Input Analysis

Dave Goldsman

Georgia Institute of Technology, Atlanta, GA, USA

10/19/17

Outline

Intro to Input Analysis

You've made your flowcharts, and you have a pretty good idea of all of the processes that customers have to undergo as they move through the system. You've even programmed your model in your favorite simulation language.

But there's one little hurdle left — proper simulation *input analysis*.

What distributions do you use to model interarrival times, service times, breakdown times, etc.?

For instance, let's think about interarrival times. Here are some natural questions:

- Are the interarrival times exponential? Weibull? Stationary? Independent?
- What if you think they're constant but they're actually exponential? You may never see a line form! BAD model!
- What if you think the arrival rate stays constant over the day but it doesn't? BAD!
- What if you correctly model the interarrivals as exponential but get the arrival rate wrong? BAD!

Moral: Proper input analysis can save you from GIGO — Garbage-In-Garbage-Out!

So whatever shall we do?

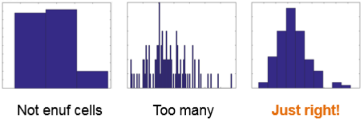
High-Level Game Plan

- 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.

Easy Data Analysis Ideas

You should always plot out your data before doing anything else. Example: *Histograms.*



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

Another Old Friend

Stem-and-leaf diagrams (which are sort of sideways histograms with numbers), etc.

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

So 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?

So Which Distribution, II?

Discrete?

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

So 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.

Outline

Intro to Estimation

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

Examples: $\bar{X}$, S^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.

Let $X_1, \ldots, X_n$ be i.i.d. 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.

Unbiased Estimators

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

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

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

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

Example/Theorem: Suppose $X_1, \ldots, X_n$ are i.i.d. 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*.

Proof: First, some algebra gives

$$
\begin{aligned}
S^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \\
&= \frac{\sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2)}{n-1} \\
&= \frac{\sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + n\bar{X}^2)}{n-1} \\
&= \frac{\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2)}{n-1} \\
&= \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1}
\end{aligned}
$$

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

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

Remark: S is *not* unbiased for the standard dev σ.

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

We'll look at three unbiased estimators:

$Y_1 = 2\bar{X}$

$Y_2 = \frac{n+1}{n} \max_{1 \le i \le n} X_i$

$$Y_3 = \begin{cases} 12\bar{X} & \text{w.p. } 1/2 \\ -8\bar{X} & \text{w.p. } 1/2 \end{cases}$$

<u>Easy Estimator</u>: $Y_1 = 2\bar{X}$.

Proof (it's unbiased): $\mathrm{E}[Y_1] = 2\mathrm{E}[\bar{X}] = 2\mathrm{E}[X_i] = \theta$.

<u>Harder Estimator</u>: $Y_2 = \frac{n+1}{n} \max_{1 \le i \le n} X_i$.

Why might this estimator for θ make sense?

Proof: $\mathrm{E}[Y_2] = \frac{n+1}{n}\mathrm{E}[\max_i X_i] = \theta$ iff

$$\mathrm{E}[\max X_i] \;=\; \frac{n\theta}{n+1} \quad \text{(what we'll show).}$$

First, let's get the c.d.f. of $M \equiv \max_i X_i$:

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

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

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

and this implies that

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

Whew! This finally shows that $Y_2 = \frac{n+1}{n} \max_{1\le i\le n} X_i$ is unbiased for θ.

Finally, let's look at…

Stupid Estimator:

$$Y_3 = \begin{cases} 12\bar{X} & \text{w.p. } 1/2 \\ -8\bar{X} & \text{w.p. } 1/2 \end{cases}$$

Ha! It's possible to get a *negative* estimate for θ, which is strange since $\theta > 0$!

Proof (it's unbiased):

$$\mathrm{E}[Y_3] = 12\mathrm{E}[\bar{X}] \cdot \frac{1}{2} - 8\mathrm{E}[\bar{X}] \cdot \frac{1}{2} = 2\mathrm{E}[\bar{X}] = \theta.$$

Usually, it's *good* for an estimator to be unbiased, but the "stupid" estimator Y_3 shows that unbiased estimators can sometimes be goofy.

Therefore, let's look at some other properties an estimator can have.

For instance, consider the *variance* of an estimator.

Big Example (cont'd): Again suppose that

$$X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \mathrm{U}(0, \theta).$$

Both $Y_1 = 2\bar{X}$ and $Y_2 = \frac{n+1}{n} M$ are unbiased for θ.

Let's find $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$.

$$\mathrm{Var}(Y_1) \;=\; 4\mathrm{Var}(\bar{X}) \;=\; \frac{4}{n} \cdot \mathrm{Var}(X_i) \;=\; \frac{4}{n} \cdot \frac{\theta^2}{12} \;=\; \frac{\theta^2}{3n}.$$

$$\begin{aligned}
\mathrm{Var}(Y_2) &= \left(\frac{n+1}{n}\right)^2 \mathrm{Var}(M) \\
&= \left(\frac{n+1}{n}\right)^2 \mathrm{E}[M^2] - \left(\frac{n+1}{n} \cdot \mathrm{E}[M]\right)^2 \\
&= \left(\frac{n+1}{n}\right)^2 \int_0^\theta \frac{ny^{n+1}}{\theta^n}\, dy - \theta^2 \\
&= \theta^2 \cdot \frac{(n+1)^2}{n(n+2)} - \theta^2 \;=\; \frac{\theta^2}{n(n+2)}.
\end{aligned}$$

Thus, both Y_1 and Y_2 are unbiased, but Y_2 has *much lower variance* than Y_1.

Mean Squared Error

Definition: The mean squared error of an estimator $T(\mathbf{X})$ of θ is

$$\mathrm{MSE}(T(\mathbf{X})) \equiv \mathrm{E}[(T(\mathbf{X}) - \theta)^2].$$

Before giving an easier interpretation of MSE, define

$$\mathrm{Bias}(T(\mathbf{X})) \equiv \mathrm{E}[T(\mathbf{X})] - \theta.$$

Remark: Easier interpretation of MSE.

$$
\begin{aligned}
\mathrm{MSE}(T(\mathbf{X})) &= \mathrm{E}[T^2] - 2\theta\mathrm{E}[T] + \theta^2 \\
&= \mathrm{E}[T^2] - (\mathrm{E}[T])^2 + (\mathrm{E}[T])^2 - 2\theta\mathrm{E}[T] + \theta^2 \\
&= \mathrm{Var}(T) + (\underbrace{\mathrm{E}[T] - \theta}_{\mathrm{Bias}})^2.
\end{aligned}
$$

So the MSE combines the bias and variance of an estimator.

The lower the MSE the better. If $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ are two estimators of θ, we'd usually prefer the one with the lower MSE — even if it happens to have higher bias.

Definition: The relative efficiency of $T_2(\mathbf{X})$ to $T_1(\mathbf{X})$ is $\mathrm{MSE}(T_1(\mathbf{X}))/\mathrm{MSE}(T_2(\mathbf{X}))$.

If this quantity is < 1, then we'd want $T_1(\mathbf{X})$.

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

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 estimators 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. $\square$

Maximum Likelihood Estimators

Definition: Consider an i.i.d. 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 $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda)$. 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. Too tedious. Need a trick. . . .

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 $\ell\mathrm{n}(L(\lambda))$!

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

This makes our job less horrible.

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

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 $\widehat{\mathrm{hat}}$ 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 $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Bern}(p)$. 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 p.m.f. as

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

Thus,

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

$\Rightarrow$

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

$\Rightarrow$

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

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

Trickier MLE Examples

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

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

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),

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

and eventually get

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

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 p.d.f. 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 $\Rightarrow$

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

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 $\hat{r}$…

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

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

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

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

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

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

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

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.

Theorem (Invariance Property): If $\hat{\theta}$ is the MLE of some parameter θ and $h(\cdot)$ is a one-to-one function, then $h(\hat{\theta})$ is the MLE of $h(\theta)$.

Example: Suppose $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Bern}(p)$.

Recall that the MLE of p is $\widehat{p} = \bar{X}$ (which also happens to be unbiased). If we consider the 1:1 function $h(\theta) = \theta^2$ (for $\theta > 0$), then the Invariance Property says that the MLE of p^2 is $\bar{X}^2$. $\square$

Remark: Recall that such a property does *not* hold for unbiasedness. E.g., $\mathrm{E}[S^2] = \sigma^2$ but $\mathrm{E}[\sqrt{S^2}] \neq \sigma$.

Example: Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Nor}(\mu, \sigma^2)$.

We saw that the MLE for σ^2 is $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. The good news is that we can still get the MLE for σ...

If we consider the 1:1 function $h(\theta) = +\sqrt{\theta}$, then the Invariance Property says that the MLE of σ is

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

Example: Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$. We saw that the MLE for λ is $\hat{\lambda} = 1/\bar{X}$.

Meanwhile, we define the *survival function* as

$$\bar{F}(x) \;=\; P(X > x) \;=\; 1 - F(x) \;=\; e^{-\lambda x}.$$

Then the Invariance Property says that the MLE of $\bar{F}(x)$ is

$$\widehat{\bar{F}(x)} \;=\; e^{-\hat{\lambda} x} \;=\; e^{-x/\bar{X}}. \quad \Box$$

This kind of thing is used all of the time in the actuarial sciences.

The Method of Moments

Recall: The kth *moment* of a random variable X is

$$\mathrm{E}[X^k] \;=\; \begin{cases} \sum_x x^k f(x) & \text{if } X \text{ is discrete} \\ \\ \int_{\mathbb{R}} x^k f(x)\,dx & \text{if } X \text{ is cts} \end{cases}$$

Definition: Suppose $X_1, \ldots, X_n$ are i.i.d. from p.m.f. / p.d.f. $f(x)$. Then the *method of moments* (MOM) estimator for $\mathrm{E}[X^k]$ is $m_k \equiv \frac{1}{n}\sum_{i=1}^n X_i^k$.

Examples:

The MOM estimator for $\mu = \mathrm{E}[X_i]$ is $m_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

The MOM estimator for $\mathrm{E}[X_i^2]$ is $m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

The MOM estimator for $\mathrm{Var}(X_i) = \mathrm{E}[X_i^2] - (\mathrm{E}[X_i])^2$ is

$$m_2 - m_1^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 \;=\; \frac{n-1}{n} S^2.$$

(Of course, it's also OK to use S^2.)

Example: Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Pois}(\lambda)$.

Since $\lambda = \text{E}[X_i]$, a MOM estimator for λ is $\bar{X}$.

But also note that $\lambda = \text{Var}(X_i)$, so another MOM estimator for λ is $\frac{n-1}{n} S^2$ (or plain old S^2).

Usually use the easier-looking estimator if you have a choice. $\square$

Example: Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$.

MOM estimators for μ and σ^2 are $\bar{X}$ and $\frac{n-1}{n} S^2$ (or S^2), respectively.

For this example, these estimators are the same as the MLEs. $\square$

Let's finish up with a less-trivial example. . .

Example: Suppose $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Beta}(a, b)$. The p.d.f. is

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1.$$

It turns out (after lots of algebra) that

$$\begin{aligned} \text{E}[X] &= \frac{a}{a+b} \quad \text{and} \\ \text{Var}(X) &= \frac{ab}{(a+b)^2(a+b+1)} = \frac{\text{E}[X]b}{(a+b)(a+b+1)}. \end{aligned}$$

Let's estimate a and b via MOM.

We have

$$\mathrm{E}[X] \;=\; \frac{a}{a+b} \;\Rightarrow\; a \;=\; \frac{b\mathrm{E}[X]}{1-\mathrm{E}[X]} \;\approx\; \frac{b\bar{X}}{1-\bar{X}}. \qquad (*)$$

Plug the following into the previous equation for $\mathrm{Var}(X)$: $\bar{X}$ for $\mathrm{E}[X]$, S^2 for $\mathrm{Var}(X)$, and $\frac{b\bar{X}}{1-\bar{X}}$ for a.

Then after lots of algebra, we can solve for b:

$$b \;\approx\; \frac{(1-\bar{X})^2\bar{X}}{S^2} - 1 + \bar{X}.$$

To finish up, plug back in $(*)$ to get the MOM estimator for a.

Example (Hayter): Suppose we take a bunch of observations from a Beta distribution and it turns out that $\bar{X} = 0.3007$ and $S^2 = 0.01966$.

Then (try it yourself and see!) the MOM estimators for a and b are 2.92 and 6.78, respectively. $\square$

Outline

Goodness-of-Fit Tests

At this point, we've guessed a reasonable distribution and then estimated the relevant parameters. Now let's conduct a formal test to see just how successful our toils have been.

In particular, test

$$H_0 : X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{ p.m.f. / p.d.f. } f(x)$$

at level of significance

$$\alpha \;\equiv\; P(\text{Reject } H_0 | H_0 \text{ true}) \;=\; P(\text{Type I error})$$

(usually, $\alpha = 0.05$ or 0.01).

High-level view of a goodness-of-fit test procedure:

1. Divide the domain of $f(x)$ into k sets, say, $A_1, A_2, \ldots, A_k$ (distinct points if X is discrete or intervals if X is continuous).
2. Tally the actual number of observations that fall in each set, say, O_i, $i = 1, 2, \ldots, k$. If $p_i \equiv P(X \in A_i)$, then $O_i \sim \text{Bin}(n, p_i)$.
3. Determine the expected number of observations that would fall in each set if H_0 were true, say, $E_i = \text{E}[O_i] = np_i$, $i = 1, 2, \ldots, k$.
4. Calculate a test statistic based on the differences between the E_i's and O_i's. The chi-squared g-o-f test statistic is

$$\chi_0^2 \equiv \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}.$$

5. A large value of χ_0^2 indicates a bad fit.

   We *reject* H_0 if $\chi_0^2 > \chi^2_{\alpha,k-1-s}$, where

   - s is the number of unknown parameters from $f(x)$ that have to be estimated. E.g., if $X \sim \text{Nor}(\mu, \sigma^2)$, then $s = 2$.
   - $\chi^2_{\alpha,\nu}$ is the $(1-\alpha)$ quantile of the χ^2_ν distribution, i.e., $P(\chi^2_\nu < \chi^2_{\alpha,\nu}) = 1 - \alpha$.

   If $\chi_0^2 \le \chi^2_{\alpha,k-1-s}$, we *fail to reject* H_0.

Remarks:

- Usual recommendation from baby stats class: For the χ^2 g-o-f test to work, pick k, n such that $E_i \geq 5$ and n at least 30.

- If the d.f. $\nu = k - 1 - s$ happens to be very big, then

$$\chi^2_{\alpha,\nu} \approx \nu \left[1 - \frac{2}{9\nu} + z_\alpha \sqrt{\frac{2}{9\nu}} \right]^3,$$

  where z_α is the appropriate standard normal quantile.

- Other g-o-f tests: Kolmogorov–Smirnov, Anderson–Darling, Shapiro–Wilk, etc.

Baby Example: Test H_0: X_i's are Unif(0,1). Suppose we have $n = 1000$ observations divided into $k = 5$ intervals.

interval	[0,0.2]	(0.2,0.4]	(0.4,0.6]	(0.6,0.8]	(0.8,1.0]
E_i	200	200	200	200	200
O_i	179	208	222	199	192

Turns out that $\chi_0^2 \equiv \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} = 5.27$.

Let's take $\alpha = 0.05$. There are no unknown parameters, so $s = 0$. Then $\chi^2_{\alpha,k-1-s} = \chi^2_{0.05,4} = 9.49$.

Since $\chi_0^2 \le \chi^2_{\alpha,k-1-s}$, we fail to reject H_0. $\square$

Discrete Example: The number of defects in printed circuit boards is hypothesized to follow a Geometric(p) distribution. A random sample of $n = 70$ printed boards has been collected, and the number of defects observed.

# defects	frequency
1	34
2	18
3	2
4	9
5	7
	70

We'll test H_0: The observations are Geom(p).

Start by estimating p via the MLE. The likelihood function is

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

$$\ell\mathrm{n}(L(p)) \;=\; \left(\sum_{i=1}^{n} x_i - n\right) \ell\mathrm{n}(1-p) + n\,\ell\mathrm{n}(p)$$

$$\frac{d\,\ell\mathrm{n}(L(p))}{dp} \;=\; \frac{-\sum_{i=1}^{n} x_i + n}{1-p} + \frac{n}{p} \;=\; 0.$$

Solving for p gives the MLE

$$\hat{p} \;=\; \frac{1}{\bar{X}} \;=\; \frac{70}{1(34)+2(18)+3(2)+4(9)+5(7)} \;=\; 0.476.$$

Let's get the g-o-f test statistic, χ_0^2. We'll make a little table, assuming $\hat{p} = 0.476$ is correct. By the Invariance Property of MLEs (this is why we learned it!), the expected number of boards having a certain value x is $E_x = nP(X = x) = n(1 - \hat{p})^{x-1}\hat{p}$. Note that I'll combine the entries in the last row (≥ 5) so the probabilities add up to one.

x	$P(X = x)$	E_x	O_x
1	0.4762	33.33	34
2	0.2494	17.46	18
3	0.1307	9.15	2
4	0.0684	4.79	9
≥ 5	0.0752	5.27	7
	1.0000	70	70

Well, we really ought to combine the last two cells too, since $E_4 = 4.79 < 5$. Let's do so to get the following improved table.

x	$P(X = x)$	E_x	O_x
1	0.4762	33.33	34
2	0.2494	17.46	18
3	0.1307	9.15	2
≥ 4	0.1436	10.06	16
	1.0000	70	70

Thus, the test statistic is

$$\chi_0^2 \;=\; \sum_{x=1}^{4} \frac{(E_x - O_x)^2}{E_x} \;=\; \frac{(33.33 - 34)^2}{33.33} + \cdots \;=\; 9.12.$$

Let $k = 4$ denote the number of cells (that we ultimately ended up with), and let $s = 1$ denote the number of parameters we had to estimate.

Suppose the level $\alpha = 0.05$. Then we compare χ^2_0 against $\chi^2_{\alpha,k-1-s} = \chi^2_{0.05,2} = 5.99$.

Since $\chi^2_0 > \chi^2_{\alpha,k-1-s}$, we reject H_0. This means that the number of defects probably isn't geometric. $\square$

Continuous Distributions: For the continuous case, let's denote the intervals $A_i \equiv (a_{i-1}, a_i]$, $i = 1, 2, \ldots, k$. For convenience, we choose the a_i's to ensure that we have equal-probability intervals, i.e.,

$$p_i \;=\; P(X \in A_i) \;=\; P(a_{i-1} < X \leq a_i) \;=\; 1/k \quad \text{for all } i.$$

In this case, we immediately have $E_i = n/k$ for all i; and then

$$\chi_0^2 \;=\; \sum_{i=1}^{k} \frac{(O_i - (n/k))^2}{n/k}.$$

The issue is that the a_i's might depend on unknown parameters.

Example: Suppose that we're interested in fitting a distribution to a series of interarrival times. Could they be *Exponential*?

$$H_0 : X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda).$$

Let's do a χ^2 g-o-f test with *equal-probability intervals*. This amounts to choosing a_i's such that

$$F(a_i) \; = \; P(X \leq a_i) \; = \; 1 - e^{-\lambda a_i} \; = \; \frac{i}{k}, \quad i = 1, 2, \ldots, k.$$

That is,

$$a_i \; = \; -\frac{1}{\lambda} \ell\text{n}\Big(1 - \frac{i}{k}\Big), \quad i = 1, 2, \ldots, k.$$

Great, but λ is unknown (so we'll estimate $s = 1$ parameter).

Good News: We know that the MLE is $\widehat{\lambda} = 1/\bar{X}$. Thus, by the Invariance Property, the MLEs of the a_i's are

$$\widehat{a}_i \;=\; -\frac{1}{\widehat{\lambda}}\,\ell\mathrm{n}\Big(1 - \frac{i}{k}\Big) \;=\; -\bar{X}\,\ell\mathrm{n}\Big(1 - \frac{i}{k}\Big), \quad i = 1, 2, \ldots, k.$$

Continue Example: Suppose that we take $n = 100$ observations and divide them into $k = 5$ equal-prob intervals. Further suppose that sample mean based on the 100 observations is $\bar{X} = 9.0$.

$$\widehat{a}_i \;=\; -9.0\,\ell\mathrm{n}\Big(1 - 0.2i\Big), \quad i = 1, \ldots, 5.$$

Suppose we determine which interval each of the 100 observations belongs to and tally them up to get the O_i's.

interval $(\hat{a}_{i-1}, \hat{a}_i]$	O_i	$E_i = \frac{n}{k}$
(0, 2.01]	25	20
(2.01, 4.60]	27	20
(4.60, 8.25]	23	20
(8.25, 14.48]	13	20
(14.48, ∞)	12	20
	100	100

$$\chi_0^2 \;=\; \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \;=\; 9.80 \quad \text{and}$$

$$\chi^2_{\alpha,k-1-s} \;=\; \chi^2_{0.05,3} \;=\; 7.81$$

So we reject H_0. These observations ain't Expo. $\square$

Example: Now let's make things more interesting.

$$H_0 : X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Weibull}(r, \lambda).$$

The Weibull has c.d.f. $F(x) = 1 - \exp\big[-(\lambda x)^r\big]$, for $x \geq 0$. Note that $r = 1$ yields the Exp(λ) as a special case.

Again do a χ^2 g-o-f test with *equal-probability intervals*. We need

$$F(a_i) \;=\; 1 - e^{-(\lambda a_i)^r} \;=\; i/k \quad \Rightarrow$$

$$a_i \;=\; \frac{1}{\lambda}\left[-\ell\text{n}\left(1 - \frac{i}{k}\right)\right]^{1/r}, \quad i = 1, 2, \ldots, k.$$

Now λ and r are unknown, so we'll get $s = 2$ MLEs.

After a little algebra (a couple of chain rules), the p.d.f. is

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

So the likelihood function for an i.i.d. sample of size n is

$$L(r, \lambda) \;=\; \prod_{i=1}^{n} f(x_i) \;=\; \lambda^{nr} r^n \prod_{i=1}^{n} x_i^{r-1} \exp\Big[-\lambda^r \sum_{i=1}^{n} x_i^r \Big].$$

$$\ell\text{n}(L) \;=\; n\,\ell\text{n}(r) + (r-1)\,\ell\text{n}\Big(\prod_{i=1}^{n} x_i \Big) + nr\,\ell\text{n}(\lambda) - \lambda^r \sum_{i=1}^{n} x_i^r.$$

Now maximize with respect to r and λ by setting

$$\frac{\partial}{\partial r}\,\ell\mathrm{n}(L) \;=\; 0 \quad \text{and} \quad \frac{\partial}{\partial \lambda}\,\ell\mathrm{n}(L) \;=\; 0.$$

After tons of algebra, get

$$\lambda \;=\; \left(\sum_{i=1}^{n} x_i^r\right)^{-1/r} \qquad \text{and}$$

$$g(r) \;=\; \frac{n}{r} + \sum_{i=1}^{n} \ell\mathrm{n}(x_i) - \frac{n\sum_i x_i^r\,\ell\mathrm{n}(x_i)}{\sum_i x_i^r} \;=\; 0.$$

The equation for λ looks easy enough, if only we could solve for r!

Recall: How to solve for a zero.

- trial-and-error — blech.
- bisection — OK.
- Newton's method — let's try it here!

To use Newton, we need…

$$g'(r) \;=\; -\frac{n}{r^2} - \frac{n\sum_i x_i^r\,[\ell\text{n}(x_i)]^2}{\sum_i x_i^r} + \frac{n\big[\sum_i x_i^r\,\ell\text{n}(x_i)\big]^2}{\big[\sum_i x_i^r\big]^2}.$$

Here's a reasonable implementation of Newton.

1. Initialize $\hat{r}_0 = \bar{X}/S$, where $\bar{X}$ is the sample mean and S^2 is the sample variance.
2. Update
$$\hat{r}_j \;\leftarrow\; \hat{r}_{j-1} - \frac{g(\hat{r}_{j-1})}{g'(\hat{r}_{j-1})}.$$
3. If $|g(\hat{r}_{j-1})| < 0.001$, then stop and set the MLE $\hat{r} = \hat{r}_j$. Otherwise, let $j \leftarrow j + 1$ and goto Step 2.

Newton usually converges pretty quickly — maybe after 3 or 4 iterations.

Given that we have the MLE $\hat{r}$, we immediately have the MLE for λ,

$$\widehat{\lambda} \;=\; \left(\sum_{i=1}^{n} x_i^{\hat{r}}\right)^{-1/\hat{r}}.$$

Then by invariance, we (finally) have the MLEs for the equal-probability points,

$$\hat{a}_i \;=\; \frac{1}{\widehat{\lambda}}\left[-\ell\mathrm{n}\Big(1-\frac{i}{k}\Big)\right]^{1/\hat{r}}, \quad i = 1, 2, \ldots, k.$$

Example (from BCNN): Suppose that we take $n = 50$ observations and divide them into $k = 8$ equal-prob intervals.

Moreover, suppose it turns out that $\hat{r} = 0.525$ and $\widehat{\lambda} = 0.161$.

Thus,

$$\hat{a}_i \;=\; 6.23 \left[-\ell\mathrm{n}\Big(1 - \frac{i}{8}\Big)\right]^{1.905}, \quad i = 1, 2, \ldots, k.$$

Further suppose that we get the following O_i's.

$(\hat{a}_{i-1}, \hat{a}_i]$	O_i	$E_i = \frac{n}{k}$
(0, 0.134]	6	6.25
(0.134, 0.578]	5	6.25
$\vdots$		
(11.54, 24.97]	5	6.25
(24.97, ∞)	6	6.25
	50	50

$$\chi_0^2 \;=\; \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \;=\; 1.20 \quad \text{and}$$

$$\chi^2_{\alpha,k-1-s} \;=\; \chi^2_{0.05,5} \;=\; 11.1$$

So we "accept" H_0. These observations are sort of Weibull. $\square$

Kolmogorov-Smirnov Goodness-of-Fit Test

There are plenty of g-o-f tests that you can use instead of a χ^2 test. The K-S test is one that works well in low-data situations.

We'll test

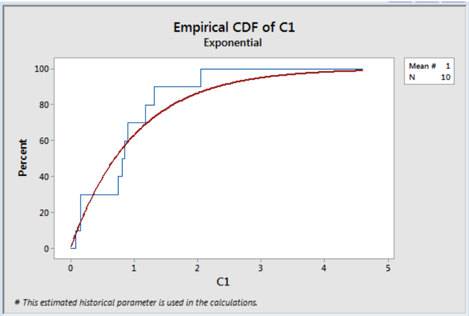
$$H_0 : X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{some distribution with c.d.f. } F(x).$$

Recall the definition of the *empirical c.d.f.* (also called the *sample c.d.f.*) of the data, $X_1, X_2, \ldots, X_n$,

$$\hat{F}_n(x) \equiv \frac{\text{number of } X_i\text{'s} \le x}{n}.$$

Note that $\hat{F}_n(x)$ is a step function with jumps of height $1/n$ (every time an observation occurs).

For example, here's the empirical c.d.f. of 10 Exp(1) observations that I generated (along with the Exp(1) c.d.f.).



The *Glivenko-Cantelli Lemma* says that $\hat{F}_n(x) \to F(x)$ for all x as $n \to \infty$. So if H_0 is true, then $\hat{F}_n(x)$ should be a good approximation to the true c.d.f., $F(x)$, for large n.

The main question: Does the empirical distribution actually support the assumption that H_0 is true?

The K-S test rejects H_0 if

$$D \equiv \max_{x \in \mathbb{R}} |F(x) - \hat{F}_n(x)| > D_{\alpha,n},$$

where α is the level of significance, and $D_{\alpha,n}$ is a K-S quantile that depends on the hypothesized c.d.f. $F(x)$.

Baby K-S Example: Let's test

$$H_0 : X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Unif}(0,1).$$

Under the Unif(0,1) assumption for $F(x)$, the K-S statistic simplifies:

$$D \;\equiv\; \max_{x \in \mathbb{R}} |F(x) - \hat{F}_n(x)| \;=\; \max_{0 \le x \le 1} |x - \hat{F}_n(x)|.$$

It's kinda easy to see that the max can only occur when x equals one of the observations, $X_1, X_2, \ldots, X_n$, i.e., at one of the jump points of $\hat{F}_n(x)$. In fact, at the ith jump point, $\hat{F}_n(x)$ increases from $\frac{i-1}{n}$ to $\frac{i}{n}$.

Before giving an easy algorithm to calculate D, let's first define the *ordered* points, $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. For example, if $X_1 = 4$, $X_2 = 1$, and $X_3 = 6$, then $X_{(1)} = 1$, $X_{(2)} = 4$, and $X_{(3)} = 6$.

Then we compute

$$D^+ \equiv \max_{1 \le i \le n} \left\{ \frac{i}{n} - X_{(i)} \right\}, \quad D^- \equiv \max_{1 \le i \le n} \left\{ X_{(i)} - \frac{i-1}{n} \right\},$$

and finally it turns out that $D = \max(D^+, D^-)$.

Numerical Example (from BCNN):

X_i	0.039	0.706	0.016	0.198	0.793
$X_{(i)}$	0.016	0.039	0.198	0.706	0.793
$\frac{i}{n}$	0.2	0.4	0.6	0.8	1.0
$\frac{i-1}{n}$	0	0.2	0.4	0.6	0.8
$\frac{i}{n} - X_{(i)}$	0.184	0.361	**0.402**	0.094	0.207
$X_{(i)} - \frac{i-1}{n}$	0.016	–	–	**0.106**	–

Thus, $D^+ = 0.402$, $D^- = 0.106$, and then $D = 0.402$.
If we go to a K-S table for the uniform distribution, we have $D_{\alpha,n} = D_{0.05,5} = 0.565$. So we fail to reject uniformity. $\square$

Remarks: K-S is conservative in the sense that it takes a lot of bad news to reject H_0.

Can easily apply K-S to other distributions.

Many other g-o-f tests: Anderson-Darling, Cramér-von Mises, Shapiro-Wilk, etc.

S-W is especially appropriate for testing normality. Can also use graphical techniques such as Q–Q plots to evaluate normality.

Outline

Nobody likes to talk about them, but every family has them. You'd think that after all of the theory we've done, we could always find good distributions to fit our data. Not exactly. Here are some cases that you have to be careful about.

1. No / little data
2. Data that doesn't look like one of the usual distributions
3. Nonstationary data (from distributions than change over time)
4. Multivariate / correlated data

1. No / Little Data

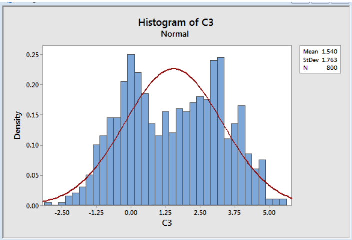
This issue turns up more often than you would expect. There could literally be no data available, or the data that you have is awful (goofy values, not cleaned properly, etc.). What to do? No great options — but here are some suggestions.

- Interview so-called "experts"
  - Try to at least get minimum, maximum, and "most likely" distribution values out of them — then you can guess uniform or triangular distributions.
  - Getting quantiles from the expert even better.
  - At least discuss the nature of the observations.

- If you have some idea about the nature of the RVs, maybe you can make a good guess as to the distribution.
  - Discrete or continuous?
  - Are observations successes / failures? Then think Bernoulli, binomial, geometric, negative binomial.
  - Do observations adhere to Poisson assumptions? Then Poisson (if you're counting arrivals) or exponential (interarrival times).
  - Are observations averages or sums? Then maybe normal.
  - Are observations bounded? Then think beta.
  - Reliability or job times? Maybe gamma, Weibull, lognormal, etc.
  - Can you think of anything else from the physical characteristics underlying the RV?

2. Nonstandard / Goofy / Mixture Distributions

Here's a forced marriage of two normals — most packages can't pick this up or fit it properly. Example: Poorly designed 6644 exam has two modes!



- Can attempt to model as a *mixture* of reasonable distributions.
- Easier: Can sample from the empirical distribution or a smoothed version of the empirical. This is a form of *bootstrapping*.

3. Nonstationary Data

Arrival rates change over time — think restaurants, traffic on the highway, call center activity, seasonal demands for a product. You *must* take this variability into account, else GIGO!

- Suggestion: Nonhomogeneous Poisson process. (Recall from RV Generation module.)
- Need to model rate function properly.
- Arena uses piecewise-constant rate function; so specify a constant arrival rate for each separate period.

4. Multivariate / Correlated Data

Data don't have to be i.i.d.! What if data is multivariate and / or serially correlated in time?

Examples:

- Multivariate RV — A person's height and weight are correlated.
- Serially correlated examples:
  - Monthly unemployment rates.
  - Arrivals to a social media site may be correlated if an interesting item appears there and the public gets wind of it.
  - A badly damaged part may require more service than usual at a series of stations.
  - If a server gets tired, his subsequent service times may be longer than usual.

So what do you need to do?

- Identify multivariate / serial correlation situations.
- Propose appropriate models. Examples:
  - Multivariate normal for heights and weights.
  - Time series models for serially correlated observations, e.g., autoregressive-moving average ARMA(p, q), EAR(1), ARTOP, which we discussed back in Module 7.
- Estimate relevant parameters (easier said than done). Examples:
  - Multivariate normal: Marginal means and variances (no big deal) plus covariances (maybe not so easy).
  - Time series: For simple models like the AR(1), estimating the coefficient ϕ is easy (just like covariance). But coefficients for more-complicated models need to be estimated using available software, e.g., Box-Jenkins technology.
- Validate to see if your estimated model is actually any good.
- Alternative: Can bootstrap samples from an empirical distribution (if you have enough data).

Demo Time!

Arena has very nice functionality that automatically fits simple distributions to your data. Just go to `Tools > Input Analyzer`.

The Input Analyzer gives you the “best” distribution from its library, along with relevant sample and goodness-of-fit statistics

ExpertFit is a specialty product that does distribution fitting with a larger library of distributions.

Minitab and R have distribution fitting functionality, though not quite convenient as the above tools.