Estimating population parameters assuming next to nothing

The CLT tells us that sample means are normal:

$$\bar{X}_n \stackrel{\bullet}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right).$$

Another type of of 'CLT' for quantiles gives for continuous X

$$\widehat{X}_p \stackrel{\bullet}{\sim} N\left(x_p, \frac{p(1-p)}{f(x_p)^2 n}\right).$$

These hold for *any* data X_i with density $f(x)$.

The kernel and histogram estimates $\hat{f}(x)$ and empirical cdf $\hat{F}(x)$ also have approximate normal distributions.

However, we can do better than this if we are willing to assume X_i comes from one of the probability models described in Chapter 2.

Chapter 8: Estimating parameters defining a parametric family of distributions

8.1: Introduction

- Most classes of pdf's or pmf's have only a few parameters. For example:

$$\text{Poisson}(\lambda),\ N(\mu,\sigma^2),\ \text{gamma}(\alpha,\lambda),\ \text{bin}(n,\pi),\ \text{geom}(\pi),\ \text{etc.}$$

- This chapter discusses two methods for estimating parameters given data

$$X_1, X_2, \ldots, X_n,$$

  method of moments and *maximum likelihood*.

- After parameters are estimated, the fitted model should be compared to the actual data to see if the fit is okay.

8.3: Parameter estimation

We will read through Examples A and B in class.

In Example A, the assumption is

$$X_1, \ldots, X_{49152} \overset{iid}{\sim} N(\mu, \sigma^2),$$

where $\boldsymbol{\theta} = (\mu, \sigma^2)$ are unknown.

In Example B, the assumption is

$$X_1, \ldots, X_{n_1} \overset{iid}{\sim} \text{gamma}(\alpha_1, \lambda_1),$$

independent of

$$Y_1, \ldots, Y_{n_2} \overset{iid}{\sim} \text{gamma}(\alpha_2, \lambda_2),$$

where $\boldsymbol{\theta} = (\alpha_1, \lambda_1, \alpha_2, \lambda_2)$ are unknown. Here we might want to test $H_0 : \alpha_1 = \alpha_2, \lambda_1 = \lambda_2$, that the two populations are the same.

For our data $\mathbf{X} = (X_1, \ldots, X_n)$, we will initially assume an *iid* probability model

$$X_1, \ldots, X_n \stackrel{iid}{\sim} f(x|\boldsymbol{\theta}),$$

where here $f(x|\boldsymbol{\theta})$ is the density for each X_i sampled, indexed by $\boldsymbol{\theta}$.

The vector $\boldsymbol{\theta}$ may be 1-dimensional (e.g. Bernoulli, Poisson, Geometric, exponential data), two-dimensional (e.g. normal, Weibull, gamma, beta data), or higher dimensional. Later on, we will consider non-*iid* data; e.g. in Example B, $\boldsymbol{\theta} = (\alpha_1, \lambda_1, \alpha_2, \lambda_2)$ is 4-dimensional.

Since $\mathbf{X} = (X_1, \ldots, X_n)$ is random, so is any function $g(\mathbf{X}) = g(X_1, \ldots, X_n)$. We will now define functions $g(\cdot)$ that estimate the unknown parameters in $\boldsymbol{\theta}$. We will often denote such functions of data by the parameter they estimate wearing a hat.

For example, for

$$X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2),$$

we will show that

$$\hat{\mu} = \hat{\mu}(X_1, \ldots, X_n) = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$\hat{\sigma}^2 = \hat{\sigma}^2(X_1, \ldots, X_n) = S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

are good estimators of μ and σ^2.

Note that both estimators are functions of the data $(X_1, \ldots, X_n)$ only, but their distributions will involve the unknown (μ, σ^2).

In fact, $\bar{X}_n \sim N(\mu, \sigma^2/n)$ and $\hat{\sigma}^2 \sim \text{gamma}\left(\frac{n-1}{2}, \frac{n}{2\sigma^2}\right)$.

def'n: A function $g(X_1, \ldots, X_n)$ of data $X_1, \ldots, X_n$ is called a *statistic*.

Both $\hat{\mu} = \bar{X}_n$ and $\hat{\sigma}^2$ are statistics commonly called the 'sample mean' and 'sample variance'.

8.4 Method of moments

Recall that $\bar{X}_n \xrightarrow{P} \mu = E(X_i)$ for *iid* data. It immediately follows that

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k = E(X_i^k),$$

for any $k = 1, 2, 3, \ldots$.

The method of moments (MOM) estimators make use of this fact my simply matching each sample moment $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ with whatever $E(X_i^k)$ ends up being as a function of $\boldsymbol{\theta}$, for $k = 1, 2, \ldots, p$ where p is the number of elements in $\boldsymbol{\theta}$. Solving the resulting system of equations for each element of $\boldsymbol{\theta}$ yields the MOM estimators.

Example: Let $X_1, \ldots, X_n \stackrel{iid}{\sim} \text{gamma}(\alpha, \lambda)$. Here there are $p = 2$ elements of $\boldsymbol{\theta} = (\alpha, \lambda)$. Then

$$E(X_i) = \alpha/\lambda \text{ and } E(X_i^2) = \frac{\alpha^2 + \alpha}{\lambda^2}.$$

We'll show this last part on the board. Setting the first $p = 2$ sample moments equal to the corresponding population moments yields

$$\begin{aligned} \hat{\mu}_1 &= \alpha/\lambda \\ \hat{\mu}_2 &= (\alpha^2 + \alpha)/\lambda^2 \end{aligned}$$

and solving gives us

$$\begin{aligned} \hat{\alpha} &= \hat{\mu}_1^2/(\hat{\mu}_2 - \hat{\mu}_1^2) \\ \hat{\lambda} &= \hat{\mu}_1/(\hat{\mu}_2 - \hat{\mu}_1^2) \end{aligned}$$

These are the MOM estimators of α and λ.

Example: Random sample if $n = 10$ *iid* data, assuming gamma distribution.

Say we collect data:
$x_1 = 21.4$, $x_2 = 87.3$, $x_3 = 37.6$, $x_4 = 54.1$, $x_5 = 42.7$, $x_6 = 35.3$, $x_7 = 35.7$, $x_8 = 21.7$, $x_9 = 41.1$, and $x_{10} = 34.6$.

Then we observe $\hat{\mu}_1 = \frac{1}{10}\sum_{i=1}^{10} x_i = 41.16$ and $\hat{\mu}_2 = \frac{1}{10}\sum_{i=1}^{10} x_i^2 = 2012.7$. Plugging in we get

$$\begin{aligned}\hat{\alpha} &= 41.16^2/(2012.7 - 41.16^2) = 5.31 \\ \hat{\lambda} &= 41.16/(2012.7 - 41.16^2) = 0.129\end{aligned}$$

So our estimate of the distribution that generated the $n = 10$ data values is $\text{gamma}(5.31, 0.129)$. In fact, the data were generated from a $\text{gamma}(5, 0.1)$ distribution.

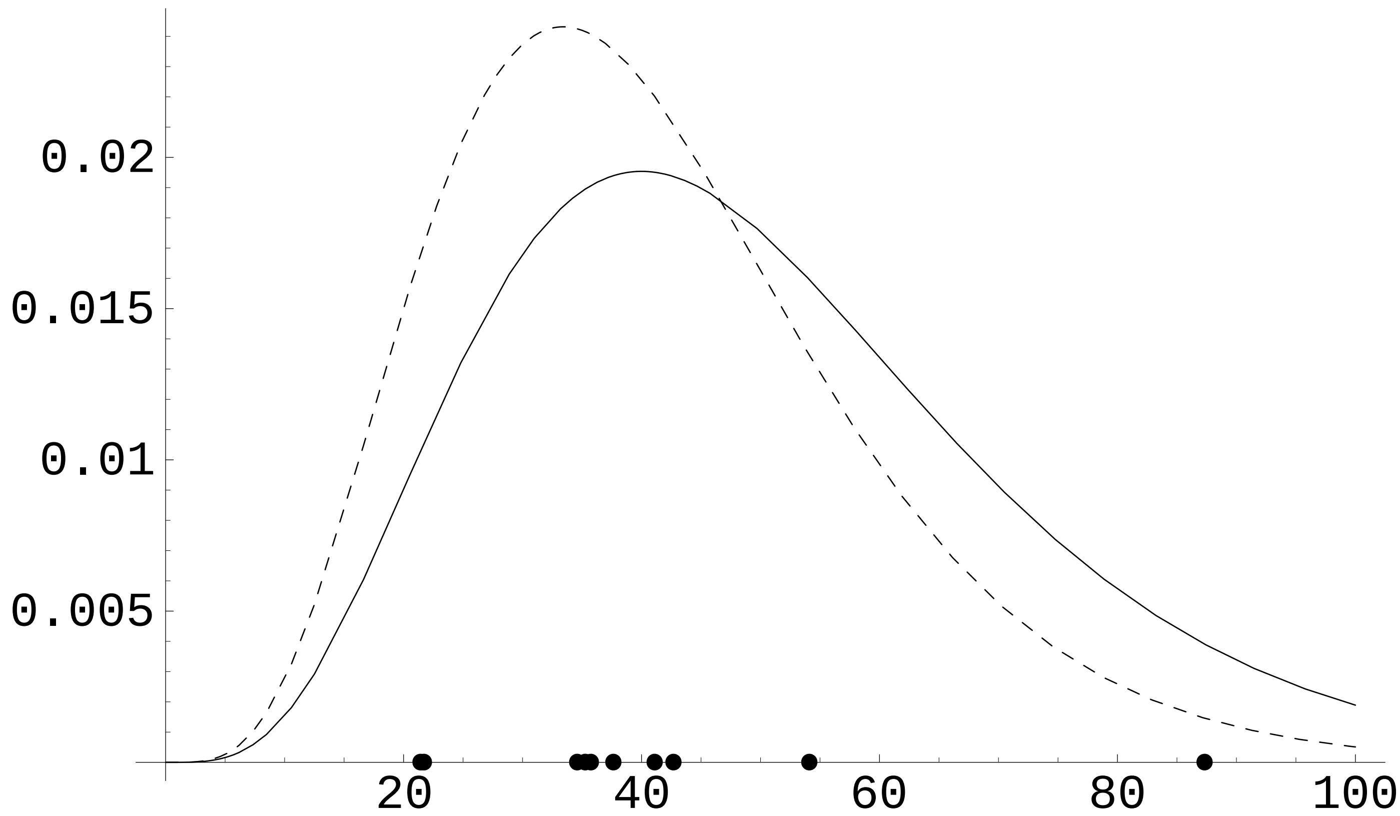

Figure 1: MOM estimated (dashed) and actual (solid) data generating distributions, $x_1, \ldots, x_{10}$ along x-axis.

Important! Before we collect the data, $\mathbf{X} = (X_1, \ldots, X_{10})$ are random, and then so must the estimators $\hat{\alpha}$ and $\hat{\lambda}$ as they are functions of $\mathbf{X}$. After we see the data $X_1 = x_1, \ldots, X_{10} = x_{10}$, the *estimators* become *estimates* and are no longer random.

We estimated the distribution of our data assuming they were gamma. What if we didn't know how the data were distributed and we instead assumed it was $U(a, b)$? Recall that for $X_i \sim U(a, b)$ that $E(X_i) = (a + b)/2$ and $\mathrm{Var}(X_i) = (b - a)^2/12$. We will partly show on the board that the MOM estimators are

$$\begin{aligned} \hat{a} &= \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} \\ \hat{b} &= \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} \end{aligned}$$

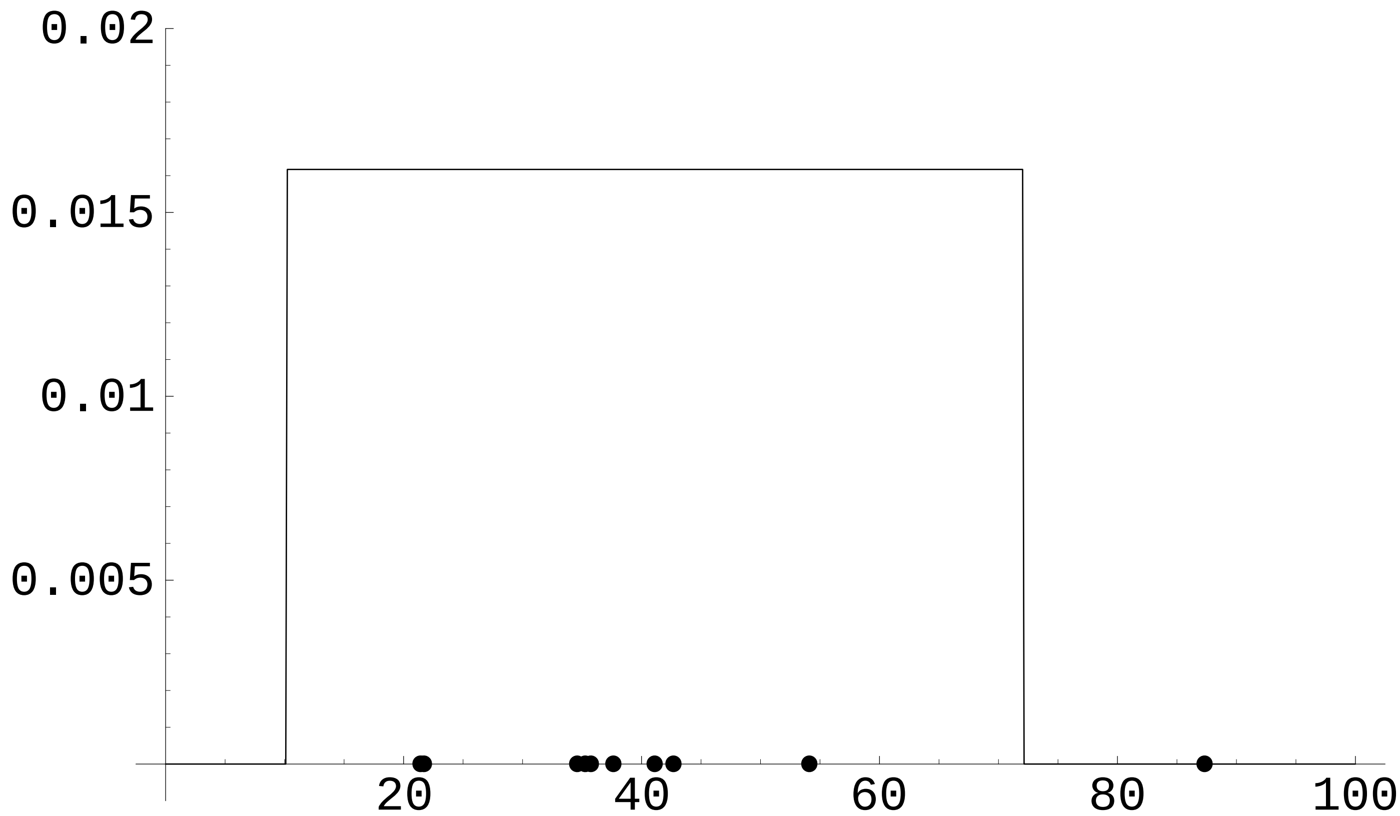

Figure 2: MOM estimated (dashed) assuming $U(a, b)$ distribution, $x_1, \ldots, x_{10}$ along x-axis. $\hat{a} = 10.2$ & $\hat{b} = 72.1$. Any problems here?

The survival times in weeks of $n = 33$ patients who died of acute myelogenous leukemia are 156, 65, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 5, 1, 1, 65, 56, 65, 17, 7, 16, 22, 3, 4, 2, 3, 8, 4, 3, 30, 4, 43. These are $x_1, \ldots, x_{33}$. A histogram of these data shows an approximate exponential density shape.

Let

$$X_1, \ldots, X_n \overset{iid}{\sim} \exp(\lambda).$$

Here, there is only one parameter, so to find the MOM estimator we set

$$\hat{\mu}_1 = \bar{X}_n = E(X_i) = \mu_1 = 1/\lambda,$$

giving us $\hat{\lambda} = 1/\bar{X}_n$.

For our data, $\bar{x}_{33} = (156 + 65 + \cdots + 4 + 43)/33 = 40.9$ and so $\hat{\lambda} = 1/40.9 = 0.0244$.

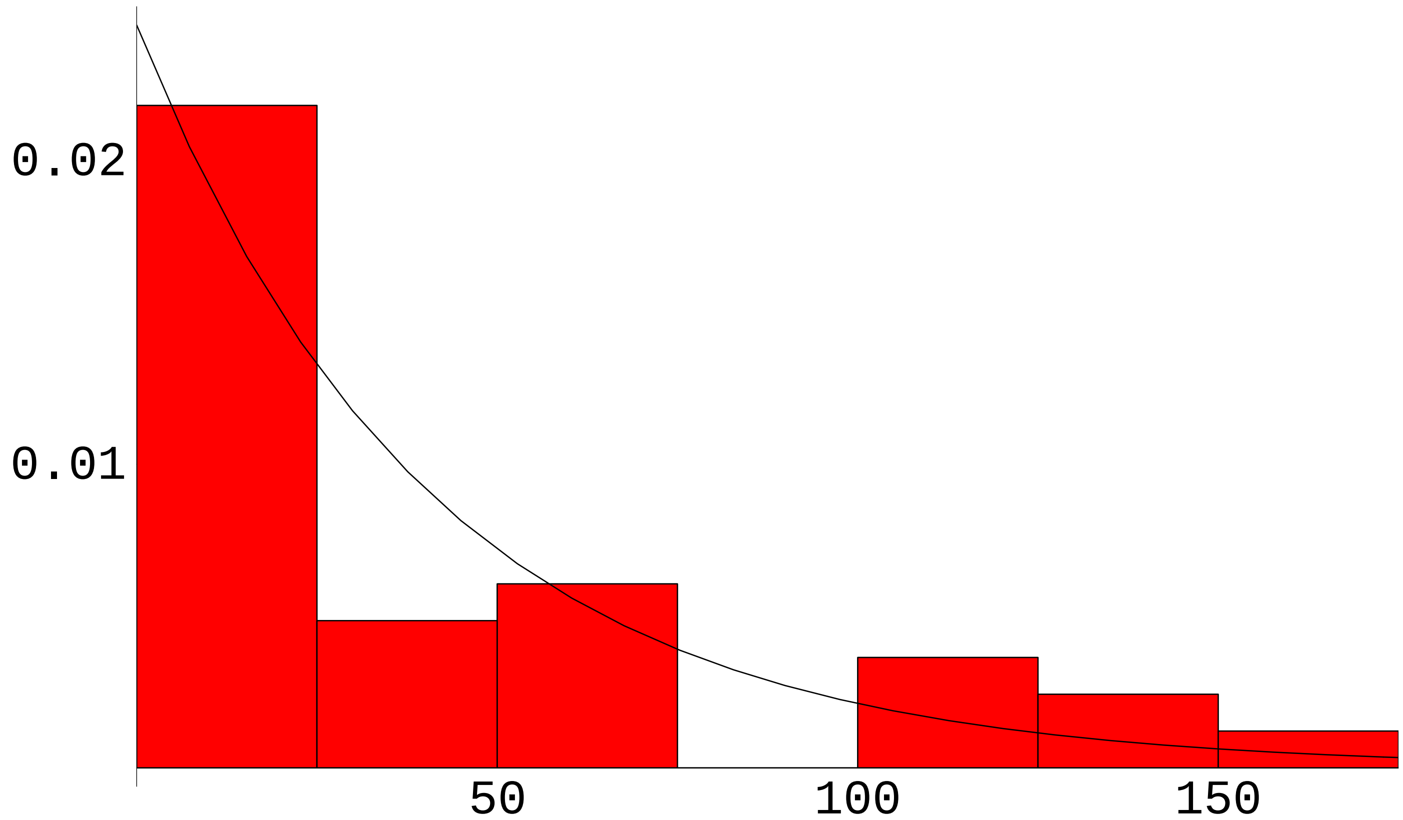

Figure 3: MOM estimated (solid line) assuming $\exp(\lambda)$ distribution, along with histogram of data $x_1, \ldots, x_{33}$.

Problems: find the MOM estimators for Poisson(λ), exp(λ), $N(\mu, \sigma^2)$, and Bern(π) data. You will find MOM estimators for two special beta distributions and the geometric distribution in your homework.

Data analysis problems...