EE 226A **Fall 2024**
Handout 4
Binary Hypothesis Testing

Likelihood ratio

Suppose first that we get a sample a real random variable X. There are two hypotheses about the distribution of the random variable being observed. If the *null hypothesis* H_0 is true, the density of X is $f(x \mid 0)$. If the *alternate hypothesis* H_1 is true, the density of X is $f(x \mid 1)$. Having observed X we wish to decide which hypothesis is true.

Definition : The *likelihood ratio* of the alternate hypothesis relative to the null hypothesis is

$$L(x) = \frac{f(x \mid 1)}{f(x \mid 0)} \,.$$

Note that this is a function of the realized observation.

Another similar scenario of interest is when we get a sample of a random variable X that takes values in a discrete set $\mathcal{X}$. Now, if the *null hypothesis* H_0 is true, the probability distribution of X is $(p(x \mid 0), x \in \mathcal{X})$. If the *alternate hypothesis* H_1 is true, the probability distribution of X is $(p(x \mid 1), x \in \mathcal{X})$. The likelihood ratio of the alternate hypothesis relative to the null hypothesis is now defined to be

$$L(x) = \frac{p(x \mid 1)}{p(x \mid 0)} \,, \;\; x \in \mathcal{X}.$$

Bayesian formulation

One assumes a prior distribution for which hypothesis is true: $\pi(0) = P(H_0)$, $\pi(1) = P(H_1)$, $\pi(0) + \pi(1) = 1$. Let $c(j \mid i), 0 \leq i, j, \leq 1$ be the cost of deciding that hypothesis H_j holds when actually hypothesis H_i holds. A decision rule is a function $\Delta : \mathbf{R} \to \{0, 1\}$. We choose Δ to minimize the overall expected cost, namely

$$\sum_{i=0}^{1} \sum_{j=0}^{1} c(j \mid i) \int_{\Delta(x)=j} f(x \mid i) \pi(i) \, dx$$

in the continuous case or

$$\sum_{i=0}^{1}\sum_{j=0}^{1} c(j \mid i) \sum_{x\in\mathcal{X}} p(x \mid i)\pi(i),$$

in the discrete case.

Theorem: Assume that $c(0 \mid 0) = c(1 \mid 1) = 0$ and $c(1 \mid 0), c(0 \mid 1) > 0$. Then the optimum decision is a threshold rule with threshold $a = \frac{c(1|0)\pi(0)}{c(0|1)\pi(1)}$. Namely, the optimal decision rule is $\Delta(x) = 1(L(x) > a)$.

Proof: Under the assumption on the costs, we can write the overall expected cost, in the continuous case, as

$$\int_{-\infty}^{\infty} [c(1 \mid 0)\pi(0)1(\Delta(x) = 1)f(x \mid 0) + c(0 \mid 1)\pi(1)1(\Delta(x) = 0)f(x \mid 1)]dx \; .$$

Substituting for $f(x \mid 1)$ by $L(x)f(x \mid 0)$, this can be written as

$$\int_{-\infty}^{\infty} [c(1 \mid 0)\pi(0)1(\Delta(x) = 1) + c(0 \mid 1)\pi(1)L(x)1(\Delta(x) = 0)]f(x \mid 0)dx \; .$$

To minimize this, it is clear that we have to set $\Delta(x) = 1$ when $c(1 \mid 0)\pi(0) < c(0 \mid 1)\pi(1)L(x)$, which is precisely the threshold rule in the statement of the theorem.

An analogous calculation works in the discrete case.

Neyman-Pearson formulation

We prescribe a maximum probability ϵ of *false alarm*, which is the probability of deciding H_1 when the true hypothesis is H_0. Subject to this, we wish to minimize the probability of *missed detection*, which is the probability of deciding H_0 when the true hypothesis is H_1. The justification for this formulation is that missed detection is typically a much more serious eventuality than false alarm. [1]

Theorem: The optimal decision rule is a *randomized threshold rule* based on the likelihood ratio. One selects the threshold Λ and $0 \leq \alpha \leq 1$ so that

$$P_0(L(x) > \Lambda) + \alpha P_0(L(x) = \Lambda) = \epsilon \; . \tag{1}$$

Here P_0 is the notation for the probability assignment to events conditioned on hypothesis H_0 being true and P_1 similarly denotes the probability assignment to

[1]However, in practice, the major problem with automated detection systems is "too high a probability of false alarm". Unless such systems are well-designed to have very few false alarms, operators monitoring such systems and charged with deploying resources in the eventuality of the alternate hypothesis being true often ignore the decision of the automated system, believing it to be a false alarm.

events conditioned on hypothesis H_1 being true. There is a unique choice of Λ and α satisfying (1). The optimal decision rule is

$$\text{Decide } H_0 \text{ if } L(x) < \Lambda \ ,$$

$$\text{Decide } H_1 \text{ if } L(x) > \Lambda \ ,$$

$$\text{Decide } H_1 \text{ with probability } \alpha \text{ and } H_0 \text{ with probability } 1-\alpha \text{ if } L(x) = \Lambda \ .$$

The optimal probability of missed detection is then

$$P_1(L(x) < \Lambda) + (1-\alpha) P_1(L(x) = \Lambda) \ .$$

Proof: Call the rule defined above Δ^*. Let Δ be any other decision rule that meets the criterion on the probability of false alarm, i.e., for which (in the continuous case)

$$\int_{\{x \,:\, \Delta(x)=1\}} f(x \mid 0) dx \leq \epsilon \ .$$

Then we can see that

$$(\Delta^*(x) - \Delta(x))(f(x \mid 1) - \Lambda f(x \mid 0)) \geq 0 \ .$$

Integrating this expression, and taking the expectation over the randomization involved in the decision rule, gives

$$P_1(\Delta^* = 1) - P_1(\Delta = 1) \geq \Lambda(P_0(\Delta^* = 1) - P_0(\Delta = 1)) \geq 0$$

where we have used $P_0(\Delta = 1) \leq \epsilon = P_0(\Delta^* = 1)$ in writing the second equality. But this gives

$$P_1(\Delta^* = 0) \leq P_1(\Delta = 0)$$

showing that Δ^* is at least as good a decision rule as Δ since its probability of missed detection is no bigger.

A similiar argument works in the discrete case.

Examples

• Under the null hypothesis X has $N(0, \sigma^2)$ distribution. Under the alternate hypothesis it has $N(\mu, \sigma^2)$ distribution. (We are trying to detect the presence of a constant signal in additive Gaussian noise. μ and σ^2 are assumed known.) Then

$$L(x) = \exp((x - \frac{\mu}{2})\frac{\mu}{\sigma^2}) \ .$$

Note that a threshold test based on $L(x)$ is equivalent to a threshold test based on x.

• The observation is an n-dimensional vector X. Under the null hypothesis X has $N(0, C)$ distribution. Under the alternate hypothesis it has $N(\mu, C)$ distribution. $\mu \in \mathbf{R}^n$ and $C \in \mathbf{R}^{n \times n}$ are assumed known. Further, C is assumed to be positive definite. This means we can write $C = U^T D U$, where U is an orthogonal matrix, and D is a diagonal matrix with strictly positive diagonal terms, say d_i, $1 \leq i \leq n$. Then

$$L(x) = \exp(\mu^T C^{-1}(x - \frac{\mu}{2})) = \exp((U\mu)^T D^{-1}(Ux) - \frac{\mu^T C^{-1} \mu}{2}) \, .$$

Let $y = Ux$ and $\nu = U\mu$. Note that a threshold test based on $L(x)$ is equivalent to a threshold test based on $\sum_{i=1}^{n} \frac{\nu_i}{d_i} y_i$.

• Under the null hypothesis X has a Poisson distribution with mean μ_0. Under the alternate hypothesis it has a Poisson distribution with mean μ_1. Then

$$L(n) = \left(\frac{\mu_1}{\mu_0} \right)^n \exp(-(\mu_1 - \mu_0)) \ , n \geq 0.$$

Note that a threshold test based on $L(n)$ is equivalent to a threshold test based on n, because

$$\log L(n) = n \log \frac{\mu_1}{\mu_0} - \mu_1 + \mu_0,$$

the logarithm being to the natural base. (Note that if $\mu_1 > \mu_0$ the alternate hypothesis will be decided on for all large enough values of n, while if $\mu_0 > \mu_1$ it will be the null hypothesis that will be decided on for all large enough values of n.)