What Are Parameters?

- Consider some probability distributions:
  - Ber(p) θ = p
  - Poi(λ) θ = λ
  - Multinomial(p_1, p_2, ..., p_m) θ = (p_1, p_2, ..., p_m)
  - Uni(α, β) θ = (α, β)
  - Normal(μ, σ^2) θ = (μ, σ^2)
  - Etc.
- Call these "parametric models"
- Given model, parameters yield actual distribution
  - Usually refer to parameters of distribution as θ
  - Note that θ that can be a vector of parameters

Why Do We Care?

- 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

Recall Sample Mean

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

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

Demo Redux

Confidence Interval for Mean

- Consider I.I.D. random variables X_1, X_2, ...
  - X_i have distribution *F* with $E[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$
  - Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\quad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$ $\quad S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$
  - For large *n*, 100(1 – α)% **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.: $\alpha = 0.05,\ \alpha/2 = 0.025,\ \Phi(z_{\alpha/2}) = 0.975,\ z_{\alpha/2} = 1.96$
  - Meaning: 100(1 – α)% of time that confidence interval is computed from sample, true μ would be in interval
    - **Not**: $\bar{X}$ or μ is 100(1 – α)% 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 $\bar{X} = 11.6$ hrs., $S^2 = 16.81$ hrs^2., so $S = 4.1$ hrs.
  - Estimate μ, mean idle hrs./CPU, with 90% conf. interval

    $$\alpha = 0.10,\ \alpha/2 = 0.05,\ \Phi(z_{\alpha/2}) = 0.95,\ z_{\alpha/2} = 1.645$$

    $$\left(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

    $$\left(11.6 - 1.645\frac{4.1}{\sqrt{225}},\ 11.6 + 1.645\frac{4.1}{\sqrt{225}}\right) = (11.15,\ 12.05)$$
  - 90% of time that such an interval computed, true μ is in it

Method of Moments

- Recall: *n*-th moment of distribution for variable X:

$$m_n = E[X^n]$$

- Consider I.I.D. random variables $X_1, X_2, ..., X_n$
  - X_i have distribution *F*
  - Let $\hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \hat{m}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \quad ... \quad \hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
  - $\hat{m}_i$ are called the "sample moments"
    - Estimates of the moments of distribution based on data
- Method of moments estimators
  - Estimate model parameters by equating "true" moments to sample moments: $m_i \approx \hat{m}_i$

Examples of Method of Moments

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

$$= \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} \bar{X}^2 = \frac{\sum_{i=1}^{n}(X_i^2 - \bar{X}^2)}{n}$$

  - Recall sample variance:

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1} = \sum_{i=1}^{n} \frac{(X_i^2 - 2X_i\bar{X} + \bar{X}^2)}{n-1} = \frac{\sum_{i=1}^{n}(X_i^2 - \bar{X}^2)}{n-1} = \frac{n}{n-1}(\hat{m}_2 - (\hat{m}_1)^2)$$

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?

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1} = \text{undefined}$$

  - i.e., don't really know variability of data
  - What is MOM estimate of variance?

$$\frac{\sum_{i=1}^{n}(X_i^2 - \bar{X}^2)}{n} = \frac{\sum_{i=1}^{1}(X_i^2 - X_i^2)}{1} = 0$$

  - i.e., have complete certainty about distribution!
    - There is no variance

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 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is unbiased estimator
  - Sample variance $S^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$ is unbiased estimator
  - MOM estimator of variance = $\frac{n-1}{n} S^2$ is biased
    - Asymptotically less biased as $n \rightarrow \infty$
  - For large *n*, either sample variance or MOM estimate of variance is fine.

Estimator Consistency

- Estimator "consistent": $\lim_{n \to \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$ for $\varepsilon > 0$
  - 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:

$$\lim_{n \to \infty} P(|\bar{X} - \mu| \geq \varepsilon) \rightarrow 0$$

  - Equivalently:

$$\lim_{n \to \infty} P(|\bar{X} - \mu| < \varepsilon) \rightarrow 1$$

  - Establishes sample mean as consistent estimate for μ
  - Generally, MOM estimates are consistent

Method of Moments with Bernoulli

- Consider I.I.D. random variables $X_1, X_2, ..., X_n$
  - $X_i \sim \mathrm{Ber}(p)$
- Estimate *p*

$$p = E[X_i] \approx \hat{m}_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{p}$$

  - Can use estimate of *p* for X ~ 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 $X_1, X_2, \ldots, X_n$
  - $X_i \sim \text{Poi}(\lambda)$
- Estimate λ

$$\lambda = E[X_i] \approx \hat{m}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\lambda}$$

  - But note that for Poisson, $\lambda = \text{Var}(X_i)$ as well!
  - Could also use method of moments to estimate:

$$\lambda = E[X_i^2] - E[X_i]^2 \approx \hat{m}_2 - (\hat{m}_1)^2 = \frac{\sum_{i=1}^{n}(X_i^2 - \overline{X}^2)}{n} = \hat{\lambda}$$

  - 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 $X_1, X_2, \ldots, X_n$
  - $X_i \sim N(\mu, \sigma^2)$
- Estimate μ

$$\mu = E[X_i] \approx \hat{m}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\mu}$$

- Now estimate σ^2

$$\sigma^2 \approx \hat{m}_2 - (\hat{m}_1)^2$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} \overline{X}^2 = \frac{\sum_{i=1}^{n}(X_i^2 - \overline{X}^2)}{n}$$

Method of Moments with Uniform

- Consider I.I.D. random variables $X_1, X_2, \ldots, X_n$
  - $X_i \sim \text{Uni}(\alpha, \beta)$
  - Estimate mean:

$$\mu \approx \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\mu}$$

  - Estimate variance:

$$\sigma^2 \approx \hat{m}_2 - (\hat{m}_1)^2 = \frac{\sum_{i=1}^{n}(X_i^2 - \overline{X}^2)}{n} = \hat{\sigma}^2$$

  - For Uni(α, β), know that: $\mu = \frac{\alpha+\beta}{2}$ and $\sigma^2 = \frac{(\beta-\alpha)^2}{12}$
  - Solve (two equations, two unknowns):
    - Set $\beta = 2\mu - \alpha$, substitute into formula for σ^2 and solve:

$$\hat{\alpha} = \overline{X} - \sqrt{3}\hat{\sigma} \quad \text{and} \quad \hat{\beta} = \overline{X} + \sqrt{3}\hat{\sigma}$$