Week3

Let $N \sim \text{Poisson}(\lambda)$. Given $N = n$, toss a fair coin n times and denote the number of heads obtained by X. What is the distribution of X?

$X \sim \text{Poisson}(\lambda/2)$

- To show X and Y are independent, verify

$$f_{XY}(t_1, t_2) = f_X(t_1) f_Y(t_2)$$

for **all** $t_1 \in T_X$, $t_2 \in T_Y$.

- To show X and Y are dependent, verify

$$f_{XY}(t_1, t_2) \neq f_X(t_1) f_Y(t_2)$$

for **some** $t_1 \in T_X$ and $t_2 \in T_Y$.

  - Special case: $f_{XY}(t_1, t_2) = 0$ when $f_X(t_1) \neq 0$, $f_Y(t_2) \neq 0$

In the case of Geometric distribution

$$P(X>n) = \sum_{k=n+1}^{\infty} (1-p)^{k-1} p = (1-p)^n p + (1-p)^{n+1} p + \cdots$$
$$= \frac{(1-p)^n p}{1-(1-p)} = (1-p)^n$$

Memoryless property of Geometric

$$P(X > m+n \mid X > m) = P(X > n)$$

Sum of n indep Bernoulli(p) = Binomial(n, p)

If X1 = Binomial (n1, p) and X2 = Binomial (n2, p), conditional probability mass function of X1 given that X1 + X2 = m is HyperGeo(n1 + n2, n1, m)

If $X \sim \text{Geometric}(p)$ and $Y \sim \text{Geometric}(p)$ are two independent random variables and $Z = X + Y$, then
$P(Z = n) = (n-1)p^2(1-p)^{n-2}$ (try derivation by yourself)

Let $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$ be independent.

$Z \sim \text{Poisson}(\lambda_1 + \lambda_2)$

Week4

Properties of Expectation and Variance of random variables

E[cX] = cE[X]

E[X + Y] = E[X] + E[Y]

E[aX + bY] = aE[X] + bE[Y]

$\text{Var}(aX) = a^2\text{Var}(X)$

$\text{Var}(X + a) = \text{Var}(X)$

Suppose X and Y are **independent** random variables.

1. $E[XY] = E[X]E[Y]$
2. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

Covariance and Correlation coefficient

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)}.$$

Properties of CDF of a continuous random variable X:

- $F_X(b) - F_X(a) = P(a < X \leq b)$
- F_X: non-decreasing function taking non-negative values
- As $x \to -\infty$, F_X goes to 0.
- As $x \to \infty$, F_X goes to 1.
- If $F(x)$ is continuous at x_0, $P(X = x_0) = 0$ (non-intuitive!)

Definition (PDF)

A continuous random variable X with CDF $F_X(x)$ is said to have a PDF $f_X(x)$ if, for all x_0,

$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)dx.$$

A function $f : \mathbb{R} \to \mathbb{R}$ is said to be a density function if

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $f(x)$ is piecewise continuous

In the case of most distributions – exponential, uniform – PDF can be obtained by integrating the CDF. However, in the case of normal distribution, it's very difficult to integrate and hence the need to standardize the PDF expression to obtain the Z-score and use the Z-table to find the PDF.

Here's an example:

Problem

$\mu=2, \sigma^2=5$

Suppose $X \sim \text{Normal}(2, 5)$. Find $P(X < 5)$, $P(X < 10)$, $P(X < -5)$, $P(X < -10)$, $P(X > 5)$, $P(X > 10)$.

$Z = \frac{X-2}{\sqrt{5}} \sim \text{Normal}(0,1)$ $\qquad X = \sqrt{5} Z + 2$

Assume: $F_Z(z)$ (CDF of Z) is known

$X < 5 \iff \sqrt{5} Z + 2 < 5 \iff Z < 3/\sqrt{5}$

$P(X<5) = P(Z < 3/\sqrt{5}) = F_Z(3/\sqrt{5})$

$X > 5 \iff \sqrt{5} Z + 2 > 5 \iff Z > 3/\sqrt{5}$

$P(X>5) = P(Z > 3/\sqrt{5}) = 1 - F_Z(3/\sqrt{5})$

Week7

Let $X_1, X_2, \ldots, X_n$ be iid samples whose distribution has a finite mean μ and variance σ^2. The sample mean $\overline{X} = \frac{X_1+X_2+\cdots+X_n}{n}$ has expected value and variance given by

$$E[\overline{X}] = \mu, Var(\overline{X}) = \frac{\sigma^2}{n}.$$

Let $X_1, X_2, \ldots, X_n$ be iid samples whose distribution has a finite variance σ^2. The sample variance $S^2 = \frac{(X_1-\overline{X})^2+\cdots+(X_n-\overline{X})^2}{n}$ has expected value given by

$$E[S^2] = \sigma^2.$$

As n increases, sample variance takes values close to distribution variance.

Let $X_1, X_2, \ldots, X_n$ be iid samples from the distribution of X. Let A be an event defined using X and let $P(A)$ be the probability of A. The sample proportion of A, denoted $S(A)$, has expected value and variance given by

$$E[S(A)] = P(A), Var(S(A)) = \frac{P(A)(1-P(A))}{n}.$$

If we take multiple samplings, by the previous result we know $E[\overline{X}] = \mu$. But each sampling can have different mean $\bar{X}$ and not necessarily equal to μ. The difference is given by WLLN. It's not exact, but a good bound.

$$P(|\overline{X} - \mu| > \delta) \leq \frac{\sigma^2}{n\delta^2} \to 0.$$

It also implies the following:

With probability more than $1 - \frac{\sigma^2}{n\delta^2}$, sample mean lies in $[\mu - \delta, \mu + \delta]$

The WLLN can be improved by using an exponential bound (Chernov)

$$P(S > n\delta/2) \leq e^{-n\delta^2/4}$$

Here S is the sum of the random variables. Note S/n = $\bar{X}$

Here's an application of the above two bounds (WLLN and Chernov) to Binomial (n, ½).

n	Event, $\delta = 0.6$	Prob	$1/n\delta^2$	$e^{-n\delta^2/4}$
10	$Y - 5 > 5 \times 0.6$	0.0107	0.278	0.407
50	$Y - 25 > 25 \times 0.6$	2.81×10^{-6}	0.056	0.011
100	$Y - 50 > 50 \times 0.6$	1.35×10^{-10}	0.028	1.23×10^{-4}
200	$Y - 100 > 100 \times 0.6$	4.16×10^{-19}	0.014	1.52×10^{-8}
400	$Y - 200 > 200 \times 0.6$	5.40×10^{-36}	0.007	2.32×10^{-16}

Notice the Chernov bound is closer to the actual Probability (third column in the table)

MGF of sum of independent random variables is product of the individual MGFs

Use this to find the distribution of the sum of random variables. Here's some examples.

$X_1, X_2 \sim$ iid X, $Y = X_1 + X_2$

- $X \in \{\overset{1/2}{-1}, \overset{1/4}{0}, \overset{1/4}{2}\}$
  - $M_X(\lambda) = 0.5e^{-\lambda} + 0.25 + 0.25e^{2\lambda}$
  - $M_Y(\lambda) = 0.25e^{-2\lambda} + 0.25e^{-\lambda} + 0.0625 + 0.25e^{\lambda} + 0.125e^{2\lambda} + 0.0625e^{4\lambda}$

The resulting distribution is given here.

X	-2	-1	0	1	2	4
P(X = x)	0.25	0.25	0.0625	0.25	.125	0.0625

MGF of a random variable X (with zero mean) is defined as $M_X(\lambda) = E[e^{\lambda X}]$

MGF of random variable X in {x1, x2, … ,xn} is defined as

- $M_X(\lambda) = f_X(x_1)e^{\lambda x_1} + f_X(x_2)e^{\lambda x_2} + \cdots$, when discrete
- $M_X(\lambda) = \int_{x \in T_X} f_X(x) e^{\lambda x}\, dx$ when continuous.

Why is it called moment-generating function?

$$E[e^{\lambda X}] = E[1 + \lambda X + \frac{\lambda^2}{2!}X^2 + \frac{\lambda^3}{3!}X^3 + \cdots]$$

$$\text{MGF} \quad = 1 + \underbrace{\lambda E[X]}_{\text{1st moment}} + \underbrace{\frac{\lambda^2}{2!}E[X^2]}_{\text{2nd moment}} + \underbrace{\frac{\lambda^3}{3!}E[X^3]}_{\text{3rd moment}} + \cdots$$

Using CLT to approximate probability: Let $X_1, X_2, \ldots, X_n \sim$ iid X with $E[X] = \mu, \mathrm{Var}(X) = \sigma^2$.
Define $Y = X_1 + X_2 + \ldots + X_n$. Then,

$$\frac{Y - n\mu}{\sqrt{n}\sigma} \approx \text{Normal}(0,1).$$

$$P\left(\frac{Y-n\mu}{\sqrt{n}\sigma} > \frac{\delta\sqrt{n}\mu}{\sigma}\right) \approx 1 - F\left(\frac{\delta\sqrt{n}\mu}{\sigma}\right)$$

Here're two examples of application of CLT.

- $X \sim \{\overset{1/6}{-3}, \overset{1/8}{-1}, \overset{1/4}{0}, \overset{1/8}{1}, \overset{1/3}{3/2}\}$
  - $\mu = 0,\ \sigma^2 = 5/2$
  - CLT: $\frac{Y}{\sqrt{5n/2}} \approx \text{Normal}(0,1)$
  - $P(Y > \delta n) = P\left(\frac{Y}{\sqrt{5n/2}} > \delta\sqrt{2n/5}\right) \approx 1 - F(\delta\sqrt{2n/5})$
    - ★ $n = 10,\ \delta = 1$: ≈ 0.0228
    - ★ $n = 100,\ \delta = 1$: $\approx 1.27 \times 10^{-10}$
- $X \sim \text{Uniform}[-1,1]$ (continuous)
  - $\mu = 0,\ \sigma^2 = 1/3$
  - CLT: $\sqrt{3}Y \approx \text{Normal}(0,1)$
  - $P(Y > 0.1\sqrt{n}) = P(\sqrt{3}Y > 0.1\sqrt{3n}) \approx 1 - F(0.1\sqrt{3n})$
    - ★ $n = 10$: ≈ 0.2919
    - ★ $n = 100$: ≈ 0.0416

Sum of n iid $\text{Exp}(\beta)$ is $\text{Gamma}(n, \beta)$

Sum of n independent $\text{Gamma}(\alpha, \beta)$ is $\text{Gamma}(n\alpha, \beta)$

Square of $\text{Normal}(0, \sigma^2)$ is $\text{Gamma}(1/2, 1/2\sigma^2)$

Suppose $X \sim \text{Gamma}(\alpha, 1/\theta)$, $Y \underset{\text{indep}}{\sim} \text{Gamma}(\beta, 1/\theta)$, then

$$\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$$

$\text{Var}(\frac{X_1+\cdots+X_n}{n}) = \frac{\text{Var}(X)}{n}$ will hold for all $X_1, \ldots, X_n \sim$ iid X.

This is because…

$$\text{Var}\left(\frac{X_1+X_2+\cdots X_n}{n}\right) = \text{Var}\frac{X_1}{n} + \text{Var}\frac{X_2}{n} + \cdots \text{Var}\frac{X_n}{n}$$

$$= \frac{1}{n^2}\text{Var}X_1 + \frac{1}{n^2}\text{Var}X_2 + \cdots + \frac{1}{n^2}\text{Var}X_n$$

$$= \frac{1}{n^2} \cdot n\text{Var}X = \frac{\text{Var}X}{n}$$

(because Variances of iid can be added)

Week8

$\text{Bias}(\hat{\theta}, \theta) = E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta.$

Since Error $= \hat{\theta} - \theta$, bias is the expected value of Error

$\text{Risk}(\hat{\theta}, \theta) = E[(\hat{\theta} - \theta)^2]$. It is called *mean squared-error* (MSE) or *second moment of error*.

$Risk(\hat{\theta}, \theta) = Bias(\hat{\theta}, \theta)^2 + Var(\hat{\theta})$

Method of moments estimation procedure

1. We'll start with the assumption that Sample moments(m1, m2..) = distribution moments(M1, M2,…)
2. Calculate sample moments. m1 = $\frac{x_1+x_2+\cdots+x_n}{n}$, m2 = $\frac{x_1^2+x_2^2+\cdots+x_n^2}{n}$. Note that m2 is needed only if you're doing computations with variance.

3. Find distribution moments from the predefined formulae for the distribution. For example, in the case of Bernoulli M1 = p, in the case of Exponential(λ) M1 = 1/λ
4. Now, equate these two. Estimate p or λ in terms of m1. If we've multiple parameters to estimate, use m2, m3 etc.

Week9

Start with a random variable p is Uniform{0.25, 0.75} (called prior). We're trying to estimate p, by looking a sample

S = 1, 0, 1, 1, 0

There are two possible Bayesian estimators

Estimator 1: Since $P(p = 0.75|S) > P(p = 0.25|S)$, we could estimate $\hat{p} = 0.75$

Estimator 2: Posterior mean, $\hat{p} = 0.25\, P(p = 0.25|S) + 0.75\, P(p = 0.75|S) = 0.625$

where,

★ $P(p = 0.25|S) = P(S|p = 0.25)P(p = 0.25)/P(S) = 0.25^3 \times 0.75^2 \times 0.5/P(S) = 0.25$

★ $P(p = 0.75|S) = 0.75^3 \times 0.25^2 \times 0.5/P(S) = 0.75$

★ $P(S) = 0.25^3 \times 0.75^2 \times 0.5 + 0.75^3 \times 0.25^2 \times 0.5 = 0.25^2 \times 0.75^2 \times 0.5$

Now, if we were to change the prior distribution, the resulting estimate of p changes.

In general, this is written as

Bayes' rule: posterior $\propto$ likelihood $\times$ prior

$$\underbrace{P(\Theta = \theta|S)}_{\text{Posterior}} = \underbrace{P(S|\Theta = \theta)}_{\text{likelihood}}\underbrace{f_\Theta(\theta)}_{\text{prior}}/\underbrace{P(S)}_{\text{Normalization}}$$

Estimating p for Bernoulli(p) samples:

1. If you choose $\mathbf{p} \sim \text{Uniform}[0, 1]$ as the prior, the posterior density is $\text{Beta}(w + 1, n - w + 1)$

and the posterior mean is $\hat{p} = \dfrac{X_1 + \cdots + X_n + 1}{n + 2}$

2. If you choose $\mathbf{p} \sim \text{Beta}(\alpha, \beta)$ as the prior, the posterior density is $\text{Beta}(w + \alpha, n - w + \beta)$ and the posterior mean is $\frac{x_1+x_2+\cdots+x_n+\alpha}{n+\alpha+\beta}$

NOTE1: if α = 1 and β = 1, you get the same equation as the first one where prior is uniform.

NOTE2: if α = 0 and β = 0, you get posterior mean as the maximum likelihood.

Estimating μ for Normal (μ, σ^2) samples, when σ^2 is known.

If you choose $M \sim \text{Normal}(\mu_0, \sigma_0^2)$ as the prior, the posterior mean is $\overline{X}\frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2} + \mu_0\frac{\sigma^2}{n\sigma_0^2+\sigma^2}$

Here's a sample problem of estimation using MME, ML and Bayesian methods:

Consider n iid samples from a Geometric(p) distribution.

- Find the method of moments estimate.
- Find the MLE.
- Using a Uniform[0, 1] prior, find the posterior distribution and mean.

$X \sim \text{Geometric}(p)$

Samples: $x_1, x_2, \dots, x_n$

$P(X_i = x_i) = (1-p)^{x_i-1}p$

Method of moments

$E[X] = 1/p$

$\bar{x} = 1/p$

$p = \frac{1}{\bar{x}} = \frac{n}{x_1+\dots+x_n}$

$\hat{p}_{mm} = \frac{n}{x_1+\dots+x_n}$

Maximum Likelihood

$L = (1-p)^{x_1-1}p \cdot (1-p)^{x_2-1}p \cdots (1-p)^{x_n-1}p$

$= p^n (1-p)^{x_1+\dots+x_n-n}$

$\log L = n \log p + (x_1+\dots+x_n-n)\log(1-p)$

$\frac{n}{p} + (x_1+\dots+x_n-n)\frac{1}{1-p}(-1) = 0$

$p = \frac{n}{x_1+\dots+x_n}$

$\hat{p}_{ML} = \frac{n}{x_1+\dots+x_n}$

Bayesian Prior: $p \sim \text{Uniform}[0,1]$

$$\text{Posterior} \propto \underbrace{p^{n}(1-p)^{x_1+\dots+x_n-n}}_{\text{likelihood}} \cdot \underbrace{1}_{\text{prior}}$$

$\text{Beta}(\alpha,\beta)$: Expected value $= \dfrac{\alpha}{\alpha+\beta}$

$$\sim \text{Beta}(n+1,\ x_1+\dots+x_n-n+1)$$

$$\hat{p}_{\text{Bayes}} = \text{Posterior mean} = \frac{n+1}{\cancel{n}+1+x_1+\dots+x_n-\cancel{n}+1} = \frac{n+1}{x_1+\dots+x_n+2}$$

8)

Let a coin be tossed 10 times and number of heads, n, be recorded. This process is repeated 100 times and results are provided in the following frequency table.

n	frequency	n	frequency
0	2	6	10
1	2	7	4
2	15	8	1
3	16	9	0
4	25	10	0
5	25		

We know that mean of the binomial distribution is given by $np = 10p$.

Therefore, $10p = \dfrac{X_1 + X_2 + \ldots + X_{100}}{100}$

$\Rightarrow p = \dfrac{X_1 + X_2 + \ldots + X_{100}}{1000}$

$\Rightarrow p = \dfrac{\sum n(\text{frequency})}{1000}$

$\Rightarrow p = 0.40$

So, the given data fits to Binomial$(10, 0.4)$.

How to find α, β for which $\Pr(|X - \mu| < \alpha) = \beta$?

- Suppose X is continuous and has CDF F_X
  - $P(X \leq x) = F_X(x)$

$$P(|X - \mu| < \alpha) = P(\mu - \alpha < X < \mu + \alpha) = F_X(\mu + \alpha) - F_X(\mu - \alpha)$$

If distribution of X is symmetric about its mean, then

$$F_X(\mu + \alpha) - F_X(\mu - \alpha) = 1 - 2F_X(\mu - \alpha)$$

For normal samples,

$$\hat{\mu} \sim \text{Normal}(\mu, \sigma^2/n)$$

$$P(|\hat{\mu} - \mu| < \alpha) = \beta \leftrightarrow P\left(\left|\frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}}\right| < \frac{\alpha}{\sigma/\sqrt{n}}\right) = \beta$$

Here is the table we use to arrive at the alpha value, from a given beta.

β	$\frac{\alpha}{\sigma/\sqrt{n}}$
0.68	0.99
0.90	1.64
0.95	1.96

This is explained in *Miscellanous Problems.docx* under *Solve with us-8* section.

Week10

Problem 1

Consider 100 tosses of a coin, which could be either authentic with probabiltiy of heads equal to 0.5, or counterfeit with probability of heads 0.6. Suppose T is the number of heads seen. Consider a test that rejects H_0 if $T > c$ for some constant c. What is the significance level of the test? What is the power of the test?

$A = \{\text{outcomes}: T \le c\}$ $\quad T \sim \text{Binomial}(100, P(H))$

$$\alpha = P(\underbrace{\text{Reject } H_0}_{\text{outcome} \in A^c} \mid \underbrace{H_0}_{P(H)=0.5}) = P(A^c \mid P(H)=1/2) = \sum_{k=c+1}^{100} \binom{100}{k} \left(\tfrac{1}{2}\right)^k \left(\tfrac{1}{2}\right)^{100-k}$$

→ Fix an α and find c

$$1-\beta = P(\underbrace{\text{Reject } H_0}_{\text{outcome} \in A^c} \mid \underbrace{H_A}_{P(H)=0.6}) = P(A^c \mid P(H)=0.6) = \sum_{k=c+1}^{100} \binom{100}{k} 0.6^k \, 0.4^{100-k}$$

→ lower c ⇒ higher power.

→ Fix α. Find lowest possible c.

Problem 4

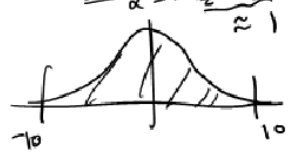
Consider 100 samples $X_1, \ldots, X_{100} \sim$ iid Normal$(\mu, 1)$. Let the null and alternative hypothesis be $H_0 : \mu = -1$ and $H_A : \mu = 1$. Suppose $T = (X_1 + \cdots + X_{100})/100$. Consider a test that rejects H_0 if $T > c$ for some constant c. What is the significance level of the test? What is the power of the test?

$T \sim N(\mu, \frac{1}{100})$ $\quad A = \{T \le c\}$

$$\alpha = P(T > c \mid \mu = -1) = P\left(\underbrace{\frac{N(-1,\frac{1}{100})+1}{1/10}}_{Z \sim N(0,1)} > \frac{c+1}{1/10}\right) = P(Z > 10(c+1)) = 1 - F_Z(10(c+1))$$

$$\overset{c=0}{=} \alpha = 1 - F_Z(10) \approx 1$$

$$1-\beta = P(T > c \mid \mu = 1) = P(Z > 10(c-1)) = 1 - F_Z(10(c-1))$$



$$\overset{c=0}{=} 1-\beta = 1 - \underbrace{F_Z(-10)}_{\approx 0}$$

p-value or alpha?

- Samples are given, and there is some hypothesis that needs to be tested
- Step 1: Decide on the null and alternative hypotheses H_0 and H_A
- Step 2: Decide on the test statistic T
- Step 3: "Philosophy" of testing
  - Choice 1: Pick a significance level first
    - ★ Probability of Type I error can be fixed in some applications. In those cases, significance level is easy to fix
    - ★ Historically, in many applications, 0.05 or 0.01 is accepted as a common significance level
    - ★ Find rejection region (find the *critical value* c and reject H_0 if $T > c$, for example)
  - Choice 2: Use P-value
    - ★ Report the P-value
    - ★ If P-value is *low enough*, choose to reject H_0; otherwise, accept H_A
    - ★ How low is low enough? Depends on applications and other information

---

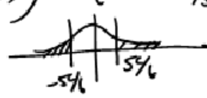
Suppose $X \sim \text{Normal}(\mu, 36)$. For $n = 25$ iid samples of X, the observed sample mean is 6.2. What conclusion would a z-test reach if the null hypothesis assumes $\mu = 4$ (against an alternative hypothesis $\mu \neq 4$) at a signifiance level of $\alpha = 0.05$? What if the null hypothesis assumes $\mu = 8$ (against an alternative hypothesis $\mu < 8$)?

$H_0: \mu = 4$, $H_A: \mu \neq 4$, Test: Reject H_0, if $|\bar{X} - 4| > c$

$\bar{X} \sim N\left(\mu, 36/25\right)$ $\quad \frac{\bar{X}-\mu}{(6/5)} \sim Z$ $\quad \bar{X} > 4 + c$ (or) $\bar{X} < 4 - c$

$$\alpha = P(|\bar{X} - 4| > c \mid \mu = 4) = P\left(\underbrace{\left|\frac{\bar{X}-4}{6/5}\right|}_{\sim |Z|} > \frac{c}{(6/5)}\right) = P\left(|Z| > \frac{c}{6/5}\right) = 2 F_Z\left(-\frac{5c}{6}\right)$$

$$c = -\frac{6}{5} F_Z^{-1}\left(\frac{0.05}{2}\right) = 2.352$$



Since $|6.2 - 4| = 2.2 < 2.352$, Accept H_0

$H_0: \mu = 8$, $H_A: \mu < 8$, Test: Reject H_0 if $\bar{X} < c$

$$\alpha = P(\bar{X} < c \mid \mu = 8) = F_Z\left(\frac{c-8}{(6/5)}\right) = 0.05 \Rightarrow c = 8 + \frac{6}{5} F_Z^{-1}(0.05) = 6.02$$

Since $\bar{X} = 6.2 > 6.02$, z-test @ $\alpha = 0.05$: Accept H_0