Consider some probability distributions:
- \( \text{Ber}(p) \quad \theta = p \)
- \( \text{Poi}(\lambda) \quad \theta = \lambda \)
- \( \text{Multinomial}(p_1, p_2, \ldots, p_n) \quad \theta = (p_1, p_2, \ldots, p_n) \)
- \( \text{Uni}(\alpha, \beta) \quad \theta = (\alpha, \beta) \)
- \( \text{Normal}(\mu, \sigma^2) \quad \theta = (\mu, \sigma^2) \)
- Etc.

Call these “parametric models”

Given model, parameters yield actual distribution

- Usually refer to parameters of distribution as \( \theta \)
- Note that \( \theta \) that can be a vector of parameters

In real world, don’t know “true” parameters

- But, we do get to observe data
  - E.g., number of times coin comes up heads, lifetimes of disk drives produced, number of visitors to web site per day, etc.

Need to estimate model parameters from data

“Estimator” is random variable estimating parameter

Want “point estimate” of parameter

- Single value for parameter as opposed to distribution
- Estimate of parameters allows:
  - Better understanding of process producing data
  - Future predictions based on model
  - Simulation of processes

Why Do We Care?

**Recall Sample Mean**

- Consider \( n \) I.I.D. random variables \( X_1, X_2, \ldots, X_n \)
  - \( X_i \) have distribution \( F \) with \( E[X_i] = \mu \) and \( \text{Var}(X_i) = \sigma^2 \)
  - We call sequence of \( X_i \) a **sample** from distribution \( F \)
  - Recall sample mean: \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \) where \( E[\bar{X}] = \mu \)
  - Recall variance of sample mean: \( \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \)
  - Clearly, sample mean \( \bar{X} \) is a random variable

Why Do We Care?

**Sampling Distribution**

- Note that sample mean \( \bar{X} \) is random variable
  - “Sampling distribution of mean” is the distribution of the random variable \( \bar{X} \)
  - Central Limit Theorem tells us sampling distribution of \( \bar{X} \) is approximately normal when sample size, \( n \), is large
    - Rule of thumb for “large” \( n \): \( n > 30 \), but larger is better (\( > 100 \))
    - Can use CLT to make inference about sample mean

**Confidence Interval for Mean**

- Consider I.I.D. random variables \( X_1, X_2, \ldots, X_n \)
  - \( X_i \) have distribution \( F \) with \( E[X_i] = \mu \) and \( \text{Var}(X_i) = \sigma^2 \)
  - Let \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \) \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \)
  - For large \( n \), \( 100(1 - \alpha)\% \) **confidence interval** is:
    \[
    \left( \bar{X} - z_{\alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{S}{\sqrt{n}} \right)
    \]
    where \( \phi(z_{\alpha/2}) = 1 - (\alpha/2) \)
    - E.g.: \( \alpha = 0.05 \), \( \alpha/2 = 0.025 \), \( \phi(z_{0.025}) = 0.975 \), \( z_{0.025} = 1.96 \)
    - Meaning: \( 100(1 - \alpha)\% \) of time that confidence interval is computed from sample, true \( \mu \) would be in interval
      - **Not** \( \bar{X} \) or \( \mu \) is \( 100(1 - \alpha)\% \) likely to be in this particular interval

**Example of Confidence Interval**

- Idle CPUs are the bane of our existence
  - Large (unnamed) company wants to estimate average number of idle hours per CPU
  - 225 computers are monitored for idle hours
  - Say \( \bar{X} = 11.6 \text{ hrs.}, \quad S^2 = 16.81 \text{ hrs}^2 \), so \( S = 4.1 \text{ hrs.} \)
  - Estimate \( \mu \), mean idle hrs./CPU, with 90% conf. interval
    \[
    \left( \bar{X} - 1.645 \frac{S}{\sqrt{225}}, \bar{X} + 1.645 \frac{S}{\sqrt{225}} \right) = (11.15, 12.05)
    \]
    - 90% of time that such an interval computed, true \( \mu \) is in it
Method of Moments

- Recall: $n$-th moment of distribution for variable $X$: $m_k = E[X^k]$
- Consider I.I.D. random variables $X_1, X_2, ..., X_n$:
  - $X_i$ have distribution $F$
  - Let $\hat{m}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$, $\hat{m}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$, ..., $\hat{m}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$
- $\hat{m}_k$ are called the “sample moments”
- Methods of moments estimators
  - Estimate model parameters by equating “true” moments to sample moments: $m_k \approx \hat{m}_k$

Examples of Method of Moments

- Recall the sample mean: $\hat{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \hat{m}_1 = E[X]$
- This is method of moments estimator for $E[X]$
- Method of moments estimator for variance
  - Estimate second moment: $\hat{m}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$
  - Variance: $\text{Var}(X) = E[X^2] - (E[X])^2$
  - Estimate: $\text{Var}(X) = \hat{m}_2 - (\hat{m}_1)^2$
    - Recall sample variance:
      \[ S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{X})^2 = \frac{1}{n-1} \sum_{i=1}^{n} X_i^2 - \frac{n}{n-1} (\hat{m}_1 - (\hat{m}_1)^2) \]

Small Samples = Problems

- What is difference between sample variance and MOM estimate for variance?
  - Imagine you have a sample of size $n = 1$
- What is sample variance?
  \[ S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{X})^2 \] - undefined
  - i.e., don’t really know variability of data
- What is MOM estimate of variance?
  \[ \sum_{i=1}^{n} (X_i - \hat{X})^2 \] - $n$
  - i.e., have complete certainty about distribution!

Estimator Consistency

- Estimator “consistent”: $\lim_{n \to \infty} P[|\hat{\theta} - \theta| < \epsilon] = 1$ for $\epsilon > 0$
- As we get more data, estimate should deviate from true value by at most a small amount
- This is actually known as “weak” consistency
- Note similarity to weak law of large numbers:
  \[ \lim_{n \to \infty} P[|X - \mu| \geq \epsilon] = 0 \]
- Equivalently:
  \[ \lim_{n \to \infty} P[|X - \mu| < \epsilon] = 1 \]
- Establishes sample mean as consistent estimate for $\mu$
- Generally, MOM estimates are consistent

Examples of Method of Moments

- Recall the sample mean: $\hat{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \hat{m}_1 = E[X]$
- This is method of moments estimator for $E[X]$
- Method of moments estimator for variance
  - Estimate second moment: $\hat{m}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$
  - Variance: $\text{Var}(X) = E[X^2] - (E[X])^2$
  - Estimate: $\text{Var}(X) = \hat{m}_2 - (\hat{m}_1)^2$
    - Recall sample variance:
      \[ S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{X})^2 = \frac{1}{n-1} \sum_{i=1}^{n} X_i^2 - \frac{n}{n-1} (\hat{m}_1 - (\hat{m}_1)^2) \]

Estimator Bias

- Bias of estimator: $E[\hat{\theta}] - \theta$
  - When bias = 0, we call the estimator “unbiased”
  - A biased estimator is not necessarily a bad thing
  - Sample mean $\hat{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is unbiased estimator
  - Sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{X})^2$ is unbiased estimator
  - MOM estimator of variance $= \frac{n-1}{n} S^2$ is biased
    - Asymptotically less biased as $n \to \infty$
    - For large $n$, either sample variance or MOM estimate of variance is fine.

Method of Moments with Bernoulli

- Consider I.I.D. random variables $X_1, X_2, ..., X_n$
  - $X_i \sim \text{Ber}(p)$
- Estimate $p$
  \[ p = E[X_i] = \hat{m}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \hat{p} \]
  - Can use estimate of $p$ for $X \sim \text{Bin}(n, p)$
  - If you know what $n$ is, you don’t need to estimate that
Method of Moments with Poisson

- Consider I.I.D. random variables $X_1, X_2, ..., X_n$
- $X_i \sim \text{Poi}(\lambda_i)$
- Estimate $\lambda$:
  $$\hat{\lambda} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \lambda$$
- But note that for Poisson, $\lambda = \text{Var}(X_i)$ as well!
- Could also use method of moments to estimate:
  - Usually, use first moment estimate
  - More generally, use the one that’s easiest to compute

Method of Moments with Normal

- Consider I.I.D. random variables $X_1, X_2, ..., X_n$
- $X_i \sim \text{N}(\mu, \sigma^2)$
- Estimate $\mu$:
  $$\hat{\mu} = E[X_i] = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \mu$$
- Now estimate $\sigma^2$:
  $$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} X_i \right)^2$$

Method of Moments with Uniform

- Consider I.I.D. random variables $X_1, X_2, ..., X_n$
- $X_i \sim \text{Unif}(\alpha, \beta)$
- Estimate mean:
  $$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i = \mu$$
- Estimate variance:
  $$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2 = \sigma^2$$
- For $\text{Unif}(\alpha, \beta)$, know that: $\mu = \frac{\alpha + \beta}{2}$ and $\sigma^2 = \frac{(eta - \alpha)^2}{12}$
- Solve (two equations, two unknowns):
  - Set $\beta = 2\mu - \alpha$; substitute into formula for $\sigma^2$ and solve:
    $$\hat{\alpha} = \bar{X} - \sqrt{3}\hat{\sigma} \quad \text{and} \quad \hat{\beta} = \bar{X} + \sqrt{3}\hat{\sigma}$$