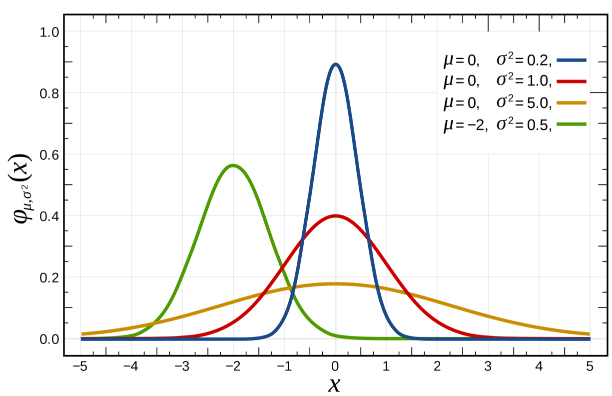
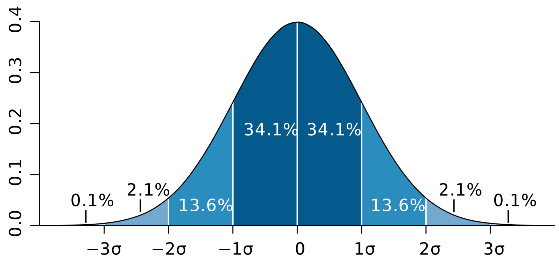
CAS CS 365
Lec 5

1. Review of Chebyshev theorem
   a. Pr(|S_n - n*pi/4| >= ε*n) <= (pi/4)*(1-pi/4)/(n*ε^2) <= beta
   b. E[S_n] = n*pi/4
   c. Var[S_n] = pi/4 * (1-pi/4)
   d. We can use this inequality in various ways
      i. Example) If ε, n are given we can compute the confidence 1- beta
2. Lindeberg-Levy Central Limit Theorem
   a. Let X_1, X_2,... be a sequence of iid RV with mean μ, variance o^2. Consider the sum S_n = X_1 + … + X_n and normalize it to obtain the RV Z_n with zero mean, and unit variance as follows:
   b. $$\boxed{Z_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathbb{V}\mathrm{ar}[S_n]}} = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu)}$$
      E[S_n] = nμ
      Var[S_n] = n * o^2
   c. As n approaches infinity, Z_n ⟶ N(0,1)
   d. In other words,
      $$\Pr(Z_n \geq t) \to \Pr(g \geq t) = \frac{1}{\sqrt{2\pi}}\int_t^{\infty} e^{-\frac{x^2}{2}}dx \text{ as } n \to \infty.$$
3. Normal distribution
   a. 

   b. As sigma increases, the graph is more spread out
   c. The graph is symmetric about the mean
   d. X ~ N(μ, o^2)
      Y = X - μ
      E[Y] = E[x-μ] = E[x] - μ = 0
      Var[Y/o] = 1/(o^2) * Var[Y] = 1

Z = (X-μ)/o ~ N[0,1]

4. CLT - confidence intervals
    a. 

    b. CLT conveniently provides us with confidence intervals tighter than WLLN
    c. If X is a gaussian RV with mean μ and std o, then we know

$$\Pr(-\sigma < X - \mu < \sigma) \approx 0.68$$
$$\Pr(-2\sigma < X - \mu < 2\sigma) \approx 0.95$$
$$\Pr(-3\sigma < X - \mu < 3\sigma) \approx 0.99$$

    d. Thus, when the CLT applies, we can use it in combination with the knowledge of properties of the Normal distribution to get confidence intervals as the following example shows

5. Tails of the normal distribution
    a. Let g → N(0,1). Then for all t > 0, we have

$$\boxed{\frac{1}{\sqrt{2\pi}} e^{-t^2/2} \le \Pr(g \ge t) \le \frac{1}{t} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}}$$

    b. For t >= 1, we get the bound $$\Pr(g \ge t) \le \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

6. Bayes' theorem
    a. Bayes' rule is a direct consequence of the definition of conditional probability / conditional density
    b. Discrete RVs
    c. $$\boxed{Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)}, \text{if } Pr(B) > 0}$$
    d. Continuous RVs
    e. $$\boxed{f_{X|Y=y}(x) = \frac{f_{Y|X=x}(y) f_X(x)}{f_Y(y)}}$$

7. Note: Bayes' rule mixed case
    a. Mixed discrete/continuous Z~p_z discrete, X~f_x continuous

$$p_{Z|X}(z \mid x) = \frac{f_{X|Z}(x \mid z)p_Z(z)}{f_X(x)} = \frac{f_{X|Z}(x \mid z)p_Z(z)}{\sum_{z'} f_{X,Z}(x,z')} = \frac{f_{X|Z}(x \mid z)p_Z(z)}{\sum_{z'} f_{X|Z}(x \mid z')p_Z(z')}$$

$$f_{X|Z}(x \mid z) = \frac{p_{Z|X}(z \mid x)f_X(x)}{p_Z(z)} = \frac{p_{Z|X}(z \mid x)f_X(x)}{\int_{-\infty}^{+\infty} p_{Z|X}(z \mid x')f_X(x')dx'}$$

    b. $$\int f_{X,Z}(x,z)dx = p_Z(z), \ \sum_z f_{X,Z}(x,z) = f_X(x).$$
8. Example
    a. Kahneman-Tversky Taxi accident
        i. A cab was involved in a hit and run accident at night
        ii. 85% Green, 15% Blue cabs in the city
        iii. A witness identified the cab as Blue
        iv. The witness correctly identified each one of the two colors 80% of the time
        v. What is the probability that the cab involved in the accident was Blue rather than Green?
            1. Let G be the event of the delinquent being Green
            2. Let B be the event of the delinquent being Blue
            3. Finally, let W be the witness' report
            4. $$Pr(B|W=b) = \frac{\Pr(W=b \mid B)\Pr(B)}{\Pr(W=b)} = \frac{\Pr(W=b \mid B)\Pr(B)}{\Pr(W=b \mid B)\Pr(B) + \Pr(W=b \mid G)\Pr(G)} = \frac{.8 \times .15}{.8 \times .15 + .2 \times .85} = .41$$
            5. Pr(W=b | B) * Pr(B) is when the car is blue and the witness said that the car is blue
            6. Pr(W=b| G) * Pr(G) is when the car is green but the witness said that the car is blue
9. Inference using Bayes' rule
    a. $$Pr(H|D) = \frac{Pr(D|H)Pr(H)}{Pr(D)}, \text{and } Pr(D) > 0$$
    b. Hypothesis given data → posterior
    c. Posterior is proportional to "data given hypothesis" (known as likelihood) multiplied by prior (Pr(H))
10. Bayes rule inference example
    a. Fair coin or biased coin → 1000 times. What is the probability of 700 heads?

Pr(D) = Pr(D, fair coin) + Pr(D, biased coin)

= Pr(D| fair coin) * Pr(fair coin) + Pr(D | biased coin) * Pr(biased coin)

P(Fair coin gives 700 heads | D) = Pr(D | fair coin gives 700 Heads) * Pr(heads) / Pr(D)

$$= \frac{{}_{1000}C_{700} * (½)^{1000} * ½}{{}_{1000}C_{700} * (½)^{1000} * ½ + {}_{1000}C_{700} * (0.8)^{700} * (0.2)^{300} * ½}$$

$$= \frac{(½)^{1000}}{(½)^{1000} + (0.8)^{700} * (0.2)^{300}}$$