Module 2.2
Statistical Estimation
Basic Concepts

S. Lakshmivarahan

School of Computer Science
University of Oklahoma
Norman, OK, 73071
USA

A Simple Estimation Problem: Gaussian Distribution

- Let $p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] = N(\mu, \sigma^2)$.
- Assume μ and/or σ^2 are *not* known.
- We have access to a generator for $p(x)$, from which we can draw **IID** samples $x_1, x_2, \ldots x_n$.
- Goal is to estimate the unknowns from the random samples.

- An estimator $\hat{x}$ of μ is a function of the known observations: $\hat{x}(n) = \phi(x_1, x_2, \ldots, x_n)$
- A natural choice for ϕ is the numerical average: $\hat{x}(n) = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Since $\phi(x_1, x_2, \ldots, x_n)$ is a function of the random numbers $x_1, x_2, \ldots, x_n$, clearly $\hat{x}(n)$ is also a **random variable**.

Sampling Distribution

- The probability distribution of $\hat{x}(n)$ is called the sampling distribution and it depends on the choice of the function ϕ and the distribution of x_i
- The behavior of the mean and the variance of the sampling distribution as the number of samples $n \to \infty$, characterize the properties and quality of the estimate $\hat{x}(n)$ - see Problem at the end of the module

Mean of the Sampling Distribution - Unbiased Estimate

- An estimate is said to be unbiased, if the mean of the sampling distribution coincides with the true mean.
- $\mathbf{E}(\hat{x}(n)) = \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}(x_i) = \mu.$
- Hence, the estimate $\hat{x}(n)$ is an unbiased estimate of the unknown, μ called the population mean.

Consistency of Estimates

- The estimate $\hat{x}(n)$ is said to be consistent if

$$\textbf{Prob}\left(|\hat{x}(n) - \mu| > \delta\right) \rightarrow 0 \text{ as } n \rightarrow \infty \text{ for } \delta > 0 \quad (1)$$

- That is, the probability measure defined by the sampling distribution tends to concentrate on an arbitrarily small interval around the true mean μ as $n \rightarrow \infty$.

Chebyshev Inequality

- Recall:

$$\mathbf{E}(x^2) = \int_{-\infty}^{\infty} x^2 p(x)dx \geq \int_{|x|>\sigma} p(x)dx = \sigma^2 \,\mathbf{Prob}[|x| > \sigma] \tag{2}$$

- $\mathbf{Prob}[|x| > \sigma] \leq \frac{\mathbf{E}(x^2)}{\sigma^2}$ - Chebyshev inequality

- $$\begin{aligned}\textbf{Var}\left(\hat{x}(n)\right) = \textbf{E}\left(\hat{x}(n) - \mu\right)^2 &= \textbf{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \mu\right]^2 \\ &= \textbf{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)\right]^2 = \frac{1}{n^2}\sum_{i=1}^{n}\textbf{E}\left(x_i - \mu\right)^2 = \frac{\sigma^2}{n}\end{aligned}$$

- By Chebyshev inequality:

$$\begin{aligned}\textbf{Prob}\left[|\hat{x}(n) - \mu| > \sigma\right] \quad &\leq \frac{\textbf{E}\left(\hat{x}(n) - \mu\right)^2}{\sigma^2} = \frac{\textbf{Var}\left(\hat{x}(n)\right)}{\sigma^2} \\ &= \frac{\sigma^2}{n\sigma^2} \to 0 \;\text{ as }\; n \to \infty\end{aligned}$$

$\therefore \hat{x}(n)$ is consistent

Relative Efficiency

- Let $\hat{x}^{(1)}$ and $\hat{x}^{(2)}$ be the two estimates of the unknown.
- Say that $\hat{x}^{(1)}$ is relatively more efficient than $\hat{x}^{(2)}$ if

$$\mathbf{Var}(\hat{x}^{(1)}) < \mathbf{Var}(\hat{x}^{(2)}) \tag{3}$$

- Clearly, for $n_1 > n_2$, $\hat{x}(n_1)$ is more efficient than $\hat{x}(n_2)$ since

$$\mathbf{Var}\left(\hat{x}(n_1)\right) = \frac{\sigma^2}{n_1} \leq \frac{\sigma^2}{n_2} = \mathbf{Var}(\hat{x}(n_2)) \tag{4}$$

- That is the relative efficiency of $\hat{x}(n)$ increases as $n \to \infty$.

- The standard error (SE) of an estimate is the square root of its variance
- $$SE \text{ for } \hat{x}(n) = \frac{\sigma}{\sqrt{n}} = \mathbf{O}\left(\frac{1}{\sqrt{n}}\right) \tag{5}$$

Central Limit Theorem for $\hat{x}(n)$

- Recall: $\mathbf{E}[\hat{x}(n)] = \mu$, $\mathbf{Var}[\hat{x}(n)] = \frac{\sigma^2}{n}$
- Claim: The distribution of the centered and normalized random variable $\frac{\hat{x}(n)-\mu}{(\sigma/\sqrt{n})} \rightarrow N(0,1)$, the standard normal distribution as $n \rightarrow \infty$, called the Central limit theorem (CLT)

Consequence of CLT

- For large n:

$$\textbf{Prob}\left[-2 \leq \frac{\sqrt{n}(\hat{x}(n) - \mu)}{\sigma} \leq 2\right] = 0.95 \qquad (6)$$

- That is, with 0.95 probability, the centered and normalized estimate will lie in the interval $[-2, 2]$.

CASE 2: μ is known, σ^2 not known

- Let $\hat{\sigma}^2$ be an estimate for σ^2
- A natural choice is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{7}$$

- $\mathbf{E}(\hat{\sigma}^2) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}(x_i - \mu)^2 = \frac{n\sigma^2}{n} = \sigma^2$
- That is, $\hat{\sigma}^2$ is unbiased

Variance of $\hat{\sigma}^2$

- $\mathbf{Var}(\hat{\sigma}^2) = \mathbf{E}\left[\hat{\sigma}^2 - \sigma^2\right]^2$
- Let $y_i = x_i - \mu$, $\mathbf{E}(y_i) = 0$, $\mathbf{E}({y_i}^2) = \sigma^2$
- $\mathbf{Var}(\hat{\sigma}^2) = \mathbf{E}\left[\left(\frac{1}{n}\sum_{i=1}^n {y_i}^2\right) - \sigma^2\right]^2 = \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^n ({y_i}^2 - \sigma^2)\right]^2$
- Recall: $(a_1 + a_2 + \cdots + a_n)^2 = \sum_{i=1}^n a_i^2 + 2\sum_{i<j}^n a_i a_j$

Variance of $\hat{\sigma}^2$ - continued

- $$\begin{aligned}\mathbf{Var}(\hat{\sigma}^2) &= \frac{1}{n^2}\mathbf{E}\left\{\sum_{i=1}^{n}({y_i}^2-\sigma^2)^2 + 2\sum_{i<j}(y_i^2-\sigma^2)(y_j^2-\sigma^2)\right\}\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{E}({y_i}^2-\sigma^2)^2 + \frac{2}{n^2}\sum_{i<j}\mathbf{E}(y_i^2-\sigma^2)\mathbf{E}(y_j^2-\sigma^2)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{E}({y_i}^2-\sigma^2)^2 \text{ since } y_i \text{ are } \mathbf{IID} \text{ and } \mathbf{E}({y_i}^2-\sigma^2)=0\end{aligned}$$

Variance of $\hat{\sigma}^2$ - continued

- $\mathbf{E}({y_i}^2 - \sigma^2)^2 = \mathbf{E}(y_i^4) - 2\mathbf{E}({y_i}^2)\sigma^2 + \sigma^4$
- Recall: $y_i \sim N(0, \sigma^2)$
  $\therefore \mathbf{E}({y_i}^3) = 0$ - skewness
  $\mathbf{E}({y_i}^4) = 3\sigma^4$ - kurtosis
- Substituting :
  $\mathbf{E}({y_i}^2 - \sigma^2)^2 = 3\sigma^4 - 2\sigma^4 + \sigma^4 = 2\sigma^4$

Variance of $\hat{\sigma}^2$ - continued

- $\textbf{Var}(\hat{\sigma}^2) = \frac{2}{n^2}\sum_{i=1}^{n}\sigma^4 = \frac{2\sigma^4}{n}$
- The standard error is the estimate $\hat{\sigma}^2$ of the variance σ^2 is $\mathbf{O}\left(\frac{1}{\sqrt{n}}\right)$
- This estimate $\hat{\sigma}^2$ is an unbiased and asymptotically consistent estimate of σ^2

χ^2 - (Chi-Squared) Distribution

- A nonnegative random variable z is said to be $\chi^2(k)$ distributed with degree of freedom $k(\geq 1)$ if its density function is given by

$$p(z) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} z^{\frac{k}{2}-1} \exp\left(-\frac{z}{2}\right) \quad \text{for} \quad z > 0 \tag{8}$$

- $\Gamma(r)$ is the standard Gamma function
  - $\Gamma(1) = 1$
  - $\Gamma(1/2) = \sqrt{\pi}$
  - $\Gamma(r+1) = r\Gamma(r)$ if r is a positive integer.

Mean/Variance of $\chi^2(k)$ - (Chi-Squared) Distribution

- Let $z \sim \chi^2(k)$
- Then $\mathbf{E}(z) = k$, $\mathbf{Var}(z) = 2k$
- Verify these claims - Home Work

Relation between Normal and χ^2 - (Chi-Squared) Distribution

- Let $x_1, x_2, \ldots, x_n \sim \textbf{IID} N(0, 1)$, the standard normal
- Let $z = \sum_{i=1}^{n} x_i^2 \sim \chi^2(n)$
- Then $\mathbf{E}(z) = n, \mathbf{Var}(z) = 2n$
- If $z_1 \sim \chi^2(n_1)$, $z_2 \sim \chi^2(n_2)$ and z_1, z_2 are independent, then $z = z_1 + z_2 \sim \chi^2(n_1 + n_2)$

$\mathbf{Var}(\hat{\sigma}^2)$ - An Alternative Method

- Recall $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$
- Consider
  $\left(\frac{n}{\sigma^2}\right)\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{(\sigma/\sqrt{n})}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} z_i^2, z_i \sim \mathbf{IID}\, N(0,1)$
- Hence $\left(\frac{n}{\sigma^2}\right)\hat{\sigma}^2 = \eta \sim \chi^2(n)$
- $\mathbf{E}(\eta) = n$ and $\mathbf{Var}(\eta) = 2n$
- $\mathbf{E}(\hat{\sigma}^2) = \mathbf{E}\left[\frac{\sigma^2}{n}\eta\right] = \frac{\sigma^2}{n}\mathbf{E}(\eta) = \sigma^2$
- $\mathbf{Var}(\hat{\sigma}^2) = \mathbf{Var}(\frac{\sigma^2}{n}\eta) = \frac{\sigma^4}{n^2}2n = \frac{2\sigma^4}{n}$

CASE 3: Both μ and σ^2 are not known

- Let

$$\begin{aligned} \hat{x} &= \frac{1}{n} \sum\nolimits_{i=1}^{n} x_i, \\ s^2 &= \frac{1}{n-1} \sum\nolimits_{i=1}^{n} (x_i - \hat{x})^2, \end{aligned} \tag{9}$$

  be the estimates of μ, and σ^2 respectively
- Recall: $\hat{x}$ is an unbiased estimate of μ

s^2 is unbiased

- Consider with $\hat{x} = \hat{x}(n)$:

$$\begin{aligned} &\mathbf{E}\left[\sum_{i=1}^{n}(x_i - \hat{x})^2\right] \\ &= \mathbf{E}\left[\sum_{i=1}^{n}\left(x_i^2 - 2x_i\hat{x} + \hat{x}^2\right)\right] \\ &= \mathbf{E}\left[\sum_{i=1}^{n} x_i^2 - 2\hat{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\hat{x}^2\right] \\ &= \mathbf{E}\left[\sum_{i=1}^{n} x_i^2 - n\hat{x}^2\right] \left(\text{since } \sum_{i=1}^{n} x_i = n\hat{x}\right) \\ &= \sum_{i=1}^{n}\mathbf{E}(x_i^2) - n\mathbf{E}(\hat{x}^2) \end{aligned} \tag{10}$$

Unbiasedness of s^2

- By definition:

$$\begin{aligned} \mathbf{E}(x_i^2) &= \mathbf{Var}(x_i) + (\mathbf{E}(x_i))^2 = \sigma^2 + \mu^2 \\ \therefore \sum_{i=1}^{n} \mathbf{E}(x_i^2) &= n(\sigma^2 + \mu^2) \end{aligned} \tag{11}$$

- $n\,\mathbf{E}(\hat{x}^2)$

$$\begin{aligned} & n\,\mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]^2 \\ &= \frac{1}{n}\mathbf{E}\left[\sum_{i=1}^{n} x_i\right]^2 (\because x_i \text{ are } \mathbf{IID}) \\ &= \frac{1}{n}\mathbf{E}\left[\sum_{i=1}^{n}\left(x_i^2\right) + 2\sum_{i<j}\mathbf{E}(x_i)\mathbf{E}(x_j)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\left(x_i^2\right) + \frac{2}{n}\binom{n}{2}\mu^2 \\ &= (\sigma^2+\mu^2)+(n-1)\mu^2 = \sigma^2 + n\mu^2 \end{aligned} \tag{12}$$

Unbiasedness of s^2 - continued

- Taking expectations on both sides of (9), substituting (10)–(12) and simplifying

$$\begin{aligned}\mathbf{E}(s^2) &= \frac{1}{n-1}\left[\sum_{i=1}^{n} \mathbf{E}\left(x_i\right)^2 - n\,\mathbf{E}\left(\hat{x}^2\right)\right] \\ &= \frac{1}{n-1}\left[n\left(\sigma^2+\mu^2\right) - \sigma^2 - n\,\mu^2\right] \\ &= \sigma^2\end{aligned}$$

$\therefore\ s^2$ is unbiased

Var(s^2)

- Recall

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

- Hence

$$\left(\frac{n-1}{\sigma^2}\right)s^2 = \eta = \sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{\sigma}\right)^2 \sim \chi^2(n-1)$$

$\mathbf{Var}(s^2)$

- Recall $\mathbf{E}(\eta) = (n-1)$ and $\mathbf{Var}(\eta) = 2(n-1)$
- $\mathbf{E}(s^2) = \mathbf{E}\left[\frac{\sigma^2}{n-1}\eta\right] = \sigma^2$

  $\mathbf{Var}(s^2) = \mathbf{Var}\left[\frac{\sigma^2}{n-1}\eta\right] = \frac{2(n-1)\sigma^4}{(n-1)^2} = \frac{2\sigma^4}{(n-1)}$
- Hence, standard error in s^2 is $\mathbf{O}\left(\frac{1}{\sqrt{n-1}}\right)$ and s^2 is a consistent estimate

Problem

- Set $n = 10^2$ and generate $N = 10^4$ different sets of n **IID** random variables $x_1, x_2, \ldots, x_n$ from the standard normal distribution
    - (a) Compute N different values of $\hat{x}(n)$ and plot them on a real line
    - (b) Compute the histogram, mean and variance of the sampling distribution