Maximum Likelihood II

STAT 24400
Lecture 10
2/9/16

Issues:

(1) Finding MLE
→ $\frac{d}{d\theta} L(\theta) = 0$ solve
→ $\frac{d}{d\theta} \log L(\theta) = 0$ solve
→ numerical methods
→ algebraic ingenuity

(Next time:)

(2) Distribution of MLE
→ find exactly
→ Central Limit Theorem
→ Fisher's App

(3) Properties of MLE
→ unbiased? Not usually
→ Approximate Var MSE (Fisher)
→ consider exact distribution

13

We were here

θ "state of Nature" <u>parameter</u>

X (or $X_1, X_2, .. X_n$) <u>data</u>

$f(x|\theta)$ (or $f(x_1, \ldots x_n|\theta)$) <u>model</u>

We estimate θ by the estimator $\hat{\theta}$, a random var.

<u>But</u> Before we go on to the distribution of $\hat{\theta}$, need a few new results.

The Distribution of Sums

We've discussed $E(\hat{\Theta})$ and $Var(\hat{\Theta})$, but for detailed assessments of $\hat{\Theta}$'s accuracy, we need to know its distribution.

But! $\hat{\Theta} = \hat{\Theta}(X) = \hat{\Theta}(X_1, X_2, \ldots X_n)$ is a transformation of the data, its distribution can be VERY complicated.

Some cases are easy, though.

I. Binomial Estimators

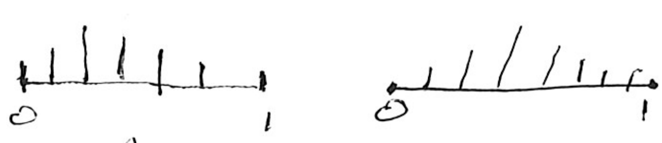
For the estimator $\hat{\Theta}(X) = \frac{X}{n}$

of the parameter Θ in a binomial distribution, $h(x) = \frac{x}{n}$, $h^{-1}(y) = g(y) = nx$

$$p_{\hat{\Theta}}(y) = P_r(\hat{\Theta}_1 = y)$$
$$= p_X(ny)$$
$$= Bin(ny; n, \Theta)$$

$$\hat{\Theta}^*(X) = \frac{(X+1)}{(n+2)}, \text{ so } g(y) = (n+2)y - 1$$

$$p_{\hat{\Theta}^*} = Bin((n+2)y - 1; n, \Theta)$$



$\hat{\Theta} = \frac{X}{n}$ $\qquad$ $\hat{\Theta}^* = \frac{X+1}{n+2}$

The distributions of $\hat{\Theta}$ and $\hat{\Theta}^*$, for $n=6$ $\Theta=0.4$

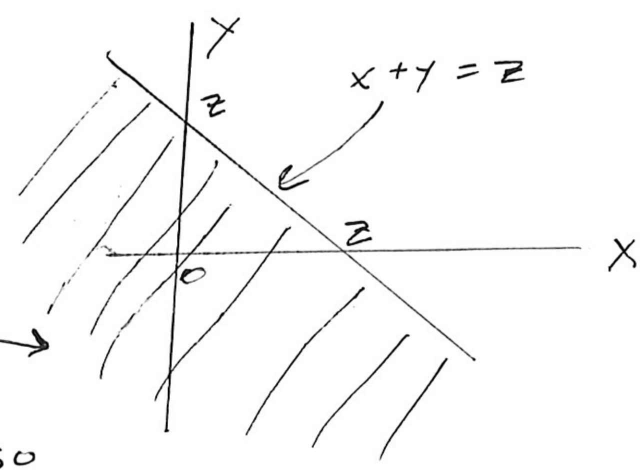
II. $Z = X + Y$

We've already seen $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, and in general sums of random variables are quite frequent. If (X, Y) are a bivariate random variable with density $f(x,y)$, then the density of

$$Z = X + Y$$

is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f(z-y, y)\,dy.$$



Pf

$$F_Z(z) = P_r(Z \le z)$$
$$= P_r(X + Y \le z)$$

the shaded region.

We integrate over that region, so

$$F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f(x,y)\,dx\,dy$$

but

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} \frac{d}{dz}\left(\int_{-\infty}^{z-y} f(x,y)\,dx\right) dy$$

$$= \int_{-\infty}^{\infty} f(z-y, y)\,dy \quad \therefore$$

III $Z = X + Y$, X and Y normal and independent.

The "reproductive property" of normal distributions.

Say $X \sim N(\mu, \sigma^2)$, $Y \sim N(\theta, \tau^2)$.

Then $$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(z-y-\mu)^2} \cdot \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{1}{2\tau^2}(y-\theta)^2} dy$$

$$= \frac{1}{2\pi\sigma\tau} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[(z-y-\mu)^2/\sigma^2 + (y-\theta)^2/\tau^2\right]} dy$$

The part of the exponent in brackets can be written
$$A(z-B)^2 + C(y-Dz)^2 + E$$

(trust me... complete the squares)

but now:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\,\sigma\tau\sqrt{C}} e^{-\frac{E}{2}} \cdot e^{-\frac{A}{2}(z-B)^2} \underbrace{\int_{-\infty}^{\infty} \sqrt{\frac{C}{2\pi}}\, e^{-\frac{C}{2}(y-Dz)^2} dy}_{\text{integral of } N(Dz, \frac{1}{C}) = 1 \,!}$$

so now

$$f_Z(z) \propto e^{-\frac{A}{2}(z-B)^2},$$ a $N(B, \frac{1}{A})$ density

we now have

$$Z \sim \mathcal{N}(B, 1/A)$$

$$B = E(Z) \text{ and } 1/A = Var(Z)$$

but

$$E(Z) = E(X) + E(Y) = \mu + \theta$$

and since X and Y are indep,

$$Var(Z) = Var(X) + Var(Y)$$
$$= \sigma^2 + \psi^2$$

hence $Z \sim \mathcal{N}(\mu + \theta, \sigma^2 + \psi^2)$.

Note: it hence follows that if $X_1, X_2, .. X_n$, each distributed

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

if they are all distributed $\mathcal{N}(\mu, \sigma^2)$, then

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2)$$

and $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

[Note: no limits involved here, cf CLT]

4. The Chi-Square Distribution

Way back in Lecture 3, pp. 13–14, we found the Chi-square (χ^2) distribution with one degree of freedom, namely the dist of U^2, where $U \sim N(0,1)$. It had density

$$f_{U^2}(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} \quad \text{for } y > 0.$$

The chi-square dist for n degrees of freedom is: $\chi^2(n) = U_1^2 + U_2^2 + \cdots + U_n^2$, U's indep and $\sim N(0,1)$.

Then

$$f_{\chi^2(n)}(x) = \frac{1}{2^{n/2}\,\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad \text{for } x > 0.$$

(0 otherwise)

Why? Well, for $n=1$, $\Gamma(\frac{n}{2}) = \sqrt{\pi}$, so for the $n=1$ case: [illegible] follows

$$\frac{1}{2^{n/2}\,\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} = \frac{1}{\sqrt{2\pi x}} \qquad \text{(same as lec 3)}$$

Let's continue by induction.

want to prove

$$f_{\chi^2(n)}(x) = \frac{1}{2^{n/2}\,\Gamma(\frac{n}{2})}\, x^{\frac{n}{2}-1}\, e^{-x/2} \qquad x>0$$

We have the case $n=1$ taken care of. Now let's assume the above formula holds for $n = k-1$. Let

$$X = U_1^2 + \cdots + U_{k-1}^2$$
$$Y = U_k^2$$

$U_1, \ldots U_k$ indep, $\sim N(0,1)$.

So X and Y are independent, and $\chi^2(k) = X + Y$ by definition.

By hypothesis,

$$f_X(x) = \frac{1}{2^{\frac{k-1}{2}}\,\Gamma(\frac{k-1}{2})}\, x^{\frac{k-1}{2}-1} e^{-\frac{x}{2}} \qquad x>0$$

and

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}}\, e^{-y/2} \qquad y>0$$

(both 0 otherwise)

So,

$$f_{\chi^2(k)}(z) = \int_{-\infty}^{\infty} f_X(z-y)\, f_Y(y)\,dy,$$

and since $f_X(z-y) = 0$ iff $y \geq z$,

$$= \int_0^z \frac{1}{2^{\frac{k-1}{2}}\,\Gamma(\frac{k-1}{2})} \cdot (z-y)^{\frac{k-1}{2}-1} e^{-\frac{(z-y)}{2}} \cdot \frac{1}{\sqrt{2\pi y}}\, e^{-y/2}\,dy$$

Now, extract all terms not containing y...

want to prove

$$\underbrace{f_{\chi^2(n)}(x) = \frac{1}{2^{n/2}\,\Gamma(\frac{n}{2})}\, x^{\frac{n}{2}-1}\, e^{-\frac{x}{2}}}_{\text{"}}$$
(continued..)

$$f_{\chi^2(K)}(z) = \frac{1}{2^{\frac{K-1}{2}}\,\Gamma(\frac{K-1}{2})\,\sqrt{2\pi}}\, e^{-\frac{z}{2}} \int_0^z (z-y)^{\frac{K-3}{2}}\, y^{-1/2}\, dy$$

The above engenders hope, but how to do the integral?

The spirit of how to procede is to note that we are proving something about $f(z)$, but the integral is with respect to y. How can we extract all z terms to the left of the integral sign?

Answer: if $y = zu$, $(z-y) \to (z - zu) = z(1-u)$

in detail: $dy = z\,du$, $(z-y)^{\frac{K-3}{2}} = z^{\frac{K-3}{2}} (1-u)^{\frac{K-3}{2}}$

$y^{-1/2} = z^{-\frac{1}{2}} u^{-\frac{1}{2}}$, so

$$\int_0^z (z-y)^{\frac{K-3}{2}}\, y^{-\frac{1}{2}}\, dy = z^{\frac{K-3}{2}} \cdot z^{-\frac{1}{2}} \cdot z \underbrace{\int_0^1 (1-u)^{\frac{K-3}{2}}\, u^{-\frac{1}{2}}\, du}_{\text{const!}}$$

Hence

$$f_{\chi^2(K)}(z) = C\, z^{\frac{K}{2}-1}\, e^{-\frac{z}{2}} \quad \text{for } z>0$$

... the desired functional form.

What about C? We have, at this point,

$$f_{\chi^2(k)} = \frac{1}{2^{\frac{k-1}{2}}\Gamma\left(\frac{k-1}{2}\right)\sqrt{2\pi}} z^{\frac{k}{2}-1} e^{-\frac{z}{2}} \underbrace{\int_0^1 (1-u)^{\frac{k-3}{2}} u^{-\frac{1}{2}}\,du}_{\text{a Beta function!}}$$

$$\to B\left(\tfrac{1}{2}, \tfrac{k-1}{2}\right) = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{k-1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$$

$$= \frac{\sqrt{\pi}\,\Gamma\left(\frac{k-1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$$

$$f_{\chi^2(k)} = \frac{\Gamma\left(\frac{k-1}{2}\right)\sqrt{\pi}}{2^{\frac{k-1}{2}}\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{k-1}{2}\right)\sqrt{2}\sqrt{\pi}} z^{\frac{k}{2}-1} e^{-\frac{z}{2}}$$

$$= \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} z^{\frac{k}{2}-1} e^{-\frac{z}{2}}$$

$$\therefore$$

Because the estimator for σ^2 is $\hat{s}^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i-\bar{x})^2$ [For $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$],

$$(n-1)\hat{s}^2/\sigma^2 \sim \chi^2(n-1)$$

and hence χ^2 is the distribution of a multiple of the sample variance of a normally distributed sample.

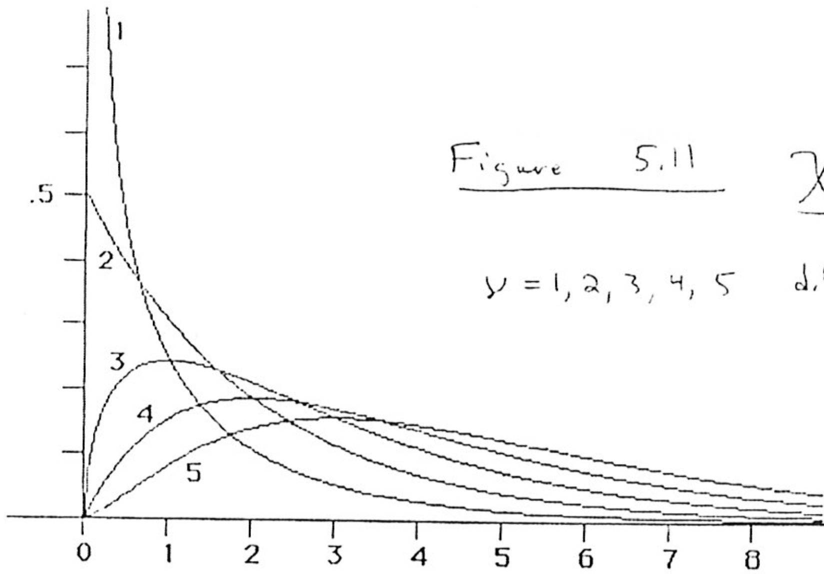
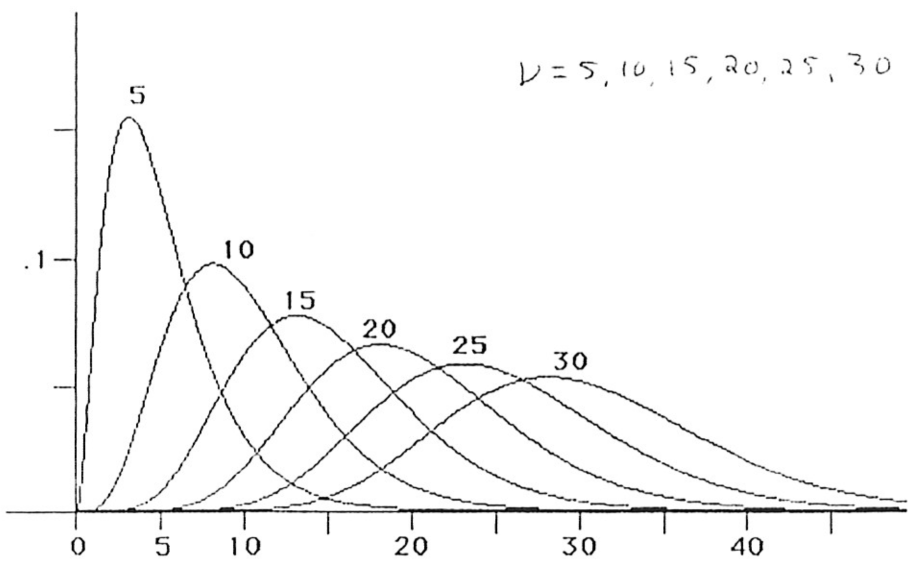


Figure 5.11 χ^2 densities

$\nu = 1, 2, 3, 4, 5$ d.f.

$\nu = 5, 10, 15, 20, 25, 30$ d.f.



Some MLE's are sums or averages.

Central Limit Theorem (in Ch. 5)

$X_1, \ldots, X_n$ indep. $E(X_i) = \mu$, $Var(X_i) = \sigma^2$

Then: $\bar{X}$ approx* $N(\mu, \frac{\sigma^2}{n})$

$\sum_{i=1}^{n} X_i$ approx* $N(n\mu, n\sigma^2)$

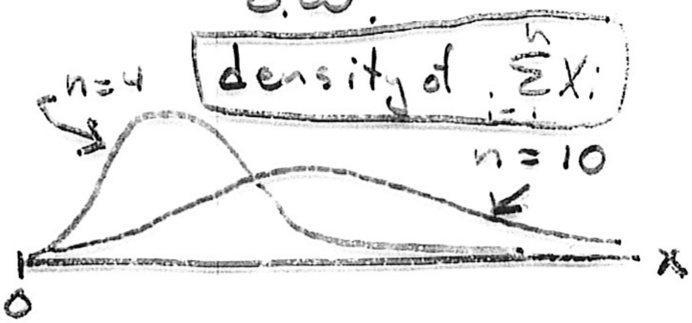
*Best if n large.

[Helps explain normal distribution in nature, of aggregates ($\equiv$ sums)]

Ex: Chi-Square large d.f.

Ex: $X_1, X_2, \ldots, X_n$ indep exponential

$$f(x_i \mid \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & 0 < x < \infty \\ 0 & \text{o.w.} \end{cases} \quad \Big| \quad \hat{\theta}(x_1, \ldots, x_n) = \bar{X}$$



Now - Back At Last to the MLE

Review

Θ "State of Nature", parameter

X (or $X_1, X_2, \ldots X_n$) Data

$f(x|\Theta)$ (or $f(x_1, \ldots, x_n|\Theta)$) Model

An Estimator = a tool for estimating

$$\hat{\Theta} = \hat{\Theta}(X) \text{ or } \hat{\Theta}(X_1, \ldots, X_n)$$

Goal: $\hat{\Theta}$ near Θ.

A "good" estimator tends to be near Θ.

Evaluate from "before data" perspective.

X random, dist $f(x|\Theta)$

$\hat{\Theta} = \hat{\Theta}(X)$ random $f_{\hat{\Theta}}(u|\Theta)$

Want $f_{\hat{\Theta}}(u|\Theta)$ concentrated near Θ

One measure: $MSE_{\hat{\Theta}}(\Theta) = E(\hat{\Theta} - \Theta)^2$

$$= Var(\hat{\Theta}) + (Bias_{\hat{\Theta}}(\Theta))^2$$

Example: "Serial Number" Problem
(from WWII, simplified)

Observe serial number = X

Largest possible = Θ

$P(X=1) = P(X=2) = \cdots = P(X=\theta) = \frac{1}{\theta}$

Model: $p(x|\theta) = \frac{1}{\theta}$, for $x = 1, 2, \ldots, \theta$

$= 0$ otherwise

Maximum Liklihood:

$L(\theta) = \frac{1}{\theta}$ for $\theta \geq X$

MLE: $\hat{\theta}_1 = X$ (If $X = 15$, estimate 15)

Biased $E(\hat{\theta}_1) = E(X) = \frac{1+\theta}{2}$

Bias $= \frac{1+\theta}{2} - \theta = \frac{1-\theta}{2}$.

Unbiased Est: $\hat{\theta}_2 = 2X - 1$, $E(\hat{\theta}_2) = 2EX - 1$

$= 2\left(\frac{1+\theta}{2}\right) - 1$

$= \theta$

Comparison

Calculate $Var(X)$.

$$E(X^2) = \frac{1}{\theta}\sum_{x=1}^{\theta} x^2 = \frac{1}{\theta}\cdot\frac{\theta(\theta+1)(2\theta+1)}{6}$$

$$= \frac{(\theta+1)(2\theta+1)}{6}$$

$$Var(X) = \frac{(\theta+1)(2\theta+1)}{6} - \left(\frac{\theta+1}{2}\right)^2$$

For MLE: $$MSE\,\hat{\theta}_1 = \left[\frac{(\theta+1)(2\theta+1)}{6} - \left(\frac{\theta+1}{2}\right)^2\right] + \left[\frac{(1-\theta)}{2}\right]^2$$

$$= \frac{2\theta^2 - 3\theta + 1}{6}$$

For Unbiased:

$$MSE\,\hat{\theta}_2 = 4\,Var(X) + (bias)^2$$

$$= 4[\quad] + 0^2$$

$$= \frac{\theta^2 - 1}{3}$$

θ	1	2	3	...	10	...	100
$MSE\,\hat{\theta}_1$	0	1/2	5/3	...	28.5	...	3283.5
$MSE\,\hat{\theta}_2$	0	1	8/3	...	33	...	3333