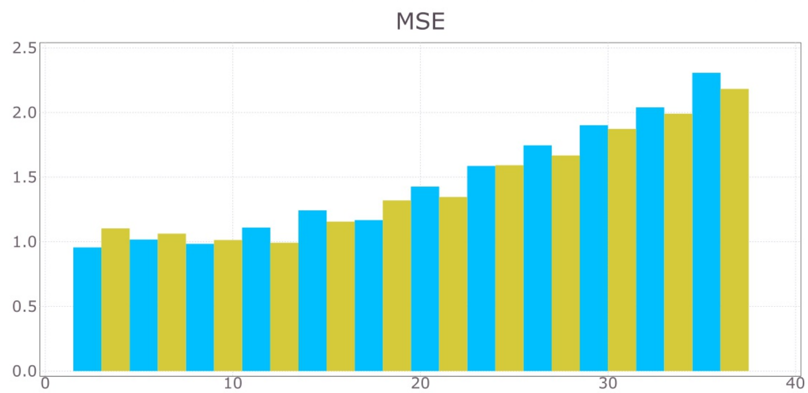
Caio Figueiredo Darshana Sunoj

Homework 1

Econometrics

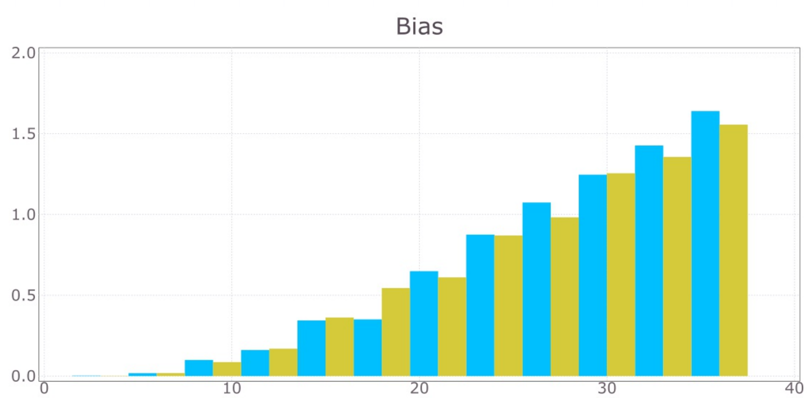
1) I believe that the main moral of this exercise is that as $d_z$ grows, i.e. the number of instruments grows, MSE also grows, as we can see:

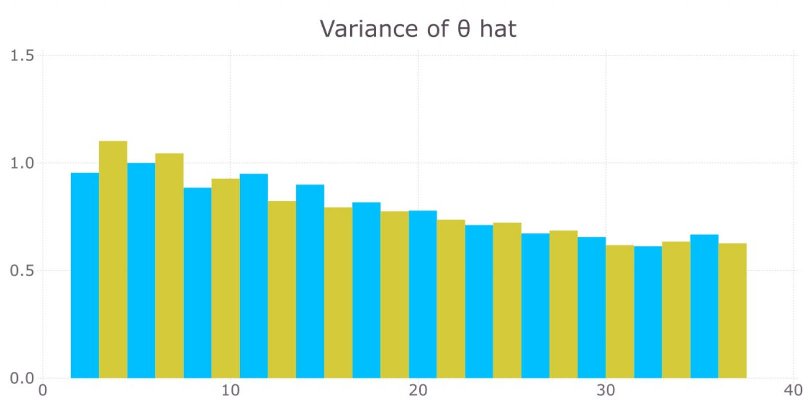



Blue: First element of $\theta$

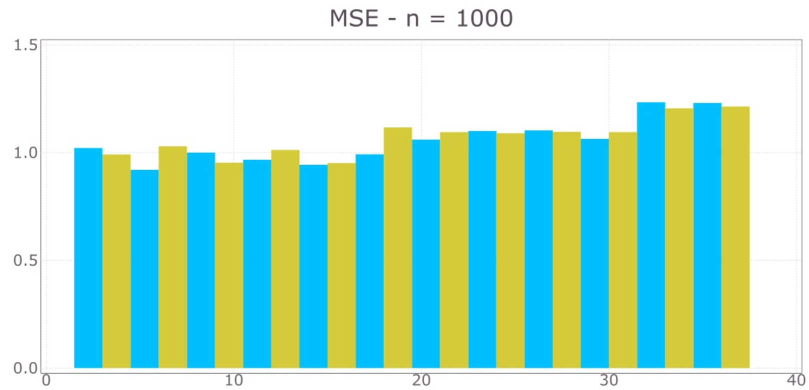
Green: Second element

This is mainly due a increase in the Square Bias of the estimator, since the Variance actualy decrease with $d_z$:



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

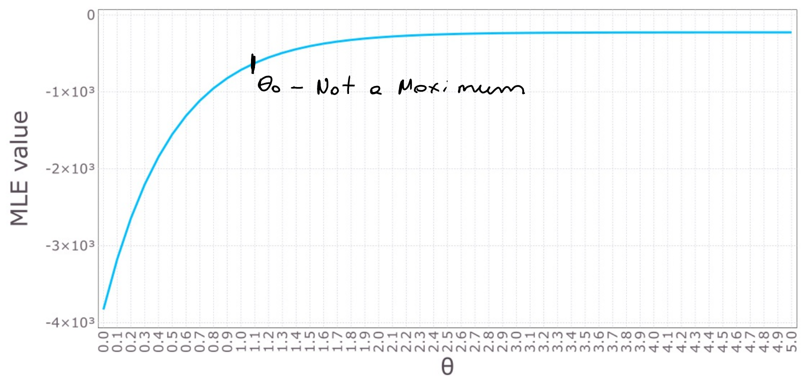
- MSE of $\hat{\theta}_{OS}$ grows as variance of $x$ decreases, specially when $(x'x)^{-1}x'y$ is being used. (More numerical imprecision??)

c) I was surprised by this results, I had to scale down a bit ($n \to 1000$) 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 $z(z'z)^{-1}z'$ 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: $z(z'z)^{-1}z'x$.

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

In fact, if we derive it:

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

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

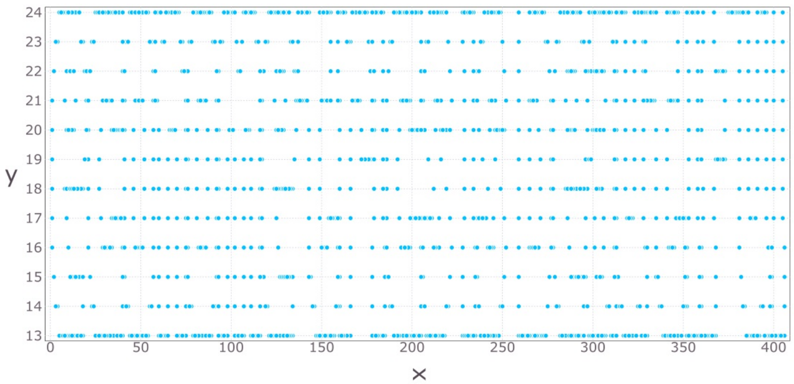
and at $\theta = \theta_0 \Rightarrow \exp(y_i-\theta_0) = u_i$

$$= \nabla_\theta Q_0(\theta_0) = -E\left[\frac{f_u'(u_i)}{f_u(u_i)}u_i\right] = -\int_0^\infty f_u'(u_i)\,u_i\,du$$

$$= -\left[\cancelto{0}{f_u(u_i)}\,u_i\right]_0^\infty + \int_0^\infty f_u(u_i)\,du_i = 1 \neq 0$$

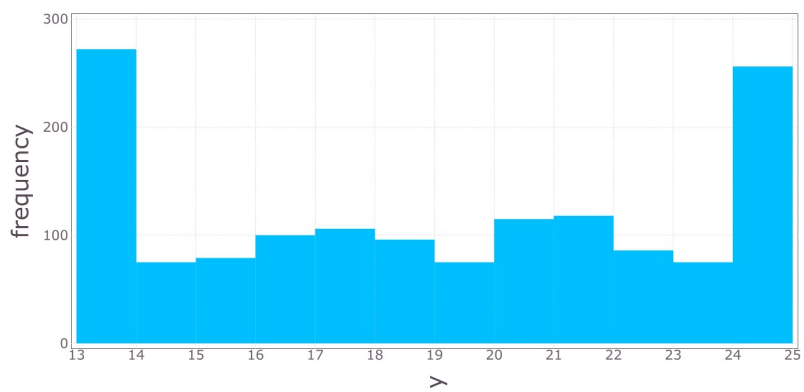
We seem to correct the problem by subtracting $\theta$ from the original MLE, which converges to $\theta_0$ as seen in the code.

4)




Oh! Dear...

Ok, it doensn't seem to have a corelation, let's dig deeper.



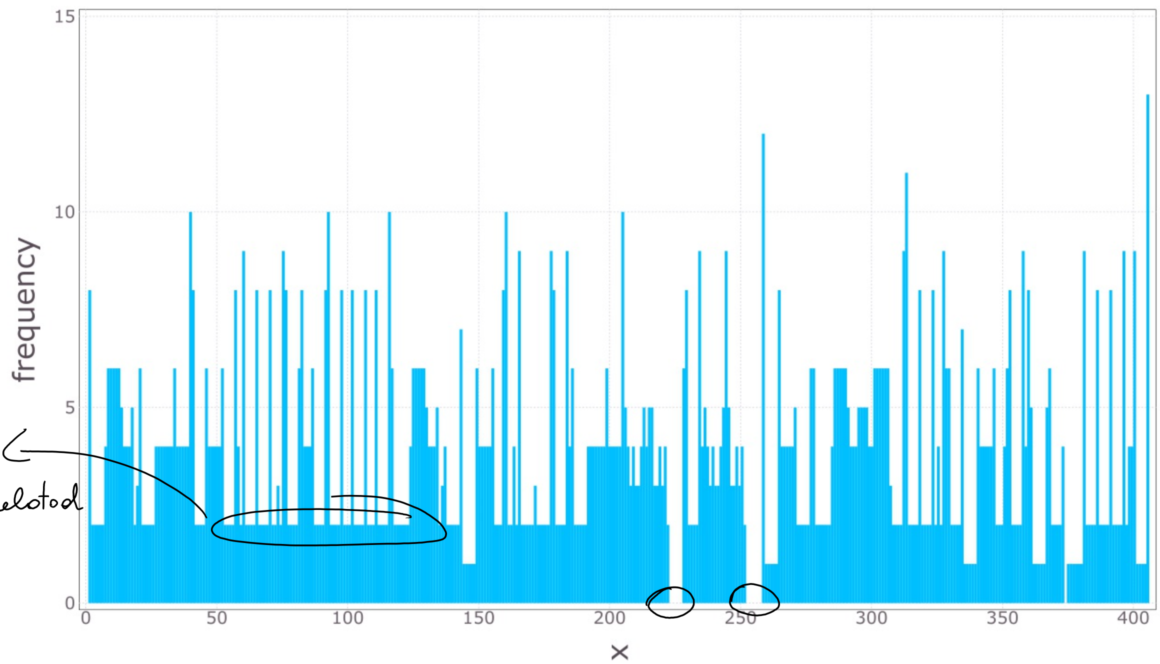

this is better, I assume that y has the follow probability function

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

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:

This seems to be autocorrelated




All ones


Hmm, it seems that given y, the same 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 $n \to \infty$. (It is not uniformly random!)

The probability of, given y, observing $X=x$ seems to be correlated to wether 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, follow a Markov Process on X.

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

$z_{xy}$ can only be 0 or 1
$z_{xy}$ is correlated to $z_{(x-1),y}$

I can then estimate:

$$P(z_{xy}=1 \mid z_{(x-1),y}=1, y=13) = 0.9$$
$$P(z_{xy}=0 \mid z_{(x-1),y}=1, y=13) = 0.1$$
$$P(z_{xy}=1 \mid z_{(x-1),y}=0, y=13) = 0.2$$
$$P(z_{xy}=0 \mid z_{(x-1),y}=0, y=13) = 0.8$$

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

5)

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 significally 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 cdf for a log normal is:

$$f(x) = \frac{1}{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)$$

where $\mu$ and $\sigma$ are such that $\ln(x) \sim N(\mu, \sigma)$

For MLE we take logs and have

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

For the maximization we ignore the nonparameters and have:

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

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

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

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

In our question we have:

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

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

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

$$Var(l) = Var(b) = \sigma_0^2 = exp\left(2\mu_l + s_l^2\right)\left(exp(s_l^2) - 1\right)$$
$$= exp\left(2\mu_b + s_b^2\right)\left(exp(s_b^2) - 1\right)$$

Finally, we have:

$$a = l \cdot b$$

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

substituting everything, we have the final MLE:

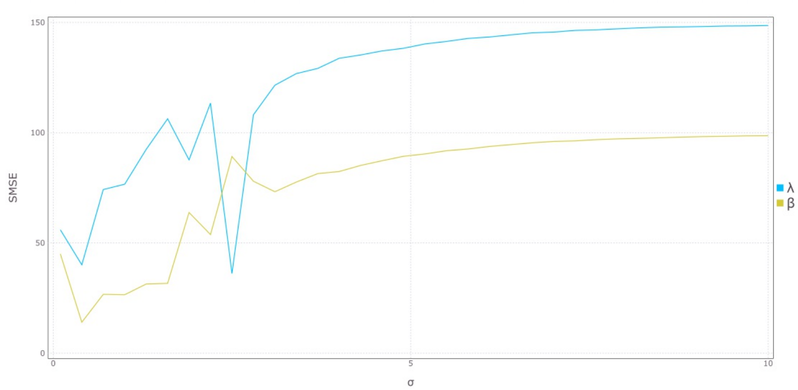
$$= -\sum_{i=1}^{n} \left\{ \ln\left(\ln\left(\left(1 + \frac{\sigma_0^2}{\lambda_0^2}\right)\left(1 + \frac{\sigma_0^2}{\beta_0^2}\right)\right)\right)^{1/2} + \frac{\left(\ln(a_i) - \ln\left[\left(\frac{\lambda_0}{\sqrt{1 + \frac{\sigma^2}{\lambda_0^2}}}\right)\left(\frac{\beta_0}{\sqrt{1 + \sigma_0^2/\beta_0^2}}\right)\right]\right)^2}{2\ln\left[\left(1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

b) Running the optimizer I found:

$\hat{\lambda} = 196.17$
$\hat{\beta} = 82.58$
$\hat{\sigma}_0 = 0.9813$

However the curve seems to be flat around this point, so the estimative may be dependant on initial values and others optimization parameters.

c)




This is odd, I used $\lambda_0 = 150$ and $\beta_0 = 100$ in this simulation, and... $SMSE_\lambda \xrightarrow[\sigma\to\infty]{} \lambda_0$ ? I don't see why this would happen.

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

$$\frac{\partial F(x,y)}{\partial x} = \varphi(x)\phi(y) - 0.5\,\phi(y)\,\phi(-y)\,(\varphi(x) - 2\phi(x)\varphi(x))$$

$$= \varphi(x)\phi(y)\left[1 - 0.5\,\phi(-y)\,(1 - 2\phi(x))\right]$$

$$= \underbrace{\varphi(x)}_{\geq 0}\underbrace{\phi(y)}_{\geq 0}\underbrace{\left[1 - \underbrace{0.5\,\phi(-y)}_{\leq 1} + \underbrace{\phi(-y)\phi(x)}_{\geq 0}\right]}_{>0}$$

$\dfrac{\partial F(x,y)}{\partial y} > 0$ 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 $\phi(-x)\phi(-y) < 1$

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

by the same logic

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

b) $$F(x) = \lim_{y\to\infty} F(x,y) = \phi(x)(1 - 0) = \phi(x)$$

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

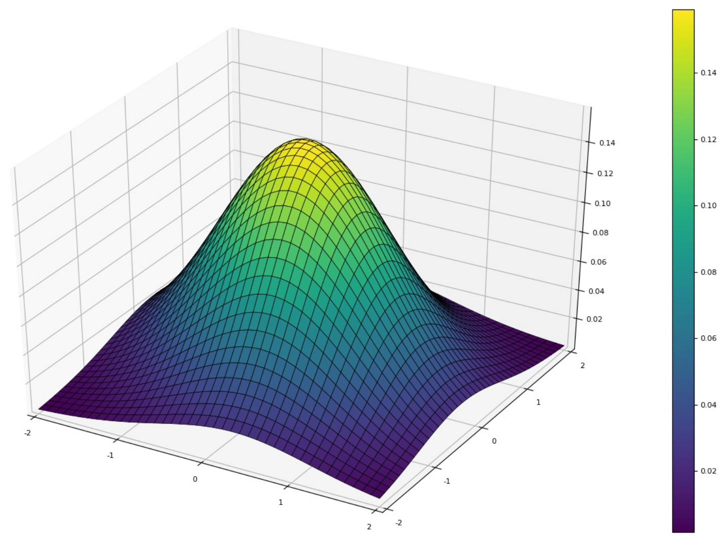
c) $$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial}{\partial y}\left(\varphi(x)\phi(y)\left[1 - 0.5(1-\phi(y)) + (1-\phi(y))\phi(x)\right]\right)$$

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

$$= \varphi(x)\Big[\varphi(y)\big[1+(1-\phi(y))(\phi(x)-0.5)\big] - \phi(y)(\phi(x)-0.5)\varphi(y)\Big]$$

$$= \varphi(x)\varphi(y)\big[1-0.5(1-2\phi(x))(1-2\phi(y))\big]$$

the pdf:




d) $$f(x \mid y=0) = \varphi(x)\varphi(0)\big[1-0.5(1-2\phi(x))(1-2\cdot 1/2)\big]$$

$$= \varphi(x)\cdot\varphi(0) \approx 0.4\,\varphi(x)$$

$$f(x \mid y=1) = \varphi(x)\varphi(1)\big[1-0.5(1-2\phi(x))(1-2\phi(1))\big]$$

$$\approx \varphi(x)\,0.24\big[1-0.5(1-2\phi(x))(1-1.68)\big]$$

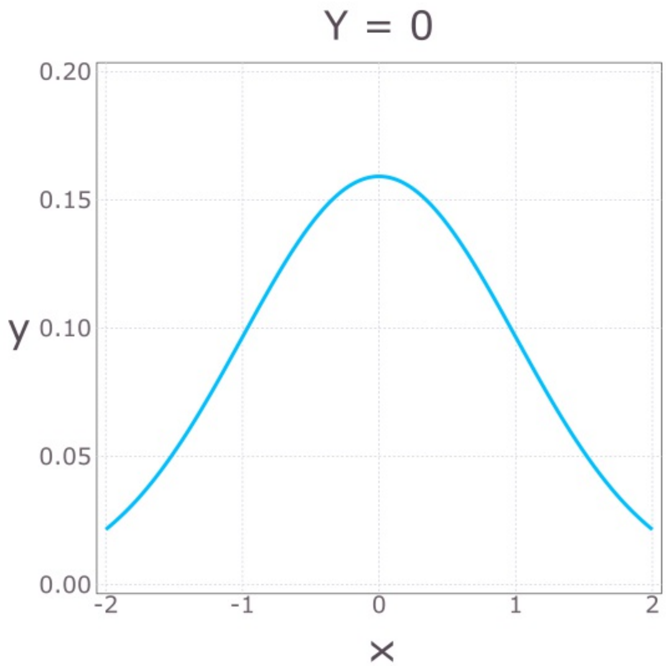
$$\approx 0.24\,\varphi(x)\big[1+0.34(1-2\phi(x))\big]$$

$$f(x \mid y=-1) \approx 0.24\,\varphi(x)\big[1-0.34(1-2\phi(x))\big]$$

$$f(x \mid y=0.5) = \varphi(x)\varphi(0.5)\big[1-0.5(1-2\phi(x))(1-2\phi(0.5))\big]$$

$$\approx 0.35\,\varphi(x)\big[1+0.19(1-2\phi(x))\big]$$

$$f(x \mid y=-0.5) \approx 0.35\,\varphi(x)\big[1-0.19(1-2\phi(x))\big]$$




given y = 0, x is normal.

given y = ±1, ±0.5 we have a normal-like imperfect curve.