Solution to Problem Set 5

Problem 1:

$$\mathrm{E}[|X-\hat{X}|]=\mathrm{E}[|X-g(Y)|]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|x-g(y)|\,f_{XY}(x,y)dxdy=$$

$$=\int_{-\infty}^{\infty}dy\ f_Y(y)\left[\int_{-\infty}^{\infty}|x-g(y)|\,f_{X|Y}(x|y)dx\right]=$$

$$=\int_{-\infty}^{\infty}dy\ f_Y(y)\left[\int_{-\infty}^{g(y)}(g(y)-x)f_{X|Y}(x|y)dx+\int_{g(y)}^{\infty}(x-g(y))f_{X|Y}(x|y)dx\right]$$

The above expression can be minimized with respect to $g(y)$ by minimizing the inner integral for every y separately. Derivation of the inner integral with respect to $g(y)$ and equalizing to zero will give a condition for the optimal value of $g(y)$ for every y separately:

$$\frac{d}{dg}\left[\int_{-\infty}^{g}(g-x)f_{X|Y}(x|y)dx+\int_{g}^{\infty}(x-g)f_{X|Y}(x|y)dx\right]=$$

$$=(g-g)f_{X|Y}(g|y)+\int_{-\infty}^{g}f_{X|Y}(x|y)dx-(g-g)f_{X|Y}(g|y)-\int_{g}^{\infty}f_{X|Y}(x|y)dx$$

$$=\int_{-\infty}^{g}f_{X|Y}(x|y)dx-\int_{g}^{\infty}f_{X|Y}(x|y)dx$$

$$\int_{-\infty}^{g_{opt}(y)}f_{X|Y}(x|y)dx-\int_{g_{opt}(y)}^{\infty}f_{X|Y}(x|y)dx=0\quad\Rightarrow\quad\int_{-\infty}^{g_{opt}(y)}f_{X|Y}(x|y)dx=\frac{1}{2}$$

In other words, the minimum is achieved when $g(y)$ is the median of the distribution of X given $Y=y$.

A further derivation of the expression with respect to $g(y)$ gives: $2f_{X|Y}(g(y)|y)\geq 0$, indeed we got a minimum.

Problem 2:

1. The optimal estimator is the conditional expectation estimator. As means of finding the conditional expectation, we will first find the conditional PDF:

$$f_{Y|X}(y|x)=\begin{cases}\dfrac{1}{x} & 0\leq y\leq x\\ 0 & o.w.\end{cases}$$

Therefore, the conditional expectation is:

$$E\{Y \mid X\} = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x) dy = \frac{1}{x}\int_0^x y dy = \frac{x}{2}$$

Namely, $\hat{Y}_{opt} = \dfrac{X}{2}$

This result can be reached directly since, given X, Y distributes uniformly in the section $[0, X]$, and, thus, the expectation is $\dfrac{X}{2}$.

2. Again, we need to find first the conditional PDF. For this cause, we will use Bayes' theorem:

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

$$f_{XY}(x, y) = f_X(x) \cdot f_{Y|X}(y \mid x) = \frac{1}{a}[u(x) - u(x-a)] \cdot \frac{1}{x}[u(y) - u(y-x)]$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \frac{1}{a}\int_0^a \frac{1}{x}[u(y) - u(y-x)] dx \qquad 0 \le y \le a$$

(the fact that $0 \le y \le a$ can be understood instantly from the givens of the problem. Another way of understanding this is by the fact that $u(y) - u(y-x) = 0, \forall y > a, y < 0$ when $0 \le x \le a$).

Furthermore, it is easy to see that $u(y-x) = 1$ only when $y - x > 0 \Rightarrow x < y$.

$$f_Y(y) = \frac{1}{a}\int_0^a \frac{1}{x} dx - \frac{1}{a}\int_0^y \frac{1}{x} dx = \frac{1}{a}\int_y^a \frac{1}{x} dx = \frac{1}{a}\ln(x)\Big|_y^a = \frac{1}{a}\ln\left(\frac{a}{y}\right) \qquad y \in [0, a]$$

That is to say, the conditional PDF is given by:

$$f_{X|Y}(x \mid y) = \frac{\frac{1}{a}[u(x) - u(x-a)] \cdot \frac{1}{x}[u(y) - u(y-x)]}{\frac{1}{a}\ln\left(\frac{a}{y}\right)} = \frac{[u(x) - u(x-a)] \cdot [u(y) - u(y-x)]}{x\ln\left(\frac{a}{y}\right)}$$

And, therefore, the conditional expectation:

$$E\{X \mid Y = y\} = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x \mid y) dx = \int_{-\infty}^{\infty} x \cdot \frac{[u(x) - u(x-a)] \cdot [u(y) - u(y-x)]}{x\ln\left(\frac{a}{y}\right)} dx =$$

$$= \frac{1}{\ln\left(\frac{a}{y}\right)}\int_y^a x \cdot \frac{1}{x} dx = \frac{a-y}{\ln\left(\frac{a}{y}\right)} \qquad y \in [0, a]$$

And from here we get: $\hat{X}_{opt} = \dfrac{a-Y}{\ln\left(\dfrac{a}{Y}\right)}$

Notice that if $Y = a$, then the only possible value of X is also a, and this is what the estimator gives (this can be checked by L'Hôpital's rule, or by looking at the graph later on).

3. The optimal estimator of Y from X is linear and, thus, is also the optimal linear estimator.

4. Let us find the estimator's constants:

$$E\{X\} = \frac{a}{2}$$
$$E\{Y\} = E\{E\{Y \mid X\}\} = E\left\{\frac{X}{2}\right\} = \frac{a}{4}$$

The second moment of a uniform random variable is given by:

$$E\{X^2\} = \frac{1}{a}\int_0^a x^2 dx = \frac{a^2}{3}$$

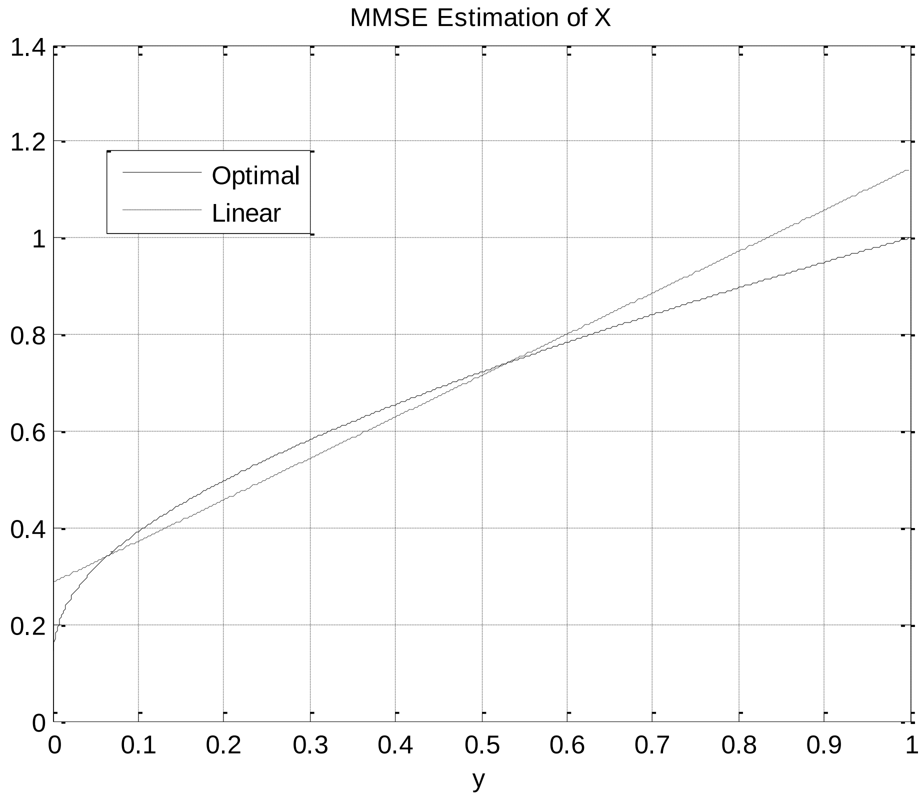
From here:

$$E\{Y^2\} = E\{E\{Y^2 \mid X\}\} = E\left\{\frac{X^2}{3}\right\} = \frac{1}{3}\frac{a^2}{3} = \frac{a^2}{9}$$
$$\Rightarrow VAR\{Y\} = E\{Y^2\} - E^2\{Y\} = \frac{a^2}{9} - \frac{a^2}{16} = \frac{7a^2}{144}$$
$$E\{XY\} = E\{E\{XY \mid X\}\} = E\{XE\{Y \mid X\}\} = E\left\{X\frac{X}{2}\right\} = \frac{1}{2}E\{X^2\} = \frac{a^2}{6}$$
$$COV(X,Y) = E\{XY\} - E\{X\}E\{Y\} = \frac{a^2}{6} - \frac{a}{2}\frac{a}{4} = \frac{a^2}{24}$$

In conclusion, the optimal linear estimator is given by:

$$\hat{X}_{BLE} = E\{X\} + \frac{COV(X,Y)}{VAR\{Y\}}(Y - E\{Y\}) = \frac{a}{2} + \frac{6}{7}\left(Y - \frac{a}{4}\right)$$

Following are graphs of both of the estimators of X from Y for $a = 1$:



Problem 3:

1. For $n<0$:

$$\mathrm{P}(Y=n)=0$$

For $n\geq 0$:

$$\mathrm{P}(Y=n)=\mathrm{P}(n\leq X<n+1)=\int_{n}^{n+1}\lambda e^{-\lambda x}dx=e^{-\lambda n}-e^{-\lambda(n+1)}=e^{-\lambda n}(1-e^{-\lambda})$$

2.

$$f_{X|Y}(x\,|\,y)=\frac{\mathrm{P}(Y=y\,|\,X=x)f_X(x)}{\mathrm{P}(Y=y)}$$

Notice that since Y is a deterministic function of X, it holds that:

$$\mathrm{P}(Y=y\,|\,X=x)=\mathrm{P}(Q(X)=y\,|\,X=x)=\begin{cases}1 & y=\lfloor x\rfloor\\ 0 & o.w.\end{cases}$$

Therefore:

$$f_{X|Y}(x\,|\,y)=\begin{cases}\dfrac{\lambda\cdot e^{-\lambda x}}{e^{-\lambda y}(1-e^{-\lambda})} & x\geq 0, y=\lfloor x\rfloor\\ 0 & o.w.\end{cases}$$

3. Since Y is a deterministic function of X, it follows that the estimator:

$$\hat{Y} = \lfloor X \rfloor$$

satisfies that the estimation error is:

$$e = \hat{Y} - Y = 0$$

Therefore, in particular, it is the optimal estimator in the sense of MMSE and in the sense of minimum probability of error.

4. The optimal estimator is the conditional expectation estimator. Let us calculate it:

$$\mathrm{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x \mid y)dx = \int_{y}^{y+1} x \frac{\lambda \cdot e^{-\lambda x}}{e^{-\lambda y}(1-e^{-\lambda})} dx = \frac{\lambda}{e^{-\lambda y}(1-e^{-\lambda})} \int_{y}^{y+1} x e^{-\lambda x} dx =$$

$$= \frac{\lambda}{e^{-\lambda y}(1-e^{-\lambda})}\left[-\frac{e^{-\lambda x}(\lambda x+1)}{\lambda^2}\right]\Bigg|_{x=y}^{x=y+1} = \frac{\lambda}{e^{-\lambda y}(1-e^{-\lambda})}\left[\frac{e^{-\lambda y}(\lambda y+1)}{\lambda^2} - \frac{e^{-\lambda(y+1)}(\lambda(y+1)+1)}{\lambda^2}\right] =$$

$$= \frac{(\lambda y+1) - e^{-\lambda}(\lambda(y+1)+1)}{\lambda(1-e^{-\lambda})} = y + \frac{1}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}}$$

$$\Rightarrow \quad \hat{X}_{opt,MMSE} = Y + \frac{1}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}}$$

5. Since the optimal MMSE estimator is linear to the measurements, it follows that it is also the optimal linear estimator. Namely:

$$\hat{X}_{opt,LMMSE} = Y + \frac{1}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}}$$

6. $Z = X - Y$ and, therefore, $f_{Z|Y}(z \mid y) = f_{X|Y}(z+y \mid y)$. Now, let us use the result from problem 3 section 2:

$$f_{X|Y}(z+y \mid y) = \begin{cases} \dfrac{\lambda \cdot e^{-\lambda(z+y)}}{e^{-\lambda y}(1-e^{-\lambda})} & z+y \geq 0, y = \lfloor z+y \rfloor \\ 0 & o.w. \end{cases}$$

Since y receives only natural values, it follows that the condition $y = \lfloor z+y \rfloor$ is equivalent to the condition $0 \leq z < 1$ and its occurrence leads to the condition $z + y \geq 0$ to hold, too. All in all:

$$f_{Z|Y}(z \mid y) = \begin{cases} \dfrac{\lambda \cdot e^{-\lambda z}}{(1-e^{-\lambda})} & , 0 \leq z < 1 \\ 0 & , \quad o.w. \end{cases}$$

7. We got that $f_{Z|Y}(z \mid y)$ is not dependent of y and, thus, Z, Y are independent and, of course, also uncorrelated.
8. Z, Y are independent and, therefore, the optimal estimator (and also the optimal linear estimator) is the expected value:

$$\hat{Z}_{opt} = E\{Z\} = \int z f_Z(z)dz = \frac{\lambda}{(1-e^{-\lambda})}\int_0^1 z e^{-\lambda z}dz = \frac{1-e^{-\lambda}(\lambda+1)}{\lambda(1-e^{-\lambda})} = \frac{1}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}}$$

Another simpler way (without integrals) of finding the expectation of Z is by using the conditional expectation that we found in problem 3:

$$E\{Z\} = E\{X-Y\} = E\{E\{X-Y \mid Y\}\} = E\{E\{X \mid Y\} - E\{Y \mid Y\}\} =$$
$$= E\left\{Y + \frac{1}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}} - Y\right\} = \frac{1}{\lambda} - \frac{e^{-\lambda}}{1-e^{-\lambda}}$$

This result gives us an intuition with regard to the optimal estimator of X from Y: the estimator takes the measurements (X rounded downwards) and adds to it the expected value of the error – namely, the expected value of the difference between the value of X and its rounded value.

Problem 4:

1. It holds that $v^T A v = X$ which is a Gaussian random variable for any ρ.
2. For any deterministic vector $(a\ b)$ and any ρ it holds that $(a\ b) \cdot A$ is a linear transformation of a Gaussian random vector, since

$$(a\ b) \cdot A = (a\ b)\begin{pmatrix} X & \rho Y \\ \rho Y & X \end{pmatrix} = (aX + b\rho Y\ \ bX + a\rho Y) = \begin{pmatrix} a & b\rho \\ b & a\rho \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix}$$

3. In this case we have $\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} X - Y \\ Y - X \end{pmatrix}$, hence Z_2 is deterministically known given Z_1, so $\hat{Z}_2 = -Z_1 = Z_2$ and the MSE is 0.

4. In this case we have $\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} Y + X \\ Y - X \end{pmatrix}$ where Z_2 and Z_1 are jointly Gaussian with expectations zero and are uncorrelated hence also independent, since
$$\mathbb{E}[Z_1 Z_2] = \mathbb{E}[(Y+X)(Y-X)] = \mathbb{E}[Y^2 - X^2] = 0.$$
This implies that $\hat{Z}_2 = \mathbb{E}(Z_2) = 0$ and the MSE is $Var(Y - X) = 2$.

5. In this case we have $\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} 2Y + X \\ Y + 2X \end{pmatrix}$. Due to Z_2 and Z_1 being jointly Gaussian the MMSE estimator is the linear MMSE estimator, thus:
$$\hat{Z}_2 = \frac{Cov(Z_1 Z_2)}{Var(Z_1)} Z_1 = \frac{\mathrm{E}[(2Y+X)(Y+2X)]}{Var(2Y+X)} Z_1 = \frac{4}{5} Z_1$$
$$MSE = Var(Z_2) - \frac{Cov(Z_1 Z_2)^2}{Var(Z_1)} = 5 - \frac{16}{5} = \frac{9}{5}$$

Problem 5:

$$\mathrm{E}[XYZ]=\mathrm{E}\left[\mathrm{E}\left[XYZ|X,Y\right]\right]=\mathrm{E}\left[XY\,\mathrm{E}\left[Z|X,Y\right]\right]\overset{(*)}{=}\mathrm{E}[XY(aX+bY)]=a\mathrm{E}\left[X^2Y\right]+b\mathrm{E}\left[XY^2\right]$$

when (*) is due to the fact that optimal estimator of part of a Gaussian random vector (for example Z), given any group of elements of the same vector (for example X,Y), is always linear:

$$\mathrm{E}\left[Z\,|\,X,Y\right]=aX+bY$$

Moreover:

$$\mathrm{E}\left[X^2Y\right]=\mathrm{E}\left[\mathrm{E}\left[X^2Y|X\right]\right]=\mathrm{E}\left[X^2\mathrm{E}\left[Y|X\right]\right]\overset{(**)}{=}\mathrm{E}\left[X^2cX\right]=c\mathrm{E}\left[X^3\right]=c\cdot 0=0$$

when (**) is also due to the linearity of the optimal estimator in the Gaussian case:

$$\mathrm{E}\left[Y\,|\,X\right]=cX$$

In the same way, we get $\mathrm{E}\left[XY^2\right]=0$. Therefore $\mathrm{E}[XYZ]=0$.

Problem 6:

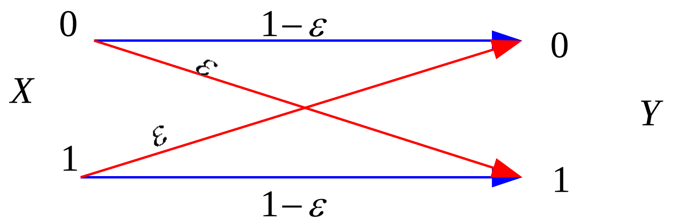
1. For $x,y\in\{0,1\}$:

$$\Pr\{Y=y\,|\,X=x\}=\Pr\{X\oplus N=y\,|\,X=x\}=\Pr\{x\oplus N=y\,|\,X=x\}=$$

$$=\Pr\{N=y\oplus x\,|\,X=x\}\underset{(1)}{=}\Pr\{N=y\oplus x\}=\begin{cases}\varepsilon & x\neq y\\ 1-\varepsilon & x=y\end{cases}$$

(1) N is independent of X

That is to say: with probability ε, the channel inverts the bit that was transmitted, and with probability $1-\varepsilon$, the channel preserves the bit that was transmitted. Symmetry exists relating to the interference that the channel causes to its input. It is accepted to describe this in the following graphical way:



where, on the arrows, the conditional probability that we calculated is written (which sometimes is named "the transition probability").

$$\Pr\{Y=y\}=\Pr\{Y=y\,|\,X=y\}\Pr\{X=y\}+\Pr\{Y=y\,|\,X\neq y\}\Pr\{X\neq y\}=$$

$$=\begin{cases}(1-\varepsilon)(1-p)+\varepsilon p & y=0\\ (1-\varepsilon)p+\varepsilon(1-p) & y=1\end{cases}$$

Finally:

$$\Pr\{X=x\mid Y=y\}=\frac{\Pr\{Y=y\mid X=x\}\Pr\{X=x\}}{\Pr\{Y=y\}}=\begin{cases}\dfrac{(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)+\varepsilon p} & x=0,y=0\\ \dfrac{\varepsilon p}{(1-\varepsilon)(1-p)+\varepsilon p} & x=1,y=0\\ \dfrac{\varepsilon(1-p)}{(1-\varepsilon)p+\varepsilon(1-p)} & x=0,y=1\\ \dfrac{(1-\varepsilon)p}{(1-\varepsilon)p+\varepsilon(1-p)} & x=1,y=1\end{cases}$$

2. The error event is translated into two disjoint events:

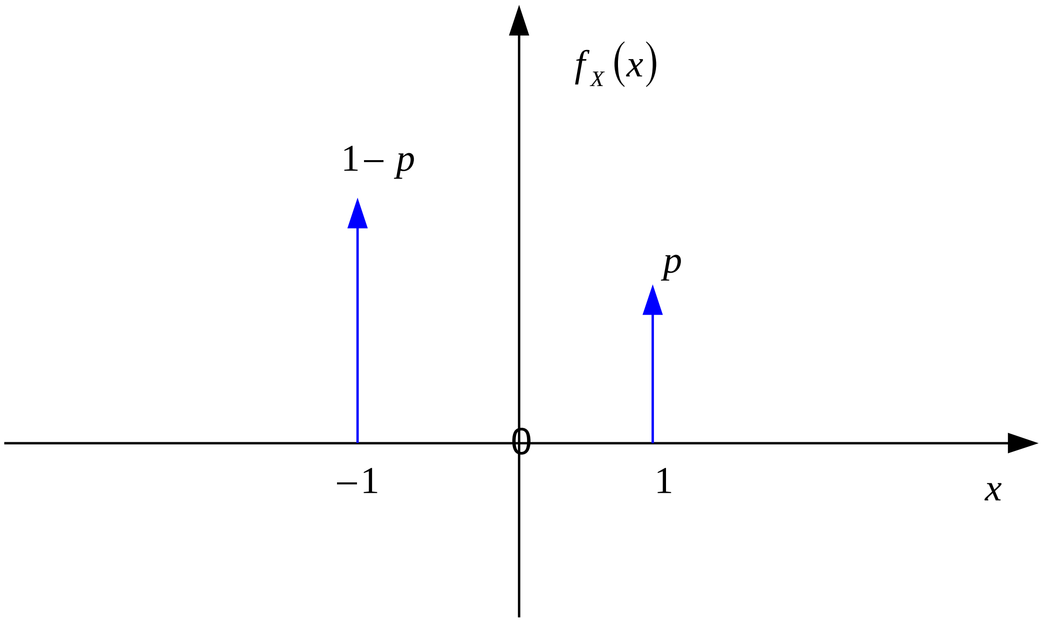
$$\{X\neq Y\}=\{X=0,Y=1\}\cup\{X=1,Y=0\}$$

$$\Pr\{X\neq Y\}=\Pr\{X=0,Y=1\}+\Pr\{X=1,Y=0\}=$$
$$=\Pr\{X=0\}\Pr\{Y=1\mid X=0\}+\Pr\{X=1\}\Pr\{Y=0\mid X=1\}=\varepsilon$$

3.

$$X=\begin{cases}-1 & w.p.\ 1-p\\ 1 & w.p.\ p\end{cases}\qquad\Rightarrow f_X(x)=(1-p)\delta(x+1)+p\delta(x-1)$$

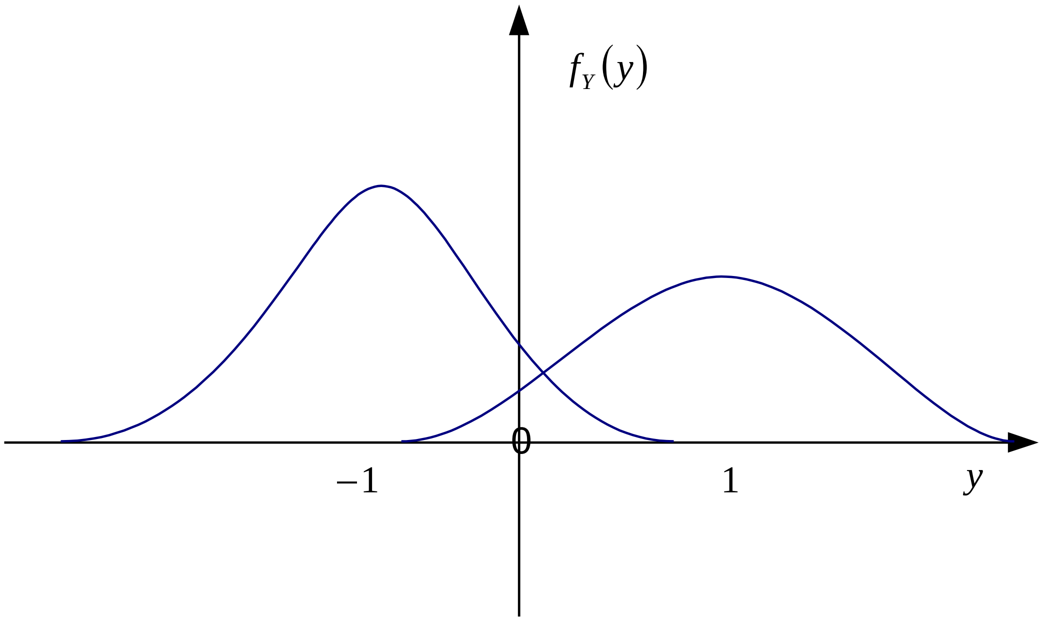
Moreover, seeing as how I is independent of N, and X is a deterministic function of I, it follows that X is independent of N.



In order to find the distribution of Y, we will make use of the formulas that deal with continuous and discrete variables:

$$f_Y(y)=f_{Y|X}(y\mid -1)\Pr\{X=-1\}+f_{Y|X}(y\mid 1)\Pr\{X=1\}=$$
$$=f_N(y+1)(1-p)+f_N(y-1)p=(1-p)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y+1)^2}{2\sigma^2}}+p\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-1)^2}{2\sigma^2}}$$

Notice that we got the weighted sum of two shifter Gaussians:



4. The error event is translated into two disjoint events:

$$\{Z \neq X\} = \{X = -1 \text{ and } Y > 0\} \cup \{X = 1 \text{ and } Y \leq 0\}$$

Therefore:

$$\begin{aligned}
&\Pr\{Z \neq X\} = \Pr\{Y > 0 \mid X = -1\}\Pr\{X = -1\} + \Pr\{Y \leq 0 \mid X = 1\}\Pr\{X = 1\} = \\
&= \Pr\{X + N > 0 \mid X = -1\}\Pr\{X = -1\} + \Pr\{X + N \leq 0 \mid X = 1\}\Pr\{X = 1\} = \\
&= \Pr\{-1 + N > 0 \mid X = -1\}\Pr\{X = -1\} + \Pr\{1 + N \leq 0 \mid X = 1\}\Pr\{X = 1\} = \\
&= \Pr\{N > 1 \mid X = -1\}\Pr\{X = -1\} + \Pr\{N \leq -1 \mid X = 1\}\Pr\{X = 1\} = \\
&\overset{(1)}{=} \Pr\{N > 1\}\Pr\{X = -1\} + \Pr\{N \leq -1\}\Pr\{X = 1\} = \\
&\overset{(2)}{=} \Pr\{N > 1\}\Pr\{X = -1\} + \Pr\{N > 1\}\Pr\{X = 1\} = \\
&= \Pr\{N > 1\}\left(\Pr\{X = -1\} + \Pr\{X = 1\}\right) = \\
&= \Pr\{N > 1\} = \Pr\left\{\frac{N}{\sigma} > \frac{1}{\sigma}\right\} \overset{(3)}{=} Q\left(\frac{1}{\sigma}\right)
\end{aligned}$$

(1) X, N are independent

(2) N distributed symmetrical around zero

(3) $\frac{N}{\sigma} \sim N(0,1)$