Contents

Chapter 6

Operations on Multiple Random Variables

6.1 Expected value of a function of random variables

The expected value of a function of N random variavles $X_1, X_2, \ldots, X_N$ can be calculated by the following integration:

$$E\left[g\left(X_1, X_2, \ldots, X_N\right)\right] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_N) dx_1 dx_2 \cdots dx_N$$

Note:
Notice that if we are interested only in calculating the mathematical expectation of $g\left(X_1, X_2, \ldots, X_N\right)$, we do NOT have to compute the joint p.d.f. $f_Y(y)$ of the newly defined random variable $Y = g(X_1, X_2, \ldots, X_N)$ and apply the definition of the mathematical expectation as:

$$E[Y] \triangleq \int_{-\infty}^{\infty} y \cdot f_Y(y) dy$$

Based on the above fact, we can define the joint moments, joint central moments, and so on, similarly to the case of single random variable. We will discuss some special cases of multiple random variables, and the generalization to the case of N random variables is left to you!!!

1. Joint moment:

Definition 6.1 For N given random variables $X_1, X_2, \ldots, X_N$, the joint moment of 4 random variables X_q, X_r, X_s and X_t is denoted and deined as follows:

$$\begin{aligned} m_{q,r,s,t}^{k,l,m,n} &\triangleq E\left[X_q^k X_r^l X_s^m X_t^n\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_q^k x_r^l x_s^m x_t^n f_4(x_q, x_r, x_s, x_t) dx_q dx_r dx_s dx_t \end{aligned}$$

In the above definition, $f_4(x_q, x_r, x_s, x_t)$ corresponds to the*marginal* probability density function of 4 random variables X_q, X_r, X_s and X_t, which could be obtained by integrating the joint p.d.f.$f_N(x_1, x_2, \ldots, x_N)$ with respect to the remaining $N-4$ random variables:

$$f_4(x_q, x_r, x_s, x_t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_N(x_1, x_2, \ldots, x_i, \ldots, x_N) dx_1 dx_2 \ldots dx_i \ldots dx_N$$

where $x_i \neq x_q, x_r, x_s, x_t$.

2. Joint central moment:

Definition 6.2 For N given random variables $X_1, X_2, \ldots, X_N$, the joint moment of two random variables X_q and X_m is denoted and deined as follows:

$$\mu_{q,m}^{k,l} \triangleq E\left[\left(X_q - m_q^1\right)^k \left(X_m - m_m^1\right)^l\right]$$

6.2 Transformation of multiple random variables

Given N random variables $X_1, X_2, \ldots, X_N$, we define a new set of N random variables $Y_1, Y_2, \ldots, Y_N$ as functions of $\{X_i\}_{i=1}^{N}$, i.e.

$$\begin{cases} Y_1 \triangleq T_1\left(X_1, X_2, \ldots, X_N\right) \\ Y_2 \triangleq T_2\left(X_1, X_2, \ldots, X_N\right) \\ \qquad\qquad \vdots \\ Y_N \triangleq T_N\left(X_1, X_2, \ldots, X_N\right) \end{cases}$$

where T_i's are *continuous* and have partial derivatives with respect to each X_i's for $i = 1, 2, \ldots, N$.

Also assume that there $\exists$ the inverses of the transformation T_i's $\ni$:

$$\begin{cases} X_1 \triangleq T_1^{-1}\left(Y_1, Y_2, \ldots, Y_N\right) \\ X_2 \triangleq T_2^{-1}\left(Y_1, Y_2, \ldots, Y_N\right) \\ \qquad\qquad \vdots \\ X_N \triangleq T_N^{-1}\left(Y_1, Y_2, \ldots, Y_N\right) \end{cases}$$

$\Longrightarrow$ This means that there is an *one-to-one* mapping between the X-space and the Y-space:

Figure 6.1: One-toone mapping between X-space and Y-space.

QUESTION: [1]
Given the joint probability density function $f_{NX}(x_1, x_2, \ldots, x_N)$ of $X_1, X_2, \ldots, X_N$, determine the corresponding joint probability density of the new set of random variables $Y_1, Y_2, \ldots, Y_N$, i.e. $f_{NY}(y_1, y_2, \ldots, y_N)$

Review of 1-dimensional case:

For a given r.v. X, let:

$$Y = g(X)$$

where $g(\cdot)$ corresponds to the 1-to-1 mapping bwteen X and Y.
Then, we have the following facts:

(i) Since $g(\cdot)$ is a one-to-one mapping, for a given value of y, we have:

$$x = g^{-1}(y)$$

(ii) ¿From the transformation $y = g(x)$, we have $dy = g^{'}(x)dx$ and thus:

$$\mathbf{dx} = \frac{dy}{g^{'}(x)} \triangleq |\mathbf{J}| \cdot \mathbf{dy} \tag{6.1}$$

where $|J|$ is called the *Jacobian*.

(iii) ¿From (i) and (ii), following relation holds:

$$\int_{y_1}^{y_2} f_Y(y)dy \equiv \int_{x_1}^{x_2} f_X(x)dx = \int_{g(x_1)}^{g(x_2)} f_X\left[g^{-1}(y)\right]|J|dy$$

¿From which we can derive the p.d.f. of Y as follows:

$$f_Y(y) = f_X\left[g^{-1}(y)\right]|J|$$

[1]Recall that for the one dimensional case where a new random variable Y is defines as $Y = g(X)$, the p.d.f. of the newly defined Y in terms of the p.d.f $f_X(x)$ is as follows:

$$f_Y(y) = \sum_{i=1}^{M} \frac{f_X(x_i)}{\left|\frac{dy}{dx}\right|_{x=x_i}}, \qquad x_i = g^{-1}(y), \quad i = 1, 2, \ldots, M$$

Now, consider the two dimensional case, extending the concepts of the above one dimensional case:

Figure 6.2: One-to-one mapping in R^2-space.

Notice that:

$$P\left\{(x_1, x_2) \in \mathcal{R}_{\S}\right\} = P\left\{(y_1, y_2) \in \mathcal{R}_{\dagger}\right\}$$

Therefore, we have:

$$\int\int_{\mathcal{R}_{\S}} f_{2X}(x_1, x_2)dx_1dx_2 = \int\int_{\mathcal{R}_{\dagger}} f_{2Y}(y_1, y_2)dy_1dy_2 \tag{6.2}$$

Here, the relation of incremental area between the X-space and the Y-space is as follows: [2]

$$dx_1dx_2 = |J|dy_1dy_2$$

where the Jacobian $|J|$ is defined as:

$$\begin{aligned} |J| = \text{Jacobian} \quad &\triangleq \frac{1}{\left|\det\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{bmatrix}\right|} \\ &\stackrel{\text{let}}{=} \frac{1}{\left|\det\left[\frac{\partial(y_1,y_2)}{\partial(x_1,x_2)}\right]\right|} \end{aligned}$$

Therefore, (6.2) becomes:

$$\int\int_{\mathcal{R}_{\dagger}} f_{2X}\left(T_1^{-1}(y_1, y_2), T_2^{-1}(y_1, y_2)\right)|J|dy_1dy_2 = \int\int_{\mathcal{R}_{\dagger}} f_{2Y}(y_1, y_2)dy_1dy_2$$

and the joint p.d.f. of Y_1 and Y_2 can be expressed as:

$$\mathbf{f_{2Y}(y_1, y_2) = f_{2X}\left(T_1^{-1}(y_1, y_2), T_2^{-1}(y_1, y_2)\right)|J|}$$

[2] The proof of this relation is beyond the scope of this course, but if you refer the one dimensional case in (6.1), the given relationship would seem reasonable...

Generalizing to the case of N dimension, the joint density of $Y_1, Y_2, \ldots, Y_N$ is in the following form:

$$\begin{aligned} & \mathbf{f_{NY}}(\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_N}) \\ = \; & \mathbf{f_{NX}}\left(\mathbf{T_1^{-1}}(\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_N}), \mathbf{T_2^{-1}}(\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_N}), \cdots, \mathbf{T_N^{-1}}(\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_N})\right) |\mathbf{J}| \end{aligned}$$

where

$$\begin{aligned} |J| = \text{Jacobian} \; & \triangleq \; \frac{1}{\left| \det \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_N} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_N} \\ & \vdots & \vdots & \\ \frac{\partial y_N}{\partial x_1} & \frac{\partial y_N}{\partial x_2} & \cdots & \frac{\partial y_N}{\partial x_N} \end{bmatrix} \right|} \\ & \stackrel{\text{let}}{=} \; \frac{1}{\left| \det \left[\frac{\partial(y_1, y_2, \ldots, y_N)}{\partial(x_1, x_2, \ldots, x_N)} \right] \right|} \end{aligned}$$

Example 6.1

Suppose two random variables X_1 and X_2 have the joint p.d.f. as follows:

$$f_{X_1X_2}(x_1, x_2) = \begin{cases} e^{-(x_1+x_2)}, & 0 \leq x_1, x_2 < \infty \\ 0, & \text{elsewhere} \end{cases}$$

A couple of new random variables Y_1 and Y_2 are given as functions of X_1 and X_2 below: [3]

$$\begin{cases} Y_1 = X_1 + X_2 \\ Y_2 = X_1 / X_2 \end{cases} \tag{6.3}$$

Determine the joint p.d.f. of Y_1 and Y_2.

[3] Notice that $f_{Y_1Y_2}(y_!, y_2) \neq 0$ only if $0 \leq y_1, y_2 < \infty$.

Solution:

Figure 6.3: One-to-one mapping from X-space to Y-space.

Before we compute the joint p.d.f. of Y_1 and Y_2, we first check that the given joing p.d.f. of X_1 and X_2 is a valid p.d.f., i.e.:

$$
\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1X_2}(x_1,x_2)dx_1dx_2 &= \int_0^{\infty}\int_0^{\infty} e^{-x_1}\cdot e^{-x_2}dx_1dx_2 \\
&= \int_0^{\infty} e^{-x_1}dx_1 \int_0^{\infty} e^{-x_2}dx_2 \\
&= 1\times 1 \\
&= 1
\end{aligned}
$$

Now, solving (6.3) for X_1 and X_2, we get:

$$
\begin{cases}
X_1 = Y_1Y_2/(1+Y_2) \\
X_2 = Y_1/(1+Y_2)
\end{cases}
\tag{6.4}
$$

and corresponding Jacobian is as follows:

$$
\begin{aligned}
|J| = \frac{1}{\left|\det\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2}\end{bmatrix}\right|} &= \frac{1}{\left|\det\begin{bmatrix} 1 & 1 \\ \frac{1}{x_2} & \frac{-x_1}{x_2^2}\end{bmatrix}\right|} \\
&= \frac{1}{\frac{x_1}{x_2^2}+\frac{1}{x_2}} = \frac{x_2^2}{x_1+x_2} \\
&= \frac{y_1^2}{(1+y_2)^2}\cdot\frac{1}{y_1} = \frac{y_1}{(1+y_2)^2}
\end{aligned}
\tag{6.5}
$$

¿From (6.4) and (6.5), we get:

$$\begin{aligned} f_{Y_1Y_2}(y_1,y_2) &= f_{X_1X_2}(x_1,x_2)\cdot|J| \\ &= e^{-(x_1+x_2)}\cdot\frac{x_2^2}{x_1+x_2} \\ &= e^{-\frac{y_1y_2+y_1}{1+y_2}}\cdot\frac{y_1}{(1+y_2)^2} \\ &= e^{-y_1}\cdot\frac{y_1}{(1+y_2)^2} \qquad \text{for } 0\le y_1,y_2<\infty \end{aligned}$$

Check: The validity of $f_{Y_1Y_2}(y_1,y_2)$ by showing the integration below:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Y_1Y_2}(y_1,y_2)dy_1dy_2 \overset{?}{=} 1$$

proof:

$$\begin{aligned} \text{LHS} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Y_1Y_2}(y_1,y_2)dy_1dy_2 &= \int_0^{\infty}\int_0^{\infty} e^{-y_1}\cdot\frac{y_1}{(1+y_2)^2}dy_1dy_2 \\ &= \underbrace{\int_0^{\infty} y_1e^{-y_1}dy_1}_{(1)}\cdot\underbrace{\int_0^{\infty}\frac{1}{(1+y_2)^2}dy_2}_{(2)} \\ &= 1\times 1 \\ &= 1 \\ &= \text{RHS} \end{aligned}$$

(cf.) Derivation of the integrals (1) and (2): see below [4]

---

[4]

(1) The 1st integration by parts:

$$\int_0^{\infty} y_1e^{-y_1}dy_1 = \left[-y_1e^{-y_1}\right]_0^{\infty} + \int_0^{\infty} e^{-y_1}dy_1 = \left[-e^{-y_1}\right]_0^{\infty} = 1$$

(2) The 2nd integration by change of variable:

$$\int_0^{\infty}\frac{1}{(1+y_2)^2}dy_2 = \left[-\frac{1}{1+y_2}\right]_0^{\infty} = 1$$

6.3 Estimation theory

Suppose we have a system (or communication channel) with an *input* and a *output*:

Figure 6.4: A communication channel.

With the observation $X = x$ on our hand, we want determine the most probable value of Y which caused X:

$$\widehat{\mathbf{Y}} = \mathbf{g}(\mathbf{X})|_{\mathbf{X}=\mathbf{x}}$$

: Best estimate of Y based on X

Criterion: *Least Mean Squared Error* (among many possible choices!)

$$\text{“ Choose } \widehat{Y} \quad \ni: \quad e = E\left[(Y - \widehat{Y})^2\right] \text{ is minimum ”, i.e.}$$

$$\widehat{Y} = \text{argmin}_{g(X)} E\left[(Y - g(X))^2\right]$$

Procedure:
First, we caculate the mean squared error as follows: [5]

$$
\begin{aligned}
e &\triangleq E\left[(Y-\widehat{Y})^2\right] \\
&= E\left[(Y-g(X))^2\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{y-g(x)\}^2\cdot f_{XY}(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{y-g(x)\}^2\cdot f_{Y|X}(y|x)\cdot f_X(x)dxdy \\
&= \int_{-\infty}^{\infty} f_X(x)\left\{\int_{-\infty}^{\infty} y^2 f_{Y|X}(y|x)dy - 2g(x)\int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy \right. \\
&\qquad \left. + g^2(x)\int_{-\infty}^{\infty} f_{Y|X}(y|x)dy\right\} dx \\
&= \int_{-\infty}^{\infty} f_X(x)\left\{E[Y^2|X=x] - 2g(x)\cdot E[Y|X=x] + g^2(x)\right\} dx \\
&= \int_{-\infty}^{\infty} f_X(x)\left\{E[Y^2|X=x] - 2g(x)\cdot E[Y|X=x] + g^2(x) + E^2[Y|X=x] \right. \\
&\qquad \left. - E^2[Y|X=x]\right\} dx \\
&= \int_{-\infty}^{\infty} f_X(x)\left[\{g(x)-E[Y|X=x]\}^2 + E[Y^2|X=x] - E^2[Y|X=x]\right] dx
\end{aligned}
$$

: What $g(x)$ makes the above mean squared error be minimum?

1. The best LMSEE(least mean squared error estimate): [6]

$$\widehat{Y} = g(x) = E[Y|X=x]$$

2. The resultant minimum mean squared error:

$$e_{min} = \int_{-\infty}^{\infty} f_X(x)\left\{E[Y^2|X=x] - E^2[Y|X=x]\right\} dx$$

: function of x

[5] Be reminded that $X = x$ is given, or observed!!!
[6] Notice that the best LMSEE is the **conditional expectation** of Y given $X = x$.

Best Linear Estimate*: special case*

The best *linear* estimate of Y given a value of X is in the following form:

$$\widehat{Y} = g(X) = aX + b \qquad \text{where } \ a, b \text{ are constants}$$

$\Rightarrow$ We want to determine the values of a and $b \ni$: $E\left[(Y - \widehat{Y})^2\right]$ is minimum:

Procedure:

Again, we first calculate the mean squared error:

$$\begin{aligned} e &\triangleq E\left[(Y - \widehat{Y})^2\right] \\ &= E\left[(Y - aX - b)^2\right] \\ &= E[Y^2] + a^2E[X^2] + b^2 - 2aE[XY] - 2bE[Y] + 2abE[X] \end{aligned} \tag{6.6}$$

Now, we take the partial derivative of (6.6) with respect to a and b:

$$\frac{\partial e}{\partial b} = 2b - 2E[Y] + 2aE[X] = 0$$

which gives [7]

$$b = E[Y] - aE[X] \tag{6.7}$$

Inserting (6.7) into (6.6), we get;

$$\begin{aligned} e &= E[Y^2] + a^2E[X^2] + \{E[Y] - aE[X]\}^2 - 2aE[XY] - 2\{E[Y] - aE[X]\}E[Y] \\ &\quad +2a\{E[Y] - aE[X]\}E[X] \\ &= \left\{E[Y^2] - E^2[Y]\right\} - 2a\{E[XY] - E[X]E[Y]\} + a^2\left\{E[X^2] - E^2[X]\right\} \\ &\triangleq \sigma_Y^2 - 2aC_{XY} + a^2\sigma_X^2 \end{aligned} \tag{6.8}$$

[7] Notice that $\frac{\partial^2 e}{\partial b^2} = 2 > 0$, which means that e is convex at the given value of b, thus providing the minimum.

Now take the partial derivative of (6.8) w.r.t. a, then we get;

$$\frac{\partial e}{\partial a} = -2C_{XY} + 2a\sigma_X^2 = 0$$

which gives [8]

$$a = \frac{C_{XY}}{\sigma_X^2} = \frac{\rho_{XY}\sigma_X\sigma_Y}{\sigma_X^2} = \rho_{XY}\frac{\sigma_Y}{\sigma_X} \tag{6.9}$$

Therefore, the best LINEAR estimator of Y given X is as follows:

$$\begin{aligned}
\widehat{Y} &= aX + b \\
&= aX + m_Y - am_X \\
&= a\left(X - m_X\right) + m_Y \\
&= \rho_{XY}\frac{\sigma_Y}{\sigma_X}\left(X - m_X\right) + m_Y
\end{aligned}$$

And the resultant minimum mean squared error is from (6.8);

$$\begin{aligned}
e_{min} &= \sigma_Y^2 - 2\rho_{XY}\frac{\sigma_Y}{\sigma_X}C_{XY} + \rho_{XY}^2\frac{\sigma_Y^2}{\sigma_X^2}\sigma_X^2 \\
&= \sigma_Y^2 - 2\rho_{XY}^2\sigma_Y^2 + \rho_{XY}^2\sigma_Y^2 \\
&= \sigma_Y^2\left(1 - \rho_{XY}^2\right)
\end{aligned}$$

In general, when multiple observationa are available:

Figure 6.5: A multi-channel communication system.

[8]Notice that $\frac{\partial^2 e}{\partial a^2} = 2\sigma_X^2 > 0$, which means that e is convex at the given value of a, thus providing the minimum.

The best estimator $\widehat{Y}$ based on multiple observations $X_1, X_2, \ldots, X_N$ is in the following form:

1. **Non-linear estimator:** $\widehat{Y} = g(X_1, X_2, \ldots, X_N)$

$$\begin{aligned}\widehat{Y} &= g(X_1, X_2, \ldots, X_N) \\ &= \text{argmin}_{g(\underline{X})} E\left[\{Y - g(X_1, X_2, \ldots, X_N)\}^2\right]\end{aligned}$$

   where $\underline{X} = (X_1, X_2, \ldots, X_N)$.

   Following a similar procedure as in the case of single observation, we get:

$$\begin{aligned}\widehat{Y} &= g(X_1, X_2, \ldots, X_N) \\ &= E[Y|X_1, X_2, \ldots, X_N] \\ &\quad : \text{conditional expectation}\end{aligned}$$

2. **Linear estimator:** $\widehat{Y} = \sum_{i=1}^{N} a_i X_i$

$$\begin{aligned}\widehat{Y} &= \sum_{i=1}^{N} a_i X_i \\ &= \text{argmin}_{\{a_i\}} \underbrace{E\left[\left\{Y - \sum_{i=1}^{N} a_i X_i\right\}^2\right]}_{e}\end{aligned}$$

   By taking partial derivatives of the mean squared error e w.r.t. $a_1, a_2, \ldots, a_N$ successively, we get;

$$a_i = \rho_i \frac{\sigma_Y}{\sigma_{X_i}} \quad \text{where} \quad \rho_i = \frac{C_{X_i Y}}{\sigma_{X_i}\sigma_Y}$$

   Thus, the best linear estimator becomes:

$$\widehat{Y} = \sum_{i=1}^{N} \rho_i \frac{\sigma_Y}{\sigma_{X_i}} \cdot X_i$$

proof: assignment

Example 6.2

Consider the following communication channel, where the input signal Y and the channel noise N are assumed to be statistically independent. Suppose we know the means and the variances of Y and N, i.e. m_Y, m_N, σ_Y^2 and σ_N^2.

Then, what is the best linear estimate of Y based on the observation X?

Figure 6.6: A communication channel with additive noise.

Solution:

We know that the best linear estimator is in the following form:

$$\widehat{Y} = \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - m_X) + m_Y$$

All we need is computing m_X, σ_X, and ρ_{XY}:

(i) Mean:

$$m_X = E[X] = E[Y+N] = E[Y] + E[N] = m_Y + m_N$$

(ii) Variance: [9]

$$\begin{aligned}
\sigma_X^2 &= E[X^2] - E^2[X] \\
&= \left\{E[Y^2] - m_Y^2\right\} + 2m_Y m_N + \left\{E[N^2] - m_N^2\right\} - 2m_Y m_N \\
&= \sigma_Y^2 + \sigma_N^2
\end{aligned}$$

[9] Note that $E[X^2] = E[Y^2 + 2YN + N^2] = E[Y^2] + 2E[Y]E[N] + E[N^2]$.

(iii) Covariance:

$$\begin{aligned}
C_{XY} &= E\left[(X - m_X)(Y - m_Y)\right] \\
&= E[XY] - m_X m_Y \\
&= E[(Y + N)Y] - (m_Y + m_N)m_Y \\
&= \left\{E[Y^2] - m_Y^2\right\} + E[Y]E[N] - m_Y m_N \\
&= \sigma_Y^2
\end{aligned}$$

From which we get the correlation coefficient as:

$$\rho_{XY} = \frac{C_{XY}}{\sigma_X \sigma_Y} = \frac{\sigma_Y}{\sigma_X}$$

Therefore, the best linear estimator becomes:

$$\begin{aligned}
\widehat{Y} &= \frac{\sigma_Y^2}{\sigma_X^2}(X - m_X) + m_Y \\
&= \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_N^2}(X - m_Y - m_N) + m_Y
\end{aligned}$$

The corresponding minimum mean squared error is :

$$\begin{aligned}
e_{min} &= \sigma_Y^2(1 - \rho_{XY}^2) \\
&= \sigma_Y^2\left(1 - \frac{\sigma_Y^2}{\sigma_X^2}\right) \\
&= \sigma_Y^2\left(1 - \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_N^2}\right) \\
&= \frac{\sigma_Y^2 \sigma_N^2}{\sigma_Y^2 + \sigma_N^2}
\end{aligned}$$