Recap

- (Ch 4) Random variables
  - Weak law of large numbers
  - Simulation examples
  - Continuous random variables

Today

- (Ch 4) Continuous random variables
- (Ch 5) Useful probability distributions

Probability density function (pdf)

- For a continuous random variable X, the probability that $X = x$ is essentially zero for all (or most) x, so we can't define $P(X = x)$

- Instead, we define the **probability density function (pdf)** over an infinitesimally small interval dx

$$p(x)dx = P(X \in [x, x + dx])$$

- For $a < b$

$$\int_a^b p(x)\, dx = P(X \in [a, b])$$

Properties of the probability density function

- $p(x)$ is a bit like a discrete random variable's probability distribution
  - $p(x) \geq 0$ for all x
  - The probability of X taking some value is 1

$$\int_{-\infty}^{\infty} p(x)\, dx = 1$$

- $p(x)$ is **not** like a discrete random variable's probability distribution
  - $p(x)$ is not the probability that $X = x$
  - $p(x)$ can exceed 1

Probability density function: height example

- Suppose we heard that Napoleon was 62.5 inches tall, rounded up to the nearest half inch. What is the pdf of his height H?

- Assume that H is equally likely to be any value in $(62, 62.5]$ inches

$$p(h) = \begin{cases} \overset{2}{\cancel{c}} & \text{if } \; h \in (62, 62.5] \\ 0 & \text{if } \; h \notin (62, 62.5] \end{cases}$$ where c is a constant

- Then

$$1 = \int_{-\infty}^{\infty} p(h)\, dh = \int_{62}^{62.5} c\, dh = \frac{c}{2} \Longrightarrow c = 2$$

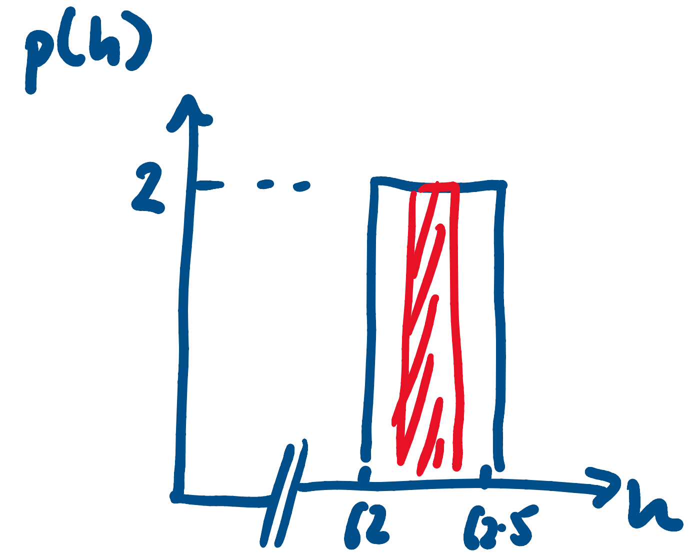
Probability density function: height example

- What is the probability that $62.1 \leq H \leq 62.2$?

$$P(62.1 \leq H \leq 62.2) = \int_{62.1}^{62.2} p(h)\,dh = \int_{62.1}^{62.2} 2\,dh = 2h\Big|_{62.1}^{62.2} = \frac{1}{5}$$

- What is the probability that $H = 62.1$?

$$P(H = 62.1) = \int_{62.1}^{62.1} 2\,dh = 0$$



Expected value

$$E[x] = \sum_{x} x\,p(x)$$

- Expected value of a continuous random variable X

$$E[X] = \int_{-\infty}^{\infty} xp(x)\,dx$$

- Expected value of function of continuous random variable $Y = f(X)$

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

Expected value: height example

- We have that Napoleon's height H has probability density function

$$p(h) = \begin{cases} 2 & \text{if } \; h \in (62, 62.5] \\ 0 & \text{if } \; h \notin (62, 62.5] \end{cases}$$

- Then the expected value of his height is

$$E[H] = \int_{-\infty}^{\infty} hp(h)\, dh = \int_{62}^{62.5} 2h\, dh = [h^2]_{62}^{62.5} = 62.25$$

$62.5^2 - 62^2$

Useful probability distributions

- Many common processes generate data with probability distributions that belong to families with known properties

- Even if the data are not distributed according to a known probability distribution, it is sometimes useful in practice to approximate

Discrete uniform distribution

- A discrete random variable X is **uniform** if it takes k different values and

$$P(X = x_i) = \frac{1}{k} \qquad \text{for all } x_i \text{ that are allowable values}$$

- Examples
  - Rolling a **fair** k-sided die
  - Tossing a **fair** coin $(k = 2)$

Bernoulli distribution

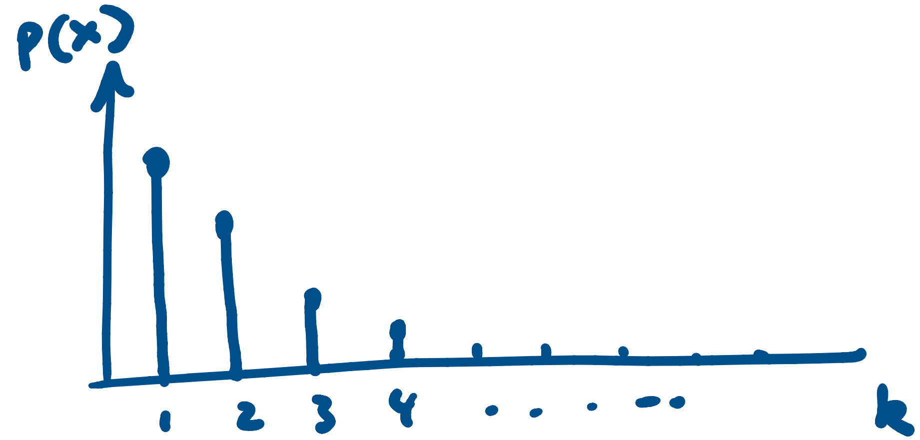
- A random variable X is **Bernoulli** if it takes on two values 0 and 1 such that

$$P(X = 1) = p \quad \text{and} \quad P(X = 0) = 1 - p$$

$$\text{with} \quad E[X] = p \quad \text{and} \quad \text{var}[X] = p(1 - p)$$

- Examples
  - Tossing a biased (or fair) coin
  - Making a free throw
  - Rolling a six-sided die and checking if it comes up 6 or not
  - **Any indicator function of a random variable**

Geometric distribution



- Examples
  - How many rolls of a six-sided die will it take to see the first 6?
  - How many free throws must I attempt to score my first point?
  - How many Bernoulli trials must take place before the first 1 occurs?

- A discrete random variable X is **geometric** if

$$P(X = k) = (1-p)^{k-1}p \qquad \text{for integer } k \geq 1$$

with $E[X] = \frac{1}{p}$ and $\mathrm{var}[X] = \frac{1-p}{p^2}$

Derivation of geometric expected value

$$E[X] = \sum_{k=1}^{\infty} k\,(1-p)^{k-1}p$$

$$= 1p + 2(1-p)p + 3(1-p)^2p + \cdots$$

$$= \underbrace{\left(p + (1-p)p + (1-p)^2p + \cdots\right)}_{1} + \underbrace{\left((1-p)p + 2(1-p)^2p + 3(1-p)^3p + \cdots\right)}_{(1-p)E[X]}$$

$$E[X] = 1 + (1-p)E[X] \Rightarrow E[X] = \frac{1}{p}$$

Geometric distribution: die example

- Let X be the number of rolls of a fair six-sided die needed to see the first 6. What is $P(X = k)$ for $k = 1, 2, 3$?

$p = \frac{1}{6}$

$P(X=1) = (1-p)^0 p = p = \frac{1}{6} \approx 0.167$

$P(X=2) = (1-p)^1 p = \frac{5}{6} \cdot \frac{1}{6} \approx 0.139$

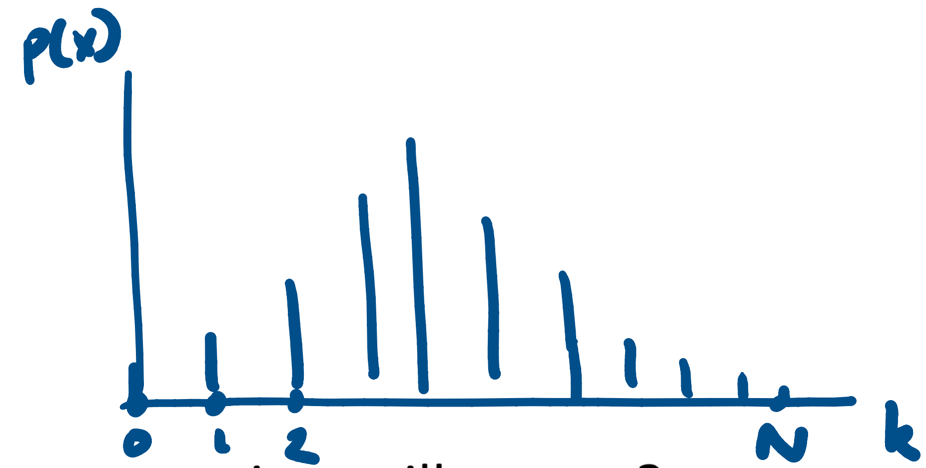
$P(X=3) = (1-p)^2 p = \left(\frac{5}{6}\right)^2 \frac{1}{6} \approx 0.116$

- Calculate $E[X]$ and $\text{std}[X]$

$E[X] = \frac{1}{p} = \frac{1}{1/6} = 6$

$\text{std}[X] = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{5/6}{(1/6)^2}} = \sqrt{30} \approx 5.48$

Binomial distribution



- Examples
  - If we roll a six-sided die N times, how many sixes will we see?
  - If I attempt N free throws, how many points will I score?
  - What is the sum of N independent and identically distributed Bernoulli trials?

- A discrete random variable X is **binomial** if

$$P(X = k) = \binom{N}{k} p^k (1-p)^{N-k} \qquad \text{for integer } 0 \le k \le N$$

with $E[X] = Np$ and $\mathrm{var}[X] = Np(1-p)$

Binomial distribution: die example

- Let X be the number of sixes in 36 rolls of a fair six-sided die. What is $P(X = k)$ for $k = 5, 6, 7$?

$P(X=5) = \binom{36}{5}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^{31} \approx 0.170$

$P(X=7) = \binom{36}{7}\left(\frac{1}{6}\right)^7\left(\frac{5}{6}\right)^{29} \approx 0.151$

$P(X=6) = \binom{36}{6}\left(\frac{1}{6}\right)^6\left(\frac{5}{6}\right)^{30} \approx 0.176$

- Calculate $E[X]$ and $\text{std}[X]$

$E[X] = (36)\left(\frac{1}{6}\right) = 6$

$\text{std}[X] = \sqrt{Np(1-p)} = \sqrt{36 \cdot \frac{1}{6} \cdot \frac{5}{6}} = \sqrt{5} \approx 2.24$

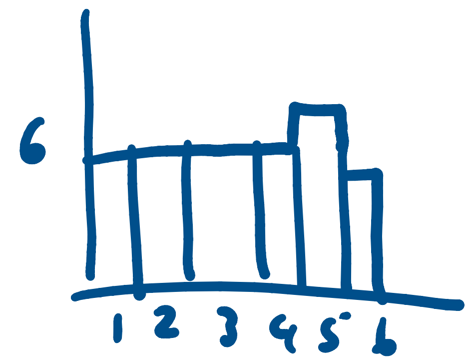
Betting brainteaser

- What would you rather bet on?
  - How many rolls of a fair six-sided die will it take to see the first 6? Bet on 1
  - ✓ How many sixes will appear in 36 rolls of a fair six-sided die? Bet on 6
- Why?

Geometric : $P(1) \approx 0.167$

Binomial : $P(6) \approx 0.176$ ✓

Multinomial distribution



- Examples
  - If we roll a six-sided die N times, how many of each value will we see?
  - What are the counts of N independent and identically distributed trials?

- A discrete k-tuple random variable $(X_1, X_2, \ldots, X_k)$ is **multinomial** if

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{N!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

ILLINOIS $\frac{8!}{3!2!1!1!1!}$

where $N = n_1 + n_2 + \cdots + n_k$

Multinomial distribution: die example

$0! = 1$

What is the probability of seeing 1 one, 2 twos, 3 threes, 4 fours, 5 fives and 0 sixes in 15 rolls of a fair six-sided die?

$$P(x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5, x_6 = 0)$$

$$= \frac{15!}{1!\,2!\,3!\,4!\,5!\,0!}\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^3\left(\frac{1}{6}\right)^4\left(\frac{1}{6}\right)^5\left(\frac{1}{6}\right)^0$$

$$= \frac{15!}{2!\,3!\,4!\,5!}\left(\frac{1}{6}\right)^{15}$$