Testing Hypotheses

MIT 18.443

Dr. Kempthorne

Spring 2015

Outline

Hypothesis Testing: Bernoulli Trials

Statistical Decision Problem

- Two coins: Coin 0 and Coin 1

$$P(Head \mid Coin\ 0) = 0.5$$
$$P(Head \mid Coin\ 1) = 0.7$$

- Choose one coin, toss it 10 times and report number of Heads
- Decide which coin was chosen.

Hypothesis Testing Framework

- Data: X = number of heads in 10 tosses of coin
- Probability Model

$$X \sim Binomial(n = 10, prob = \theta)$$
$$P(X = x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x},\ x = 0, 1, \ldots, 10$$

- Hypotheses:

$$H_0 : \theta = 0.5$$
$$H_1 : \theta = 0.7$$

Outline

1. Hypothesis Testing
   - Bernoulli Trials
   - Bayesian Approach
   - Neyman-Pearson Framework
   - P-Values

Hypothesis Testing: Bernoulli Trials

Bayesian Approach to Hypothesis Testing

- Specify prior distribution on Hypotheses (θ)

$$\begin{aligned} P(H_0) &= P(\theta = 0.5) &= \pi_0 \\ P(H_1) &= P(\theta = 0.7) &= 1 - \pi_0. \end{aligned}$$

- Observe $X = x$ (count of heads on 10 tosses), which specifies the *likelihood* function.

$$lik(\theta) = P(X = x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

- Compute posterior probabilities

$$\begin{aligned} P(H_0 \mid x) &= \frac{P(H_0 \cap x)}{P(x)} &= \frac{P(H_0)P(X = x \mid H_0)}{P(x)} \\ P(H_1 \mid x) &= \frac{P(H_1 \cap x)}{P(x)} &= \frac{P(H_1)P(X = x \mid H_1)}{P(x)} \end{aligned}$$

Note: $P(x) = P(H_0 \cap x) + P(H_1 \cap x)$ (Law of Total Probability)

- Decision rule: $\delta(x) = 0$ if $P(H_0 \mid x) > 1/2$.

Decision Rule Based on Posterior Odds Ratio

- Posterior Odds Ratio:

$$
\begin{aligned}
\frac{P(H_0 \mid x)}{P(H_1 \mid x)} &= \frac{P(H_0)P(X = x \mid H_0)/P(x)}{P(H_1)P(X = x \mid H_1)/P(x)} \\
&= \left[\frac{P(H_0)}{P(H_1)}\right] \times \left[\frac{P(X = x \mid H_0)}{P(X = x \mid H_1)}\right] \\
&= [\text{Prior Odds}] \times [\text{Likelihood Ratio}]
\end{aligned}
$$

- Decision rule: $\delta(x) = 0$ if $\dfrac{P(H_0 \mid x)}{P(H_1 \mid x)} > 1$.
- Decision rule equivalent to $\delta(x) = 0$ if

$$[\text{Likelihood Ratio}] > c$$
$$(= P(H_1)/P(H_0))$$

- Likelihood Ratio measures evidence of x in favor of H_0
  Stronger evidence $\equiv$ Higher Likelihood Ratio (smaller x)

Bayes Decision Rules

Bayes Decision Rule:

- Given prior: $P(H_0) = \pi_0$ and $P(H_1) = 1 - \pi_0$,
- Accept H_0 if

  $$[\text{Likelihood Ratio}] > \frac{P(H_1)}{P(H_0)} = \frac{(1-\pi_0)}{\pi_0}$$

- Reject H_0 if

  $$[\text{Likelihood Ratio}] \leq \frac{P(H_1)}{P(H_0)} = \frac{(1-\pi_0)}{\pi_0}$$

Example Cases:

- $\pi_0 = 1/2$: Accept H_0 if [Likelihood Ratio] > 1
- $\pi_0 = 1/11$: Accept H_0 if [Likelihood Ratio] > 10.
  (Stronger evidence required to accept H_0
- $\pi_0 = 5/6$: Accept H_0 if [Likelihood Ratio] $> 1/5$.
  (H_0 accepted with weaker evidence)

Outline

1. Hypothesis Testing
   - Bernoulli Trials
   - Bayesian Approach
   - Neyman-Pearson Framework
   - P-Values

Neyman-Pearson Framework: Components

- Hypotheses
  - Null Hypothesis: H_0.
  - Alternative Hypothesis: H_1.
- Decision rule $\delta = \delta(X)$: accepts/rejects H_0 based on data X
  $$\delta(x) = 0 \text{ (accept } H_0\text{) and } \delta(x) = 1 \text{ (reject } H_0\text{)}$$
- Evaluate performance of decision rules using probabilities of two types of errors:
  - Type I Error: Rejecting H_0 when H_0 is true.
    $$P(\mathit{Type\ I\ Error}) = P(\delta = 1 \mid H_0)$$
  - Type II Error: Accepting H_0 when H_1 is true.
    $$P(\mathit{Type\ II\ Error}) = P(\delta = 0 \mid H_1)$$
- Optimal decision rule:
  $$\text{Minimizes: } \quad P(\mathit{Type\ II\ Error})$$
  $$\text{Subject to: } \quad P(\mathit{Type\ I\ Error}) \leq \alpha$$
  where $\alpha : 0 < \alpha < 1$ is the **significance level**.

- Consider "Risk Set" $\mathcal{R}$
  $$\mathcal{R} = \{(x, y) : x = P(\delta = 1 \mid H_0), \quad y = P(\delta = 0 \mid H_1), \text{ for a decision rule } \delta\}$$
  Note that:
  $$\mathcal{R} = \{(x, y) : x = P(\textit{Type I Error for } \delta), \quad y = P(\textit{Type II Error for } \delta), \quad \text{for a decision rule } \delta\}$$
- The Risk Set $\mathcal{R}$ is convex on the space of all decision rules
  $\mathcal{D} = \{\delta\}$ (including *randomized* decision rules)
- Apply convex optimization theory to solve for optimal δ using
  $h(\delta) = P(\delta = 1 \mid H_0)$ and $g(\delta) = P(\delta = 0 \mid H_1)$

Constrained Optimization:

Minimize: $g(\delta)$, subject to: $h(\delta) \leq \alpha$

Unconstrained Optimization of Lagrangian:

Minimize: $q(\delta, \lambda) = g(\delta) + \lambda[h(\delta) - \alpha]$

Solving the Optimization:

- Fix λ and minimize
  $$q^*(\delta) = g(\delta) + \lambda h(\delta) \text{ for } \delta \in \mathcal{D}$$
- For the solution δ^*, set $K^* = q^*(\delta^*)$
- The risk point for δ^* lies on the line
  $$\{(x, y) : K^* = y + \lambda x\}$$
  which is equivalent to
  $$\{(x, y) : y = K^* - \lambda x\}$$
- δ^* corresponds to the tangent point of $\mathcal{R}$ with slope$= -\lambda$.
- Specify λ to solve $h(\delta^*) = \alpha$.
  If $h(\delta^*) > \alpha$, then increase λ
  If $h(\delta^*) < \alpha$, then decrease λ

Nature of Solution: For given λ, the solution δ^* minimizes

$$\begin{aligned} q^*(\delta) &= g(\delta) + \lambda h(\delta) = P(\delta = 0 \mid H_1) + \lambda P(\delta = 1 \mid H_0) \\ &= \int_{\mathcal{X}} [(1 - \delta(x)) f_1(x) + \lambda \delta(x) f_0(x)] dx \\ &= 1 + \int_{\mathcal{X}} [\delta(x) \times [\lambda f_0(x) - f_1(x)] dx \end{aligned}$$

Nature of Solution: For given λ, the solution δ^* minimizes

$$\begin{aligned} q^*(\delta) &= g(\delta) + \lambda h(\delta) = P(\delta = 0 \mid H_1) + \lambda P(\delta = 1 \mid H_0) \\ &= \int_{\mathcal{X}} [(1 - \delta(x)) f_1(x) + \lambda \delta(x) f_0(x)] dx \\ &= 1 + \int_{\mathcal{X}} \delta(x) \times [\lambda f_0(x) - f_1(x)] dx \end{aligned}$$

To minimize $q^*(\delta)$:

- Note that $\delta(x) : 0 \leq \delta(x) \leq 1$ for all tests δ
- Set $\delta^*(x) = 0$ when $[\lambda f_0(x) - f_1(x)] > 0$
- Set $\delta^*(x) = 1$ when $[\lambda f_0(x) - f_1(x)] < 0$

The test δ^* accepts H_0, $\delta^*(x) = 0$, when

$$\frac{f_0(x)}{f_1(x)} > 1/\lambda$$

and rejects H_0, $\delta^*(x) = 1$, when

$$\frac{f_1(x)}{f_0(x)} > \lambda$$

The signficance level of δ^* is

$$\alpha = E[\delta^*(X) \mid H_0] = P[f_1(x)/f_0(x) > \lambda \mid H_0]$$

Neyman-Pearson Lemma:

- H_0 and H_1 are simple hypotheses.
- Define the test δ^* of significance level α using the Likelihood Ratio:

  $$\delta^*(X) = 1 \text{ when } \mathit{LikelihoodRatio} < c, \text{ and } c \text{ is}$$

  chosen such that:

  $$P(\delta^*(X) = 1 \mid H_0) = \alpha.$$

Then δ^* is the most powerful test of size α. For any other test δ':

$$\text{If } P(\delta' = 1 \mid H_0) \leq \alpha, \text{ then}$$
$$P(\delta'(X) = 1 \mid H_1) \leq P(\delta^*(X) = 1 \mid H_1)$$

Connection To Bayes Tests:

- Consider the Likelihood Ratio Test δ^* corresponding to c
- δ^* is the Bayes test corresponding to

  $$\frac{P(H_1)}{P(H_0)} = c = 1/\lambda.$$

Additional Terminology

- The **power** of a test rule δ is
  $$\beta = P(\textit{reject } H_1 \mid H_1) = 1 - P(\textit{Type II Error}).$$
- The **acceptance region** of a test rule δ is
  $$\{x : \delta(x) = 0\}$$
- The **rejection region** of a test rule δ is
  $$\{x : \delta(x) = 1\}$$

Additional Terminology

- A **test statistic** $T(X)$ is often associated with a decision rule δ, e.g.,
$$T(X) > t^* \iff \delta(X) = 1$$
- The distribution of $T(X)$ given H_0 is the **null distribution**.
- An hypothesis is a **simple hypothesis** if it completely specifies the distribution of X, and of $T(X)$.
E.g.,
$$X \sim f(x \mid \theta), \theta \in \Theta$$
$$H_0 : \theta = \theta_0 \text{ (simple)}$$
$$H_1 : \theta = \theta_1 \text{ (simple)}$$
- An hypothesis is a **composite hypothesis** if it does not completely specify the probability distribution.
E.g., $H_0 : X \sim Poisson(\theta)$ for some $\theta > 0$.

Additional Terminology

- **Uniformly Most Powerful Tests**.
    - Suppose

      $H_0 : \theta = \theta_0$ is **simple**

      $H_1 : \theta > \theta_0$ is **composite**

      (The value θ_0 is fixed and known.)
    - If the most powerful level-α test of H_0 versus a simple alternative $\theta = \theta_1 > \theta_0$ is the same for all alternatives $\theta_1 > \theta_0$, then it is the **Uniformly Most Powerful Test** of H_0 versus H_1.
- **One-sided Alternative**: $H_1 : \theta > \theta_0$, or, $H_1 : \theta < \theta_0$
- **Two-sided Alternative**: $H_1 : \theta \neq \theta_0$

Outline

P-Values

Neyman-Pearson Hypothesis-Testing Framework

- $X \sim f(x \mid \theta)$, $\theta \in \Theta$ (pdf or pmf)
- Test Hypotheses:

  $H_0 : \theta = \theta_0$ versus an alternative H_1

  (θ_0 is a fixed value, so H_0 is **simple**)
- **Test Statistic:**

  $T(X)$, defined so that large values are evidence against H_0
- The **rejection region** is

  $\{x : T(X) > t_0\}$ where t_0 is chosen to that $P(T \geq t_0 \mid H_0) = \alpha$, (the **significance level** of test)

Definition: Given $X = x$ is observed, the **P-value** of the test statistic $T(x)$ is

$$P\text{-}Value = P(T(X) > t(x) \mid H_0).$$

What P-Values Are:

- The P-Value is the smallest significance level at which H_0 would be rejected.
- The P-Value is the chance of observing evidence as extreme or more extreme than $T(x)$ under the probability model of H_0.
- The P-Value measures how unlikely (surprising) the data are if H_0 is true.

What P-Values Are Not:

- The P-value is **not** the probability H_0 is true.