Introduction to Computational Probability Modelling

Lecture 1

The Unfinished Game

- Game win 3 out of 5
- Alice now 2 wins
- Bob now 1 win
- In the following 2 games,
  - C(Alice)=C(Alice=2)+C(Alice=1)=$\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$
  - C(Bob)=C(Alice=0)=$\frac{1}{4}$

Categories of Probability

- **Frequentist** Probability
  - *Empirical or experimental* probability: on data
  - *Theoretical* probability
- **Conditional** Probability - *Monty Hall Problem*
  - 3 doors
  - Door 1 chosen
  - Door 3 revealed without award
  - P(Door 2)=$\frac{2}{3}$
  - P(Door 1)=$\frac{1}{3}$
- **Subjective** Probability - *Bayes' Guessing Game*
  - Update prior belief by relative position

Permutations and Combinations

Example: *World Series*

- Game Win 4 out of 7
- The following enumerations only list the results where A wins
- $P(4)=2\times\frac{1}{2^4}=\frac{1}{8}$
  - `AAAA`
- $P(5)=2\times\frac{1}{2^4}\times\frac{1}{2}\times C_4^1=\frac{1}{4}$
  - `BAAAA  ABAAA  AABAA  AAABA`
- $P(6)=2\times\frac{1}{2^4}\times\frac{1}{2^2}\times C_5^2=\frac{5}{16}$
  - `BBAAAA  BABAAA  BAABAA  BAAABA`
  - `ABBAAA  ABABAA  ABAABA`
  - `AABBAA  AABABA`
  - `AAABBA`
- $P(7)=2\times\frac{1}{2^4}\times\frac{1}{2^3}\times C_6^3=\frac{5}{16}$

- or, $P(7) = 1 - P(4) - P(5) - P(6) = \frac{5}{16}$
- or, $P(7) = P(\text{3-3 tie at 6 games}) = \frac{C_6^3}{2^6} = \frac{5}{16}$

Frequency Distributions

- Grouped Data: grouped into intervals
- Class Limits & Frequency

Lecture 2

Terminology

- **Sample Space**: the set of all possible outcomes of an experiment, denoted by **S**
- Any two outcomes in the sample space must be **mutually exclusive**
- Type of Sample Space:
    - Discrete sample space
        - Finite sample space
        - Countable infinite sample space (e.g. natural numbers)
    - Continuous sample space (e.g. points in a line)
- **Event**: a subset of sample space
    - Event can be empty
    - Event is a set
    - Union $A \cup B$, Intersection $AB = A \cap B$, Complement $\bar{A} = S - A$
    - DeMorgan's Law:
        - $\overline{A \cup B} = \bar{A} \cap \bar{B}$
        - $\overline{A \cap B} = \bar{A} \cup \bar{B}$

Axiom of Probability

- for each event A in S, $0 \le P(A) \le 1$
- $P(S) = 1$, $P(\emptyset) = 0$
- if $AB = \emptyset$, then $P(A \cup B) = P(A) + P(B)$

Deduction

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

$$\begin{aligned} P(A \cup B \cup C) &= P(A \cup (B \cup C)) \\ &= P(A) + P(B \cup C) - P(A \cap (B \cup C)) \\ &= P(A) + [P(B) + P(C) - P(BC)] - P((AB) \cup (AC)) \\ &= P(A) + P(B) + P(C) - P(BC) - [P(AB) + P(AC) - P((AB) \cap (AC))] \\ &= P(A) + P(B) + P(C) - P(BC) - P(AC) - P(AC) + P(ABC) \end{aligned}$$

Unfinished Game, Continued

Expected Value

- Denote $e(2,3)$ as the chance of the first player to win, with the first player $m=2$ more rounds to win and the second player $n=3$ more rounds
- $e(2,3)=\frac{1}{2}e(1,3)+\frac{1}{2}e(2,2)$
- $e(1,3)=\frac{1}{2}e(0,3)+\frac{1}{2}e(1,2)$
- $e(1,2)=\frac{1}{2}e(0,2)+\frac{1}{2}e(1,1)$
- $e(0,n)=1$
- $e(m,m)=\frac{1}{2}$
- $\Rightarrow e(2,3)=\frac{11}{16}$

Pascal's Triangle

The chance of the first player to win in the $m+n-1=4$ rounds is: $\frac{C_4^2+C_4^3+C_4^4}{2^4}=\frac{11}{16}$

Lecture 3

Terminology

- **Random experiments**: experiment with random outcome that can be **repeated** many times
- **Outcome**: possible result of an experiment, *unique, mutually exclusive*
- **Sample space**: the set of all possible outcomes
- **Event**: a set of outcomes, also a subset of sample space
- **Probability**: P(E) = (number of outcomes in E) / (total number of outcomes)
  - Function of E
  - Input domain: Sample space
  - Output domain: real number in $[0,1]$
- Math definition
  - $0 \leq P(A) \leq 1$
  - for sample space S, $P(S)=1$
  - for sequences of **disjoint** (mutually exclusive) events $P(\cup A_i)=\sum_i P(A_i)$
- Math properties
  - $P(\emptyset)=0$
  - $P(\bar{A})=1-P(A)$
  - $P(A\cup B)=P(A)+P(B)-P(AB)$

Sample Proof

$$\begin{aligned}
P(\emptyset\cup\emptyset) &= P(\emptyset)+P(\emptyset)=2P(\emptyset) \\
P(\emptyset\cup\emptyset) &= P(\emptyset) \\
P(\emptyset) &= 0
\end{aligned}$$

$$\begin{aligned}
P(\bar{A}\cup A) &= P(\bar{A})+P(A)=P(S)=1 \\
P(\bar{A}\cap A) &= P(\emptyset)=0 \\
P(\bar{A}) &= P(\bar{A}\cup A)-P(A)+P(\bar{A}\cap A) \\
&= 1-P(A)+0=1-P(A)
\end{aligned}$$

Computation of Probability

Rules of Probability

- **Addition: mutually exclusive** $P(A \cup B) = P(A) + P(B)$ if $AB = \emptyset$
- **Multiplication: independent** $P(AB) = P(A)P(B)$
- independent events: do not influence each other

Two Children Problem (Boy or Girl Paradox)

Lecture 4

Culminative Distribution Function

$P(X \leq x) = \sum_{X=x_0}^{x} P(X)$

False Positive

$TPR = \frac{TP}{TP+FN}, FNR = \frac{FN}{TP+FN}, TNR = \frac{TN}{TN+FP}, FPR = \frac{FP}{TN+FP}, IR = \frac{TP+FN}{TP+FN+TN+FP}$

Joint Distribution

$P(x, y) = P(X = x \wedge Y = y)$

- $P(x, y) > 0$
- $\sum_{x_i} \sum_{y_i} P(x_i, y_i) = 1$
- $P(x) = \sum_{y_i} P(x, y_i)$
- $P(y) = \sum_{x_i} P(x_i, y)$

Bernoulli Distribution

$P(X = 0) = 1 - p, P(X = 1) = p$

Binomial Distribution

$P(X = R \text{ out of } N) = C_N^R p^R (1-p)^{N-R}$

Lecture 5

Expectation of Random Variables

- If x, y are independent, $P(x = X, y = Y) = P(x = X) \cdot P(y = Y)$
- $E(nx) = nE(x)$
- Binomial Distribution: $E(X) = Np$
- $E(X) = \sum_i i \cdot P(X = i) = \sum_i P(X \geq i)$

> Example: Let X be the smaller number of two randomly rolled dice.
>
> $E(X) = P(X \geq 1) + P(X \geq 2) + \cdots + P(X \geq 6) = \frac{6}{6}^2 + \frac{5}{6}^2 + \cdots + \frac{1}{6}^2$

Markov's Inequality

If $\mathbf{X \geq 0}$ (**important**!), $P(X \geq a) \leq \frac{E(X)}{a}$ for $\forall a > 0$

Proof:

> $$\begin{aligned} \text{Let } Y &= \begin{cases} a, \ X \geq a \\ 0, \ 0 \leq X \leq a \end{cases} \\ X \geq Y &\Rightarrow E(X) \geq E(Y) \\ P(Y = a) &= P(X \geq a) \\ P(Y = 0) &= P(X < a) \\ \Rightarrow E(Y) &= a \, P(X \geq a) \\ \therefore E(X) &\geq a \, P(X \geq a) \\ \Leftrightarrow P(X \geq a) &\leq \frac{E(X)}{a} \end{aligned}$$

Bernoulli's Utility

- *Diminishing Utility*: change in utility decreases as wealth increases - **concave function**, where $\forall x_1, x_2 \in X, 0 \leq \alpha \leq 1, f(\alpha x_1 + (1-\alpha)x_2) \geq \alpha f(x_1) + (1-\alpha) f(x_2)$
- **Expected Utility Change**: $P(+)[f(x + \Delta x) - f(x)] + P(-)[f(x - \Delta x) - f(x)]$

Lecture 6

Multiplication Theorem of Independent Variables

If X, Y are independent variables, $E(XY) = E(X)E(Y)$

> Example: $E\left((X-Y)^2\right) = E(X^2) - 2E(X)E(Y) + E(Y^2)$

Variance and Standard Derivation

$Var(X) = E([X - \mu]^2)$, where $\mu = E(X)$

$SD(X) = \sqrt{Var(X)}$

Computational Formula for Variance

$Var(X) = E[X^2] - [E(X)]^2$

Proof:

> $$\begin{aligned} Var(X) &= E([X-\mu]^2) \\ &= E(X^2 - 2X\mu + \mu^2) \\ &= E[X^2] - 2E(X)E(\mu) + E(\mu^2) \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - [E(X)]^2 \end{aligned}$$

Properties of Variance

- $Var(X) = E[X - \mu]^2] \geq 0$
- $Var(X) = 0 \Leftrightarrow P(X = E(X)) = 1$

> Given: $P(X = 1) = p, P(X = 0) = 1 - p$

> $\therefore E(X)=p, E(X^2)=p, Var(X)=E(X^2)-[E(X)]^2=p-p^2$

> Given: $P(X=X_i, X_i \in \{1,2,3,4,5,6\})=\frac{1}{6}$
>
> $\therefore E(X)=\frac{7}{2}, E(X^2)=\frac{91}{6}, Var(X)=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{35}{12}$

Chebyshev's Inequality

$P(|X-E(X)|<k\cdot SD(X))\geq 1-\frac{1}{k}^2$, X not necessarily positive

at least $1-\frac{1}{k^2}$ of the data lies in $E(X)-k\cdot SD(X)\leq X\leq E(X)+k\cdot SD(X)$

Variance of Sum of Independent Variables

$Var(\sum_i X_i)=\sum_i Var(X_i)$

Proof:

> $$\begin{aligned} Var(X+Y) &= E[(X+Y)^2]-[E(X+Y)]^2 \\ &= [E(X^2)+2E(X)E(Y)+E(Y^2)]-[E(X)+E(Y)]^2 \\ &= \{E(X^2)-[E(X)]^2\}+\{E(Y^2)-[E(Y)]^2\} \\ &= Var(X)+Var(Y) \end{aligned}$$

Variance of Binomial Distribution

Let $X\sim B(n,p)$

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

Variance of Linearity

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

For random variable X, $E(X)=\mu$, $SD(X)=\sigma$

standardized random variable for X: $X^*=\frac{X-\mu}{\sigma}$

$E(X^*)=0, SD(X^*)=1$

Lecture 8

Sn and $\overline{X}_n$

Definition

$$S_n=\sum_{i=1}^{n} X_i$$
$$\overline{X}_n=\frac{S_n}{n}$$

- S_n and $\overline{X}_n$ **are random variables**

Properties

- $\mathrm{E}(S_n) = n\mathrm{E}(X)$
- $\mathrm{Var}(S_n) = n\mathrm{Var}(X)$
- $\mathrm{SD}(S_n) = \sqrt{n} \cdot \mathrm{SD}(X)$
- $\mathrm{E}(\overline{X}_n) = \mathrm{E}(X)$
- $\mathrm{Var}(\overline{X}_n) = \frac{\mathrm{Var(X)}}{n}$
- $\mathrm{SD}(\overline{X}_n) = \frac{\mathrm{SD(X)}}{\sqrt{n}}$

> $\mathrm{Var}(\overline{\mathrm{X}}_\mathrm{n}) = \mathrm{Var}\left(\frac{\mathrm{S_n}}{\mathrm{n}}\right) = \left(\frac{1}{n}\right)^2 \mathrm{Var}(S_n) = \frac{\mathrm{Var(X)}}{n}$

Standard Normal Distribution

Standard Normal Density Function (PDF)

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

Properties

- $\phi(x) = \phi(-x) > 0$
- $\lim_{|x|\to\infty} \phi(x) = 0$
- $\int_{-\infty}^{+\infty} \phi(x)\mathrm{d}x = 1$

Standard Normal Distribution Variable

$X \sim \mathcal{N}(0, 1)$

Properties

- $\mathrm{E}(X) = \int_{-\infty}^{+\infty} x \cdot \phi(x)\mathrm{d}x = 0$
- $\mathrm{Var}(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 \cdot \phi(x)\mathrm{d}x = \int_{-\infty}^{+\infty} x^2\phi(x)\mathrm{d}x = 1$
- $\mathrm{SD}(X) = 1$

Cumulative Distribution Function (CDF)

$F(z) = \int_{-\infty}^{z} \phi(x)\mathrm{d}x$

Properties

- $F(1) - F(-1) = 0.68$
- $F(2) - F(-2) = 0.95$
- $F(3) - F(-3) = 0.997$
- $P(a < x < b) = \int_a^b \phi(x)\mathrm{d}x = F(b) - F(a)$

Generalized Normal Distribution

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

$X \sim \mathcal{N}(\mu, \sigma^2)$

Properties

- $\mathrm{E(X)} = \mu$
- $\mathrm{Var(X)} = \sigma^2, \mathrm{SD}(X) = \sigma$
- $P(a < x < b) = \int_a^b f(x)\mathrm{d}x$

Central Limit Theorem

- Given $E(X) = \mu, SD(X) = \sigma$
- Let S_n^* be **standardized random variable** for S_n:
  $S_n^* = \frac{S_n - \mathrm{E}(S_n)}{\mathrm{SD}(S_n)} = \frac{S_n - n\mu}{\sqrt{n}\sigma}$
- then $\lim_{n\to\infty} P(a < S_n^* < b) = \int_a^b \phi(x)\mathrm{d}x = F(b) - F(a)$

Lecture 10

Terminology

- **Hypothesis** (H): Assumption about event
- **Initial (Prior) Odds** ($P(H) : P(\neg H)$)
- **Evidence** (E): Observation of an event outcome
- **Likelihood Ratio** ($P(E|H) : P(E|\neg H)$)

Identities

- $P(EH) = P(H)P(E|H) = P(E)P(H|E)$
- $P(E) = P(E|H)P(H) + P(E|\neg H)P(\neg H)$

Bayes Method

- **Update (Posterior) Odds** = **likelihood ratio** $\times$ **initial odds**
- **Posterior Probability** $P(H|E) = P(H) \times P(E|H)/P(E) \quad \cdots (1)$
- $P(\neg H|E) = P(\neg H) \times P(E|\neg H)/P(E) \quad \cdots (2)$
- (1) and (2) $\rightarrow$ **Posterior Odds**
  $P(H|E) : P(\neg H|E) = P(H) \times P(E|H) : P(\neg H) \times P(E|\neg H)$

Bayes Inference

- The exact value x is unknown
- Observation value y is corrupted by additive Gaussian noise $y = x + n$
- Find the most possible x given y ($\hat{x} = \arg\max P(x|y)$, i.e., the x that maximize $P(x, y)$)
- Given: Bayes' rule $P(x|y) = \frac{P(y|x)P(x)}{P(y)}$
- Given: Gaussian noise $P(y|x) = P(y - x = n|x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{y-x^2}{2\sigma^2}\right)$

Lecture 11

Terminology

- **Regression**: best-fit mathematical equation, used to predict output variable as a function of input variable
- **Linear Regression**: $y = \alpha + \beta x + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \delta)$ is a small random variable
- **Residual**: **vertical distance** of the point from the line $= y - \hat{y}$ (!= projective distance)

Principle of Least Squares

- $\hat{y} = a + \beta x, e_i = y_i - \hat{y}_i$, optimal regression minimizes $\sum e_i^2$
- $\sum e_i^2 = \sum [y_i - (\alpha + \beta x_i)]^2$

> **take derivative** of $\sum e_i^2$ with respect to α and β to set them to zero
>
> Define: $\bar{x} = \frac{\sum x_i}{n}, y = \frac{\sum y_i}{n}$
>
> Define: $S_{xx} = \sum (x_i - \bar{x})^2, S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$
>
> Equivalently: $S_{xx} = (\sum x_i^2) - n\bar{x}^2, S_{xy} = (\sum x_i y_i) - n\bar{x}\bar{y}$

Least Square Estimators:

$$\beta = \frac{S_{xy}}{S_{xx}}, \alpha = \bar{y} - \beta \bar{x}$$

Correlation Coefficient

- test how strong the linear relationship between two variables
- $r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{S_x} \right) \left(\frac{y_i - \bar{y}}{S_y} \right)$
- where $S_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}, S_y = \sqrt{\frac{1}{n-1} \sum (y_i - \bar{y})^2}$ (**Sample deviation**)
- $-1 \leq r \leq 1$; the closer $|r|$ to 1 is, the stronger the correlation is
- **Coefficient of Determination** $0 \leq r^2 \leq 1$

Maximum Likelihood Estimation

Method: find **parameter values that maximize probability**

$L_{\text{Data}}(p) = \Pr(\text{Data}; p) = p^T (1-p)^F$

log-likelihood: $l_{\text{Data}}(p) = \log[L_{\text{Data}}(p)] = T \log(p) + F \log(1 - P)$

taking derivative and make derivative zero: $\frac{T}{P} - \frac{F}{1-P} = 0$

Conclusion:

$$\hat{p} = \frac{T}{T+F}$$

> Example:
>
> 10 tests, 6 positives, 4 negatives
>
> Probability of all the data $\Pr(\text{Data}; p) = p^6 (1-p)^4$ (combined probability)
>
> treat this as a function of $p (0 \leq p \leq 1)$: $L_{\text{Data}}(p) = \Pr(\text{Data}; p) = p^6 (1-p)^4$, which maximizes at $p = \frac{6}{6+4} = 0.6$

MLE for Gaussian Distribution

- $X \sim \mathcal{N}(\mu, \sigma^2)$
- **assume observed data is random samples of $\mathcal{N}(\mu, \sigma^2)$**
- Solution: $\hat{\mu} = \frac{1}{n}\sum x_k = \bar{x}, \sigma^2 = \frac{1}{n}\sum(x_k - \hat{\mu})^2 = \mathrm{Var}(x)$

Lecture 12 and 13

Poisson Distribution

Presets

- X = the number of outcomes per time interval (thus only integers)
- $E(X) = \lambda$
- all the occurrences are independent and only one outcome at the same time

Expression

$X \sim \pi(\lambda) : P(X = k) = \frac{\lambda^k \exp(-\lambda)}{k!}$, where $k = 0, 1, 2, \cdots$

Identities

- $E(X) = \lambda$

> $$\begin{aligned} \exp(\lambda) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \\ E(X) &= \sum_{k=0}^{\infty} k \cdot P(X = k) \\ &= \sum_{k=0}^{\infty} \frac{k \cdot \lambda^k \exp(-\lambda)}{k!} \\ &= k \cdot \exp(-\lambda) \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= k \cdot \exp(-\lambda) \cdot \exp(\lambda) = k. \end{aligned}$$

- $\mathrm{Var}(X) = \lambda, \mathrm{SD}(X) = \sqrt{\lambda}$

> $$\begin{aligned} \text{Let } Y &= X(X-1) \\ E(Y) &= \sum_{k=0}^{\infty} k(k-1) P(X = k) \\ &= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k \exp(-\lambda)}{k!} \\ &= \lambda^2 \exp(-\lambda) \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} \\ &= \lambda^2 \exp(-\lambda) \exp(\lambda) = \lambda^2 \\ E(Y) &= E(X^2 - X) \\ \to E(X^2) &= E(X + Y) \\ \mathrm{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= (\lambda^2 + \lambda) - (\lambda)^2 = \lambda \end{aligned}$$

Approximation of Binomial by Poisson

When n is very large and p is very small, let $np = \lambda$

$P(X = k) \approx \pi(np) = \frac{(np)^k \exp(-np)}{k!}$

Discrete Distribution

- Bernoulli Distribution
  - $P(X = 1) = p, P(X = 0) = 1 - p$
  - $E(X) = p, \mathrm{Var}(X) = p(1 - p)$
- Binomial Distribution
  - $X \sim B(n, p) : P(X = k) = C_n^k \cdot p^k (1 - p)^{n-k}$
  - $E(X) = np, \mathrm{Var}(X) = np(1 - p)$
- Poisson Distribution
  - $X \sim \pi(\lambda) : P(X = k) = \frac{\lambda^k \exp(-\lambda)}{k!}$
  - $E(X) = \lambda, \mathrm{Var}(X) = \lambda$
- **Probability Mass Function** (PMF)

Continuous Distribution

- Normal Distribution
  - $X \sim \mathcal{N}(\mu, \sigma^2) : f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
  - $E(X) = \mu, \mathrm{Var}(X) = \sigma^2$
- Uniform Distribution
  - $X \sim U(x_{\min}, x_{\max}) : f(x) = \frac{1}{x_{\max} - x_{\min}} (x_{\min} \le x \le x_{\max})$
  - $E(X) = \frac{1}{2}(x_{\min} + x_{\max}), \mathrm{Var}(X) = \frac{1}{12}(x_{\max} - x_{\min})^2$

> $$\begin{aligned} E(X) &= \frac{1}{2}(x_{\min} + x_{\max}) \\ E(X^2) &= \frac{\int_{x_{\min}}^{x_{\max}} x^2 \mathrm{d}x}{x_{\max} - x_{\min}} \\ &= \frac{\frac{1}{3}(x_{\max}^3 - x_{\min}^3)}{x_{\max} - x_{\min}} \\ &= \frac{1}{3}(x_{\max}^2 + x_{\max}x_{\min} + x_{\min}^2) \\ \mathrm{Var}(X) &= E(X^2) - [E(X)^2] \\ &= \frac{1}{3}(x_{\max}^2 + x_{\max}x_{\min} + x_{\min}^2) - \frac{1}{4}(x_{\max} + x_{\min})^2 \\ &= \frac{1}{12}x_{\max}^2 - \frac{1}{6}x_{\max}x_{\min} + \frac{1}{12}x_{\min}^2 \\ &= \frac{1}{12}(x_{\max} - x_{\min})^2 \end{aligned}$$

- **Cumulative Distribution Function** (CDF)
  - $F(x) = P(X \le x)$
  - $P(a \le X \le b) = F(b) - F(a) \ge 0$
- **Probability Density Function** (PDF)
  - $F(x_0) = \int_{-\infty}^{x_0} f(x)\mathrm{d}x$
  - $f(x) \ge 0\,(f(x) > 1$ **possible in PDF** but not in PMF)

- $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$
- $P(a \le X \le b) = \int_a^b f(x)\mathrm{d}x$

Monte Carlo Simulation

Q2

$$
\begin{aligned}
P\left(|X-Y| \ge \frac{1}{2}\right) &= 2P\left(X-Y \ge \frac{1}{2}\right) = 2P\left(0 < Y < X+\frac{1}{2}\right) = 2P\left(\frac{1}{2} < Y+\frac{1}{2} < X < 2\right) \\
&= 2\int_{\frac{1}{2}}^{2}\int_{0}^{x-\frac{1}{2}} f(x,y)\mathrm{d}y\mathrm{d}x = 2\int_{\frac{1}{2}}^{2}\int_{0}^{x-\frac{1}{2}} \frac{1}{4}\mathrm{d}y\mathrm{d}x \\
&= \frac{1}{2}\int_{\frac{1}{2}}^{2}\left(x-\frac{1}{2}\right)\mathrm{d}x \\
&= \frac{1}{4}\left(x-\frac{1}{2}\right)^2\Bigg|_{\frac{1}{2}}^{2} = \frac{9}{16}
\end{aligned}
$$

Q3

$$
\begin{aligned}
X &\sim \mathcal{N}(0,1) \\
\rightarrow f(x) &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \\
E[e^x] &= \int_{-\infty}^{\infty} e^x f(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(x-\frac{x^2}{2}\right)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(\frac{1}{2}-\frac{(x-1)^2}{2}\right)\mathrm{d}x \\
&= e^{\frac{1}{2}}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-1)^2}{2}\right)\mathrm{d}x \\
&\overset{t=x-1}{=\!=\!=} e^{\frac{1}{2}}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)\mathrm{d}t \\
&\overset{t\sim\mathcal{N}(0,1)}{=\!=\!=} e^{\frac{1}{2}}E(t) = e^{\frac{1}{2}}
\end{aligned}
$$

Q4

$$
\begin{aligned}
\int_0^2 \frac{x}{1+x^2}\mathrm{d}x &\,\underset{\mathrm{d}t=2x\mathrm{d}x}{\overset{t=1+x^2}{=\!=\!=}} \int_1^5 \frac{1}{2t}\mathrm{d}t \\
&= \frac{1}{2}\ln|t|\Bigg|_1^5 \\
&= \frac{\ln 5}{2}
\end{aligned}
$$