Basics on Probability

Jingrui He

09/11/2007

Coin Flips

- You flip a coin
  - Head with probability 0.5

- You flip 100 coins
  - How many heads would you expect

Coin Flips cont.

- You flip a coin
  - Head with probability p
  - Binary random variable
  - Bernoulli trial with success probability p
- You flip k coins
  - How many heads would you expect
  - Number of heads X: discrete random variable
  - Binomial distribution with parameters k and p

Discrete Random Variables

- Random variables (RVs) which may take on only a **countable** number of **distinct** values
  - E.g. the total number of heads X you get if you flip 100 coins
- X is a RV with arity *k* if it can take on exactly one value out of $\{x_1, \ldots, x_k\}$
  - E.g. the possible values that X can take on are 0, 1, 2,…, 100

Probability of Discrete RV

- Probability mass function (pmf): $P(X = x_i)$
- Easy facts about pmf
  - $\sum_i P(X = x_i) = 1$
  - $P(X = x_i \cap X = x_j) = 0$ if $i \neq j$
  - $P(X = x_i \cup X = x_j) = P(X = x_i) + P(X = x_j)$ if $i \neq j$
  - $P(X = x_1 \cup X = x_2 \cup \ldots \cup X = x_k) = 1$

Common Distributions

- Uniform $X \sim U[1,\ldots,N]$
  - X takes values 1, 2, …, N
  - $P(X = i) = 1/N$
  - E.g. picking balls of different colors from a box
- Binomial $X \sim Bin(n, p)$
  - X takes values 0, 1, …, n
  - $P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$
  - E.g. coin flips

Coin Flips of Two Persons

- Your friend and you both flip coins
  - Head with probability 0.5
  - You flip 50 times; your friend flip 100 times
  - How many heads will both of you get

Joint Distribution

- Given two discrete RVs X and Y, their **joint distribution** is the distribution of X and Y together
  - E.g. P(You get 21 heads AND you friend get 70 heads)
- $\sum_x \sum_y \mathrm{P}\left(\mathrm{X}=x \cap \mathrm{Y}=y\right)=1$
  - E.g.

$$\sum_{i=0}^{50}\sum_{j=0}^{100}\mathrm{P}\left(\text{You get } i \text{ heads AND your friend get } j \text{ heads}\right)=1$$

Conditional Probability

- $\mathrm{P}(\mathrm{X}=x|\mathrm{Y}=y)$ is the probability of $\mathrm{X}=x$, given the occurrence of $\mathrm{Y}=y$
  - E.g. you get 0 heads, given that your friend gets 61 heads
- $\mathrm{P}(\mathrm{X}=x|\mathrm{Y}=y)=\dfrac{\mathrm{P}(\mathrm{X}=x\cap\mathrm{Y}=y)}{\mathrm{P}(\mathrm{Y}=y)}$

Law of Total Probability

- Given two discrete RVs X and Y, which take values in $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$, We have

$$
\begin{aligned}
\mathrm{P}\left(\mathrm{X} = x_i\right) &= \sum_j \mathrm{P}\left(\mathrm{X} = x_i \cap \mathrm{Y} = y_j\right) \\
&= \sum_j \mathrm{P}\left(\mathrm{X} = x_i \middle| \mathrm{Y} = y_j\right) \mathrm{P}\left(\mathrm{Y} = y_j\right)
\end{aligned}
$$

Marginalization

Marginal Probability → $P(X = x_i)$

Joint Probability → $P(X = x_i \cap Y = y_j)$

$$P\left(X = x_i\right) = \sum_j P\left(X = x_i \cap Y = y_j\right)$$

$$= \sum_j P\left(X = x_i \middle| Y = y_j\right) P\left(Y = y_j\right)$$

Conditional Probability → $P(X = x_i | Y = y_j)$

Marginal Probability → $P(Y = y_j)$

Bayes Rule

- X and Y are discrete RVs…

$$P\left(X = x \middle| Y = y\right) = \frac{P\left(X = x \cap Y = y\right)}{P\left(Y = y\right)}$$

$$\Downarrow$$

$$P\left(X = x_i \middle| Y = y_j\right) = \frac{P\left(Y = y_j \middle| X = x_i\right) P\left(X = x_i\right)}{\sum_k P\left(Y = y_j \middle| X = x_k\right) P\left(X = x_k\right)}$$

Independent RVs

- Intuition: X and Y are independent means that $X = x$ **neither** makes it **more or less** probable that $Y = y$
- Definition: X and Y are independent iff $P(X = x \cap Y = y) = P(X = x)P(Y = y)$

More on Independence

- $$P(X = x \cap Y = y) = P(X = x)P(Y = y)$$

$$P(X = x | Y = y) = P(X = x) \qquad P(Y = y | X = x) = P(Y = y)$$

- E.g. no matter how many heads you get, your friend will not be affected, and vice versa

Conditionally Independent RVs

- Intuition: X and Y are conditionally independent given Z means that once Z is **known**, the value of X does not add any **additional** information about Y
- Definition: X and Y are conditionally independent given Z iff

$$P\left(X = x \cap Y = y \middle| Z = z\right) = P\left(X = x \middle| Z = z\right) P\left(Y = y \middle| Z = z\right)$$

More on Conditional Independence

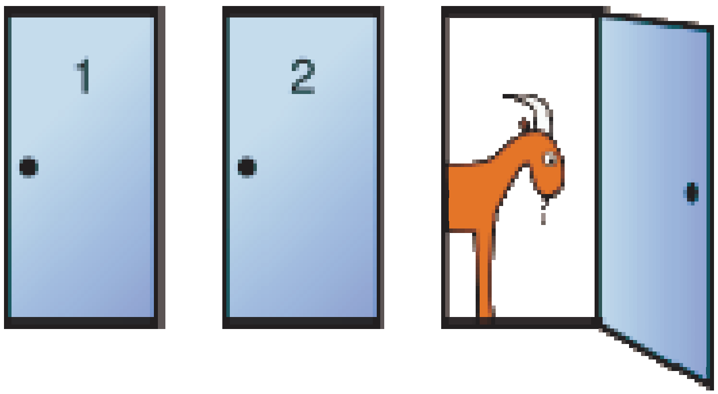
$$P\left(X = x \cap Y = y \middle| Z = z\right) = P\left(X = x \middle| Z = z\right) P\left(Y = y \middle| Z = z\right)$$

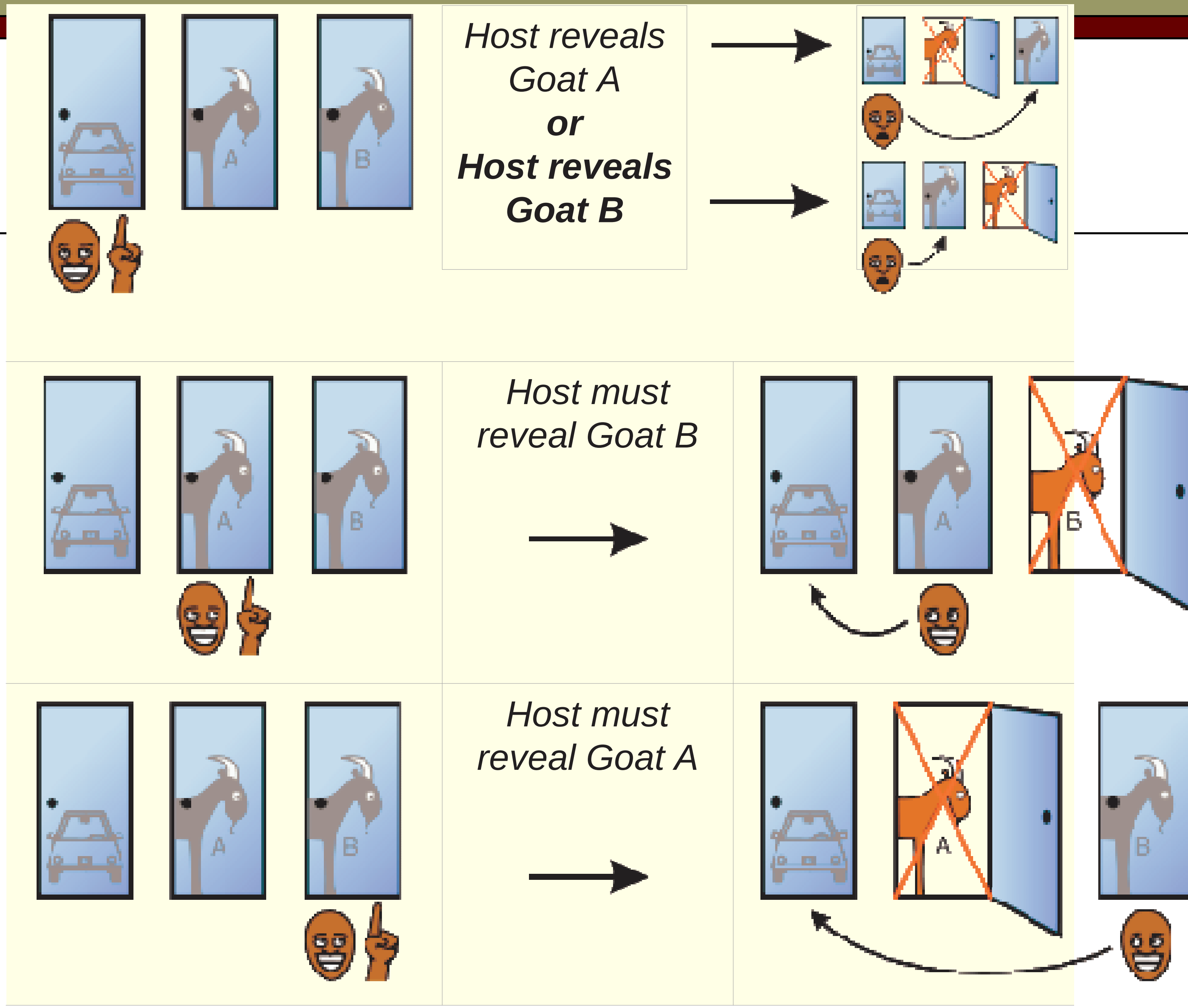
$$P\left(X = x \middle| Y = y, Z = z\right) = P\left(X = x \middle| Z = z\right)$$

$$P\left(Y = y \middle| X = x, Z = z\right) = P\left(Y = y \middle| Z = z\right)$$

Monty Hall Problem

- You're given the choice of three doors: Behind one door is a car; behind the others, goats.
- You pick a door, say No. 1
- The host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.
- Do you want to pick door No. 2 instead?





Host reveals Goat A
or
Host reveals Goat B
A
B
Host must reveal Goat B
Host must reveal Goat A

Monty Hall Problem: Bayes Rule

- C_i : the car is behind door i, i = 1, 2, 3
- $P(C_i) = 1/3$
- H_{ij} : the host opens door j after you pick door i
- $P(H_{ij} | C_k) = \begin{cases} 0 & i = j \\ 0 & j = k \\ 1/2 & i = k \\ 1 & i \neq k, j \neq k \end{cases}$

Monty Hall Problem: Bayes Rule cont.

- WLOG, i=1, j=3

- $$P\left(C_1 \middle| H_{13}\right) = \frac{P\left(H_{13} \middle| C_1\right) P\left(C_1\right)}{P\left(H_{13}\right)}$$

- $$P\left(H_{13} \middle| C_1\right) P\left(C_1\right) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

Monty Hall Problem: Bayes Rule cont.

- $P(H_{13}) = P(H_{13}, C_1) + P(H_{13}, C_2) + P(H_{13}, C_3)$

$$= P(H_{13} | C_1) P(C_1) + P(H_{13} | C_2) P(C_2)$$

$$= \frac{1}{6} + 1 \cdot \frac{1}{3}$$

$$= \frac{1}{2}$$

- $P(C_1 | H_{13}) = \frac{1/6}{1/2} = \frac{1}{3}$

Monty Hall Problem: Bayes Rule cont.

- $$P(C_1 | H_{13}) = \frac{1/6}{1/2} = \frac{1}{3}$$
- $$P(C_2 | H_{13}) = 1 - \frac{1}{3} = \frac{2}{3} > P(C_1 | H_{13})$$
- ***You should switch!***

Continuous Random Variables

- What if X is continuous?
- Probability density function (pdf) instead of probability mass function (pmf)
- A pdf is any function $f(x)$ that describes the probability density in terms of the input variable x.

PDF

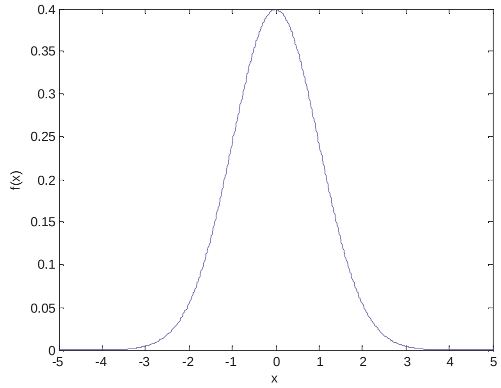
- Properties of pdf
  - $f(x) \geq 0, \forall x$
  - $\int_{-\infty}^{+\infty} f(x) = 1$
  - $f(x) \leq 1$???
- Actual probability can be obtained by taking the integral of pdf
  - E.g. the probability of X being between 0 and 1 is $\mathrm{P}(0 \leq X \leq 1) = \int_0^1 f(x)dx$

Cumulative Distribution Function

- $F_X(v) = P(X \le v)$
- Discrete RVs
  - $F_X(v) = \sum_{v_i} P(X = v_i)$
- Continuous RVs
  - $F_X(v) = \int_{-\infty}^{v} f(x)\,dx$
  - $\frac{d}{dx} F_X(x) = f(x)$

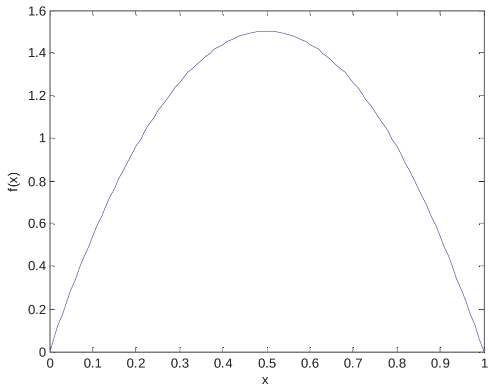
Common Distributions

- Normal $X \sim N\left(\mu, \sigma^2\right)$
  - $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}$
  - E.g. the height of the entire population



Common Distributions cont.

- Beta $X \sim Beta(\alpha, \beta)$
  - $f(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in [0,1]$
  - $\alpha = \beta = 1$: uniform distribution between 0 and 1
  - E.g. the conjugate prior for the parameter *p* in Binomial distribution



Joint Distribution

- Given two continuous RVs X and Y, the **joint pdf** can be written as $f_{\mathrm{X,Y}}(x, y)$
- $\int_x \int_y f_{\mathrm{X,Y}}(x, y)dxdy = 1$

Multivariate Normal

- Generalization to higher dimensions of the one-dimensional normal

Covariance Matrix

- $$f_{\vec{X}}(x_1, \ldots, x_d) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \cdot \exp\left\{ -\frac{1}{2} (\vec{x} - \mu)^T \Sigma^{-1} (\vec{x} - \mu) \right\}$$

Mean

Moments

- Mean (Expectation): $\mu = E(X)$
  - Discrete RVs: $E(X) = \sum_{v_i} v_i P(X = v_i)$
  - Continuous RVs: $E(X) = \int_{-\infty}^{+\infty} xf(x)\,dx$
- Variance: $V(X) = E(X - \mu)^2$
  - Discrete RVs: $V(X) = \sum_{v_i} (v_i - \mu)^2 P(X = v_i)$
  - Continuous RVs: $V(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)\,dx$

Properties of Moments

- Mean
  - $E(X+Y)=E(X)+E(Y)$
  - $E(aX)=aE(X)$
  - If X and Y are independent, $E(XY)=E(X)\cdot E(Y)$
- Variance
  - $V(aX+b)=a^2V(X)$
  - If X and Y are independent, $V(X+Y)=V(X)+V(Y)$

Moments of Common Distributions

- Uniform $X \sim U[1,\ldots,N]$
  - Mean $(1+N)/2$; variance $(N^2-1)/12$
- Binomial $X \sim Bin(n,p)$
  - Mean np; variance np^2
- Normal $X \sim N(\mu,\sigma^2)$
  - Mean μ; variance σ^2
- Beta $X \sim Beta(\alpha,\beta)$
  - Mean $\alpha/(\alpha+\beta)$; variance $\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Probability of Events

- X denotes an event that could possibly happen
  - E.g. X="you will fail in this course"
- P(X) denotes the **likelihood** that X happens, or X=true
  - What's the probability that you will fail in this course?
- Ω denotes the entire event set
  - $\Omega = \{X, \overline{X}\}$

The Axioms of Probabilities

- $0 <= P(X) <= 1$
- $P(\Omega) = 1$
- $P(X_1 \cup X_2 \cup \ldots) = \sum_i P(X_i)$, where X_i are disjoint events
- Useful rules
  - $P(X_1 \cup X_2) = P(X_1) + P(X_2) - P(X_1 \cap X_2)$
  - $P(\overline{X}) = 1 - P(X)$

Interpreting the Axioms