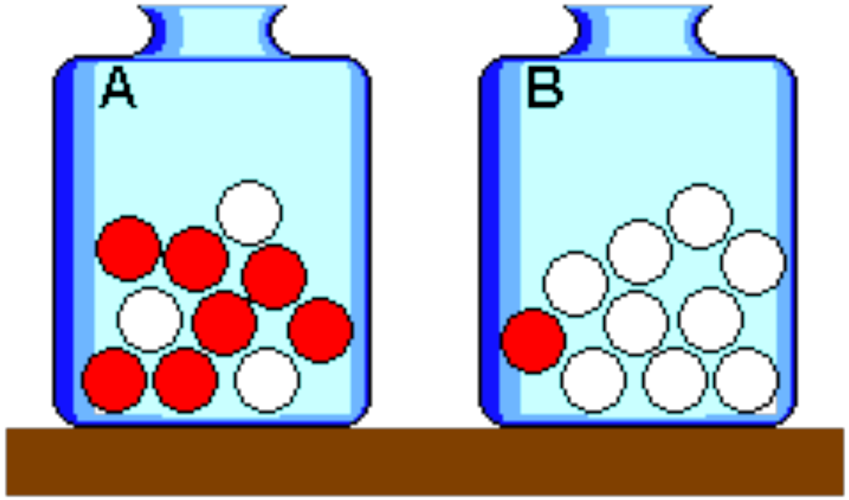
- Given:
  - PDF: $p(x, y) = c, x \in [-1, 1], y \in [-1, 1]$

- Task:
  - Find c such that it's a valid PDF
  - Compute the probability that x is larger than 0, $P(x > 0)$

- Tip:
  - $P(x > a) = \int_a^\infty p(x)dx$

Exercise 2

- Given:
  - We have two jars (A, B) with red (R) and white (W) marbles.
  - Drawing a marble from A or B is equally likely: $P(A) = P(B) = \frac{1}{2}$



- Task:
  - Write out Bayes Rule with the appropriate random variables for $P(\, A \mid R \,)$
  - Compute the normalization factor $P(R)$
  - Compute $P(\, A \mid R \,)$

Exercise 3

- Task:
  - What could the expectation value and the variance roughly correspond to in something that you have seen in mechanics?

Exercise 4

- Task:
  - Compute the expectation value of a dice
  - Compute the varianceof a dice

Exercise 5

- Task:
  - When and why do we use sampling?
  - Given and example of where you could use sampling?

Homework

HW 1

- Given:
  - $p(\text{„}gold\text{“} \mid spam) = 0.6$
  - $p(\text{„}gold\text{“} \mid no\ spam) = 0.01$
  - $p(spam) = 0.2$

- Task:
  - Find the probability that an email containing the word „gold“ is spam, ie. $p(spam \mid \text{“}gold\text{“})$.

HW 2

- Given:
  - We want to sample from a uniform distribution $p(x) = \frac{1}{2}, x \in [-1, 1]$
  - But can only generate samples from $q(x) = 1/\sqrt{\pi} \cdot \exp -x^2$ (normal dist)

- Task:
  - Determine an optimal L such that $p(x) \leq L \cdot q(x)$, use $L_{optimal} = \left\| \frac{p(x)}{q(x)} \right\|_{\infty}$
  - You draw $y = 1.1$ and $u = 0.9$ – is the sample accepted or rejected?

HW 3

- Given:
  - The target distribution is $p(x) = 2x, x \in [0, 1]$
  - You have access to the following values:
    - $x_1 = 0.5$
    - $s_1 = 0.1$ and $s_2 = -0.2$ both drawn from $s \sim \mathcal{N}(0, \sigma^2)$
    - $u_1 = 0.9$ and $u_2 = 0.3$ both drawn from $u \sim unif(0,1)$

- Task:
  - Perform MCMC for two steps

- Tip:
  - $y = \mu + s$ then $y \sim \mathcal{N}(\mu, \sigma^2)$ if $s \sim \mathcal{N}(0, \sigma^2)$

HW 4

This problem was first stated in 1777 by Comte de Buffon

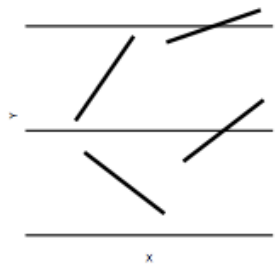


Figure 10: Buffon's needles

Let a needle of length 1 *be thrown at random onto a horizontal plane ruled with parallel straight lines spaced by a distance* 1 *from each other. What is the probability* p *that the needle will intersect one of these lines? (see Figure 10)* Write a pseudo-code which estimate the probability p.