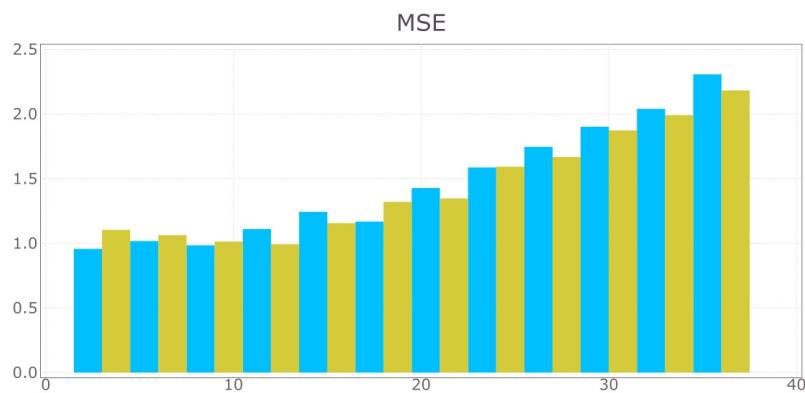


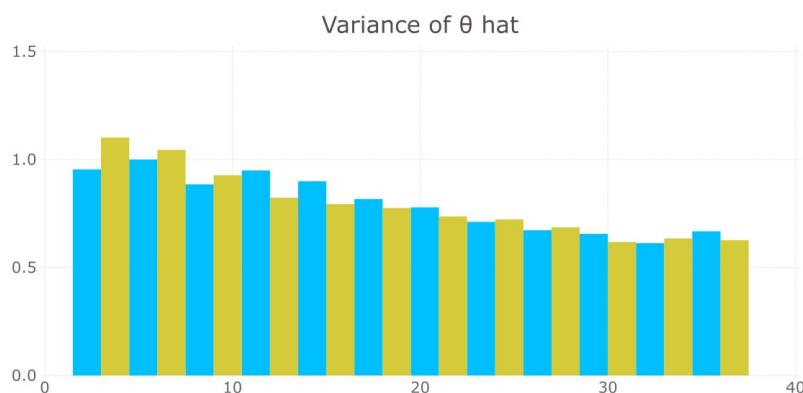
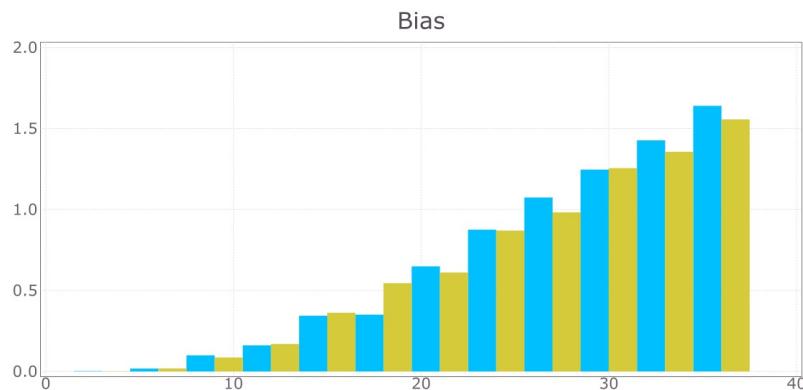
1) I believe that the main moral of this exercise is that as  $d_2$  grows, i.e. the number of instruments grows,  $MSE$  also grows, as we can see:



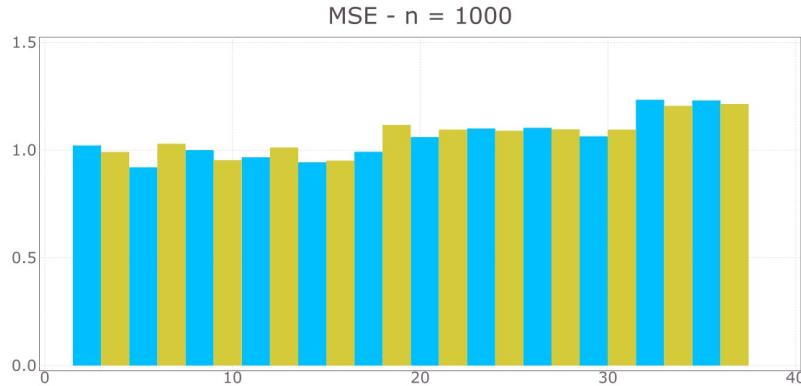
Blue: First element of  $\theta$

Green: Second element

This is mainly due to an increase in the Square Bias of the estimator, since the Variance actually decreases with  $d_2$ :



The effects can be mitigated by a higher  $n$ , as we can see with  $n = 1000$ :



2 - Even though  $\log(1 - \phi(y)) = \log(\phi(-y)) \approx -\log(y) - y^2/2$

The results shows that the functions performs very differently with  $\log(1 - \phi(y))$  being the first to become infinity with Julia being unable to differentiate  $1 - \phi(\exp(2.5))$  and zero.  $\log(\phi(-y))$  performs a little better, which indicates that some precision is lost when we perform the subtraction.

The approximation though, don't have to deal with near-zero numbers and perform better.

b)  $X^T y$  have a lower MSE than  $(X^T X)^{-1} X^T y$ , probably due to numerical imprecisions avoided by Julia

- MSE of  $\hat{\theta}_{03}$  grows as variance of  $x$  decreases, specially when  $(X^T X)^{-1} X^T y$  is being used. (More numerical imprecision?)

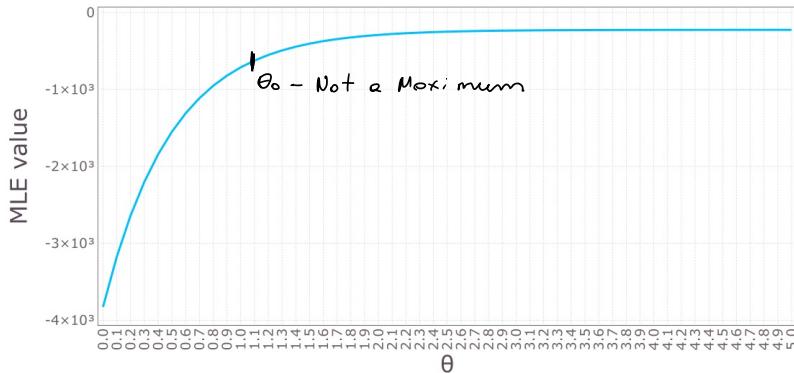
c) I was surprised by this results, I had to scale down a bit ( $n \rightarrow 1000$ ) because my small laptop wasn't handling it well.

But i, iii had a speed that was several orders of magnitude faster than ii, iv. (And I almost wrote i as iv to save keystrokes).

I believe that  $Z(Z^T Z)^{-1}$  ends up to be such a big matrix ( $n \times n$ ) that saving in memory and handling it gets time consuming.

To a lesser degree this also happens to:  $Z(Z^T Z)^{-1} Z^T X$ .

3) To my surprise the original MLE function explodes to the upper bound of the optimization process. In fact, it clearly don't have a maximum:



as  $\theta \rightarrow \infty$ ,  $\exp(y-\theta) \rightarrow 0$   
 $\forall y$ , since  $f(0) > f(u)$   
 $\forall u$  the only maximum  
at infinity.

In fact, if we derive it:

$$Q_0(\theta) = E \left[ \log(f_a(\exp(y_i - \theta))) \right]$$

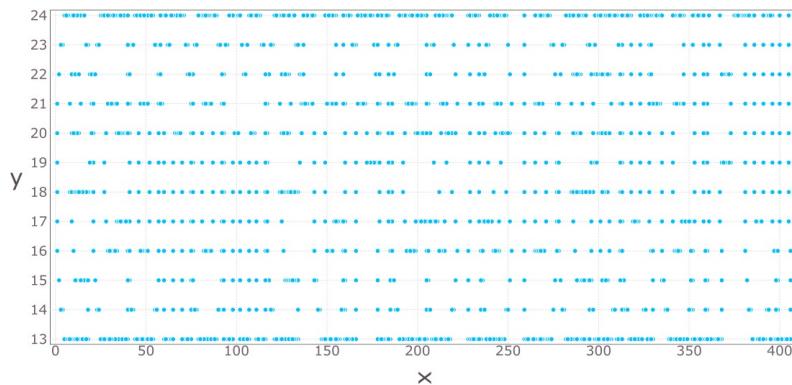
$$\Rightarrow \nabla_{\theta} Q_0(\theta) = -E \left[ \frac{f'_a(\exp(y_i - \theta))}{f_a(\exp(y_i - \theta))} \exp(y_i - \theta) \right]$$

$$\text{and at } \theta = \theta_0 \Rightarrow \exp(y_i - \theta_0) = c_i$$

$$\begin{aligned} \nabla_{\theta} Q_0(\theta_0) &= -E \left[ \frac{f'_a(c_i)}{f_a(c_i)} c_i \right] = - \int_0^\infty f'_a(u_i) c_i du \\ &= - \left[ f_a(u_i) c_i \right]_0^\infty + \int_0^\infty f'_a(u_i) c_i du = 1 \neq 0 \end{aligned}$$

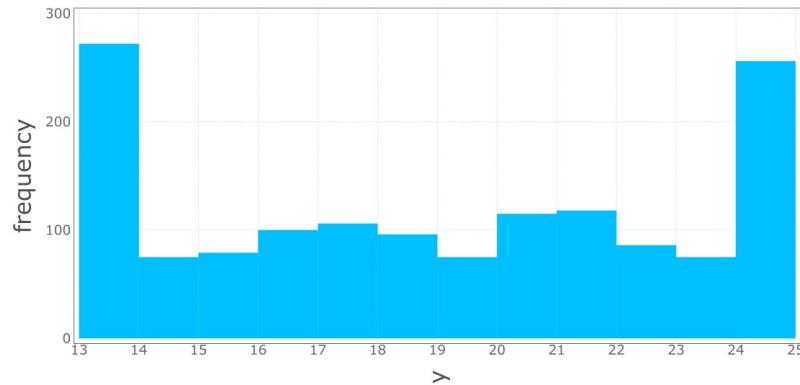
We seem to correct the problem by subtracting  $\theta$  from the original MLE, which converges to  $\theta_0$  as seen in the code.

4)



Oh ! Dear ...

Ok, it doesn't seem to have a correlation, let's dig deeper.

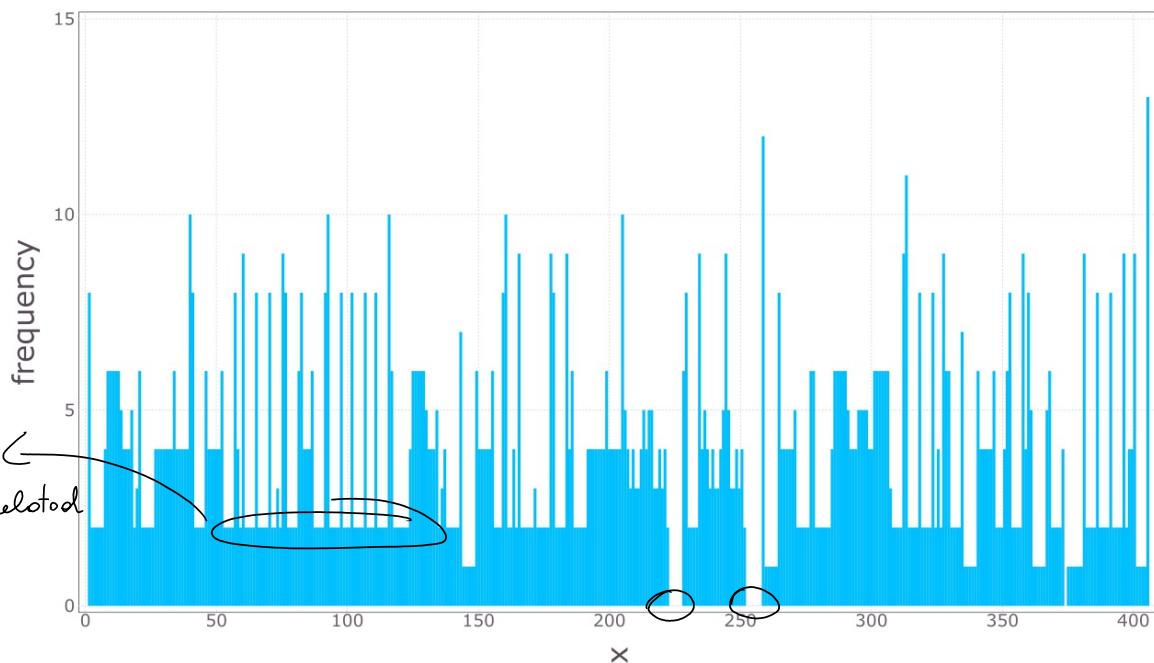


this is better, I assume that  $y$  has the following probability function

$$P(y = y) = \begin{cases} p & \text{if } 14 \leq y \leq 23 \\ \frac{(1-10p)}{2} & \text{if } y = 13 \text{ or } 24 \\ 0 & \text{otherwise} \end{cases} * y \in \mathbb{N}$$

Assuming this I can use MLE to estimate  $p$ , which was estimated to be 6,3% which imply that  $P(y = 13)$  and  $P(y = 24)$  is 18,1% which fits the data nicely.

let's check  $x$ :



All ones



Hmm, it seems that given  $y$ , the some  $x$  can only appear once, and therefore each pair  $(x, y)$  is unique.

If  $x$  was, in fact, uniformly random the probability of never drawing the same  $x$  twice in  $n$  draws is:  $\frac{406!}{(406-n)! \cdot 406^n}$ , which goes to zero very fast as  $n \rightarrow \infty$ . (It is not uniformly random!)

The probability of, given  $y$ , observing  $X = x$  seems to be correlated to whether we have observed  $X = x - 1$ .

I might be overthinking this, but I propose that is another random variable,  $z \in \{0, 1\}$ , that, given  $y$ , follows a Markov Process on  $X$ .

We would then treat  $x$  as if it was a time index, and a observation  $(x, y)$  as "happened on  $y$  at  $x$ "!

$z_{xy}$  can only be 0 or 1

$z_{xy}$  is connected to  $z_{(x-1),y}$

I can then estimate:

$$P(z_{xy} = 1 \mid z_{(x-1),y} = 1, y = 13) = 0.9$$

$$P(z_{xy} = 0 \mid z_{(x-1),y} = 1, y = 13) = 0.1$$

$$P(z_{xy} = 1 \mid z_{(x-1),y} = 0, y = 13) = 0.2$$

$$P(z_{xy} = 0 \mid z_{(x-1),y} = 0, y = 13) = 0.8$$

estimates for  $y > 13$   
are on the code.

5)

# 1 Q5

## 1.1 a

Assume  $\theta$  is not identifiable. That is, for  $\theta \neq \tilde{\theta}$ , given that the conditional expectation of  $u_i$  given  $a_i$  and  $x_i$  is zero,

$$\begin{aligned}\theta x_i + \theta^2 a_i &= \tilde{\theta} x_i + \tilde{\theta}^2 a_i \\ \implies (\theta - \tilde{\theta})x_i &= -(\theta^2 - \tilde{\theta}^2)a_i \\ \implies x_i &= -(\theta^2 - \tilde{\theta}^2)a_i\end{aligned}$$

But this contradicts the fact that  $\text{corr}(x_i, a_i) = 0.5$

## 1.2 b

The optimal instrument is given by

$$\begin{aligned}E_i^T(\theta)|z_i)[E(b_i(\theta)b_i^T(\theta)|z_i)]^{-1} \\ = E(-x_i - 2\theta a_i|z_i)E((y_i - \theta x_i - \theta^2 a_i)|z_i)^{-1} \\ = -(x_i + 2\theta a_i)\end{aligned}$$

The last equation follows from the fact that the conditional variance of  $u_i$  is 1.

## 1.3 c

From the formula for  $z$  above, we can see that equation 8 and 9 are cubic functions of  $\theta$  and so have more than one root. We thus lose identification. We might use MLE to solve this problem.

6) As we can see in the code result,  $\hat{\theta}$  has a lower MSE than  $\hat{\mu}$ , this is mainly a result of a lower variance of  $\hat{\theta}$ , the bias of  $\hat{\theta}$  is significantly higher than that of  $\hat{\mu}$ , suggesting this to be a biased estimator.

- Since  $\hat{\theta}$  is consistent and have a lower variance than  $\hat{\mu}$  it results in a better estimator.

7) the usual pdf for a log normal is:

$$f(x) = \frac{1}{x} \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) \quad \text{where } \mu \text{ and } \sigma \text{ are such that } \ln(x) \sim N(\mu, \sigma^2)$$

For MLE we take logs and have

$$\log(f(x)) = -\left(\ln(x) + \frac{\ln(2\pi)}{2} + \ln(\sigma) + \frac{(\ln(x) - \mu)^2}{2\sigma^2}\right)$$

For the maximization we ignore the non parameters and have:

$$\text{MLE}(\mu, \sigma) = -\sum_{i=1}^n \frac{\ln(x_i) - \mu}{2\sigma^2} + \ln(\sigma)$$

However,  $\mu$  and  $\sigma^2$  are not observed, we can see only the moments of the lognormal, which are:

$$E(x) = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$$

$$\text{Var}(x) = \exp(2\mu + \sigma^2) (\exp(\sigma^2) - 1)$$

In our question we have:

$$l \sim \log N(\mu_e, s_e^2) ; b \sim \log N(\mu_b, s_b^2)$$

$$\text{where: } E(l) = \lambda_0 = \exp\left(\mu_e + \frac{1}{2}s_e^2\right) \Rightarrow \mu_e = \ln \lambda_0 - \frac{1}{2}s_e^2$$

$$E(b) = \beta_0 = \exp\left(\mu_b + \frac{1}{2}s_b^2\right) \Rightarrow \mu_b = \ln \beta_0 - \frac{1}{2}s_b^2$$

$$\begin{aligned}\text{Var}(l) &= \text{Var}(b) = \sigma_0^2 = \exp\left(2\mu_e + s_e^2\right)\left(\exp(s_e^2) - 1\right) \\ &= \exp\left(2\mu_b + s_b^2\right)\left(\exp(s_b^2) - 1\right)\end{aligned}$$

Finally, we have:

$$a = l \cdot b$$

$$\Rightarrow \ln(a) = \ln(l) + \ln(b) \sim N(\mu_e + \mu_b, \sigma_e^2 + \sigma_b^2)$$

Substituting everything, we have the final MLE:

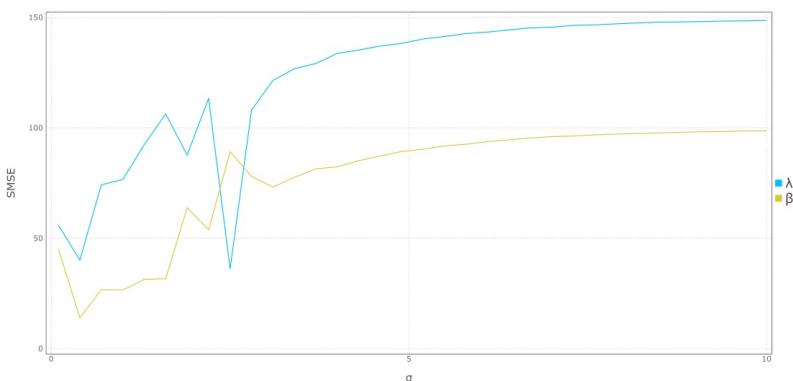
$$= -\sum_{i=1}^n \left\{ \ln \left( \ln \left( \left( 1 + \frac{\sigma_0^2}{\lambda_0^2} \right) \left( 1 + \frac{\sigma_0^2}{\beta_0^2} \right) \right) \right)^{1/2} + \frac{\left( \ln(a_i) - \ln \left[ \left( \frac{\lambda_0}{\sqrt{1 + \frac{\sigma_0^2}{\lambda_0^2}}} \right) \left( \frac{\beta_0}{\sqrt{1 + \frac{\sigma_0^2}{\beta_0^2}}} \right) \right] \right)}{2 \ln \left[ \left( 1 + \frac{\sigma_0^2}{\lambda_0^2} \right) \left( 1 + \frac{\sigma_0^2}{\beta_0^2} \right) \right]} \right\}$$

b) Running the optimizer I found:

$$\begin{aligned}\hat{\lambda} &= 196.17 \\ \hat{\beta} &= 82.53 \\ \hat{\sigma}_0 &= 0.9813\end{aligned}$$

However the curve seems to be flat around this point, so the estimate may be dependent on initial values and other optimization parameters.

c)



This is odd, I used  $\lambda_0 = 150$  and  $\beta_0 = 100$  in this simulation, and ...  $\text{SMSE}_{\lambda} \rightarrow \lambda_0$ ? I don't see why this would happen.

8) a)  $F$  is monotonically increasing:  $\Rightarrow \phi(-x) = (1 - \phi(x))$

$$\begin{aligned} \frac{\partial F(x,y)}{\partial x} &= \psi(x)\phi(y) - 0.5\phi(y)\phi(-y)(\psi(x) - 2\phi(x)\varphi(x)) \\ &= \psi(x)\phi(y) \left[ 1 - 0.5\phi(-y)(1 - 2\phi(x)) \right] \\ &= \underbrace{\psi(x)\phi(y)}_{\geq 0} \underbrace{\left[ 1 - 0.5\phi(-y) + \frac{\phi(-y)\phi(x)}{2} \right]}_{\geq 0} \\ &\quad > 0 \end{aligned}$$

$\frac{\partial F(x,y)}{\partial y} > 0$  by the same argument since it is symmetrical to  $\nabla_x F(x,y)$ .

(ii)  $F$  is continuous since  $\phi$  is continuous.

(iii)  $0 \leq F(x,y) \leq 1$  since  $0 < \phi(x)\phi(y) < 1$ .

Suppose  $F(x,y) < 0$

$\Rightarrow \phi(x)\phi(y) < 0$  (not possible)  
 or  $(1 - 0.5\phi(x)\phi(y)) < 0$   
 also not possible since  $\phi(-x)\phi(-y) < 1$

$$\text{iv) } \lim_{x,y \rightarrow \infty} F(x,y) = \lim_{x,y \rightarrow \infty} \underbrace{\phi(x)\phi(y)}_{=1} \underbrace{(1 - 0.5\phi(-x)\phi(-y))}_{=0} = 1$$

by the same logic

$$\lim_{x,y \rightarrow -\infty} F(x,y) = 0$$

$$\text{v) } F(x) = \lim_{y \rightarrow \infty} F(x,y) = \phi(x)(1 - 0) = \phi(x)$$

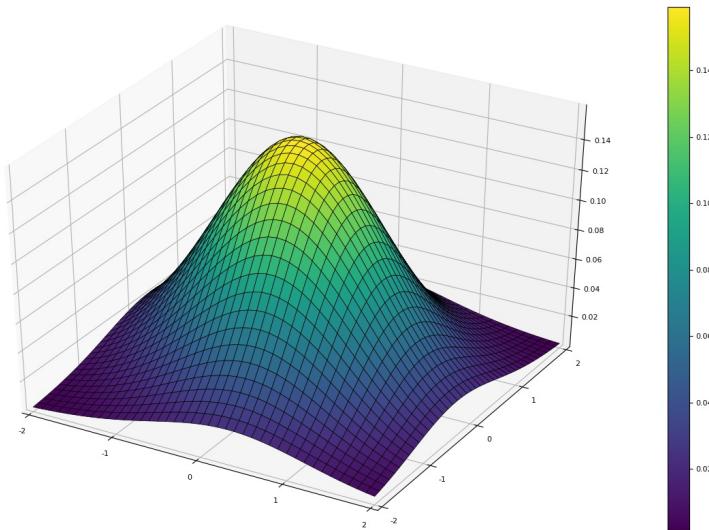
$$F(y) = \lim_{x \rightarrow \infty} F(x,y) = \phi(y)$$

$$\text{vi) } f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial}{\partial y} (\psi(x)\phi(y) [1 - 0.5(1 - \phi(y)) + (1 - \phi(y))\phi(x)])$$

$$= \frac{\partial}{\partial y} (\psi(x)\phi(y) [1 + (1 - \phi(y))(\phi(x) - 0.5)])$$

$$\begin{aligned}
&= \varphi(x) \left[ \varphi(y) \left[ 1 + (1 - \varphi(y)) (\varphi(x) - 0.5) \right] - \varphi(y) (\varphi(x) - 0.5) \varphi(y) \right] \\
&= \varphi(x) \varphi(y) \left[ 1 - 0.5 (1 - 2\varphi(x)) (1 - 2\varphi(y)) \right]
\end{aligned}$$

the pdf:



$$\begin{aligned}
d) f(x | y=0) &= \varphi(x) \varphi(0) \left[ 1 - 0.5 (1 - 2\varphi(x)) (1 - 2 \cdot 0.5) \right] \\
&= \varphi(x) \cdot \varphi(0) \approx 0.4 \varphi(x)
\end{aligned}$$

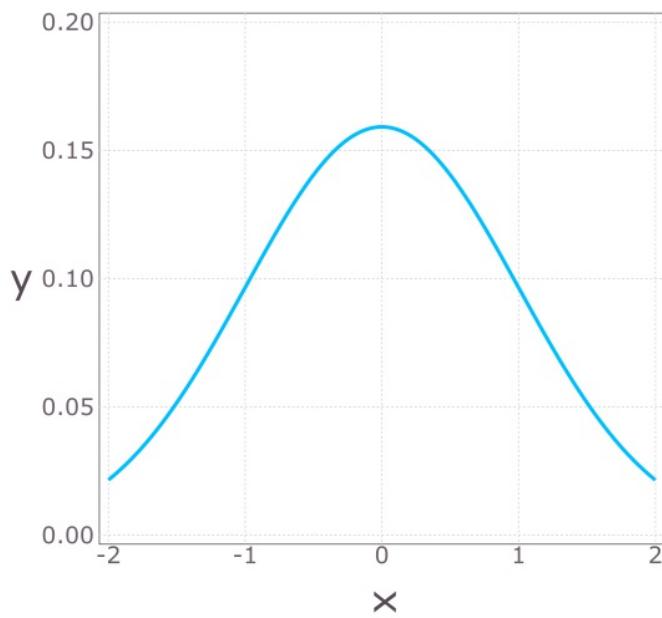
$$\begin{aligned}
f(x | y=1) &= \varphi(x) \varphi(1) \left[ 1 - 0.5 (1 - 2\varphi(x)) (1 - 2\varphi(1)) \right] \\
&\approx \varphi(x) 0.24 \left[ 1 - 0.5 (1 - 2\varphi(x)) (1 - 1.68) \right] \\
&\approx 0.24 \varphi(x) \left[ 1 + 0.34 (1 - 2\varphi(x)) \right]
\end{aligned}$$

$$f(x | y=-1) \approx 0.24 \varphi(x) \left[ 1 - 0.34 (1 - 2\varphi(x)) \right]$$

$$\begin{aligned}
f(x | y=0.5) &= \varphi(x) \varphi(0.5) \left[ 1 - 0.5 (1 - 2\varphi(x)) (1 - 2\varphi(0.5)) \right] \\
&\approx 0.35 \varphi(x) \left[ 1 + 0.19 (1 - 2\varphi(x)) \right]
\end{aligned}$$

$$f(x | y=-0.5) \approx 0.35 \varphi(x) \left[ 1 - 0.19 (1 - 2\varphi(x)) \right]$$

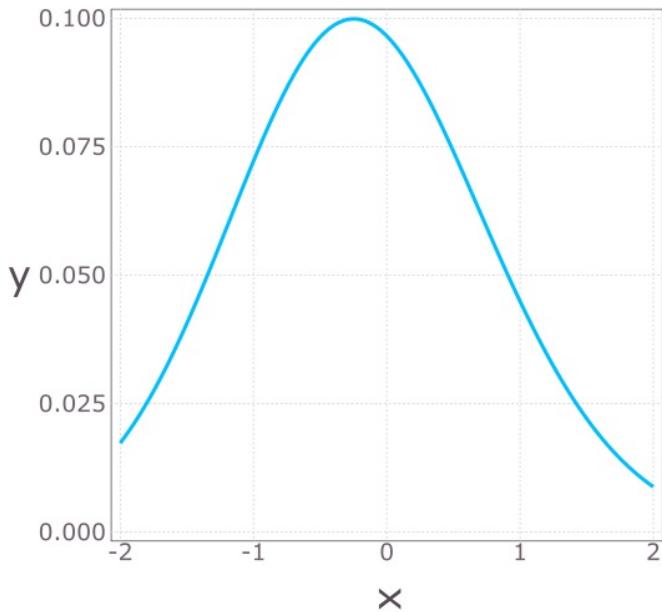
$\gamma = 0$



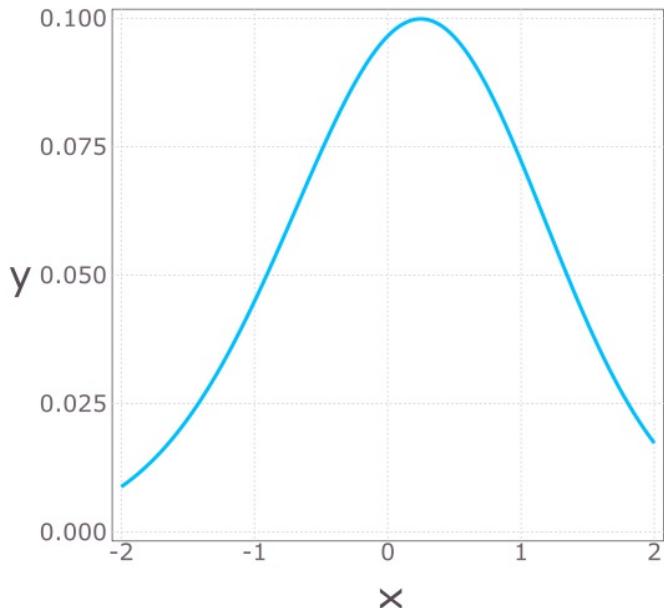
given  $\gamma = 0$ ,  $x$  is normal.

given  $\gamma = \pm 1$ ,  $\pm 0.5$   
we have a normal-like  
imperfect curve.

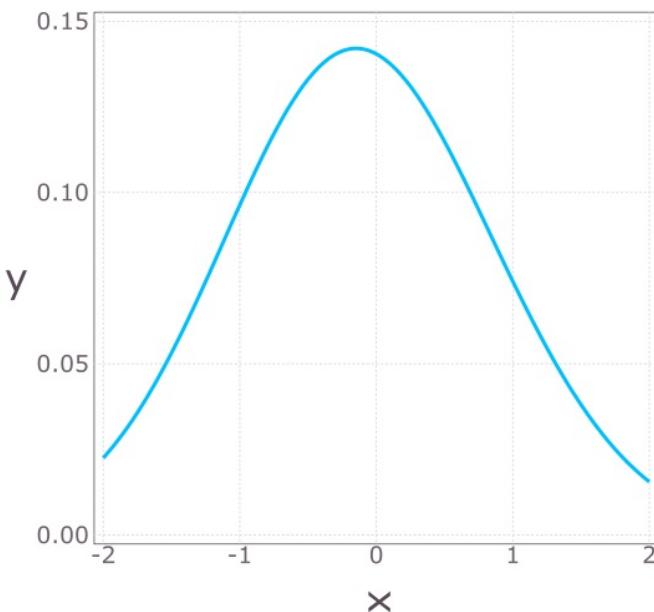
$\gamma = 1$



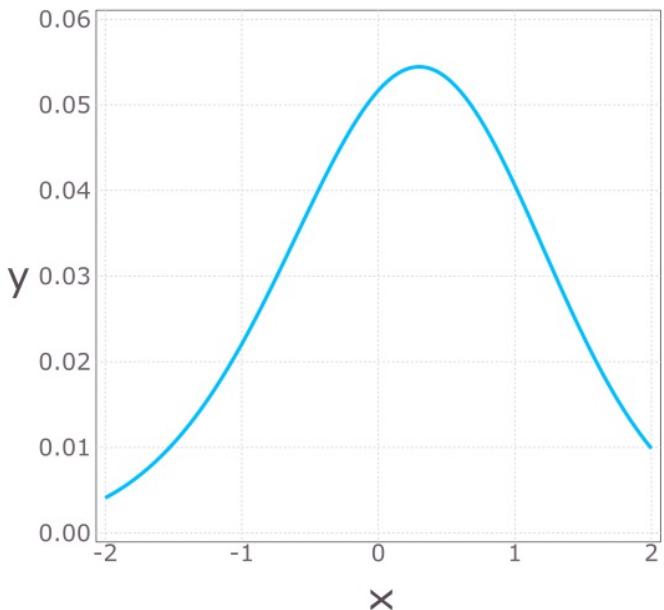
$\gamma = -1$



$\gamma = 0.5$



$\gamma = -0.5$



8f) with the given  $f(x, y)$ ,

the marginal distribution of  $x$  (analogously  $y$ ) is

$$f(x) = k \phi(x) + K |x| \int_{|y|=1} \phi(y) dy \quad |x| < 1$$
$$= k \phi(x) \quad |x| > 1$$

The conditional distribution given  $y=0$

is

$$f(x) = \frac{k \phi(x) \phi(0)}{K \phi(0)} = \phi(x) \quad -\infty < x < \infty$$

When  $y = \pm 1$ ,

$$f(x) = \frac{k \phi(x) \phi(\pm 1)}{k \phi(\pm 1)} = \phi(x) \quad -\infty < x < \infty$$

When  $y = \pm 0.5$  (analogously  $-0.5$ )

$$f(x) = \frac{k(\phi(x) \phi(0.5) + 0.5|x|)}{k \phi(0.5) + k 0.5 \int_{|x|=1} \phi(x) dx} \quad |x| < 1$$

$$= \frac{k \phi(x) \phi(0.5)}{k \phi(0.5) + 0.5 k \int_{|x| < 1} \phi(x) |x|}$$

where  $k$  is s.t.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x) \phi(y) \delta(|x| + |y|) \max\{|x|, |y|\} dx dy = 1$$