EECS 126: Probability & Random Processes
Fall 2021

Elementary Probability

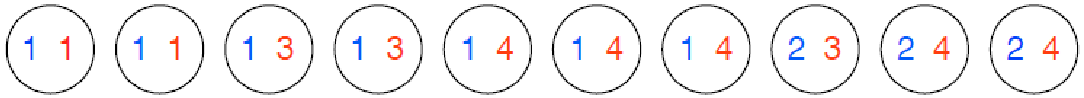
Shyam Parekh

Course Organization

- https://inst.eecs.berkeley.edu/~ee126/fa21/

Definition

Fig. A.1

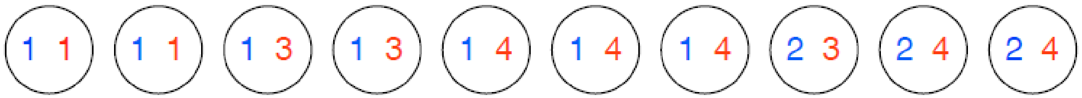


Equally likely marbles marked with a number in blue (B) and a number in red (R)

- Probability = # of favorable outcomes/# of possible outcomes
- Experiment outcome and random variables
- Probability is additive: $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

Conditional Probability

Fig. A.1



Equally likely marbles marked with a number in blue (B) and a number in red (R)

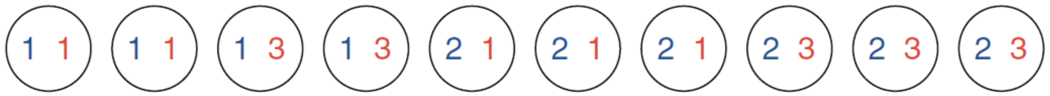
- $P[C|D] = P(C \cap D)/P(D)$
- Maximum A Posteriori (MAP) estimate of B given $\{R = r\} = \underset{b}{argmax}\, P[B = b|R = r]$
  - Which value of B most likely given R = r?
- Maximum Likelihood Estimate (MLE) of B given $\{R = r\} = \underset{b}{argmax}\, P[R = r|B = b]$
  - Which value B makes R = r most likely?
- MAP vs MLE for Ebola vs Flu

Examples of Common Confusion

- Definition of the experiment and its outcomes
  - Bill has two children and one of them is named Isabelle. What’s the probability Bill has two daughters?
  - Bill has two children and the younger one is named Isabelle. What’s probability Bill has two daughters?
- Regression to the mean

Independence

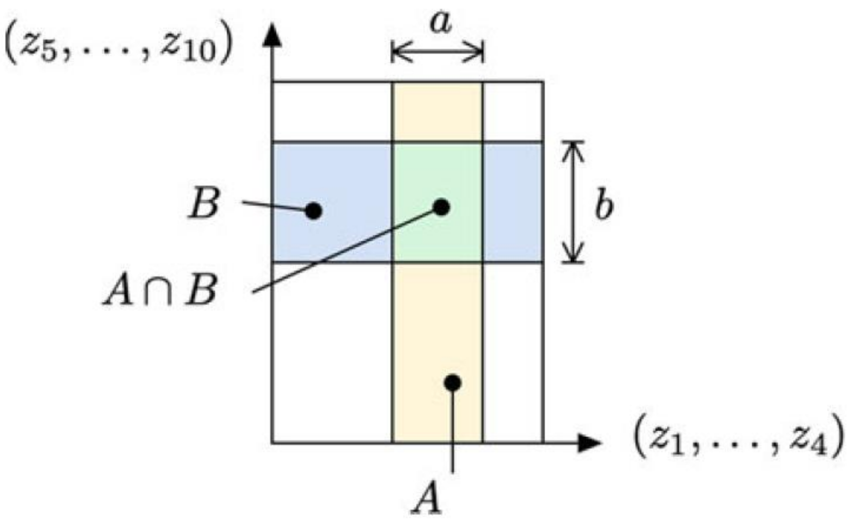
Fig. A.2



Equally likely marbles marked with a number in blue (B) and a number in red (R)

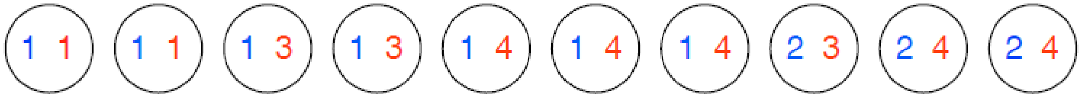
- B and R are independent: For $\forall r, b,\ P[B=b|R=r] = P(B=b)$, or $P[R=r|B=b] = P(R=r)$, or $P(B=b, R=r) = P(B=b)P(R=r)$
- 10 flips of a fair coin: X = # of heads in first 4 flips, Y = # of heads in the last 6 flips. Show X and Y are independent.

Fig. A.3



Expectation

Fig. A.1



Equally likely marbles marked with a number in blue (B) and a number in red (R)

- $E(B), E(R)$
- $E(X) = \sum_{m=1}^{M} x_m P(X = x_m) = \sum_{m=1}^{M} x_m p_m$
- $E(a) = a$
- $E(X) = \sum_{n=1}^{N} X(n) \frac{1}{N} = \sum_{m=1}^{M} (p_m N) x_m \frac{1}{N}$
- Expectation is linear:
  - $E(X+Y) = E(X) + E(Y), E(5 + 3X^2 + 5Y^3) = 5 + 3E(X^2) + 5E(Y^3)$
- Expectation is monotone: $X \leq Y \Rightarrow E(X) \leq E(Y)$.
- X and Y are independent $\Rightarrow \mathrm{E(XY) = E(X)E(Y)}$.
  - $E(XY) = \sum_i \sum_j x_i y_j P(X = x_i, Y = y_j) = (\sum_i x_i P(X = x_i))(\sum_j y_j P(Y = y_j))$
  - In the example, $E(BR) \neq E(B)E(R) \rightarrow$ B and R are not independent.
  - $E(XY) = E(X)E(Y) \nRightarrow$ X and Y are independent.

Variance

- $var(X) = E((X - E(X))^2)$
- $var(X) \geq 0$
- $var(X) = E(X^2) - [E(X)]^2$
- For a constant a, $var(aX) = a^2 var(X)$
- X and Y are independent $\Rightarrow var(X + Y) = var(X) + var(Y)$
- Standard Deviation = Square Root of Variance
  - $\sqrt{var(X)} = \sigma_X$

Inequalities

- Markov's Inequality: RV $X \geq 0$. Then, $P(X \geq a) \leq \frac{E(X)}{a}$, $\mathrm{a} > 0$.
- Chebyshev's Inequality: $P(|X - E(X)| \geq \epsilon) \leq \frac{var(X)}{\epsilon^2}$, $\epsilon > 0$.

Weak Law of Large Numbers

- Assume $X_1, X_2, X_3, \ldots$ are independent RVs with the same expected value μ and the same variance σ^2, and define $Y = (X_1 + X_2 + \ldots + X_n)/n$.
  - $P(|Y - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$

Covariance and Linear Regression

- $cov(X,Y) = E(XY) - E(X)E(Y) = E\big((X - E(X))(Y - E(Y))\big)$
- X and Y are independent $\Rightarrow cov(X,Y) = 0$.
  - X and Y are uncorrelated RVs if $cov(X,Y) = 0$.
  - X and Y are uncorrelated RVs $\nRightarrow$ X and Y are independent RVs.
- Observe X and want to estimate Y by $\hat{Y}$.
  - Restrict $\hat{Y}$ to linear functions of X such that $E((Y - \hat{Y})^2)$ is minimized.
  - Linear Least Squares Estimate (LLSE): $\hat{Y} = E(Y) + \frac{cov(X,Y)}{var(X)}(X - E(X))$
  - Linear Regression of Y against X: Same as the LLSE assuming the observed (X, Y) pairs are from equally likely samples

Need for a Sophisticated Framework

- Extend to RVs that can assume infinite or even uncountably many possible values.
- Classical Probability Theory due to Kolmogorov.
  - Relies on Axiom of Choice.
  - $A_1 \subset A_2 \subset A_3 \ldots, A = \bigcup_n A_n \Rightarrow P(A) = \lim_n P(A_n)$.
  - SLLN