EE 631: Estimation and Detection
Part 3

Dr. Mehrdad Soumekh

Maximum Likelihood (ML) decision

Minimum probability of error or MAP decision rule is given by:
($C_{ii} = 0$ and $C_{ij} = 1$, $i \neq j$)

$$P_m \Lambda_m(\vec{R}) \gtrless^{\text{not } H_k}_{\text{not } H_m} P_k \Lambda_k(\vec{R})$$

- decision: $\max_{[k]} P_k \Lambda_k(\vec{R})$.
If a priori probabilities i.e. P_i's are not known, one option would be to assume equally probable hypotheses; i.e. $P_i = \frac{1}{M}$, $i = 0, 1, \cdots, M-1$.
For this choice of a priori probabilities, the test becomes

$$\Lambda_m(\vec{R}) \gtrless^{\text{not } H_k}_{\text{not } H_m} \Lambda_k(\vec{R})$$

$$\text{or } p_M(\vec{R}) \gtrless^{\text{not } H_k}_{\text{not } H_m} p_K(\vec{R})$$

- decision: $\max_{[k]} p_k(\vec{R})$ i.e. choose the hypothesis that yields the largest $p_k(\vec{R})$ or $\Lambda_k(\vec{R})$. This receiver is known as the maximum likelihood (ML) receiver or detector.
We showed that the decision threshold in the binary hypothesis testing, i.e.

$$\Lambda(\vec{R}) \gtrless^{\text{select } H_1}_{\text{select } H_0} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})} \triangleq \eta$$

depends on the a priori probability P_0.

Thus the statistic used for the decision making, i.e. $\Lambda(\vec{R})$ is invariant in a priori probability P_0. This implies that the sufficient statistic is unaffected for the ML detector. Only the threshold η varies with P_0.

Minimax detector

This is also used for the case of unknown a priori probabilities.
For a given channel, the risk is a function of P_i $\forall$ $i = 0, 1, \cdots, M-1$.
We can inspect all possible risk functions $\Re(\vec{P})$ in the domain of $\vec{P}$. We then choose the $\vec{P}$ values that maximize the risk function (worst case scenario), and set up the decision based on that a priori vector.

- decision: $\min[max_{\vec{P}} \Re(\vec{P})] = \min[\vec{\Re}(\vec{P}_{max})]$.
- example: Binary decision

$$\Lambda(\vec{R}) \gtrless^{\text{select } H_1}_{\text{select } H_0} \eta = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

For minimax detector, use

$$\Lambda(\vec{R}) \gtrless^{\text{select } H_1}_{\text{select } H_0} \eta^* = \frac{P_0^*(C_{10} - C_{00})}{P_1^*(C_{01} - C_{11})}$$

where $P_1^* = (1 - P_0^*)$

Consider:

$$H_0 : r_i \overset{\text{iid}}{\sim} \mathbb{N}(0, \sigma_0^2)$$
$$H_1 : r_i \overset{\text{iid}}{\sim} \mathbb{N}(0, \sigma_1^2)$$

where $i = 1, \cdots, N$ and $\sigma_1 > \sigma_0$. Observation pdf under H_K, $K = 0, 1$ is:

$$p(\vec{R}\big|H_K) = \frac{1}{(\sqrt{2\pi}\sigma_K)^N} exp\Big[- \frac{\sum_{i=1}^N r_i^2}{2\sigma_K^2}\Big]$$

Likelihood function:

$$\begin{aligned} \lambda(\vec{R}) &= \tfrac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)} \\ &= \big(\tfrac{\sigma_0}{\sigma_1}\big)^N exp\Big[\tfrac{1}{2}\big(\tfrac{1}{\sigma_0^2} + \tfrac{1}{\sigma_1^2}\big)\textstyle\sum_{i=1}^N r_i^2\Big] \gtrless^{\text{select } H_1}_{\text{select } H_0} \eta \end{aligned}$$

Log-likelihood function:

$$\ln \lambda(\vec{R}) = N \ln \big(\frac{\sigma_0}{\sigma_1}\big) + \Bigg[\frac{1}{2}\bigg(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2}\bigg)\sum_{i=1}^N r_i^2\Bigg] \gtrless^{\text{select } H_1}_{\text{select } H_0} \ln \eta$$

Thus the sufficient statistic can be expressed via:

$$l(\vec{R}) = \underbrace{\frac{1}{N}\sum_{i=1}^N r_i^2}_{\overline{R}^2} \gtrless^{\text{select } H_1}_{\text{select } H_0} \frac{\ln \eta + N \ln \frac{\sigma_1}{\sigma_0}}{\frac{N}{2}\big(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\big)} \triangleq \gamma$$

Test: General Bayes decision

$$\overline{R}^2 \gtrless^{\text{select } H_1}_{\text{select } H_0} \gamma$$

The r.v. $l(\vec{R}) = \overline{R}^2$ is central chi-square distributed with N degrees of freedom. The distribution is central since r_i's have zero mean.
$E(\overline{R}^2\big|H_K) = \sigma_K^2$, for $K = 0, 1$. Variance of $\overline{R}^2$ decreases as N increases.

Minimum probability of error as a function of the a priori probabilities

In general, probability of error P_e is given as:

$$\begin{aligned} P_e &= P_0.P_F + P_1.P_M \\ &= P_0.P_F + (1 - P_0).P_M \\ &= P_M + P_0(P_F - P_M) \end{aligned}$$

This expression indicates that the probability of error P_e is a linear function of P_0 if both P_F and P_M are invariant in P_0.
Note that in a Bayes decision rule, the threshold and as a result P_F and P_M are functions of P_0 and nonlinear as shown in figure 1.
In a general Bayes rule problem, P_F and P_M are defined as:

$$P_F = \textstyle\int_{\gamma(P_0)}^{\infty} p(l\big|H_0)dl$$

$$P_M = \textstyle\int_{-\infty}^{\gamma(P_0)} p(l\big|H_1)dl$$

Minimum probability of error as a function of a priori probabilities is given by:

$$P_e = P_0.P[l > \gamma(P_0)\big|H_0] + (1 - P_0)P[l < \gamma(P_0)\big|H_1]$$

where $\gamma(P_0) \triangleq g[\eta(P_0)] = g(\frac{P_0}{1-P_0})$.
$\eta(P_0) = \frac{P_0}{1-P_0}$ with $C_{00} = C_{11} = 0$ and $C_{10} = C_{01} = 1$.

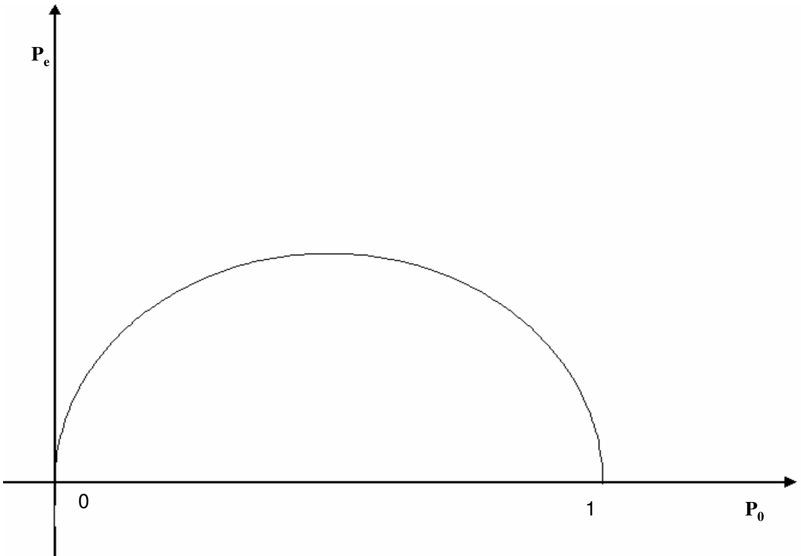

Figure 1. Plot of P_0 vs. P_e for Bayes decision rule

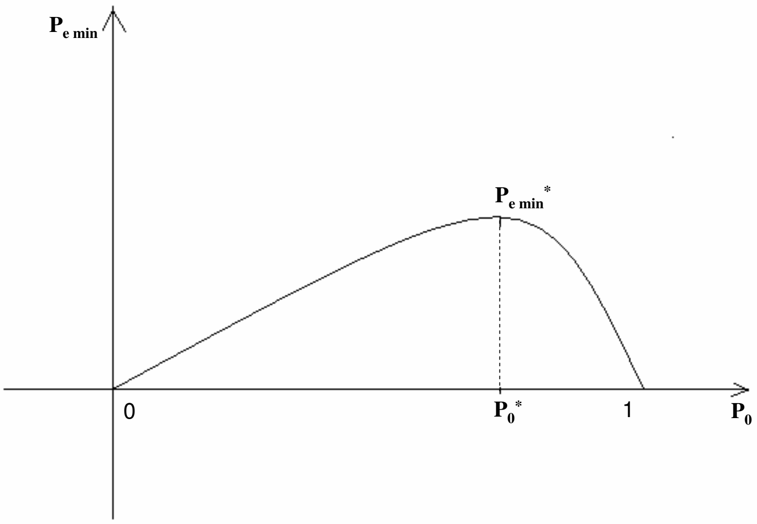

Figure 2. Plot of P_0 vs. $P_{e_{min}}$

Suppose $P_{e_{min}}$ takes on its maximum value at $P_0 = P_0^*$ (i.e. minimax detector with unitary cost) as shown in figure (2)
Select the the threshold based on $P_0 = P_0^*$.
Therefore, $\eta(P_0^*) = \frac{P_0^*}{1-P_0^*}$ and $\gamma(P_0^*) = g(\frac{P_0^*}{1-P_0^*})$. Once a threshold is selected based on $P_0 = P_0^*$, then the resultant probability of error for the actual P_0 is:

$$P_e^*(P_0) = P_0.P_F^* + (1 - P_0).P_M^*$$
$$= P_M^* + P_0(P_F^* - P_M^*)$$

which is a linear function of P_0.

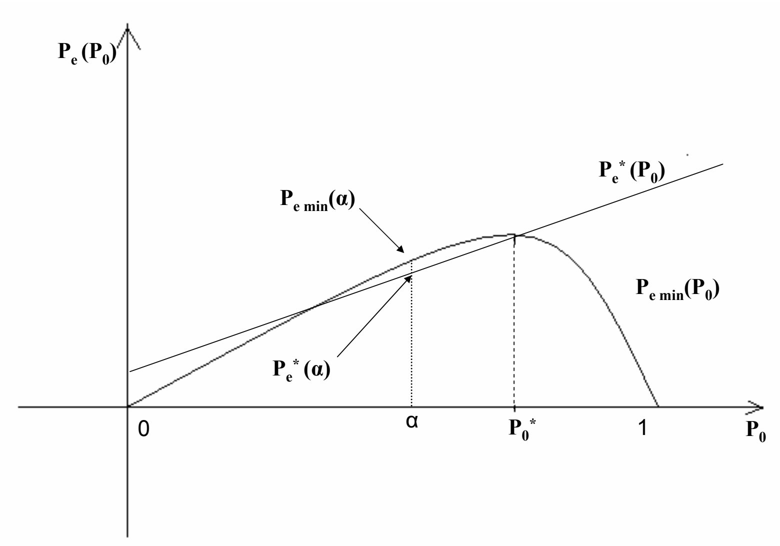

Figure 3.

Consider the curve for P_e^* to be as shown in figure 3. We know that $P_{e_{min}}(P_0)$ should always yield the minimum probability of error. But from figure 3, we have $P_{e_{min}}(\alpha) > P_e^*(\alpha)$ which is not possible. Thus the distribution of the line $P_e^*(P_0)$ should be above the curve $P_{e_{min}}(P_0)$ at every point except one where the two curves touch each other. In other words, the line for $P_e^*(P_0)$ is a tangent to the latter curve as shown in figure 4.

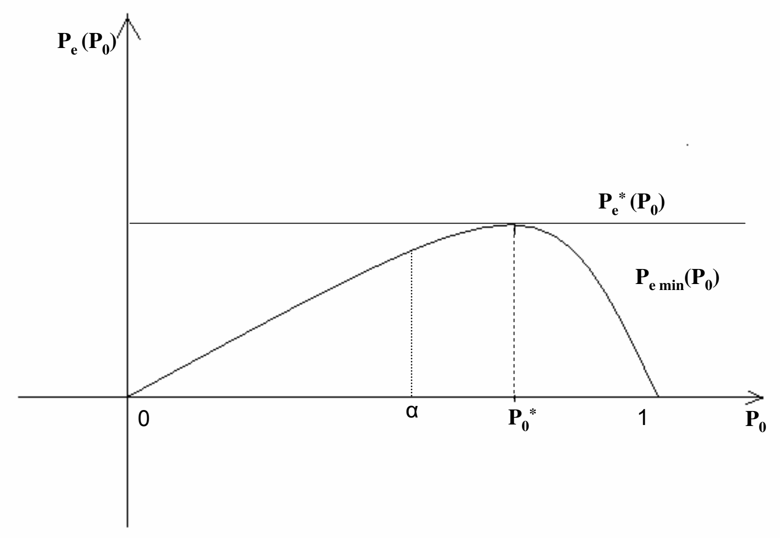

Figure 4.

This implies that $P_e^*(P_0)$=constant, i.e. not varying with P_0
Therefore

$$P_e^*(P_0) = P_M^* + P_0 \underbrace{(P_F^* - P_M^*)}_{=0}$$

$$\Rightarrow P_F^* = P_M^*$$

at the minimax point.

Neyman Pearson criterion

We wish to construct the decision problem based on the detection and false alarm probabilities.
- Criterion: Fix $P_F = \alpha$ and maximize P_D.
- Solution: Define the Lagrange:

$$\mathfrak{L} = P_D - \lambda(P_F - \alpha)$$

where λ is the Lagrange multiplier. - Substitute for P_D and P_F based on the pdf's and the decision threshold.

$$\mathfrak{L} = \int_{\mathbb{Z}_1} p(\vec{R}|H_1)d\vec{R} - \lambda\big[\int_{\mathbb{Z}_1} p(\vec{R}|H_0)d\vec{R} - \alpha\big]$$

$$= \lambda\alpha + \int_{\mathbb{Z}_1} \big[p(\vec{R}|H_1) - \lambda(\vec{R}|H_0)\big]d\vec{R}$$

Case 1: $\lambda < 0$
In this case, the integrand $\big[p(\vec{R}|H_1) - \lambda(\vec{R}|H_0)\big]$ is always positive. Thus the function $\mathfrak{L}$ is maximized when the integration is done over the largest possible region for $\mathbb{Z}$, i.e $\mathbb{Z}_1 = \mathbb{Z}$.
$\Rightarrow$ choose H_1 for all $\vec{R}$, which is not an acceptable solution.
Case 2: $\lambda > 0$
To achieve maximum $\mathfrak{L}$, we should integrate over the region in $\mathbb{Z}$ where the integrand is positive. i.e. $\big[p(\vec{R}|H_1) - \lambda(\vec{R}|H_0)\big] > 0$ for $\vec{R} \in \mathbb{Z}_1$.
This yields the following decision rule:

$$p(\vec{R}|H_1) \gtrless^{\text{select } H_1}_{\text{select } H_0} \lambda p(\vec{R}|H_0)$$

$$\text{or } \Lambda(\vec{R}) = \frac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)} \gtrless^{\text{select } H_1}_{\text{select } H_0} \lambda$$

This is the Bayes decision rule with

$$\lambda \triangleq \eta = \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}$$

To find λ (the threshold), we use the fact that $P_F = \alpha$, implying

$$P_F = \int_\lambda^\infty p(\Lambda|H_0)d\Lambda$$

Therefore, we vary λ until $P_F = \alpha$ is achieved.

Receiver Operating Characteristics (ROC)

ROC is simply a plot of P_D vs P_F for a given receiver (decision rule) as a function of the parameter of interest.
Likelihood ratio test (LRT):

$$\lambda = \frac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)} \gtrless^{\text{select } H_1}_{\text{select } H_0} \eta$$

yields P_F and P_D as functions of parameters such as additive noise variance or average SNR, signaling type (unipolar, bipolar) etc.

Example

$$H_0: \quad r_i = A_0 + n_i \quad i = 1, 2, ..., N$$
$$H_1: \quad r_i = A_1 + n_i$$

where (A_0, A_1) are constants and $n_i \overset{\text{i.i.d}}{\sim} (0, \sigma_n^2)$.
We showed that

$$\Lambda(\vec{R}) = \frac{p(\vec{R}|H_1)}{p(\vec{R}|H_0)} = exp\left[\frac{\sum_{i=1}^{N}(r_i - A_0)^2 - (r_i - A_1)^2}{2\sigma_n^2}\right]$$

For general ASK signaling, we have:

$$l(\vec{R}) = \underbrace{\sum_{i=1}^{N}(S_{1i} - S_{0i})r_i}_{(\vec{S_1} - \vec{S_0})^T.\vec{R}} \gtrless^{\text{select } H_1}_{\text{select } H_0} \ln\eta.\sigma_n^2 + \frac{1}{2}\sum_{i=1}^{N}(S_{1i}^2 - S_{0i}^2)$$

For this example, we have $S_{0i} = A_0$ and $S_{1i} = A_1$, $\forall\ i = 1, 2, ..., N$. Thus the test becomes:

$$l(\vec{R}) = \sum_{i=1}^{N}(A_1 - A_0)r_i \gtrless^{\text{select } H_1}_{\text{select } H_0} \ln\eta.\sigma_n^2 + \frac{N}{2}(A_1^2 - A_0^2)$$

Redefine the sufficient statistic by normalizing it as:

$$l(\vec{R}) \triangleq \frac{\sum_{i=1}^{N} r_i}{\sqrt{N}\sigma_n} \gtrless^{\text{select } H_1}_{\text{select } H_0} \frac{\ln\eta.\sigma_n}{\sqrt{N}(A_1 - A_0)} + \frac{\sqrt{N}(A_1 + A_0)}{2\sigma_n}$$

We have the following distributions which are also shown in figure 5:

$$l|H_0 \sim N\left(\frac{\sqrt{N}A_0}{\sigma_n}, 1\right)$$
$$l|H_1 \sim N\left(\frac{\sqrt{N}A_1}{\sigma_n}, 1\right)$$

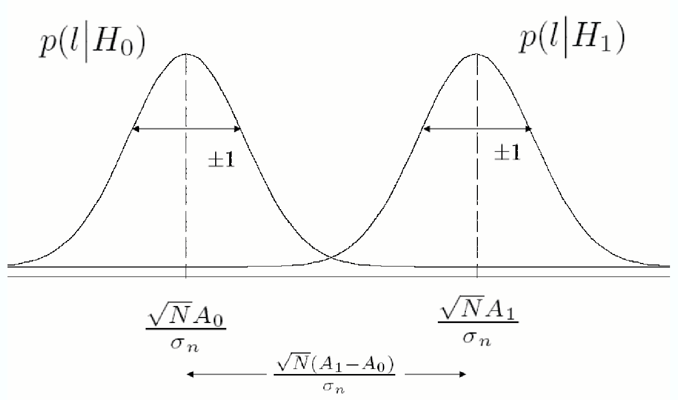

Figure 5.

Define $d \triangleq \frac{\sqrt{N}(A_1 - A_0)}{\sigma_n}$ and $D \triangleq \frac{\sqrt{N}(A_1 + A_0)}{\sigma_n}$. Therefore, the LRT becomes:

$$l(\vec{R}) \triangleq \frac{\sum_{i=1}^{N} r_i}{\sqrt{N}\sigma_n} \gtrless^{\text{select } H_1}_{\text{select } H_0} \frac{\ln \eta}{d} + \frac{D}{2}$$

Note: For unipolar ASK where $A_0 = 0$ and $D = d$, the quantity representing average bit energy under $P_0 = P_1 = \frac{1}{2}$ becomes:

$$E_b = \frac{1}{2}(A - 1^2 + A_0^2) = \frac{\sigma_n^2}{4N}(D_2 + d^2)$$

Thus, with (d, D), we identify two important features of this form of data transmission.
a) Separation of the two hypotheses in the l domain;
b) Average energy of the transmitter with respect to the noise power.

Performance probabilities:

$$P_F = \int_{\frac{\ln \eta}{d} + \frac{D}{2}}^{\infty} p(l \big| H_0) dl$$

$$= \int_{\frac{\ln \eta}{d} + \frac{D}{2}}^{\infty} \left[\frac{1}{2\pi} e^{\left\{ - \frac{(l - \frac{\sqrt{N}A_0}{\sigma_n})^2}{2} \right\}} \right] dl$$

$$= erfc^* \left(\frac{\ln \eta}{d} + \frac{D}{2} - \frac{\sqrt{N}A_0}{\sigma_n} \right)$$

Similarly

$$P_D = \int_{\frac{\ln \eta}{d} + \frac{D}{2}}^{\infty} p(l \big| H_1) dl$$

$$= erfc^* \left(\frac{\ln \eta}{d} + \frac{D}{2} - \frac{\sqrt{N}A_1}{\sigma_n} \right)$$

Property: The threshold η for the likelihood ratio test is given by:

$$\eta = \frac{dp_D}{dP_F}$$

at the operating point of interest. Proof:

$$P_D = \int_{\eta}^{\infty} p(\Lambda \big| H_1) d\Lambda$$

$$= \int_{\mathbb{Z}_1} p(\vec{R} \big| H_1) d\vec{R}$$

But, we know:

$$\Lambda(\vec{R}) = \frac{p(\vec{R} \big| H_1)}{p(\vec{R} \big| H_0)}$$

or

$$p(\vec{R} \big| H_1) = \Lambda(\vec{R}).p(\vec{R} \big| H_0)$$

We substitute this in the expression for P_D:

$$P_D = \int_{\mathbb{Z}_1} \Lambda(\vec{R}).p(\vec{R} \big| H_0) d\vec{R}$$

Rewrite the above in terms of $\Lambda \big| H_0$:

$$P_D = \int_{\eta}^{\infty} \Lambda.p(\Lambda \big| H_0) d\Lambda$$

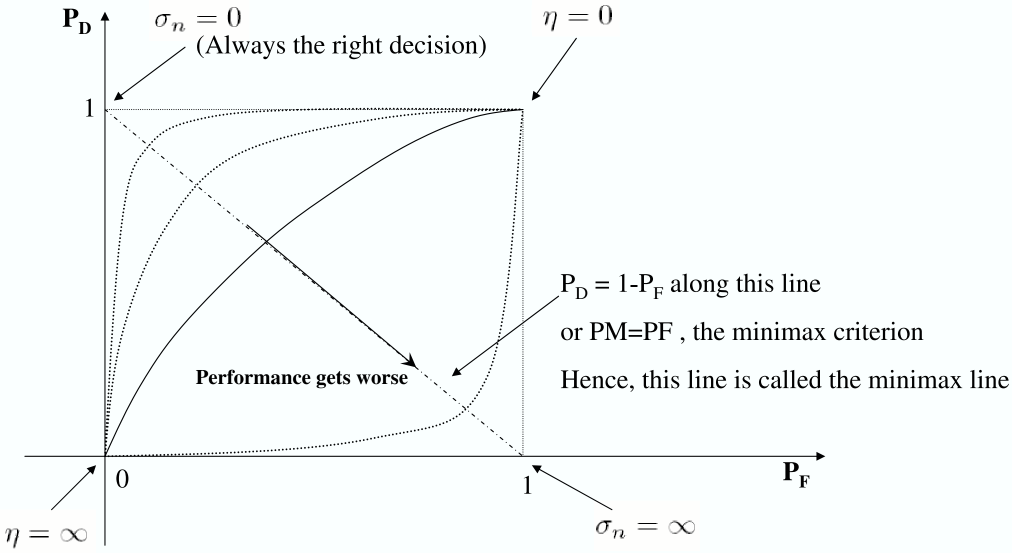

Figure 6.

We also have:

$$P_F = \int_{\eta}^{\infty} p(\Lambda \big| H_0) d\Lambda$$

Therefore, using Leibnitz rule of differentiation (refer to Papoulis & Pillai, page: 181)

$$\frac{dP_D}{d\eta} = -\eta p(\eta \big| H_0)$$

$$\frac{dP_F}{d\eta} = -p(\eta \big| H_0)$$

Finally:

$$\frac{dP_D/d\eta}{dP_F/d\eta}$$

$$= \frac{-\eta p(\eta \big| H_0)}{-p(\eta \big| H_0)} = \eta$$

$$\Rightarrow \frac{dP_D}{dP_F} = \eta$$