Probability and Stochastic Processes: Decision/Detection

Zhenyu Liao

School of Electronic Information and Communications, HUST

May, 23, 2024

Outline

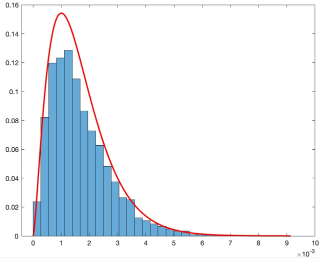
Recap on stochastic convergence

Example: if your friend cheating you in dice?
Statistical modeling:

- for $n = 500$ trials, $Y \sim \text{Multinomial}(n, p)$
- compute the squared difference as a test $T_n = \sum_{i=1}^{6} \left(\frac{Y_i}{n} - \frac{1}{6} \right)^2$

```
close all; clear; clc

n_trials = 10000;
T = zeros(n_trials,1);
n = 500;
p = ones(1,6)/6;
for i = 1:n_trials
    X = mnrnd(n,p);
    T(i) = sum( (X/n - p).^2 );
end
figure
histogram(T,30,'Normalization','probability')
```



Example: if your friend cheating you in dice?
Analysis:

- test $T_n = \sum_{i=1}^{6} \left(\frac{Y_i}{n} - \frac{1}{6} \right)^2$
- multivariate view: $(Y_1, \ldots, Y_6) = \sum_{i=1}^{n} \mathbf{X}_i$, $\mathbf{X}_i = (X_1, \ldots, X_6)$ with

$$\mathbb{E}[X_i] = \frac{1}{6}, \quad \mathrm{Var}[X_i] = \frac{1}{6} - \left(\frac{1}{6}\right)^2 = \frac{5}{36}, \quad \mathrm{Cov}[X_i, X_j] = -\frac{1}{36}, \tag{1}$$

- by the LLN, as $n \to \infty$, we have $(Y_1, \ldots, Y_6) \to \left(\frac{1}{6}, \ldots, \frac{1}{6}\right)$ a.s. or in probability, and by the CLT

$$\sqrt{n} \left(\frac{Y_1}{n} - \frac{1}{6}, \ldots, \frac{Y_6}{n} - \frac{1}{6} \right) \to \mathcal{N}(0, \Sigma) \tag{2}$$

in distribution, with

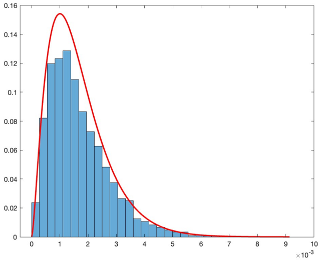
$$\Sigma = \frac{1}{6}\mathbf{I}_6 - \frac{1}{36}\mathbf{1}_6\mathbf{1}_6^\mathsf{T} \in \mathbb{R}^{6\times 6}. \tag{3}$$

Example: if your friend cheating you in dice?
Analysis:

- for $nT_n = n \sum_{i=1}^{6} \left(\frac{Y_i}{n} - \frac{1}{6} \right)^2$ with $\sqrt{n} \left(\frac{Y_1}{n} - \frac{1}{6}, \ldots, \frac{Y_6}{n} - \frac{1}{6} \right) \to \mathcal{N}(0, \Sigma)$
- the function $g(x_1, \ldots, x_6) = \sum_{i=1} x_i^2$ is continuous (continuous mapping!), so

$$nT_n \to \sum_{i=1}^{6} Z_i^2, \tag{4}$$

  in distribution as $n \to \infty$, with $(Z_1, \ldots, Z_6) \sim \mathcal{N}(0, \Sigma)$
- then, for n large, the distribution of T_n approximately the **same** as that of $\frac{1}{n} \sum_{i=1}^{6} Z_i^2$ (has distribution $\frac{1}{6}\chi_5^2$)



Today

- **Testing a Simple Null Hypothesis**: null hypothesis, type-I error, significant level, p-value
- **Hypothesis Testing**:
  - Bayesian Hypothesis Testing: average error probability, Likelihood Ratio Test (LRT), Maximum A-Posteriori Probability (MAP), etc.
  - Neyman-Pearson Hypothesis Testing

Null distribution and type I error

- A **hypothesis test** is a binary question about the "data" distribution
- our goal is to either accept a **null hypothesis** H0 (i.e., some specifications about the distribution), or to reject it in favor of an (known or unknown) **alternative hypothesis** H1

Suppose that we've chosen our test statistics T_n, how large (or small) should T_n be, before we can "confidently" assert that the hypothesis H0 is **false**?

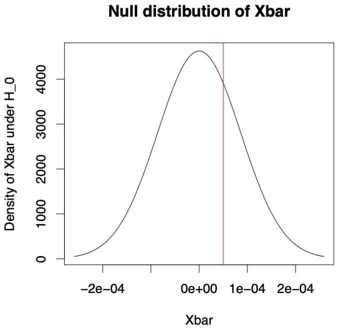
Example:

$$H_0 \colon X_1, \ldots, X_n \sim \mathcal{N}(0, 2.23 \times 10^{-7}), \tag{5}$$

so that under H_0, we have $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathcal{N}(0, 2.23 \times 10^{-7}/n)$.

Example: case I

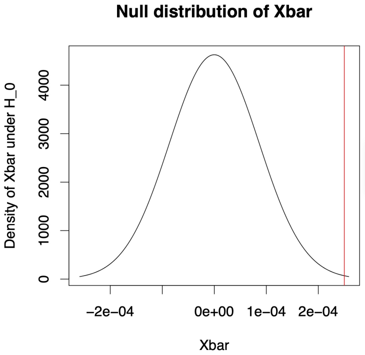
For $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathcal{N}(0, 2.23 \times 10^{-7}/n)$, here's the PDF for $n = 30$.



- if we observe $\bar{X} = 0.5 \times 10^{-4}$, this does **NOT** provide strong evidence against (i.e., to reject) H_0
- we might accept H_0 in this case

Example: case II

For $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(0, 2.23 \times 10^{-7}/n)$, here's the PDF for $n = 30$.



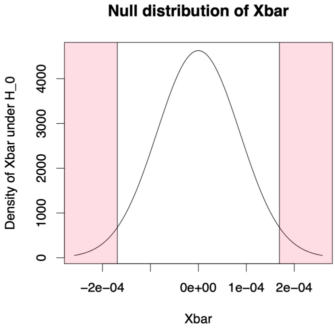
- if we observe $\bar{X} = 2.5 \times 10^{-4}$, this **does** provide strong evidence against (i.e., to reject) H_0
- we might reject H_0 in this case

Null distribution and type I error

- **rejection regime** is the set of values of T_n for which we choose to reject H_0
- **acceptance regime** is the set of values of T_n for which we choose to accept H_0
- choose the rejection regime so as to control the probability of the **type I error**

$$\alpha = \mathbb{P}(\text{reject } H_0 | H_0) \tag{6}$$

  value α also the **significance level** of the test T_n
- if, under its null distribution H0, T_n belong to the rejection region with probability α, the test T_n is said to be level-α



P-values

- **p-value**: smallest significance level at which the test T_n would have rejected H_0
- for a one-sided test that rejects for large T_n, let t_{obs} denote the observed value of T_n, the p-value is $\mathbb{P}(T_n > t_{\text{obs}} | H_0)$
- for two-sided test, the p-value is 2 times the smaller of
- p-values provide a **quantitative** measure of the extent to which the observations supports (or against) H0

Example: testing the fairness of a coin

- null hypothesis H0: the coin is fair, with $\mathbb{P}(\text{heads}) = 0.5$
- test statistics T_n: number of heads after $n = 20$ flips
- α-level: 0.05 (what does this mean?)
- observation: 14 out of 20 flips
- two-sided p-values of the observation = $2 \times 0.058 = 0.115 > 0.05$
- what does this mean?: meaning that the observation falls within the range of what would happen 95% of the time, if the coin were fair (H0)
- **decision**: not to reject H0
- however, **if one more head**, resulting p-value = $0.0414 < 0.05$, then **decide to reject** H0

Bayes' Theorem

Theorem

Bayes' Theorem For two events A, B with $\mathbb{P}(B) \neq 0$, we have

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}, \tag{7}$$

with conditional probabilities $\mathbb{P}(A|B), \mathbb{P}(B|A)$, and marginal probabilities $\mathbb{P}(A), \mathbb{P}(B)$.

Bayesian Hypothesis Testing

- We observe some data y, which we assume to be produced as the realization of some RV Y.
- However, we do not know the distribution of Y.
- We only know that Y may be distributed according to two possible distributions: F_0 or F_1.
- These two hypotheses, referred to as H0 and H1, are known to occur with prior probabilities p_0, p_1, with obviously $p_1 = 1 - p_0$.
- We have $H_0 : Y \sim F_0$ and $H_1 : Y \sim F_1$
- Bayes Risk: for a given decision rule $g : \mathbb{R} \to \{0, 1\}$, we define the Bayes risk

$$r(g) = p_0 r_0(g) + p_1 r_1(g)$$

where we define the conditional risks

$$\begin{aligned} r_0(g) &= c_{00}\mathbb{P}(g(Y) = 0|\mathrm{H0}) + c_{10}\mathbb{P}(g(Y) = 1|\mathrm{H0}) \\ r_1(g) &= c_{01}\mathbb{P}(g(Y) = 0|\mathrm{H1}) + c_{11}\mathbb{P}(g(Y) = 1|\mathrm{H1}) \end{aligned}$$

- c_{ij} is the cost associated to deciding for hypothesis Hi when Hj is true.
- We should have $c_{00} < c_{10}$ and $c_{11} < c_{10}$
- The optimal Bayesian decision rule (or Bayesian hypothesis test) is a rule minimizing the Bayes risk:

$$g^* = \arg\min r(g)$$

Example: average error probability

- A typical Bayes risk function (design of cost c) is the **average error probability**.
- This is obtained by letting $c_{00} = c_{11} = 0$ and $c_{10} = c_{01} = 1$, i.e.,

$$\begin{aligned} P_e(g) &= \mathbb{P}(g(Y) = 1 | \mathrm{H0}) p_0 + \mathbb{P}(g(Y) = 0 | \mathrm{H1}) p_1 \\ &= \mathbb{P}(g(Y) = 1, \mathrm{H0}) + \mathbb{P}(g(Y) = 0, \mathrm{H1}) \\ &= \mathbb{P}(g(Y) \neq \mathrm{H}) \end{aligned}$$

where H denotes a binary RV taking on the hypothesis value H0 and H1 with probabilities p_0 and p_1, respectively.

Decision regions

- Any function $g : \mathbb{R} \to \{0,1\}$ is defined by the two decision regions

$$D_0 = \{y \in \mathbb{R} : g(y) = 0\}, \quad D_1 = \{y \in \mathbb{R} : g(y) = 1\}$$

- We can write:

$$\mathbb{P}(g(Y) = i | \mathrm{H}j) = \int_{D_i} dF_j(y)$$

- We can write

$$\begin{aligned} r(g) &= \sum_{j=0}^{1} p_j \left(c_{0j} \left(1 - \int_{D_1} dF_j(y) \right) + c_{1j} \int_{D_1} dF_j(y) \right) \\ &= \sum_{j=0}^{1} p_j c_{0j} + \int_{D_1} \sum_{j=0}^{1} p_j (c_{1j} - c_{0j}) dF_j(y) \end{aligned}$$

Optimal decision regions (continuous)

- Suppose that Y is continuous with respect to all hypotheses, and let $dF_j(y) = f_j(y)dy$.
- $r(g)$ is minimized by letting (we are in fact determining the decision rule $g(y)$)

$$D_1 = \left\{ y \in \mathbb{R} : \sum_{j=0}^{1} p_j (c_{1j} - c_{0j}) f_j(y) \leq 0 \right\}$$

- Explicitly, we find the following threshold rule

$$D_1 = \{ y \in \mathbb{R} : L(y) \geq \tau \}$$

where $L(y) = \frac{f_1(y)}{f_0(y)}$ is called Likelihood Ratio, and the threshold is given by

$$\tau = \frac{p_0(c_{10} - c_{00})}{p_1(c_{01} - c_{11})}.$$

- This is known under the name of **Likelihood Ratio Test (LRT)**
- Of course, if $L(y) < \tau$, then y is allocated to D_0.
- Notice that the boundary region $L(y) = \tau$, i.e., $f_1(y) = \tau f_0(y)$ does not contribute to the Bayes risk, and therefore it can be arbitrarily allocated to D_0 or to D_1.
- Note: For discrete RVs, just replace the conditional pdfs with the conditional pmfs.

Example

- Consider X taking values in $\mathcal{X} = \{+1, -1\}$ with equal probability, and the observation
$$Y = aX + Z$$
where $a \in \mathbb{R}_+$ and $Z \sim \mathcal{N}(0, \sigma^2)$.
- We let H0 and H1 denote the hypotheses of $X = -1$ and $X = +1$, respectively, and we wish to find the optimal decision rule (i.e., the regions D_0, D_1), such that the **average error probability** $P_e(g)$ is minimized ($c_{00} = c_{11} = 0$ and $c_{10} = c_{01} = 1$).
- In this case $\tau = p_0/p_1$, and the two conditional pdfs are
$$f_0(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|y+a|^2}{2\sigma^2}}, \quad f_1(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|y-a|^2}{2\sigma^2}}$$
with Likelihood ratio
$$L(y) = e^{\frac{2a}{\sigma^2} y}$$

- The decision region D_1 can be expressed as

$$D_1 = \left\{ y \in \mathbb{R} : y \geq \frac{\sigma^2}{2a} \log \frac{p_0}{p_1} \right\}$$

- More explicitly, we notice that the regions are the two intervals

$$D_1 = \left[\frac{\sigma^2}{2a} \log \frac{p_0}{p_1}, +\infty \right), \quad D_0 = \left(-\infty, \frac{\sigma^2}{2a} \log \frac{p_0}{p_1} \right)$$

- The resulting minimum average error probability is given by

$$P_e = p_0 \int_{D_1} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|y+a|^2}{2\sigma^2}} dy + p_1 \int_{D_0} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|y-a|^2}{2\sigma^2}} dy$$

- As a consequence, we obtain

$$P_e = p_0 \left(1 - \Phi \left(\frac{\sigma}{2a} \log \frac{p_0}{p_1} + \frac{a}{\sigma} \right) \right) + p_1 \Phi \left(\frac{\sigma}{2a} \log \frac{p_0}{p_1} - \frac{a}{\sigma} \right)$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$.

- In the equiprobable hypothesis case, $\log(p_0/p_1) = 0$, and we can use the fact that $\Phi(-x) = 1 - \Phi(x)$, such that

$$P_e = 1 - \Phi \left(\frac{a}{\sigma} \right) = Q \left(\frac{a}{\sigma} \right)$$

where we define the Gaussian complementary CDF as
$Q(x) = 1 - \Phi(x) = \int_{x}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$ (also known as the Gaussian Q function).

- The same idea can be generalized to the case where Y can be distributed according to M possible hypotheses, where Hj is $Y \sim F_j(y)$ for $j = 1, 2, \ldots, M$, with a priori probabilities $p_1, p_2, \ldots, p_M$.
- In this case $g : \mathbb{R} \to \{1, 2, \ldots, M\}$ is defined by M decision regions.
- In the case of **average error probability**, we have the simple explicit characterization

$$D_j = \left\{ y \in \mathbb{R} : p_j f_j(y) \geq p_k f_k(y) \quad \forall k \neq j,\ k = 1, \ldots, M \right\}$$

Maximum a posteriori probability (MAP)

- The a posteriori conditional probability of hypothesis Hj given the observation $Y = y$ is given by:

$$\mathbb{P}(\mathrm{H}j|Y = y) = \frac{p_j f_j(y)}{\sum_{k=1}^{M} p_k f_k(y)}$$

- The decision region D_j can be equivalently expressed as

$$D_j = \{y \in \mathbb{R} : \mathbb{P}(\mathrm{H}j|Y = y) \geq \mathbb{P}(\mathrm{H}k|Y = y) \ \ \forall\, k \neq j,\ k = 1, \ldots, M\}$$

- For this reason, the Bayesian M-ary hypothesis test that minimizes the average error probability is called Maximum A-Posteriori Probability (MAP) decision rule.

Exact expression of P_e

- In general, the probability of error is given by

$$
\begin{aligned}
P_e &= \sum_{i=1}^{M} p_i \mathbb{P}\left(Y \in \bigcup_{j \neq i} D_j \middle| \mathrm{H}i \right) \\
&= 1 - \sum_{i=1}^{M} p_i \mathbb{P}(Y \in D_i | \mathrm{H}i) \\
&= 1 - \sum_{i=1}^{M} p_i \int_{D_i} f_i(y) dy
\end{aligned}
$$

Upper and lower bounds on P_e

- We can find a simpler and general upper bound to P_e as follows: for any given pair of hypotheses i, j we define the pairwise error event

$$\{i \to j\} = \left\{ y \in \mathbb{R} : p_i f_i(y) \leq p_j f_j(y) \right\}$$

- The corresponding pairwise error probability (PEP) is given by

$$P(i \to j) = \mathbb{P}\left(Y \in \{i \to j\} \,|\, \mathrm{H}i \right) = \int_{\{i \to j\}} f_i(y) dy$$

$$\begin{aligned} P_e &= \sum_{i=1}^{M} p_i \mathbb{P}\left(Y \in \bigcup_{j \neq i} D_j \,\middle|\, \mathrm{H}i \right) \\ &= \sum_{i=1}^{M} p_i \mathbb{P}\left(Y \in \bigcup_{j \neq i} \{i \to j\} \,\middle|\, \mathrm{H}i \right) \\ &\leq \sum_{i=1}^{M} p_i \sum_{j \neq i} \mathbb{P}\left(Y \in \{i \to j\} \,|\, \mathrm{H}i \right) \\ &= \sum_{i=1}^{M} \sum_{j \neq i} p_i \int_{\{i \to j\}} f_i(y) dy = \sum_{i=1}^{M} \sum_{j \neq i} p_i \boxed{P(i \to j)} \end{aligned}$$

- Next, we consider a lower bound on the error probability:

$$
\begin{aligned}
P_e &= \sum_{i=1}^{M} p_i \mathbb{P}\left(Y \in \bigcup_{j \neq i} D_j \middle| \mathrm{H}i \right) \\
&= \sum_{i=1}^{M} p_i \mathbb{P}\left(Y \in \bigcup_{j \neq i} \{i \to j\} \middle| \mathrm{H}i \right) \\
&\geq \sum_{i=1}^{M} p_i \max_{j \neq i} \mathbb{P}\left(Y \in \{i \to j\} \middle| \mathrm{H}i \right) \\
&= \sum_{i=1}^{M} p_i \max_{j \neq i} \boxed{P(i \to j)}
\end{aligned}
$$

Example: M-QAM modulation

- Digital modulation (discrete-time complex baseband equivalent model)

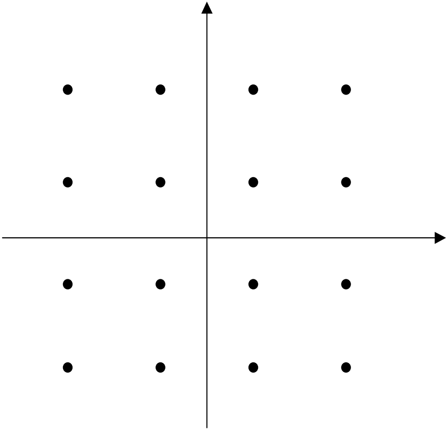
$$Y = X + Z$$

with $Z \sim \mathcal{CN}(0, N_0)$ and $X \sim$Uniform on $\mathcal{X}$.

- $\mathcal{X}$ is a squared QAM (Quadrature-Amplitude Modulation) signal constellation, of the form

$$\mathcal{X} = \left\{ \frac{\Delta}{2} [(2m - \sqrt{M} + 1) + \jmath(2n - \sqrt{M} + 1)] : m, n = 0, \ldots, \sqrt{M} - 1 \right\}$$

Example: 16-QAM constellation



- EXERCISE: check that $E_s = \frac{1}{M} \sum_{x \in \mathcal{X}} |x|^2 = \frac{\Delta^2}{6}(M-1)$, i.e., for given energy per symbol E_s, the minimum squared distance between the constellation points is

$$d_{\min}^2 = \Delta^2 = \frac{6E_s}{M-1}$$

- EXERCISE: find the MAP rule for this problem.
- EXERCISE: find upper and lower bounds and an exact closed-form expression of P_e in terms of the Signal to Noise Ratio $\mathrm{SNR} = E_s/N_0$.

Neyman-Pearson Hypothesis Testing

- In this context we have no a priori probabilities for the hypotheses.
- H0 and H1 are fundamentally asymmetric. Example in radar detection:
  H0 is "there is no enemy bomber"
  H1 is "there is an enemy bomber".
- Type I error (false alarm or false positive): falsely reject H0.
- Type II error (miss or false negative): falsely reject H1.
- This is in **sharp contrast** with the **average error probability** framework (with $c_{00} = c_{11} = 0$ and $c_{01} = c_{10} = 1$)
- For a given decision rule g, we have the false alarm probability and the successful detection probability

$$P_f(g) = \mathbb{P}(g(Y) = 1 | \mathrm{H0}), \quad P_d(g) = \mathbb{P}(g(Y) = 1 | \mathrm{H1})$$

- REMARK: the same **type I error** discussed in null hypothesis, which defines the **significance level** of a test; here, for the type II error, let $\beta = \mathbb{P}(\text{reject } H_1 | H_1)$, $1 - \beta$ is called the **power** of the test.

- Neyman-Pearson Hypothesis Testing Problem: for some fixed $\alpha \in [0,1]$ find g^* solution of:

$$\boxed{\max_g \; P_d(g) \; \text{ subject to } \; P_f(g) \leq \alpha}$$

Any rule solving this constrained maximization problem is called an **α-NP rule**.

- Randomized decision rule: for $g(y) \in [0,1]$ we can interpret $g(y)$ as the probability of accepting H1 given $Y = y$. Hence,

$$P_d(g) = \mathbb{E}[g(Y)|\mathrm{H1}] = \int g(y) f_1(y) dy$$

and

$$P_f(g) = \mathbb{E}[g(Y)|\mathrm{H0}] = \int g(y) f_0(y) dy$$

Neyman-Pearson Lemma

Theorem (Neyman-Pearson)

In the problem setting defined above, the optimal decision rule (Neyman-Pearson rule) is given by

$$g^*(y) = \begin{cases} 1 & \textit{if } L(y) > \eta^* \\ \gamma^* & \textit{if } L(y) = \eta^* \\ 0 & \textit{if } L(y) < \eta^* \end{cases}$$

where $L(y) = \frac{f_1(y)}{f_0(y)}$ is the Likelihood Ratio, the threshold $\eta^* \geq 0$ and the probability $\gamma^* \in [0,1]$ are chosen such that $P_f(g^*) = \alpha$.

Implications:

- let the likelihood ratio test $g(y)$ be designed so that H0 is rejected with significance level α
- then, for any other test of H0 with significance level at most α, its power against H1 is at most the power of this likelihood ratio test (**optimality**)

Proof of Neyman-Pearson

Form of the NP rule

- By construction, for any randomized decision rule $g(y)$ with $P_f(g) \leq \alpha$, we have

$$(g^*(y) - g(y))(f_1(y) - \eta^* f_0(y)) \geq 0, \quad \forall\, y \in \mathbb{R}$$

- Integrating over $\mathbb{R}$ and separating terms we get

$$\int g^*(y) f_1(y) dy - \int g(y) f_1(y) dy \geq \eta^* \left(\int g^*(y) f_0(y) dy - \int g(y) f_0(y) dy \right)$$

- Using the definitions of P_f and P_d, we can write

$$P_d(g^*) - P_d(g) \geq \eta^*(P_f(g^*) - P_f(g)) = \eta^*(\alpha - P_f(g)) \geq 0$$

Explicit expression of the NP-rule (existence)

- Let η^* denote the smallest number such that $\mathbb{P}(L(Y) > \eta^* | \mathrm{H0}) \leq \alpha$.
- If the inequality is strict, then let

$$\gamma^* = \frac{\alpha - \mathbb{P}(L(Y) > \eta^* | \mathrm{H0})}{\mathbb{P}(L(Y) = \eta^* | \mathrm{H0})}$$

- Then, the NP-rule g^* with threshold η^* and boundary randomization γ^* achieves

$$P_f(g^*) = \mathbb{P}(L(Y) > \eta^* | \mathrm{H0}) + \gamma^* \mathbb{P}(L(Y) = \eta^* | \mathrm{H0}) = \alpha$$

Example: Likelihood Ratio Test

Consider data $X_1, \ldots, X_n$ and the following null and alternative hypotheses:

$$H_0 \colon X_1, \ldots, X_n \sim \mathcal{N}(0, 1),$$
$$H_1 \colon X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1),$$

for some known $\mu \neq 0$ (which may be positive or negative).
The joint PDF of $(X_1, \ldots, X_n)$ under H_0 and H_1:

$$f_{H_0}(x_1, \ldots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{x_1^2 + \ldots + x_n^2}{2} \right)$$
$$f_{H_1}(x_1, \ldots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2}} = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{(x_1 - \mu)^2 + \ldots + (x_n - \mu)^2}{2} \right),$$

so likelihood ratio test:

$$L(X_1, \ldots, X_n) = \exp\left(\frac{-2\mu(\sum_{i=1}^n X_i) + n\mu^2}{2} \right) = \exp\left(-n\mu\bar{X} + \frac{n\mu^2}{2} \right), \tag{8}$$

that is, for $\mu > 0$, a strictly decreasing function of $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

Example: Likelihood Ratio Test

$$\boxed{L(X_1, \ldots, X_n) = \exp\left(-n\mu\bar{X} + \frac{n\mu^2}{2}\right).} \tag{9}$$

- reject small values of $L(X_1, \ldots, X_n)$ is equivalent, for $\mu > 0$, to reject for large values of $\bar{X} \sim \mathcal{N}(0, \frac{1}{n})$ under H_0
- Neyman–Pearson lemma tells us that the most **powerful** (1 - type II) test should reject when $\bar{X} > c$ for some threshold c chosen so that the **significant level** (type I) is α under H_0
- NOTE: the most powerful test against the alternative $H_1 \colon X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1)$, is the **SAME** for any $\mu > 0$, and neither the (expression of the) test statistic nor the rejections region depend on the parameter μ
- that is, this test is **uniformly most powerful** against the (one-sided) composite (i.e., combination of) alternative

$$H_1 \colon X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1), \text{ for some } \mu > 0. \tag{10}$$

- Question: what happens if $\mu < 0$? Reject large positive values or "large" negative values? And, is there a **single** most powerful test for the **two-sided** composite alternative

$$H_1 \colon X_1, \ldots, X_n \sim \mathcal{N}(\mu, 1), \text{ for some } \mu \neq 0? \tag{11}$$

Thank you! Q & A?

Exercises

Location testing with Gaussian error

- The two hypotheses are:
$$\begin{cases} \text{H0}: & Y = \mu_0 + Z \\ \text{H1}: & Y = \mu_1 + Z \end{cases}$$
where $Z \sim \mathcal{N}(0, \sigma^2)$ and the two hypotheses are equiprobable.
- We wish to find the optimal decision regions such that the average error probability $P_e(g)$ is minimized.

Location testing with Gaussian error: Neyman-Pearson

- The two hypotheses are:
$$\begin{cases} \text{H0}: & Y = \mu_0 + Z \\ \text{H1}: & Y = \mu_1 + Z \end{cases}$$
where $Z \sim \mathcal{N}(0, \sigma^2)$ and the two hypotheses are equiprobable.
- We wish to find the α-NP rule and expression the corresponding optimal probability P_d as a function of $\alpha \in [0, 1]$.