EE 126 — Probability and Random Processes

April 05, 2018 — Hypothesis Testing

1 Lecture

Hypothesis Testing

We observe a random variable Y (e.g. caloric intake). We say that

$$\begin{cases} Y \sim f(y \mid 0) & \text{under hypothesis } H_0 \text{ (e.g. no cancer)} \\ Y \sim f(y \mid 1) & \text{under hypothesis } H_1 \text{ (e.g. cancer)} \end{cases}$$

and define the **decision rule** as $r : \mathbb{R} \mapsto \{0, 1\}$ ("takes an observation as input and returns one of two hypotheses"). The decision will not always be 100% accurate. There are two types of errors:

- If $r(y) = 0$ but it's actually H_1, it's a false negative.
- If $r(y) = 1$ but it's actually H_0, it's a false positive.

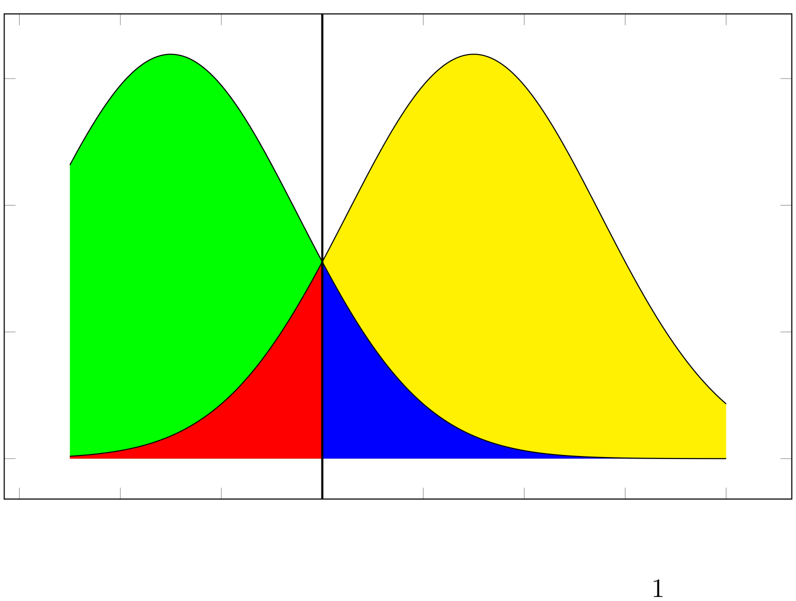
Bayesian Formulation

We have prior probabilities $\pi(0)$ for H_0 and $\pi(1)$ for H_1. We can model each hypothesis as a Gaussian, centered at 0 for H_0 and 1 for H_1. Our task is to decide whether an observation comes from the H_0 Gaussian or the H_1 Gaussian.

In other words, H_0 produces observations $Y \sim \mathcal{N}(0, \sigma^2)$ and H_1 produces observations $Y \sim \mathcal{N}(1, \sigma^2)$. We will assume uniform priors, i.e. $\pi(0) = \frac{1}{2}$ and $\pi(1) = \frac{1}{2}$.

Since the Gaussians intersect at $\frac{1}{2}$, we can formalize our decision rule as

$$r(y) = \begin{cases} 0 & \text{if } y < 1/2 \\ 1 & \text{if } y \geq 1/2 \end{cases}$$



The red region signifies false negative error, while the blue region signifies false positive error.

It turns out that this is the same as MLE:

$$
\begin{aligned}
r(y) &= \text{MLE}[X \mid Y = y] \\
&= \begin{cases} 0 & \text{if } L(y) < 1 \\ 1 & \text{if } L(y) \geq 1 \end{cases}
\end{aligned}
$$

The likelihood is

$$
L(y) = \frac{f(y \mid 1)}{f(y \mid 0)} = \frac{\exp\left(-\frac{(y-1)^2}{2\sigma^2}\right)}{\exp\left(-\frac{y^2}{2\sigma^2}\right)} = \exp\left(\frac{2y-1}{2\sigma^2}\right)
$$

so $L(y) > 1$ iff $y > 1/2$.

Minimum Cost Criterion

Imagine we pay one unit for every error (i.e. for a false positive or a false negative). To find the optimal decision rule r, we can minimize the expected cost:

$$
\begin{aligned}
&\min_{r:\mathbb{R}\mapsto\{0,1\}} \int_{-\infty}^{\infty} \mathbb{1}\{r(y) = 1\} \cdot \pi(0) \cdot f(y \mid 0)\ dy \\
= &\min_{r:\mathbb{R}\mapsto\{0,1\}} \int_{-\infty}^{\infty} \mathbb{1}\{r(y) = 0\} \cdot \pi(1) \cdot f(y \mid 1)\ dy
\end{aligned}
$$

Theorem. The minimizer r^* is

$$
r^*(y) = \text{MAP}[X \mid Y = y] = \begin{cases} 0 & \text{if } L(y) < \frac{\pi(0)}{\pi(1)} \\ 1 & \text{if } L(y) \geq \frac{\pi(0)}{\pi(1)} \end{cases}
$$

Proof. Total cost is

$$
\int_{-\infty}^{\infty} \left(\mathbb{1}\{r(y) = 1\} \cdot \pi(0) \cdot f(y \mid 0) + \mathbb{1}\{r(y) = 0\} \cdot \pi(1) \cdot f(y \mid 1)\right) dy
$$

For each y, we will have either the left term or the right term. Therefore, to minimize this cost we want to choose $r(y)$ such that it selects the lower of the two terms.

$$
\begin{aligned}
r^*(y) &= \begin{cases} 0 & \text{if } \pi(1)f(y \mid 1) < \pi(0)f(y \mid 0) \\ 1 & \text{if } \pi(1)f(y \mid 1) \geq \pi(0)f(y \mid 0) \end{cases} \\
&= \begin{cases} 0 & \text{if } \frac{f(y|1)}{f(y|0)} < \frac{\pi(0)}{\pi(1)} \\ 1 & \text{if } \frac{f(y|1)}{f(y|0)} \geq \frac{\pi(0)}{\pi(1)} \end{cases} \\
&= \begin{cases} 0 & \text{if } L(y) < \frac{\pi(0)}{\pi(1)} \\ 1 & \text{if } L(y) \geq \frac{\pi(0)}{\pi(1)} \end{cases}
\end{aligned}
$$

Hence the MAP estimate is optimal according to this formulation of hypothesis testing. (MAP: here the decision rule returns the hypothesis that has the largest product of prior and density.) But this is not the only formulation...

Neyman-Pearson Formulation

We can also find the optimal decision rule according to a different criterion. In this formulation, we do not have priors and can only have two hypotheses. We'd like to specify a decision rule $r : \mathbb{R} \mapsto \{0, 1\}$ that minimizes the false negative error, i.e.

$$
\min_{r:\mathbb{R}\mapsto\{0,1\}} P(r(y) = 0 \mid X = 1)
$$

s.t.

$$P(r(y) = 1 \mid X = 0) \leq \beta$$

If we don't constrain the false positive error to be less than a constant, the optimal decision rule will always declare 1.

Theorem. The optimizer r^* is a randomized threshold rule given by

$$r^*(y) = \begin{cases} 1 & \text{if } L(y) > \lambda \\ 1 \text{ with probability } \gamma & \text{if } L(y) = \lambda \\ 0 \text{ with probability } 1-\gamma & \text{if } L(y) = \lambda \\ 0 & \text{if } L(y) < \lambda \end{cases}$$

where γ, λ are calculated as per

$$\begin{aligned} \text{require } P(r^*(y) = 1 \mid X = 0) = \beta \\ P(L(y) > \lambda \mid X = 0) + \gamma P(L(y) = \lambda \mid X = 0) = \beta \end{aligned}$$

In other words, the the false positive probability must be equal to β.

Note: we do this because the minimum constrained r will maximize the false positive error, i.e. the false positive error will be exactly equal to β.

Example: Gaussian again. We have that H_0 has observations distributed as $y \sim \mathcal{N}(0, \sigma^2)$, and H_1 has observations distributed as $y \sim \mathcal{N}(1, \sigma^2)$. The likelihood $L(y) = \exp\left(\frac{2y-1}{2\sigma^2}\right)$ is an increasing function s.t.

$$L(y) > \lambda \iff y > \sigma^2 \ln \lambda + \frac{1}{2} = y_0$$

$P(L(y) = \lambda \mid X = 0) = 0$ (since $L(y)$ is continuous), so we can pick anything for γ. We will pick $\gamma = 1$.

Then

$$P(Y > y_0) = \beta \iff P(\mathcal{N}(0,1) \geq y_0) = \beta$$

and

$$r^*(y) = \begin{cases} 1 & \text{if } y \geq y_0 \\ 0 & \text{if } y < y_0 \end{cases}$$

Example: no observation y. We have two hypotheses H_0 and H_1, but no observations. Our decision rule r will either be a constant or a random variable. If r is a constant, then there is only one possibility s.t. it satisfies the constraint: $r = 0$. Here the objective is $P(r = 0 \mid X = 1) = 1$. But 1 is not a very good value for the objective (in fact it is the worst value possible). So we will choose to make r a random variable:

$$r = \begin{cases} 1 \text{ with probability } \beta & \text{meaning } P(r = 1 \mid X = 0) = P(r = 1) = \beta \\ 0 \text{ with probability } 1-\beta & \text{meaning } P(r = 0 \mid X = 1) = P(r = 0) = 1 - \beta \leq 1 \end{cases}$$

Proof: r^* is the optimizer of Neyman-Pearson. Let r be any feasible decision rule that satisfies the constraints (i.e. s.t. $P(r(y) = 1 \mid X = 0) \leq \beta$). We want to show that

$$P(r^*(y) = 0 \mid X = 1) \leq P(r(y) = 0 \mid X = 1)$$

i.e. that the objective value for r^* is less than or equal to the objective value for any other r.

By the definition of r^*,

$$(r^*(y) - r(y)) \cdot (L(y) - \lambda) \geq 0$$

since $y, r^*(y) = 1$ and $r(y) \in \{0, 1\}$.

We can verify the equality by stepping through the cases for comparing $L(y)$ and λ. First we'll take expectations under H_0, $X = 0$:

$$\begin{aligned}
\mathbb{E}[r^*(y)L(y) \mid X=0] - \mathbb{E}[r(y)L(y) \mid X=0] &\geq \lambda\left(\mathbb{E}[r^*(y) \mid X=0] - \mathbb{E}[r(y) \mid X=0]\right) \\
&= \lambda\left(P(r^*(y)=1 \mid X=0) - P(r(y)=1 \mid X=0)\right) \geq 0
\end{aligned}$$

Note that $P(r^*(y)=1 \mid X=0) = \beta$ and $P(r(y)=1 \mid X=0) \leq \beta$.

Therefore

$$\begin{gathered}
\mathbb{E}[r^*(y)L(y) \mid X=0) \geq \mathbb{E}[r(y)L(y) \mid X=0] \\
\int r^*(y)\frac{f^*(y \mid 1)}{f^*(y \mid 0)}f^*(y \mid 0)\; dy \geq \int r(y)\frac{f(y \mid 1)}{f(y \mid 0)}f(y \mid 0)\; dy \\
\mathbb{E}[r^*(y) \mid X=1] \geq \mathbb{E}[r(y) \mid X=1] \\
P(r^*(y)=1 \mid X=1) \geq P(r(y)=1 \mid X=1) \\
P(r^*(y)=0 \mid X=1) \leq P(r(y)=0 \mid X=1)
\end{gathered}$$

■

Example: exponential.

$$\begin{aligned}
H_0 &: Y_1, ..., Y_n \sim \text{Exp}(\lambda_0) \\
H_1 &: Y_1, ..., Y_n \sim \text{Exp}(\lambda_1)
\end{aligned}$$

for $\lambda_1 > \lambda_0$ and hence $\frac{1}{\lambda_0} > \frac{1}{\lambda_1}$.

The scenario is as follows: a factory is producing lightbulbs. If the factory is working properly, the lightbulbs have a lifetime that is exponentially distributed with parameter λ_0. If it is not, we'll end up with defective lightbulbs that have lifetimes exponentially distributed with parameter λ_1.

$$\begin{aligned}
L(y_1, ..., y_n) &= \frac{\prod_{i=1}^n \lambda_1 e^{-\lambda_1 y_i}}{\prod_{i=1}^n \lambda_0 e^{-\lambda_0 y_i}} \\
&= \left(\frac{\lambda_1}{\lambda_0}\right)^n \exp\left(-(\lambda_1 - \lambda_0)\sum_{i=1}^n y_i\right)
\end{aligned}$$

Let $S = \sum_{i=1}^n y_i$. Then

$$L(S) = \left(\frac{\lambda_1}{\lambda_0}\right)^n \exp\left(-(\lambda_1 - \lambda_0)S\right) \qquad \text{which is decreasing.}$$

$$r^*(y) = \begin{cases} 0 & \text{if } S > S_0 \\ 1 & \text{if } S < S_0 \end{cases}$$

where S_0 is determined by solving

$$\beta = P(S < S_0 \mid X = 0)$$

which can be approximated with the CLT. Note that S, as a sum of exponentials, is actually Erlang(λ_0, n).

$$\begin{aligned}
&\frac{(y_1 + ... + y_n) - n\lambda_0^{-1}}{\sqrt{n} \cdot \lambda_0^{-1}} \approx \mathcal{N}(0,1) \\
&\rightarrow P(S < S_0 \mid X = 0) \approx P\left(\mathcal{N}(0,1) < \frac{S_0 - n\lambda_0^{-1}}{\sqrt{n} \cdot \lambda_0^{-1}}\right)
\end{aligned}$$

We can use the standard normal table to obtain the actual value.