Computer Simulation
Module 8: Input Data Analysis

Dave Goldsman, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

Introduction

Module Overview

Last Module: We looked at a bunch of random variate generation techniques.

This Module: But which RVs do we use as inputs to the simulation?

Have to be very careful about GIGO (garbage-in-garbage out) – Bad input can ruin your nice model!

Module Overview

1. Introduction ← This lesson
2. Identifying Distributions
3. Unbiased Point Estimators
4. Mean Squared Error
5. Maximum Likelihood Estimators
6. MLE Examples
7. Invariance Property of MLEs
8. Method of Moments

Overview (cont’d)

9. Goodness-of-Fit Tests
10. Exponential Example
11. Weibull Example
12. More Goodness-of-Fit Tests
13. Problem Children
14. Demos

Input Analysis

Goal: Use random variables in your simulation that adequately approximate what's going on in the real world. Examples:

- Interarrival times
- Service times
- Breakdown times

These RVs don't just come out of thin air… you have to specify them… accurately!

Why Worry? GIGO!

Warning: If you specify improper RVs, this can easily result in Garbage-In-Garbage-Out…

…which can ruin your entire model and invalidate any results you obtain!

GIGO Example

Consider a single-server queuing system.

Suppose that we have constant service times of 10 min.

What if the simulation assumes constant 12 min interarrivals, yet in reality they're exponential with mean 12?

Then the simulation never sees the line that actually occurs!

So What To Do?

High-Level Game Plan

- Collect data for analysis.
- Determine / estimate the underlying distribution (along with associated parameters), e.g., Nor(30,8).
- Conduct a statistical test to see if your distribution is "approximately" correct.

And this is what we'll try…

Summary

This Time: Discussed what’s coming up in this module on input data analysis.

Next Time: We’ll look at baby methods for identifying distributions.

Computer Simulation

Module 8: Input Data Analysis

Dave Goldsman, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

Identifying Distributions

Lesson Overview

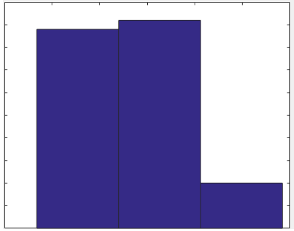
Last Module: Gave a brief intro to simulation input analysis and cautioned everyone to take it seriously.

This Lesson: Some high-level, baby methods for looking at data and guessing at distributions.

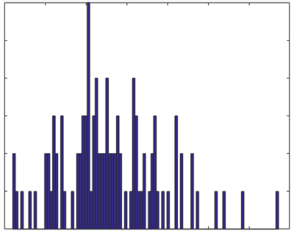
It's Weibull, and please change my diaper.

Three Little Bears

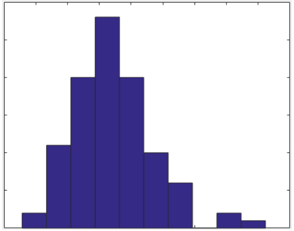
You can always present data in the form of a histogram and see if it's interesting…



Not enuf cells



Too many



Just right!

Fun Fact: If you take enough observations, the histogram will eventually converge to the true distribution.

Stem-and-Leaf

Turn the histogram on its side, and add some numerical information…

10	000
9	998764422110
8	9764544321
7	965432100
6	7532
5	433
4	8

Which Distribution?

How to even start?

- Discrete vs. Continuous?
- Univariate / Multivariate?
- How much data available?
- Are experts around to ask about nature of the data?
- What if you don't have much / any data – can you at least guess at a good distribution?

Which Distribution, II?

Discrete?

- Bernoulli(p) (success prob p)
- Binomial(n,p) (# successes in n Bern(p) trials)
- Geometric(p) (# Bern(p) trials until 1st success)
- Negative Binomial
- Poisson(λ) (counts # arrivals over time)
- Empirical ("sample" distribution)

Which Distribution, III?

Continuous?

- Uniform (not much known from the data, except maybe min and max possible values)
- Triangular (know min, max, “most likely” values)
- Exponential(λ) (interarrival times from a Poisson process)
- Normal (good model for heights, weights, IQs, sample means, etc.)
- Beta (good for specifying bounded data)
- Gamma, Weibull, Gumbel, lognormal (reliability data)
- Empirical (“sample” distribution)

Game Plan

Choose a “reasonable” distribution, and then test to see if it’s not too ridiculous.

Example: If you hypothesize that some data is normal, then the data should…

Fall approximately on a straight line when you do graph it on a normal probability plot. It should also pass goodness-of-fit tests for normality.

Summary

This Time: High-level discussion on the types of distributions that might be useful for modeling input random variables for simulations.

Next Time: Before we can conduct hypothesis tests to see if a certain distributions are any good, we have to estimate relevant parameters.

The next few lessons will be concerned with various kinds of estimators.

Computer Simulation

Module 8: Input Data Analysis

Dave Goldsman, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

Unbiased Point Estimation

Lesson Overview

Last Time: A high-level chat on useful input distributions.

This Time: Now let's narrow down the parameters of the distributions. We'll begin a series of lessons on point estimation. Cover unbiased estimation first.

Some of this stuff was covered in Module 2, but we'll do more now.

Definition: A *statistic* is a function of the observations $X_1, \ldots, X_n$, and not explicitly dependent on any unknown parameters.

Examples of statistics: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Statistics are *random variables*. If we take two different samples, we'd expect to get two different values of a statistic.

A statistic is usually used to estimate some unknown *parameter* from the underlying probability distribution of the X_i's.

Examples of parameters: μ, σ^2.

Let $X_1, \ldots, X_n$ be iid RV's and let $T(\mathbf{X}) \equiv T(X_1, \ldots, X_n)$ be a statistic based on the X_i's. Suppose we use $T(\mathbf{X})$ to estimate some unknown parameter θ. Then $T(\mathbf{X})$ is called a *point estimator* for θ.

Examples: $\bar{X}$ is usually a point estimator for the mean $\mu = \mathrm{E}[X_i]$, and S^2 is often a point estimator for the variance $\sigma^2 = \mathrm{Var}(X_i)$.

It would be nice if $T(\mathbf{X})$ had certain properties:

* Its expected value should equal the parameter it's trying to estimate.

* It should have low variance.

Definition: $T(\mathbf{X})$ is *unbiased* for θ if $\mathrm{E}[T(\mathbf{X})] = \theta$.

Example/Theorem: Suppose $X_1, \ldots, X_n$ are iid anything with mean μ. Then

$$\mathrm{E}[\bar{X}] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] = \mathrm{E}[X_i] = \mu.$$

So $\bar{X}$ is always unbiased for μ. That's why $\bar{X}$ is the *sample mean*.

Baby Example: In particular, suppose $X_1, \ldots, X_n$ are iid Exp(λ). Then $\bar{X}$ is unbiased for $\mu = \mathrm{E}[X_i] = 1/\lambda$.

But be careful.... $1/\bar{X}$ is *biased* for λ in this exponential case, i.e., $\mathrm{E}[1/\bar{X}] \neq 1/\mathrm{E}[\bar{X}] = \lambda$.

Example/Theorem: Suppose $X_1, \ldots, X_n$ are iid anything with mean μ and variance σ^2. Then

$$\mathrm{E}[S^2] = \mathrm{E}\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right] = \mathrm{Var}(X_i) = \sigma^2.$$

Thus, S^2 is always unbiased for σ^2. This is why S^2 is called the *sample variance.*

Baby Example: Suppose $X_1, \ldots, X_n$ are iid $\mathrm{Exp}(\lambda)$. Then S^2 is unbiased for $\mathrm{Var}(X_i) = 1/\lambda^2$.

Proof (of general result): First, some algebra gives

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

Since $\mathrm{E}[X_1] = \mathrm{E}[\bar{X}]$ and $\mathrm{Var}(\bar{X}) = \mathrm{Var}(X_1)/n = \sigma^2/n$, we have

$$\begin{aligned}
\mathrm{E}[S^2] &= \frac{\sum_{i=1}^n \mathrm{E}[X_i^2] - n\mathrm{E}[\bar{X}^2]}{n-1} = \frac{n}{n-1}\Big(\mathrm{E}[X_1^2] - \mathrm{E}[\bar{X}^2]\Big) \\
&= \frac{n}{n-1}\Big(\mathrm{Var}(X_1) + (\mathrm{E}[X_1])^2 - \mathrm{Var}(\bar{X}) - (\mathrm{E}[\bar{X}])^2\Big) \\
&= \frac{n}{n-1}(\sigma^2 - \sigma^2/n) = \sigma^2. \quad \square
\end{aligned}$$

Remark: S is *biased* for the standard deviation σ.

Summary

This Time: Began a series of lessons on point estimation. Covered unbiased estimation.

Next Time: Mean squared error, an estimator performance measure that combines bias and variance.

I've always thought that "nice squared error" is a much better name.

Computer Simulation

Module 8: Input Data Analysis

Dave Goldsman, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

Mean Squared Error

Lesson Overview

Last Time: Talked about unbiased point estimators, which aimed to get estimators with good expected values.

This Time: Mean squared error – which combines an estimator's bias and variance.

Bias and Variance

Good Estimator Properties:

- Low bias (difference between the estimator's expected value and the true parameter value)
- Low variance
- Actually need to have BOTH. 
  - Low bias + high var = bad (meaningless noisy estimator)
  - Low var + high bias = bad (confident of wrong answer)

Big Example: Suppose that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Unif}(0, \theta)$, i.e., the p.d.f. is $f(x) = 1/\theta$, $0 < x < \theta$.

Consider two estimators: $Y_1 \equiv 2\bar{X}$ and $Y_2 \equiv \frac{n+1}{n} \max_{1 \le i \le n} X_i$

Since $\mathrm{E}[Y_1] = 2\mathrm{E}[\bar{X}] = 2\mathrm{E}[X_i] = \theta$, we see that Y_1 is unbiased for θ.

It's also the case that Y_2 is unbiased, but it takes a little more work to show this. As a first step, let's get the cdf of $M \equiv \max_i X_i \ldots$

GTx

$$
\begin{aligned}
P(M \le y) &= P(X_1 \le y \text{ and } X_2 \le y \text{ and } \cdots \text{ and } X_n \le y) \\
&= \prod_{i=1}^{n} P(X_i \le y) = [P(X_1 \le y)]^n \quad (X_i\text{'s are iid}) \\
&= \left[\int_0^y f_{X_1}(x)\,dx\right]^n = \left[\int_0^y 1/\theta\,dx\right]^n = (y/\theta)^n.
\end{aligned}
$$

This implies that the p.d.f. of M is

$$
f_M(y) \equiv \frac{d}{dy}(y/\theta)^n = \frac{ny^{n-1}}{\theta^n}.
$$

Then

$$E[M] = \int_0^\theta y f_M(y)\,dy = \int_0^\theta \frac{ny^n}{\theta^n} = \frac{n\theta}{n+1}.$$

Whew! So we see that $Y_2 = \frac{n+1}{n} \max_{1 \le i \le n} X_i$ is unbiased for θ.

So both Y_1 and Y_2 are unbiased for θ, but which is better?

Let's now compare *variances*. After similar algebra, we have

$$\mathrm{Var}(Y_1) = \frac{\theta^2}{3n} \quad \text{and} \quad \mathrm{Var}(Y_2) = \frac{\theta^2}{n(n+2)}.$$

Thus, Y_2 has *much lower variance* than Y_1. $\square$

Definition: The *bias* of an estimator $T(\mathbf{X})$ is $\text{Bias}(T) \equiv \text{E}[T] - \theta$.

The *mean squared error* of $T(\mathbf{X})$ is $\text{MSE}(T) \equiv \text{E}[(T - \theta)^2]$.

Remark: After some algebra, we get an easier expression for MSE that combines the bias and variance of an estimator

$$\text{MSE}(T) = \text{Var}(T) + (\underbrace{\text{E}[T] - \theta}_{\text{Bias}})^2.$$

Lower MSE is better — even if there's a little bias.

Definition: The *relative efficiency* of T_2 to T_1 is $\mathrm{MSE}(T_1)/\mathrm{MSE}(T_2)$. If this quantity is < 1, then we'd want T_1.

Example: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Unif}(0, \theta)$.

Two estimators: $Y_1 = 2\bar{X}$ and $Y_2 = \frac{n+1}{n} \max_i X_i$.

Showed before $\mathrm{E}[Y_1] = \mathrm{E}[Y_2] = \theta$ (so both are unbiased).

Also, $\mathrm{Var}(Y_1) = \frac{\theta^2}{3n}$ and $\mathrm{Var}(Y_2) = \frac{\theta^2}{n(n+2)}$.

Thus, $\mathrm{MSE}(Y_1) = \frac{\theta^2}{3n}$ and $\mathrm{MSE}(Y_2) = \frac{\theta^2}{n(n+2)}$, so Y_2 is better.

Summary

We continued our Stats Attack with a discussion on MSE – an estimator performance measure combining bias and variance.

Next Time: We'll study maximum likelihood point estimators, which are often quite flexible, even if they're occasionally a little bit biased.

Computer Simulation
Module 8: Input Data Analysis

Dave Goldsman, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

Maximum Likelihood Estimators

Lesson Overview

Last Time: Continued our discussion of unbiased estimators and MSE.

This Time: Maximum likelihood estimation – perhaps the most popular point estimation method.

Very flexible technique that many software packages use to help estimate distributions.

Definition: Consider an iid random sample $X_1, \ldots, X_n$, where each X_i has pdf/pmf $f(x)$. Further, suppose that θ is some unknown parameter from X_i. The *likelihood function* is $L(\theta) \equiv \prod_{i=1}^{n} f(x_i)$.

Definition: The *maximum likelihood estimator* (MLE) of θ is the value of θ that maximizes $L(\theta)$. The MLE is a function of the X_i's and is a RV.

Example: Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$. Find the MLE for λ.

$$L(\lambda) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \exp\Big(-\lambda \sum_{i=1}^{n} x_i\Big).$$

Now maximize $L(\lambda)$ with respect to λ.

Could take the derivative and plow through all of the horrible algebra.

Useful Trick: Since the natural log function is one-to-one, it's easy to see that the λ that maximizes $L(\lambda)$ also maximizes $\ell\text{n}(L(\lambda))$!

$$L(\lambda) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n \exp\Big(-\lambda \sum_{i=1}^{n} x_i\Big).$$

$$\ell\mathrm{n}(L(\lambda)) = \ell\mathrm{n}\left(\lambda^n \exp\Big(-\lambda \sum_{i=1}^{n} x_i\Big)\right) = n\ell\mathrm{n}(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

This makes our job less horrible.

$$\frac{d}{d\lambda}\ell\mathrm{n}(L(\lambda)) = \frac{d}{d\lambda}\Big(n\ell\mathrm{n}(\lambda) - \lambda \sum_{i=1}^{n} x_i\Big) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \equiv 0.$$

This implies that the MLE is $\hat{\lambda} = 1/\bar{X}$. $\square$

Remarks: (1) $\hat{\lambda} = 1/\bar{X}$ makes sense since $\mathrm{E}[X] = 1/\lambda$.

(2) At the end, we put a little $\widehat{\text{hat}}$ over λ to indicate that this is the MLE.

(3) At the end, we make all of the little x_i's into big X_i's to indicate that this is a RV.

(4) Just to be careful, you probably ought to perform a second-derivative test, but I won't blame you if you don't.

And now, another example to get the blood flowing…

Example: Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$. Find the MLE for p.

Useful trick for this problem: Since

$$X_i = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases},$$

we can write the pmf as

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1.$$

Thus,

$$L(p) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

$\Rightarrow$

$$\ell\text{n}(L(p)) = \sum_{i=1}^{n} x_i\,\ell\text{n}(p) + (n - \sum_{i=1}^{n} x_i)\ell\text{n}(1-p)$$

$\Rightarrow$

$$\frac{d}{dp}\ell\text{n}(L(p)) = \frac{\sum_i x_i}{p} - \frac{n - \sum_i x_i}{1-p} \equiv 0.$$

Solving, we get $\hat{p} = \bar{X}$, which makes sense since $\mathrm{E}[X] = p$. $\square$

Summary

Went over some basics on Maximum Likelihood Estimators, along with a couple of easy examples.

Next Time: We'll do a bunch of additional, more-interesting examples involving MLEs.

Computer Simulation
Module 8: Input Data Analysis

Dave Goldsman, Ph.D.
Professor
Stewart School of Industrial and Systems Engineering

MLE Examples

Lesson Overview

Last Time: Defined MLEs and gave a couple of easy examples to start things off.

This Time: We'll dive a little deeper with some additional MLE examples.

This material will prove to be very useful when we eventually carry out goodness-of-fit tests.

Trickier MLE Examples

Example: Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$. Find the *simultaneous* MLE's for μ and σ^2.

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n f(x_i) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right\} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^2}\right\}. \end{aligned}$$

GTx

This $\Rightarrow$

$$\ell\text{n}(L(\mu,\sigma^2)) \;=\; -\frac{n}{2}\ell\text{n}(2\pi) - \frac{n}{2}\ell\text{n}(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Thus,

$$\frac{\partial}{\partial\mu}\ell\text{n}(L(\mu,\sigma^2)) \;=\; \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) \;\equiv\; 0,$$

and so $\hat{\mu} = \bar{X}$ (which makes sense).

Now do the same thing for σ^2...

Similarly, take the partial w/rt σ^2 (*not* σ),

$$\frac{\partial}{\partial \sigma^2} \ell\text{n}(L(\mu, \sigma^2)) \; = \; -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \hat{\mu})^2 \; \equiv \; 0,$$

and eventually get

$$\widehat{\sigma^2} \; = \; \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}. \quad \Box$$

Hmmm… Have we seen this before???

Remark: Notice how close $\widehat{\sigma^2}$ is to the (unbiased) sample variance

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

$\widehat{\sigma^2}$ is a little bit biased, but it has slightly less variance than S^2.

Anyway, as n gets big, S^2 and $\widehat{\sigma^2}$ become the same.

Example: The pdf of the Gamma distrn w/parameters r and λ is

$$f(x) \;=\; \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x > 0.$$

Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Gam}(r, \lambda)$. Find the MLE's for r and λ.

$$L(r, \lambda) \;=\; \prod_{i=1}^{n} f(x_i) \;=\; \frac{\lambda^{nr}}{[\Gamma(r)]^n} \left(\prod_{i=1}^{n} x_i \right)^{r-1} e^{-\lambda \sum_i x_i}.$$

This $\Rightarrow$

$$\ell\mathrm{n}(L) \; = \; rn\,\ell\mathrm{n}(\lambda) - n\,\ell\mathrm{n}(\Gamma(r)) + (r-1)\ell\mathrm{n}\Big(\prod_i x_i\Big) - \lambda\sum_i x_i.$$

Let's get the MLE of λ...

$$\frac{\partial}{\partial\lambda}\ell\mathrm{n}(L) \; = \; \frac{rn}{\lambda} - \sum_{i=1}^{n} x_i \; \equiv \; 0,$$

so that $\hat{\lambda} = \hat{r}/\bar{X}$.

The trouble is, we need to find $\hat{r}\ldots$

Similar to the above work, we get

$$\frac{\partial}{\partial r}\ell\mathrm{n}(L) \;=\; n\,\ell\mathrm{n}(\lambda) - \frac{n}{\Gamma(r)}\,\frac{d}{dr}\Gamma(r) + \ell\mathrm{n}\Big(\prod_i x_i\Big) \;\equiv\; 0.$$

Note that $\Psi(r) \equiv \Gamma'(r)/\Gamma(r)$ is the *digamma* function.

At this point, substitute in $\hat{\lambda} = \hat{r}/\bar{X}$, and use a *computer* (bisection, Newton's method, etc.) to search for the value of r that solves

$$n\,\ell\mathrm{n}(r/\bar{X}) - n\Psi(r) + \ell\mathrm{n}\Big(\prod_i x_i\Big) \;\equiv\; 0.$$

We actually did a bit of this in Module 2. If you try this yourself and can't find an expression for $\Gamma'(r)$, then you can choose your favorite small h and use

$$\Gamma'(r) \;\approx\; \frac{\Gamma(r+h) - \Gamma(r)}{h}.$$

GTx

Example: Suppose $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Unif}(0, \theta)$. Find the MLE for θ.

The pdf is $f(x) = 1/\theta, 0 < x < \theta$ (beware of the funny limits).

$$L(\theta) = \prod_{i=1}^{n} f(x_i) = \begin{cases} 1/\theta^n & \text{if } 0 \le x_i \le \theta, \forall i \\ 0 & \text{otherwise} \end{cases}$$

In order to have $L(\theta) > 0$, we must have $0 \leq x_i \leq \theta, \forall i$. In other words, we must have $\theta \geq \max_i x_i$.

Subject to this constraint, $L(\theta) = 1/\theta^n$ is maximized at the smallest possible θ value, namely, $\hat{\theta} = \max_i X_i$.

Remark: This makes sense in light of the similar (unbiased) estimator, $Y_2 = \frac{n+1}{n} \max_i X_i$, that we looked at previously.

Summary

We went over a number of (tricky) MLE examples, some of which took a little work, eh?

Next Time: We'll discuss the remarkable Invariance Property of MLEs. This is what will make our upcoming goodness-of-fit tests possible.