

# 1 Maximum likelihood review questions - Set 5

revised November 25, 2010 (Happy Thanksgiving)

See also the student questions and answers that I distributed fall 2007

1. Specify a statistical or econometric model with at least one parameter, a model of some real-world process. Be as specific as possible. Assume a random sample from the population implied by your model. Describe in words, and with functional notation, how you would derive the maximum likelihood estimate of the parameter(s). Then derive the ml parameter estimate(s). Do not choose the CLR model as your example. Briefly describe what, in general, is good and bad about maximum likelihood estimation. Then briefly describe the realism of your model
2. Consider a zoo that has only chimps and elephants where  $poop_j$  is the weight (in pounds) of  $poop$  animal  $j$  produces on a given day,  $S_j = 1$  if the animal is a elephant, and zero otherwise. The amount of  $poop$  animal  $j$  produces on a given day is a draw from a Poisson distribution (some animals on some days are constipated) with parameter  $\lambda_j$  where  $\lambda_j = \alpha + \beta S_j$ . The director of the zoo likes only chimps, but knows that chimps like to ride on the backs of elephants, so has 90 chimps and 10 elephants (lots of chimps can ride one elephant). Write down the density function for  $poop$  for this population. Explain why this is the density function. Derive the expected amount of  $poop$  produced per day. Comment briefly on the likelihood of the  $poop$  being Poisson distributed rather than normally distributed.

**answer to the first part:** The distribution function  $f(poop)$  is what is called a mixture distribution or a contagious distribution. In this case, a mixture of two densities: one for chimps and the one for elephants. For elephants, the distribution function is  $f_e(poop) = \frac{e^{-(\alpha+\beta)}(\alpha+\beta)^{poop}}{poop!}$  for  $x = 0, 1, 2, \dots$ , and 0 otherwise. For chimps, it is  $f_c(poop) = \frac{e^{-\alpha}\alpha^{poop}}{poop!}$ . At this zoo, chimps appear with probability  $p_c = .9$  and elephants appear with probability  $p_e = .1$ . So  $f(poop) = .1(\frac{e^{-(\alpha+\beta)}(\alpha+\beta)^{poop}}{p!}) + .9(\frac{e^{-\alpha}\alpha^{poop}}{p!})$  for  $x = 0, 1, 2, \dots$ , and 0 otherwise. This is an application of the following theorem. If one has a sequence  $f_o(x)f_1(x), \dots$  of density functions and a sequence  $p_o, p_1, \dots$  such that  $p_i \geq 0$  and  $\sum p_i = 1$ , then  $f(x) = \sum p_i f_i(x)$ .

The expected amount of  $poop$  is  $.9\alpha + .1(\alpha + \beta) = 1.0\alpha + 0.1\beta$ , the simple weighted average.

On a given day, an animal can't poop a negative amount. The Poisson is consistent with this restriction, the normal is inconsistent with this restriction: the normal associates positive probability with negative amounts of poop. The Poisson assumes poop is discretely distributed (animals poop in units of 0, 1, 2, .. but not 1.34 units). The normal allows all animals to

poop in other than integer amounts. The latter is more likely. On the other hand, the Poisson allows for the possibility that on a give day the animal produces no poop, the normal does not.

**Second part of question:** Assume on a randomly chosen day, Jane, the zookeeper, independently collects *poop* from two animals and the *poops* weighs 3 pounds and 1 pound. Write down the likelihood function for this sample, and explain in words why you have written down the correct function.

**answer to the second part:** We want to find the values of  $\alpha$  and  $\beta$  that maximixe the probability of a sample of 3 and 1. We know that  $\Pr[\text{poop}] = f(\text{poop}) = .1(\frac{e^{-(\alpha+\beta)}(\alpha+\beta)^p}{p!}) + .9(\frac{e^{-\alpha}\alpha^p}{p!})$ . So  $\Pr[3] = .1(\frac{e^{-(\alpha+\beta)}(\alpha+\beta)^3}{3!}) + .9(\frac{e^{-\alpha}\alpha^3}{3!})$ . And  $\Pr[1] = .1(e^{-(\alpha+\beta)}(\alpha+\beta)) + .9(e^{-\alpha}\alpha)$ . So the probability of observing this sample of two independent observations is  $\Pr[3]\Pr[1] = (.1(\frac{e^{-(\alpha+\beta)}(\alpha+\beta)^3}{3!}) + .9(\frac{e^{-\alpha}\alpha^3}{3!}))(.1(e^{-(\alpha+\beta)}(\alpha+\beta)) + .9(e^{-\alpha}\alpha)) = L(\alpha, \beta | 3, 1)$ . The maximum likelihood estimates are those values of  $a$  and  $\beta$  ( $a_{ml}$  and  $\beta_{ml}$ ) that maximizes this function,  $L(\alpha, \beta | 3, 1)$ .

**Third part of question:** Now assume that God informs you that the maximum likelihood estimates of  $\alpha, \beta$  are either 1, 1 or 1, 4. What are the maximum likelihood estimates?

**answer to part 3:**  $L(1, 1 | 3, 1) = (.1(\frac{e^{-(1+1)}(1+1)^3}{3!}) + .9(\frac{e^{-1}1^3}{3!}))(.1(e^{-(1+1)}(1+1)) + .9(e^{-1}1)) = 2.6227 \times 10^{-2}$  and  $L(1, 4 | 3, 1) = (.1(\frac{e^{-(1+4)}(1+4)^3}{3!}) + .9(\frac{e^{-1}1^3}{3!}))(.1(e^{-(1+4)}(1+4)) + .9(e^{-1}1)) = 2.3151 \times 10^{-2}$ . The second set generate a larger probability (likelihood) of observing the sample, so the ml estimates are 1 and 4. Note the maximum likelihood estimate of the expected amount of amount of *poop*, per animal, is  $1.0\alpha + 0.1\beta = 1.0 + .1(4) = 1.4$

3. Assume that the rv  $X$  has a uniform distribution

$$f_X(x : a, b) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if otherwise} \end{cases}$$

where  $b > a$ . Assume a random sample of three observations: .9, 4 and 12. Derive the maximum likelihood estimates of the parameters  $a$  and  $b$ . Show all of your work and explain all of your step.

The parameter  $b \geq 12$ , otherwise it would be impossible to observe a 12. The parameter  $a \leq .9$ , otherwise it would be impossible to observe a .9: the observations must be in the range where there is positive density.

Can we be more specific about the maximum likelihood estimates of  $a$  and  $b$ ? Yes. The likelihood function is

$$\begin{aligned} L(.9, 4, 12 : a, b) &= \left(\frac{1}{b-a}\right)\left(\frac{1}{b-a}\right)\left(\frac{1}{b-a}\right) \\ &= \left(\frac{1}{b-a}\right)^3 = (b-a)^{-3} \end{aligned}$$

And  $\ln L = -3 \ln(b - a)$ . This is maximized by minimizing  $(b - a)$  subject to the constraints that  $a \leq .9$  and  $b \geq 12$ . So,  $a_{ml} = .9$  and  $b_{ml} = 12$ .

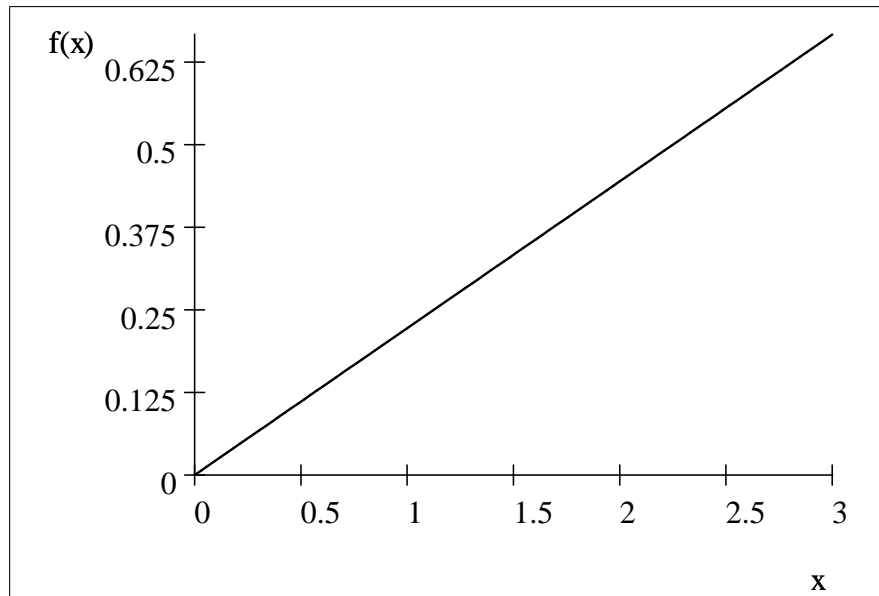
4. Consider the density function

$$f_X(x : a) = \begin{cases} \frac{2}{a^2}x & \text{if } 0 < x \leq a \\ 0 & \text{otherwise} \end{cases}$$

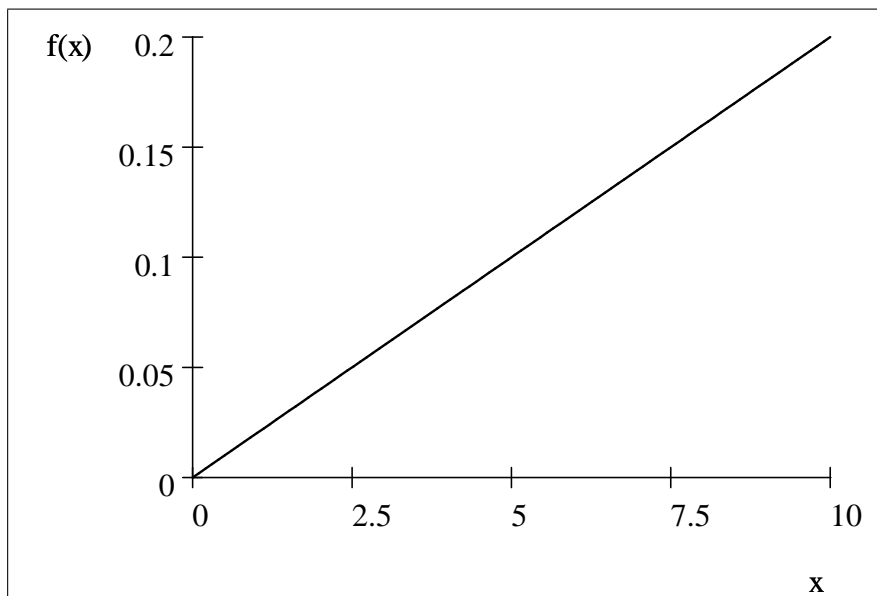
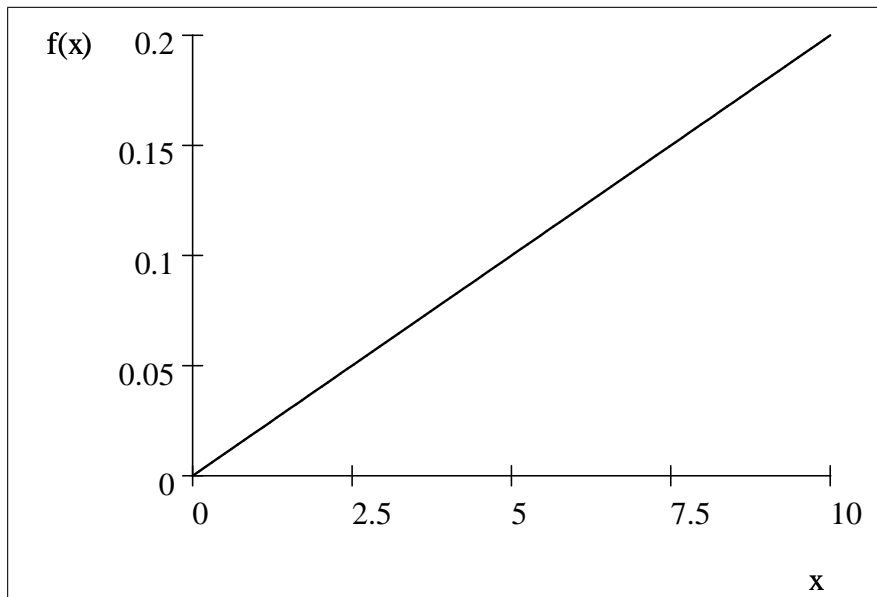
Find the maximum likelihood estimators of the mean,  $\mu$ , and variance,  $\sigma_X^2$ , of this function assuming some random sample,  $X_1, X_2, \dots, X_n$ , where  $n$  is the sample size. How do you know you have the  $ml$  estimators of  $\mu$  and  $\sigma_X^2$ ?

Note that  $\frac{2}{a^2}x > 0 \forall x > 0$  and  $\int_0^a (\frac{2}{a^2}x)dx = 1$ , so this is a density function.

Graphically, if  $a = 3$



And if  $a = 10$



One proceeds by first finding the  $ml$  estimator of the parameter  $a$  assuming some random sample,  $X_1, X_2, \dots, X_n$ . Then one plugs  $a_{ml}$  into the formula for the mean and the formula for the variance of this density function.

The first thing to note is that  $f_X(x : a)$  only has positive density in the range  $0 < x \leq a$ , so  $a_{ml}$  cannot be smaller than the largest of the  $X_i$ ;

that is  $\alpha_{ml} \geq \max(X_1, X_2, \dots, X_n) \equiv X_{\max}$ . (For example if the largest observation in the sample is 3.72, then  $\alpha_{ml}$  cannot be less than 3.72 – if it were less, an observation of 3.72 would be impossible)

Now consider the likelihood function proper:

$$\begin{aligned} L(X_1, X_2, \dots, X_n : a) &= \prod_{i=1}^n f_X(X_i : a) \\ &= \prod_{i=1}^n \left( \frac{2}{a^2} X_i \right) = \left( \frac{2}{a^2} \right)^n \prod_{i=1}^n X_i \end{aligned}$$

Or in terms of  $\ln L$

$$\ln L(X_1, X_2, \dots, X_n : a) = n \ln\left(\frac{2}{a^2}\right) + \sum_{i=1}^n \ln(X_i)$$

$a_{ml}$  is the  $a$  that maximizes  $n \ln\left(\frac{2}{a^2}\right) + \sum_{i=1}^n \ln(X_i)$  subject to the constraint

that  $\alpha_{ml} \geq X_{\max}$ . Note that only the first term in  $n \ln\left(\frac{2}{a^2}\right) + \sum_{i=1}^n \ln(X_i)$  is a function of  $a$  and it is maximized with the smallest allowable value of  $a$ , which is  $X_{\max}$ . That is,  $\alpha_{ml} = X_{\max}$ . Wow. Notice that here all that is needed to find  $\alpha_{ml}$  is  $\alpha_{ml} \geq X_{\max}$ , so  $X_{\max}$  is a *sufficient statistic* – none of the other  $X_i$  in the sample matter.

Now proceed to find the maximum likelihood estimations of  $E[X]$  and  $\sigma_X^2$ .

First determine the mean and variance of this density function as a function of  $a$ .

$$\mu = E[X] = \int_0^a x \left( \frac{2}{a^2} x \right) dx = \frac{2}{3} a$$

$$\sigma_X^2 = \text{var}(X) = E[(X - E[X])^2] = \int_0^a \left(x - \frac{2}{3}a\right)^2 \left(\frac{2}{a^2}x\right) dx = \frac{1}{18} a^2$$

Plugging in  $a_{ml}$ , the maximum likelihood estimates of  $\mu$  and  $\sigma_X$  are

$$\mu_{ml} = \frac{2}{3} a_{ml}$$

$$\sigma_{X_{ml}} = \frac{1}{18} (a_{ml})^2$$

This last step is an example of the invariance principle. if  $q = q(\beta)$  and  $\beta_{lm}$  is the maximum likelihood estimator of  $\beta$  then  $q(\beta_{ml})$  is, by the invariance principle, the maximum-likelihood estimator of  $\beta$ .

5. Assume that the random variable has the following uniform distribution

$$f(X|\alpha) = \begin{cases} 1 & \text{if } \alpha - \frac{1}{2} < X < \alpha + \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

You have drawn a random sample from this population of size 3 and the observations are 3.0, 3.1 and 3.25. Derive the ML estimate(s) of  $\alpha$ . Explain answer: the population density is a uniform density of unit length, centered on  $\alpha$ . Observations from this population must be in the range  $\alpha - \frac{1}{2} < X < \alpha + \frac{1}{2}$  so, the observations 3.0, 3.1 and 3.25 must be in this range. So, 3.0 (the min value in the sample) must be greater than or equal to  $\alpha - \frac{1}{2}$  ( $3.0 \geq \alpha - \frac{1}{2}$ ), so  $3.5 \geq \alpha$ .

The inequality  $\alpha - \frac{1}{2} < X < \alpha + \frac{1}{2}$  also implies that 3.25 must be less than or equal to  $\alpha + \frac{1}{2}$  ( $3.25 \leq \alpha + \frac{1}{2}$ ), so  $2.75 \leq \alpha$ .

Putting these two restrictions together  $2.75 \leq \alpha \leq 3.5$ . So the maximum likelihood estimate of  $\alpha$  must be in this range. The likelihood function can take only two values: 1 and 0, 1 if all of the observations are in the range, and 0 otherwise. So, any value for  $\alpha$  that puts the 3.0, 3.1 and 3.25 in the the range  $2.75 \leq \alpha \leq 3.5$  is a ml estimate of  $\alpha$ ;  $\alpha_{ml}$  is not unique' it is any number in this range. One's first guess is that there is only one ml estimate and it is the midpoint between 3.0 and 3.25, but this is not the case. See MGB page 287.

6. Acts and Desires:

Everyday every individual either does "it" twice, does it once, or does it not at all. What one does today is independent of what one did yesterday or what one will do tomorrow, and what you do is independent of what I do. What one does can vary from day to day.

Assume you observe what each of  $G$  randomly-chosen individuals do on  $M$  days. That is, you observe how many times in  $M$  days individual  $i$  doesn't do it,  $n_i$ , how many days he does it once,  $o_i$  and how many days he does it twice,  $t_i$ .

**Part 1:** Write down and explain a very simple data generation process (statistical model) for this population that could "explain" the data you have collected. Then write down and explain the likelihood function you would use to estimate the parameters in your model. Be explicit about the parameters in your model.

**Answer Part 1:** For individual  $i$ , I would assume a multinomial model

$$f_{N,O,T}(n_i, o_i, t_i : p_d, p_o) = \frac{M!}{n_i!o_i!t_i!} (p_n)^{n_i} (p_o)^{o_i} (1 - p_d - p_o)^{t_i}$$

Note that I have assumed, for simplicity, that the two parameters,  $p_n$  and  $p_o$ , do not vary across individuals -very few models allow every observation to have different parameters.

So, the likelihood function would be

$$L(n, o, t : p_n, p_o) = \prod_{i=1}^G \frac{M!}{n_i! o_i! t_i!} (p_n)^{n_i} (p_o)^{o_i} (1 - p_n - p_o)^{t_i}$$

**Part 2:** Now make the model more interesting by assuming that one's propensity to do "it" is an increasing function of one's desires. The individual does not know his own desires, he only knows that they are normally distributed with mean  $\mu$  and variance  $\sigma^2$  (he knows  $\mu$  and  $\sigma^2$ , we don't). Let  $D$  denote the random variable *desire*. Building on the model in Part 1, write down and explain a statistical model to explain your data, a model that incorporates these additional facts. Be very clear about the number of parameters in your model. Write down and explain the likelihood function for your model. Let  $\phi$  denote the density function of the normal and let  $\Phi$  denote its CDF.

**Answer Part 2:** I would assume that there are two critical levels of  $D$ :  $\mathbf{d}_n$  and  $\mathbf{d}_o$  such that if  $D \leq \mathbf{d}_n$ ,

$$\begin{aligned} p_n &= \int_{-\infty}^{\mathbf{d}_n} \phi_D(x : \mu, \sigma^2) dx \\ &= \int_{-\infty}^{\mathbf{d}_n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(x - \mu)^2/2\sigma^2] dx \\ &= \Phi_D(\mathbf{d}_n : \mu, \sigma^2) \end{aligned}$$

and that if  $\mathbf{d}_o \geq D > \mathbf{d}_n$

$$\begin{aligned} p_o &= \int_{\mathbf{d}_n}^{\mathbf{d}_o} \phi_D(x : \mu, \sigma^2) dx \\ &= \int_{\mathbf{d}_n}^{\mathbf{d}_o} \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-(x - \mu)^2/2\sigma^2] dx \\ &= \Phi_D(\mathbf{d}_o : \mu, \sigma^2) - \Phi_D(\mathbf{d}_n : \mu, \sigma^2) \end{aligned}$$

Just plug these probability functions into the multinomial function and likelihood function in part 1. This model has four parameters:  $\mu$ ,  $\sigma^2$ ,  $\mathbf{d}_n$  and  $\mathbf{d}_o$ . The  $p_o$  and  $p_n$  are not parameters.

Be aware that when one makes the probabilities a function of desire, the probabilities must be between zero and one, and sum to one. Also the monotonicity restriction must be imposed.

There was a tendency in the answer to make the probabilities continuous function of the desire level. I did not do this. It makes things more complicated.

Note that some of you made the probabilities a function of  $D$  but never accounted for the fact that  $D$  is a random variable.

**Part 3:** Now generalize the model in part 2 by allowing  $\mu$  to vary as a function of an observable characteristic of the individual.

**Answer Part 3:** For example, one might assume  $\mu_i = \mu_0 + \beta(\text{age}_i)^{1/2}$ . This model has five parameters:  $\mu_0$ ,  $\beta$ ,  $\sigma^2$ ,  $\mathbf{d}_n$  and  $\mathbf{d}_o$ . Just plug in  $\mu_0 + \beta(\text{age}_i)^{1/2}$  wherever  $\mu$  appears.

**Part 4:** Speculate, on the basis of **only** your Part 1 model, on who knows what and when; that is, is the choice individual  $i$  **will** make on day  $g$  a random variable from the investigator's perspective? A rv from individual  $i$ 's perspective? Speculate on what you assumed as to how the individual decides how many times to do it each day. (Note that your answer to Part 4 depends only on your answer to Part 1.)

**Answer Part 4:** The model in Part 1 suggest that what the individual will do on a given day is a rv from both the individual's perspective and the investigator's perspective. The individual knows  $p_n$  and  $p_o$ , and so  $p_t$ . The question is how does he translate these known probabilities into the observed choice on each day? As if he randomly draws a ball from a barrel in his brain such that if the ball has an "n" on, he doesn't do it, if the ball has an "o" on it, he does it once, and if a ball has a "t" on it, he does it twice. The proportions in the barrel reflect the probabilities. He draws anew each day. This is a bit strange but effectively what is being assumed.

**Part 5:** How would your complete model change if you assumed that each individual knows his level of  $D$  (but the investigator does not observe the individual's level of  $D$ ) and that that level does not vary from day to day? From the individual's perspective? From the researcher's perspective?

**Answer Part 5:** My model would become deterministic **from the individual's perspective** and the individual would do the same thing every day. For example, if  $d_i \leq \mathbf{d}_n$  individual  $i$  would never do it, and if  $d_i > \mathbf{d}_o$  he would do it twice everyday. From the individual's perspective, the multinomial process would collapse into the same deterministic outcome on each day. **From the investigator's perspective**, what each individual will do on the first day is a rv - once the first day is observed, the investigator knows what each individual will do on each remaining day, so the sample effectively collapses to one day for  $G$  individuals. **From the investigator's perspective**, the density function for individual  $i$ 's one observed choice would be

$$p_n^{l_{ni}} p_o^{l_{oi}} (1 - p_n - p_o)^{l_{ti}}$$

where  $l_{ni} = 1$  if individual  $i$  chooses to do nothing, and zero otherwise, etc. etc. From the investigator's perspective, the probability functions would be as before: the investigator does not know the guy's  $D$ . From

the investigator's perspective, the parameters are  $\mu, \sigma^2, \mathbf{d}_n$  and  $\mathbf{d}_o$  or  $\mu_0, \beta, \sigma^2, \mathbf{d}_n$  and  $\mathbf{d}_o$ .

7.  $X$  in the population has a  $t$  distribution with parameter  $k, k > 2$ . Assume that God tells you that  $k$  equals either 3 or 4. Assuming a random sample of 10 observations from the  $t$  distribution, derive the  $ml$  estimator of the parameter  $k$ . That is, find a function or a rule that tell you whether  $k_{ml}$  is 3 or 4 as a function of the sample. Get as far as you can get, explaining as you go. How might you tell if your estimator makes sense? Include some graphs if you think that will help to explain things.

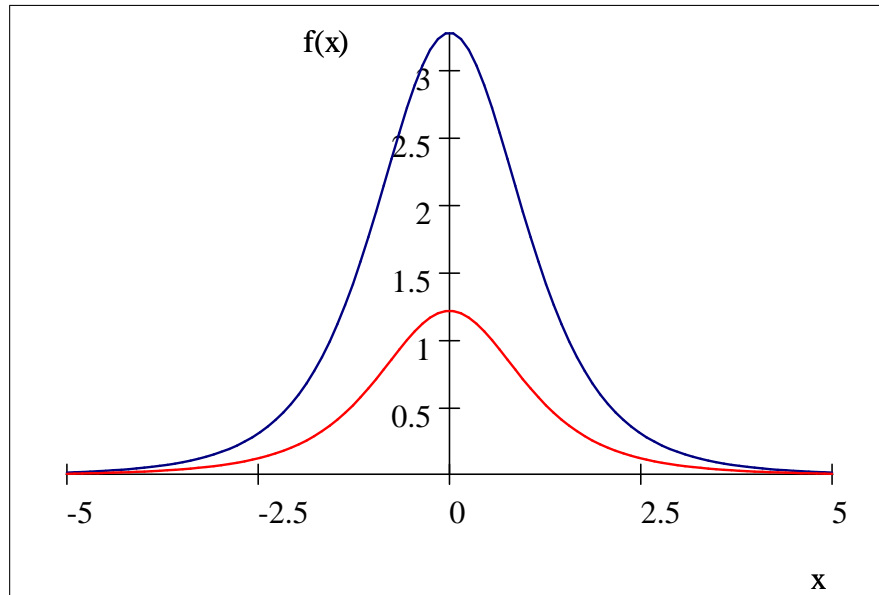
**answer:** Note that what you are estimating here is the variance of the  $t$  distribution. Recollect that  $var(x) = \frac{k}{k-2}$ , so the issue is whether, given the random sample, the best estimate of the variance is  $\frac{3}{3-2} = 3$  or  $\frac{4}{4-2} = 2$  - the larger the  $k$  the smaller the variance.

Note the phrase, "That is, find a function or a rule that tell you whether  $k_{ml}$  is 3 or 4 as a function of the sample." The estimator is a rule simply to determine whether the answer is 3 or 4. You did not need to derive the  $ml$  estimator for the case where  $k$  can take any value: God said it was 3 or 4.

The  $t$  distribution has density function

$$f_X(x) = \frac{\Gamma(k + 1/2)}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1 + x^2/k)^{(k+1)/2}}$$

The blue plot is  $k = 4$  and the red plot is  $k = 3$



When  $k = 4$ , a much larger proportion of the observations will be near zero.

So, the likelihood function is

$$\begin{aligned}
L(X_1, X_2, \dots, X_{10}) &= \left( \frac{\Gamma(k+1/2)}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+X_1^2/k)^{(k+1)/2}} \right) \\
&\times \left( \frac{\Gamma(k+1/2)}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+X_2^2/k)^{(k+1)/2}} \right) \dots \\
&\times \left( \frac{\Gamma(k+1/2)}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+X_{10}^2/k)^{(k+1)/2}} \right) \\
&= \left( \frac{\Gamma(k+1/2)}{\Gamma(k/2)} \right)^{10} \left( \frac{1}{\sqrt{k\pi}} \right)^{10} \prod_{i=1}^{10} \frac{1}{(1+X_i^2/k)^{(k+1)/2}}
\end{aligned}$$

which is a bit messy because  $k$  appears in all of the terms. The  $\ln$  likelihood function is

$$\begin{aligned}
\ln L(X_1, X_2, \dots, X_{10}) &= 10 \ln \left( \frac{\Gamma(k+1/2)}{\Gamma(k/2)} \right) + 10 \ln \left( \frac{1}{\sqrt{k\pi}} \right) \\
&+ \sum_{i=1}^{10} \ln \left( \frac{1}{(1+X_i^2/k)^{(k+1)/2}} \right) \\
&= 10 \ln(\Gamma(k+1/2)) - 10 \ln(\Gamma(k/2)) \\
&+ 10 \ln 1 - 10 \ln(\sqrt{k\pi}) \\
&+ \sum_{i=1}^{10} \left( \ln 1 - \ln \left( (1+X_i^2/k)^{(k+1)/2} \right) \right) \\
&= 10 \ln(\Gamma(k+1/2)) - 10 \ln(\Gamma(k/2)) \\
&+ 0 - 10(\ln k^{1/2} - \ln \pi) \\
&+ 0 - ((k+1)/2) \sum_{i=1}^n \ln(1+X_i^2/k) \\
&= 10 \ln(\Gamma(k+1/2)) - 10 \ln(\Gamma(k/2)) \\
&- 5 \ln k + 10 \ln \pi - ((k+1)/2) \sum_{i=1}^n \ln(1+X_i^2/k)
\end{aligned}$$

Without loss of generality, we can drop the term  $10 \ln \pi$  because it is not a function of  $k$ . We therefore want to find the  $k$  that maximizes  $10 \ln(\Gamma(k+1/2)) - 10 \ln(\Gamma(k/2)) - 5 \ln k - ((k+1)/2) \sum_{i=1}^n \ln(1+X_i^2/k)$ .

How to proceed? The first three terms do not depend on the  $X$ 's, so take the same value whether  $k$  equals 3 or 4.

Calculate these three values

$$10 \ln(\Gamma(3 + 1/2)) - 10 \ln(\Gamma(3/2)) - 5 \ln 3 = 7.7245$$

$$10 \ln(\Gamma(4 + 1/2)) - 10 \ln(\Gamma(4/2)) - 5 \ln 4 = 17.606$$

So,  $10 \ln(\Gamma(k + 1/2)) - 10 \ln(\Gamma(k/2)) - 5 \ln k$  is increasing in  $k$  over the relevant range. This alone would say choose  $k = 4$ .

Now consider the expression that depends on the  $X$ 's  $- ((k + 1)/2) \sum_{i=1}^n \ln(1 + X_i^2/k)$ .

$$- ((3 + 1)/2) \sum_{i=1}^n \ln(1 + X_i^2/3) = -2 \sum_{i=1}^n \ln(\frac{1}{3}X_i^2 + 1)$$

$$- ((4 + 1)/2) \sum_{i=1}^n \ln(1 + X_i^2/4) = -2.5 \sum_{i=1}^n \ln(\frac{1}{4}X_i^2 + 1)$$

So, if  $k = 3$ ,  $\ln L^* = 7.7245 - 2 \sum_{i=1}^n \ln(\frac{1}{3}X_i^2 + 1)$ , and if  $k = 4$ ,  $\ln L^* = 17.606 - 2.5 \sum_{i=1}^n \ln(\frac{1}{4}X_i^2 + 1)$

Which one results in the largest  $\ln L^*$  obviously depends on the  $X$ 's

To find the  $ml$  estimate of  $k$  one would plug the sample into each of these expressions and see which gives the largest value of  $\ln L^*$ .

So, the  $ml$  estimator of  $k$  is 3 if  $7.7245 - 2 \sum_{i=1}^n \ln(\frac{1}{3}X_i^2 + 1) > 17.606 - 2.5 \sum_{i=1}^n \ln(\frac{1}{4}X_i^2 + 1)$ . Otherwise it is 4. That is,

$$k_{ml} = \begin{cases} 3 & \text{if } 7.7245 - 2 \sum_{i=1}^n \ln(\frac{1}{3}X_i^2 + 1) > 17.606 - 2.5 \sum_{i=1}^n \ln(\frac{1}{4}X_i^2 + 1) \\ 4 & \text{otherwise} \end{cases}$$

Let's check this against a few samples to see if this makes sense. Compare three samples: where in one sample all the observations are 1, in one all the observations are 4, and in the last all the observations are 10.  $k_{ml}$  should either decrease or stay the same as the magnitudes of the deviations in the sample increases.

Checking: if all the observations are 1 if  $k = 3$ ,  $\ln L^* = 7.7245 - 2 \sum_{i=1}^{10} \ln(\frac{1}{3}X_i^2 + 1) = 7.7245 - 20 \ln(\frac{1}{3} + 1) = 1.9709$ , and if  $k = 4$ ,

$$\ln L^* = 17.606 - 2.5 \sum_{i=1}^{10} \ln(\frac{1}{4}X_i^2 + 1) = 17.606 - 25 \ln(\frac{1}{4} + 1) = 12.027.$$

So if the sample consists of all ones,  $k_{ml} = 4$ , and the estimated variance of  $X$  is 2.

Checking: if all the observations are 4, if  $k = 3$ ,  $\ln L^* = 7.7245 - 2 \sum_{i=1}^{10} \ln\left(\frac{1}{3}X_i^2 + 1\right) = 7.7245 - 20 \ln\left(\frac{16}{3} + 1\right) = -29.192$ , and if  $k = 4$ ,  $\ln L^* = 17.606 - 2.5 \sum_{i=1}^{10} \ln\left(\frac{1}{4}X_i^2 + 1\right) = 17.606 - 25 \ln(4 + 1) = -22.630$ .

So if the sample consists of all fours,  $k_{ml} = 4$ , and the estimated variance of  $X$  is 2 (it would have been 3 if  $k_{ml} = 3$ )

Checking: if all the observations are 10, if  $k = 3$ ,  $\ln L^* = 7.7245 - 2 \sum_{i=1}^{10} \ln\left(\frac{1}{3}X_i^2 + 1\right) = 7.7245 - 20 \ln\left(\frac{100}{3} + 1\right) = -62.998$ . And if  $k = 4$ ,  $\ln L^* = 17.606 - 2.5 \sum_{i=1}^{10} \ln\left(\frac{1}{4}X_i^2 + 1\right) = 17.606 - 25 \ln\left(\frac{100}{4} + 1\right) = -63.846$

So if the sample consists of all tens,  $k_{ml} = 3$ , and the estimated variance of  $X$  is 3 (it would have been 2 if  $k_{ml} = 4$ )

8. Let  $P_i$  be the probability that one is married  $i$  times,  $i = 1, 2, 3, 4, 5$ : no one is married more than five times, a negative number of times, or a non-integer number of times. Let  $M_i$  be the number of times one is married. Assume that  $n$  individuals are randomly sampled from the population of interest. Let  $T_i$  be the number of people in the sample married  $i$  times (for example  $T_3$  is the number of people in your sample who have been married three times). In terms of the  $T_i$ , choose and express the appropriate likelihood function for your random sample of size  $n$ .

**answer:** The density function for the population has six mutually exclusive alternatives with  $1 > P_i > 0$  and  $\sum_{i=0}^5 P_i = 1$  (a generalized Bernoulli). It can be nothing else. So the likelihood function in terms of the  $T_i$  is a multinomial.

$$L(T_0, T_1, \dots, T_5; P_0, P_1, \dots, P_5) = \frac{n!}{\prod_{i=0}^5 T_i!} \prod_{i=0}^5 (P_i)^{T_i}$$

This is the probability of observing the  $T$  vector as a function of the  $P$ . The ml estimator of  $P_i$  is  $\frac{T_i}{n}$ .

Explaining more, it can't, for example, be a Poisson, or one of the other count discrete density functions, because the Poisson associates positive probabilities with being married more than five times. Many of you assumed a Poisson; Erin conjectured that is was because your professor is, for some reason, entranced with the Poisson.

9. **(This question was asked on the Fall 2008 final and the Jan 2009 prelim.)** The City of Boulder is considering some project,  $A$ . The City knows what  $A$  will cost to produce but the City does not know how much the residents of Boulder will value the project. With this in mind, they conduct a survey asking people whether they would pay  $\$x$  to make project  $A$  happen. A hundred people answer the question with  $x = \$5$ , a hundred at  $\$25$ , a hundred at  $\$50$  and a hundred at  $\$100$ .

Assume this sample of 400 is a random sample of Boulder residents consists of four random subsamples of the population. So, each subsample of 100 at a specific bid amount is a random sample independent of the other subsamples.

Assume that if someone says they would pay  $\$x$  their value for  $A$  is at least  $\$x$ .

Assume that if someone values  $A$  at least  $\$x$ , they would say yes if they were asked whether they would pay  $\$x$  to make  $A$  happen

Note that  $A$  is the name of the project and that  $x$  is a variable denoting the bid amount. Please do not change the names of either  $A$  or  $x$ . Let  $v_i$  represents individual  $i$ 's value for project  $A$ .  $v$  is a continuously distributed rv.

Denote the probability of voting yes to  $A$ ,  $\Pr(yes)$  and the probability of voting yes to  $A$  as a function of  $x$ ,  $\Pr(yes : x)$

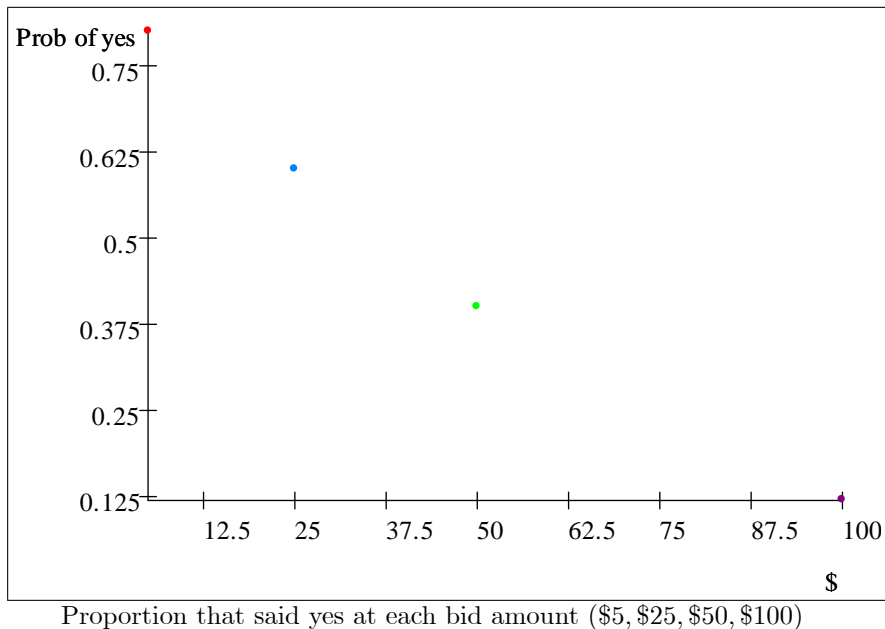
The yes and no responses at the different dollar amounts are

	yes	no
\$5	80	20
\$25	60	40
\$50	35	65
\$100	12	88

Let  $y_{x_j}$  equal the number of yes votes at bid  $x_j$  and  $n_{x_j}$  equal the number of no votes at  $x_j$ . So, for example,  $y_5 = 80$ . And let  $y_{i|x_j} = 1$  if individual

$i$  voted yes to bid amount  $x_j$ , and zero if they voted no, so  $y_{x_j} = \sum_{i=1}^{100} y_{i|x_j}$

Graphing this



10. Assume that the random variable  $X$  has a density function  $f_X(x)$ . One takes a random sample of observations from this density function. The resulting sample is
- |           |   |   |   |   |   |    |    |    |
|-----------|---|---|---|---|---|----|----|----|
| $x$       | 1 | 3 | 7 | 8 | 9 | 11 | 14 | 15 |
| # observ. | 2 | 3 | 3 | 6 | 6 | 5  | 3  | 1  |
- What is the maximum likelihood estimate of  $f_X(x)$ ? Without assuming anything more than what is specified in the question, intuit as much as you can about the m.l. estimate of  $f_X(x)$ . Explain your logic.

answer: It is

$\Pr(x)$	$\Pr(1)$	$\Pr(3)$	$\Pr(7)$	$\Pr(8)$	$\Pr(9)$	$\Pr(11)$	$\Pr(14)$	$\Pr(15)$
	$2/28$	$3/28$	$3/28$	$6/28$	$6/28$	$5/28$	$3/28$	$1/28$
	.071	.107	.107	.214	.214	.179	.107	.036

and  $\Pr(x) = 0$  for all other values of  $X$ . These estimated probabilities are the maximum likelihood probabilities. This question is a little different than you are used to in the context of maximum likelihood estimation because a form for  $f_X(x : \theta)$  was not assumed. The above density function is most likely (has the highest probability of generating our sample). In explanation, note the following:  $f_X(x)$  must have positive density at  $x = 1, 3, 7, 8, 9, 11, 14$  and  $15$ . Allowing  $f_X(x)$  to have positive density at any other values of  $X$  will decrease the probability of observing the sample, so the distribution is discrete, having positive density only over the observed values of  $X$ . The maximum likelihood estimate for the population proportions are simply the sample proportions, any other estimates would decrease the probability of observing our sample. So,

$$\begin{array}{lll}
f_X(x) = 2/28 & \text{if} & x = 1 \\
f_X(x) = 3/28 & \text{if} & x = 3 \\
f_X(x) = 3/28 & \text{if} & x = 7 \\
f_X(x) = 6/28 & \text{if} & x = 8 \\
f_X(x) = 6/28 & \text{if} & x = 9 \\
f_X(x) = 5/28 & \text{if} & x = 11 \\
f_X(x) = 3/28 & \text{if} & x = 14 \\
f_X(x) = 1/28 & \text{if} & x = 15 \\
\text{otherwise} & & f(x) = 0
\end{array}$$

Note that if one assume a continuous function for  $f_X(x : \theta)$  one would decrease the likelihood of observing the sample, because positive probability would have been associated with observing an  $X$  that is not one of the observed points. Assuming something like a Poisson would do the same.

In the answers one estimated the density function, but a number of you incorrectly called it the CDF.

One answer assumed a multinomial that could take one of eight values (the eight observed  $X$  values). This will not fit as well as the what I did, but it comes very close. It is, of course, the correct answer if God told you the distribution was multinomial.

11. In class, I told you many times that maximum likely estimation require that one specifies the functional form for the random variable in question. This is not quite true. There is something called *non-parametric maximum likelihood estimation* and that is what you are going to explore here.

The City of Boulder is considering some project,  $A$ . The City knows what  $A$  will cost to produce but the City does not know how much the residents of Boulder will value the project. With this in mind, they conduct a survey asking people whether they would pay  $\$x$  to make project  $A$  happen. A hundred people answer the question with  $x = \$5$ , a hundred at  $\$25$ , a hundred at  $\$50$  and a hundred at  $\$100$ .

Assume this sample of 400 is a random sample of Boulder residents consists of four random subsamples of the population. So, each subsample of 100 at a specific bid amount is a random sample independent of the other subsamples.

Assume that if someone says they would pay  $\$x$  their value for  $A$  is at least  $\$x$ .

Assume that if someone values  $A$  at least  $\$x$ , they would say yes if they were asked whether they would pay  $\$x$  to make  $A$  happen

Note that  $A$  is the name of the project and that  $x$  is a variable denoting the bid amount. Please do not change the names of either  $A$  or  $x$ . Let  $v_i$  represents individual  $i$ 's value for project  $A$ .  $v$  is a continuously distributed rv.

Denote the probability of voting yes to  $A$ ,  $\Pr(\text{yes})$  and the probability of voting yes to  $A$  as a function of  $x$ ,  $\Pr(\text{yes} : x)$

The yes and no responses at the different dollar amounts are

	yes	no
\$5	80	20
\$25	60	40
\$50	35	65
\$100	12	88

Let  $F_v(v)$  represent the CDF for  $v$ .

Derive the maximum likelihood estimates for  $F(5)$ ,  $F(25)$ ,  $F(50)$ , and  $F(100)$  without assuming a specific functional form for  $F(v)$ . Start generally deriving the maximum likelihood estimates as a function of the  $y_{x_j}$  and  $n_{x_j}$ . Only then get the specific estimates for this data set.

**answer:**

It is the case that

$$\Pr(v_i < x) = F_v(x) = \Pr(\text{no} | x)$$

This is the probability that individual  $i$  will vote no at bid amount  $x$ . And the probability that one votes yes is

$$\Pr(v_i > x) = 1 - F_v(x) = \Pr(\text{yes} | x)$$

Note that  $F_v(x)$  is a number. Let  $F_x$  be an abbrev. way to write this. So,  $\Pr(\text{no} | x) = F_x$  and  $\Pr(\text{yes} | x) = 1 - F_x$ . So what we want are maximum likelihood estimates of  $F_5$ ,  $F_{25}$ ,  $F_{50}$ , and  $F_{100}$ .

The likelihood function for the 100 people who got a bid amount of  $x_j$  is

$$L(F_{x_j} | n_{x_j}, y_{x_j}) = \binom{100}{n_{x_j}} F_{x_j}^{n_{x_j}} (1 - F_{x_j})^{y_{x_j}}$$

In explanation, total no votes for this subset has a binomial distribution: there are 100 trials, the probability of a no vote is  $F_{x_j}$ , and  $n_{x_j}$  no votes occur.

The likelihood function for the five independent subsamples aggregated is

$$\begin{aligned} & L(F_5, F_{10}, F_{25}, F_{50}, F_{100} | n_5, y_5, n_{10}, y_{10}, n_{25}, y_{25}, n_{50}, y_{50}, n_{100}, y_{100}) \\ &= \prod_{x_j, j=5,10,25,50,100} \binom{100}{n_{x_j}} F_{x_j}^{n_{x_j}} (1 - F_{x_j})^{y_{x_j}} \end{aligned}$$

The log likelihood function, with the constant  $\binom{100}{n_{x_j}}$  removed, is

$$\begin{aligned} & \ln L(F_5, F_{10}, F_{25}, F_{50}, F_{100} | n_5, y_5, n_{10}, y_{10}, n_{25}, y_{25}, n_{50}, y_{50}, n_{100}, y_{100}) \\ &= \sum_{x_j, j=5,10,25,50,100} n_{x_j} \ln(F_{x_j}) + y_{x_j} \ln(1 - F_{x_j}) \end{aligned}$$

Maximizing  $\ln L$  wrt each  $F_{x_j}$  for all  $x_j$

$$\frac{\partial \ln[n_{x_j} \ln(F_{x_j}) + y_{x_j} \ln(1-F_{x_j})]}{\partial F_{x_j}} = \frac{n_{x_j}}{F_{x_j}} - \frac{y_{x_j}}{(1-F_{x_j})}$$

Setting each to zero and solving for  $F_{x_j}$  one gets

$$\Pr(\text{no} | x_j) = F_{x_j} = \frac{n_{x_j}}{n_{x_j} + y_{x_j}} = \frac{n_{x_j}}{100} \quad x_j = 5, 10, 25, 50, 100$$

Simply put, the ML estimator for voting not at bid amount  $x_j$  is the proportion of no votes at this bid amount - just what intuition tells us.

The estimator  $\Pr(\text{no} | x_j) = F_{x_j} = \frac{n_{x_j}}{n_{x_j} + y_{x_j}}$  is a maximum likelihood estimator, so has all the properties of ML estimators. This is true no matter what form  $F_v(x)$  takes as long as it is monotonic in  $x$

Plugging in the data, the ML estimates for the probability of a no vote are

$$\Pr(\text{no} | 5) = \frac{20}{100} = 0.2$$

$$\Pr(\text{no} | 25) = \frac{40}{100} = 0.4$$

$$\Pr(\text{no} | 50) = \frac{65}{100} = 0.65$$

$$\Pr(\text{no} | 100) = \frac{88}{100} = 0.88$$

just what is shown in an earlier question, but in terms of no percentages rather than yes percentages.

12. Note that  $\chi_{\text{crit}=.05, df=2}^2 = 6$ . Assume that the r.v. weight,  $W$ , is distributed Extreme value ( $f_W(w) = f(w) = se^{-s(w-n)} \exp[-e^{-s(w-n)}]$  where  $s > 0$  and  $-\infty \leq n \leq \infty$ ). Assume a random sample of 3 observations from this sample: 115, 145 and 165 pounds. Using the likelihood-ratio technique outlined in the notes, find the 95% confidence interval for  $(s, n)$ . I found it easier if I expressed the weights in 100's of pounds. I expressed the confidence interval with a graph,  $n$  on the horizontal axis and  $s$  on the vertical axis. Notationally, let  $L_{s,n} \equiv \ln L(1.15, 1.45, 1.65 : s, n)$

answer: Notationally, let  $L_{s,n} \equiv \ln L(1.15, 1.45, 1.65 : s, n)$  and  $\ln L_{s_{ml}, n_{ml}} \equiv \ln L(1.15, 1.45, 1.65 : s_{ml}, n_{ml})$

Recollect that  $-2(\ln L_{s,n} - \ln L_{s_{ml}, n_{ml}})$  has a Chi-squared distribution with two degrees of freedom, and note that after estimation  $\ln L_{s_{ml}, n_{ml}}$  is a number. The 95% confidence interval for  $(s, n)$  is therefore all those combination of  $s$  and  $n$  for which  $-2(\ln L_{s,n} - \ln L_{s_{ml}, n_{ml}}) \leq 6$ . This is an answer, but not a complete answer.

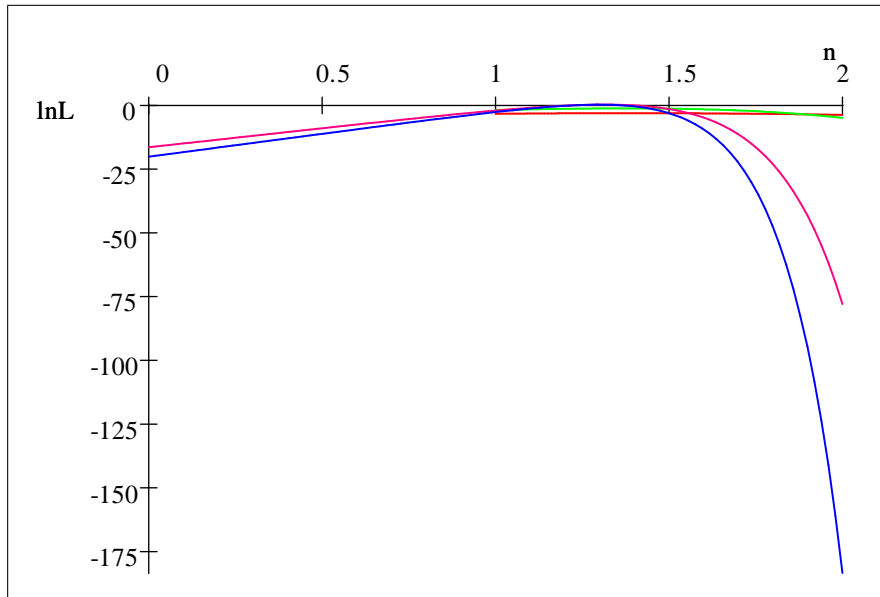
One can find the numerical value of  $\ln L_{s_{ml}, n_{ml}}$  by doing the maximum likelihood estimation. The likelihood function of our sample is, in hundreds of pounds

$$\begin{aligned}
L_{s,n} &\equiv L(1.15, 1.45, 1.65 : s, n) \\
&= se^{-s(1.15-n)} \exp \left[ -e^{-s(1.15-n)} \right] se^{-s(1.45-n)} \\
&\quad \cdot \exp \left[ -e^{-s(1.45-n)} \right] se^{-s(1.65-n)} \exp \left[ -e^{-s(1.65-n)} \right]
\end{aligned}$$

And,

$$\begin{aligned}
\ln L_{s,n} &\equiv \ln L(1.15, 1.45, 1.65 : s, n) \\
&= \ln \left( se^{-s(1.15-n)} \exp \left[ -e^{-s(1.15-n)} \right] se^{-s(1.45-n)} \right. \\
&\quad \left. \cdot \exp \left[ -e^{-s(1.45-n)} \right] se^{-s(1.65-n)} \exp \left[ -e^{-s(1.65-n)} \right] \right) \\
&= \ln se^{-s(1.15-n)} + \ln \exp \left[ -e^{-s(1.15-n)} \right] + \ln se^{-s(1.45-n)} \\
&\quad + \ln \exp \left[ -e^{-s(1.45-n)} \right] + \ln se^{-s(1.65-n)} + \ln \exp \left[ -e^{-s(1.65-n)} \right] \\
&= \ln s + \ln e^{-s(1.15-n)} + \ln \exp \left[ -e^{-s(1.15-n)} \right] + \ln s + \ln e^{-s(1.45-n)} \\
&\quad + \ln \exp \left[ -e^{-s(1.45-n)} \right] + \ln s + \ln e^{-s(1.65-n)} + \ln \exp \left[ -e^{-s(1.65-n)} \right] \\
&= \ln s - s(1.15-n) - e^{-s(1.15-n)} + \ln s + -s(1.45-n) - e^{-s(1.45-n)} + \ln s + -s(1.65-n) - \\
&= 3 \ln s - s(1.15-n) - s(1.45-n) - s(1.65-n) - e^{-s(1.15-n)} - e^{-s(1.45-n)} - e^{-s(1.65-n)} \\
&= 3 \ln s - [s(1.15-n) + s(1.45-n) + s(1.65-n) + e^{-s(1.15-n)} + e^{-s(1.45-n)} + e^{-s(1.65-n)}] \\
&= 3 \ln s - [s(1.15-n) + s(1.45-n) + s(1.65-n)] - [e^{-s(1.15-n)} + e^{-s(1.45-n)} + e^{-s(1.65-n)}] \\
&= 3 \ln s - s[(1.15-n) + (1.45-n) + (1.65-n)] - [e^{-s(1.15-n)} + e^{-s(1.45-n)} + e^{-s(1.65-n)}] \\
&= 3 \ln s - s[4.25 - 3n] - e^{-s(1.15-n)} - e^{-s(1.45-n)} - e^{-s(1.65-n)}
\end{aligned}$$

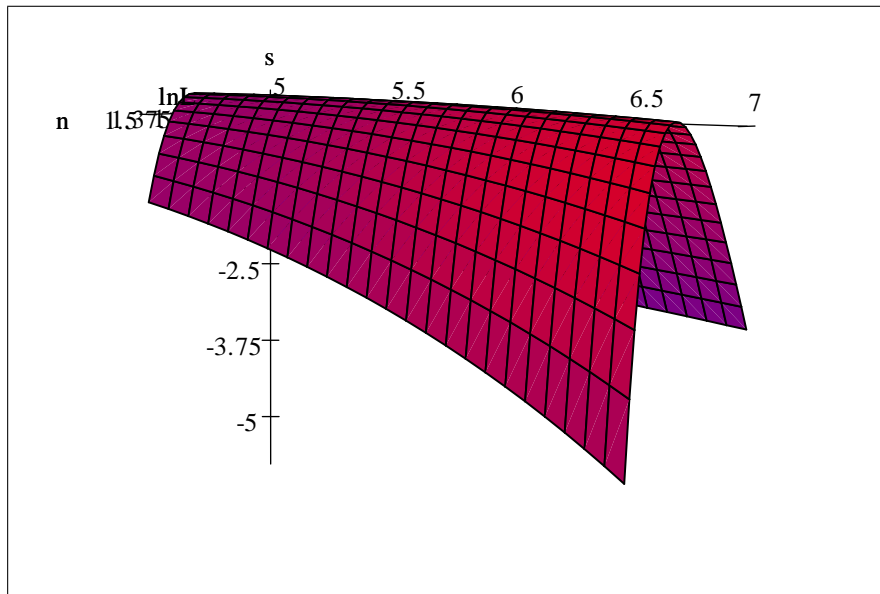
If, for example  $s = 1$ ,  $\ln L_{1,n} = 3 \ln 1 + 3n - 4.25 - e^{-(1.15-n)} - e^{-(1.45-n)} - e^{-(1.65-n)}$ . Graphing this for four different  $s$ ,



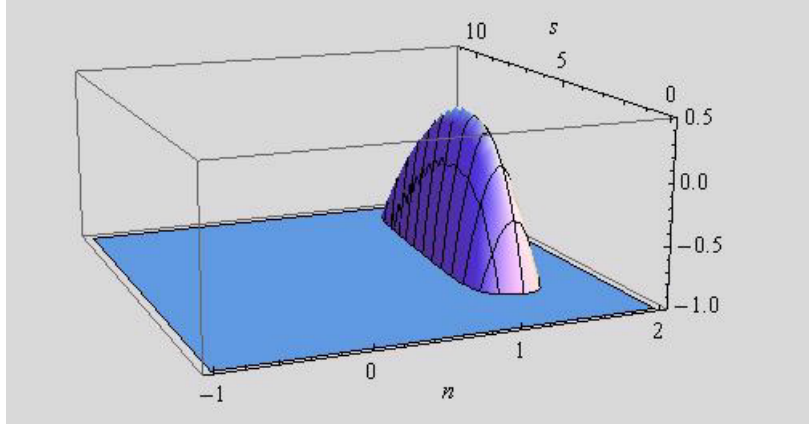
LnL if  $s=1$  (red),  $s=6$  (blue),  $s=2$  (green),  $s=5$  magenta

It looks like  $n$  is around 1.4 and  $s$  is between five and six.

Graphing this as a function of  $n$  and  $s$ ,  $\ln L_{s,n} = 3 \ln s + 3ns - 4.25s - e^{-s(1.15-n)} - e^{-s(1.45-n)} - e^{-s(1.65-n)}$



Can't tell too much from this graph. But Steven has a much better one



Most everyone who, answered this question using a grid search found  $n$  to be 1.31, and  $s$  5.26. so let's go with that. So,

$$\begin{aligned} \ln L_{5.26,1.31} &= 3 \ln(5.26114) + 3(1.31214)(5.26114) - 4.25(5.26114) \\ &\quad - e^{-5.26114(1.15-1.31214)} - e^{-5.26114(1.45-1.31214)} - e^{-5.26114(1.65-1.31214)} = \\ &= 0.33125 \end{aligned}$$

So, now that we have  $\ln L_{ml} = .33125$ , we can get more specific about the confidence interval. From above,  $-2(\ln L_{s,n} - \ln L_{s_{ml},n_{ml}}) \leq 6$  so  $-2(\ln L_{s,n} - .33125) \leq 6$  implying  $-2 \ln L_{s,n} + .66250 \leq 6$  implying  $-2 \ln L_{s,n} \leq 5.3375$  implying  $-\ln L_{s,n} \leq 2.6688$  implying  $\ln L_{s,n} \geq -2.6688$ . So the confidence interval is all those combinations of  $s$  and  $n$  for which  $\ln L_{s,n} \geq -2.6688$

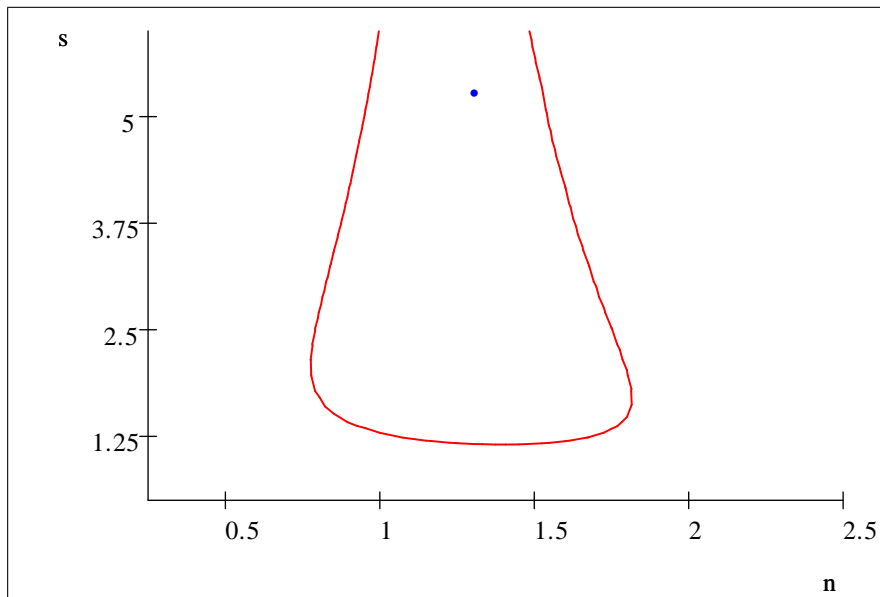
Therefore, the confidence interval is all those  $s$  and  $n$  for which

$$\ln s - s(4.25 - 3n) - e^{-s(1.15-n)} - e^{-s(1.45-n)} - e^{-s(1.65-n)} \geq -2.6688.$$

Graphing the implicit function:  $3 \ln s - s[4.25 - 3n] - e^{-s(1.15-n)} - e^{-s(1.45-n)} - e^{-s(1.65-n)} = -2.6688$ , the boundaries of the confidence interval are traced out: a contour line of the likelihood function at the level  $-2.6688$ .<sup>1</sup>

---

<sup>1</sup>It took me a while to get this plot meaningful: I kept trying to plot in ranges of  $s$  and  $n$  where the equality could not be satisfied, so no points would appear.



95% confid. interval: m.l. estim. blue

I did not draw it all out, but it will narrow to close at the top. Remember that as  $s$  increases, the variance increases. Note that this confidence interval looks large, which you would expect given that we estimated two parameters with only three observations. What exactly is a confidence interval? Note that the position and shape of the confidence interval will vary from sample to sample - it has sampling variation. 95% of these estimated confidence intervals will contain  $n$  and  $s$ . One cannot say there is a 95% chance that the  $n$  and  $s$  are in this interval.

Steven managed to demonstrate the confidence interval in three dimensions.

Trace out the  $\ln$  lik. function at the level of the blue plane. Or from Zach W.

