c. $X_1, \ldots, X_n$ independent Bernoulli trials with success probability $\pi$

i. $E[X_i] = \pi$.

ii. Hence $E[\bar{X}] = \pi$, because expectation of mean is expectation of individual values.

iii. Note that $\sum_{j=1}^n$ is the number of 1's, and so the average is the proportion of 1's.

iv. For independent observation, the CLT tells us that the estimator $\bar{X} \sim N(\pi, \pi(1 - \pi)/n)$.

d. $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$

i. Estimate $\sigma^2$ by $s^2 = \sum_{j=1}^n (X_j - \bar{X})^2 / (n - 1)$.

ii. Note that $s^2 = \sum_{j=1}^n (X_j^2 - 2X_j\bar{X} + \bar{X}^2) / (n - 1) = (\sum_{j=1}^n X_j^2 - 2\bar{X}\sum_{j=1}^n X_j + n\bar{X}^2) / (n - 1) = (\sum_{j=1}^n X_j^2 - 2n\bar{X}^2 + n\bar{X}^2) / (n - 1) = (\sum_{j=1}^n X_j^2 - n\bar{X}^2) / (n - 1)$

iii. Note that

- $E[X_j^2] = E[X_j]^2 + \text{Var}[X_j] = \mu^2 + \sigma^2$
- $E[\bar{X}^2] = E[\bar{X}]^2 + \text{Var}[\bar{X}] = \mu^2 + \sigma^2/n$

iv. Hence $E[s^2] = (\sum_{j=1}^n E[X_j^2] - nE[\bar{X}^2]) / (n - 1) = (n(\mu^2 + \sigma^2) - n(\mu + \sigma^2/n)) / (n - 1) = (n - 1)\sigma^2 / (n - 1) = \sigma^2$
v. Hence $s^2$ is unbiased for $\sigma^2$, but it wouldn’t be unbiased if we had divided by $n$ rather than $n - 1$.

vi. In the simplest case ($n=2$), let’s work out the distribution of $s^2$ and $\bar{X}$

- $\bar{X} = (X_1 + X_2)/2$, and $s^2 = (X_1 - (X_1 + X_2)/2)^2 + (X_2 - (X_1 + X_2)/2)^2 = (X_1 - X_2)^2/4 + (X_2 - X_1)^2/4 = D^2/2$
  
  for $D = X_1 - X_2$.

- $x_1 = \bar{X} + d/2$ and $x_2 = \bar{X} - d/2$

- Joint density for $X_1$ and $X_2$ is $f_{X_1, X_2}(x_1, x_2) \propto \exp\left(-\frac{(x_1 - \mu)^2}{2\sigma^2}\right) \exp\left(-\frac{(x_2 - \mu)^2}{2\sigma^2}\right) = \exp\left(-\frac{(x_1 - \mu)^2}{2\sigma^2} - \frac{(x_2 - \mu)^2}{2\sigma^2}\right) = \exp\left(-\frac{\bar{x} + d/2 - \mu)^2}{2\sigma^2} - \frac{\bar{x} - d/2 - \mu)^2}{2\sigma^2}\right) = \exp\left(-\frac{1}{4}d^2/\sigma^2 - (\bar{x} - \mu)^2/\sigma^2\right)$

- Hence the joint density for $D$ and $\bar{X}$ is also proportional to $\exp\left(-\frac{1}{4}d^2/\sigma^2 - (\bar{x} - \mu)^2/\sigma^2\right)$ since the change of variables from $(x_1, x_2)$ to $(\bar{x},)$ are constants that don’t depend on $x_1$ or $x_2$.

- Since the joint density factors, then $D$ and $\bar{X}$ are
Lecture 23

independent, and \( D = N(0, 2\sigma^2) \).

- Also, \( P \left( S^2 \geq y \right) = 2P(D \geq \sqrt{y}) = 2(1 - \Phi(\sqrt{y}/(\sigma\sqrt{2}))) \).

- Also, \( f_{S^2}(y) = -\frac{d}{dy} P \left( S^2 \geq y \right) = 2 \exp(-y/(2\sigma^2))(2\pi)^{-1/2} \frac{d}{dy} \left( \sqrt{y}/(\sigma\sqrt{2}) \right) = 2 \exp(-y/(2\sigma^2))(2\pi)^{-1/2} \frac{1}{\pi} y^{-1/2}/(\sigma) \) \( = \frac{1}{2} \exp(-y/(2\sigma^2)) \pi^{-1/2} y^{-1/2}/(\sigma) \).

- So, \( S^2 \sim \Gamma(1/2, \sigma^2/2) \), and \( S^2/\sigma^2 \sim \Gamma(1/2, 1/2) \).

- More generally, for \( S^2 \) generated from \( n \) \( N(\mu, \sigma^2) \) variables, \( (n - 1)S^2 \sim \Gamma(n/2, \sigma^2/2) \), and \( (n - 1)S^2/\sigma^2 \sim \Gamma((n - 1)/2, 1/2) \).

- This latter distribution is called the \( \chi^2_{k-1} \) distribution, read ”chi-square distribution on \( n - 1 \) degrees of freedom”.