Probability and Random Processes

Sanjit Kaul

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Parameter Estimation

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Sampling From a Finite Population

$\underline{n}$ $\quad \underbrace{k=2}$ $\qquad X_1, X_2, \ldots, X_n$

$X_i \sim X$

$X^{(1)}$

$X^{(2)}$

Are $X^{(1)}$ and $X^{(2)}$ independent.

Sampling w/o replacement

with

Definition 7.1 ***Sample Mean***

For iid random variables $X_1, \ldots, X_n$ *with PDF* $f_X(x)$*, the* sample mean *of* X *is the random variable*

$$M_n(X) = \frac{X_1 + \cdots + X_n}{n}.$$

Theorem 7.1

The sample mean $M_n(X)$ has expected value and variance

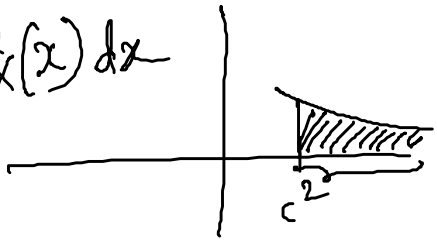
$$E\left[M_n(X)\right] = E\left[X\right], \qquad \mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}.$$

Quiz 7.1

Let X be an exponential random variable with expected value 1. Let $M_n(X)$ denote the sample mean of n independent samples of X. How many samples n are needed to guarantee that the variance of the sample mean $M_n(X)$ is no more than 0.01?

$$E[X] = \int_0^\infty x f_X(x)\,dx = \int_0^{c^2} x f_X(x)\,dx + \int_{c^2}^\infty x f_X(x)\,dx$$

$$\geq \int_{c^2}^\infty x f_X(x)\,dx \geq \int_{c^2}^\infty c^2 f_X(x)\,dx$$



Theorem 7.2 Markov Inequality

For a random variable X such that $P[X < 0] = 0$ and a constant c,

$$\int_{c^2}^\infty f_X(x)\,dx \qquad P\left[X \geq c^2\right] \leq \frac{E[X]}{c^2} \qquad Z = (Y - E[Y])^2$$

$$P\left[|Y - E[Y]| \leq c\right] \geq \frac{Var[Y]}{c^2} \qquad P\left[Z \geq c^2\right] \leq \frac{Var[Y]}{c^2}$$

$$P\left[(Y - E[Y])^2 \geq c^2\right] \leq \frac{Var[Y]}{c^2}$$

Theorem 7.3 Chebyshev Inequality

For an arbitrary random variable Y and constant $c > 0$,

$$P\left[|Y - \mu_Y| \geq c\right] \leq \frac{\mathrm{Var}[Y]}{c^2}.$$

$$|Y-\mu_Y| \geq c \iff (Y-\mu_Y)^2 \geq c^2$$

$$-(Y-\mu_Y) \geq c$$

Quiz 7.2

Elevators arrive randomly at the ground floor of an office building. Because of a large crowd, a person will wait for time W in order to board the third arriving elevator. Let X_1 denote the time (in seconds) until the first elevator arrives and let X_i denote the time between the arrival of elevator $i-1$ and i. Suppose X_1, X_2, X_3 are independent uniform $(0, 30)$ random variables. Find upper bounds to the probability W exceeds 75 seconds using

$E[W] = 45$

0 75

$P[W > 75] \leq \frac{E[W]}{75}$

(1) the Markov inequality,

$|W-45| > 30$

$-(W-45) > 30$

$P[W - E[W] > 75 - E[W]]$

(2) the Chebyshev inequality.

$W-45$

< -30

$W < -75$

$\boxed{P[|W-\mu_W| \geq c] \leq \frac{Var[W]}{c^2}}$

$P[|W-45| > 30]$

$W-45 > 30$

$p = \frac{1}{36}$ $\quad P[R=r] = {}_{r-1}C_2\, p^2 (1-p)^{r-3} p \quad P[R=r] = \sum$

Problem 7.2.4

$R = X_1 + X_2 + X_3 \qquad E[R] = 3E[X] = \frac{3}{p}$

In a game with two dice, the event *snake eyes* refers to both dice showing one spot. Let R denote the number of dice rolls needed to observe the third occurrence of *snake eyes*. Find $\quad P[R \geq (\sqrt{250})^2] \leq \frac{3/p}{250}$

(a) the upper bound to $P[R \geq 250]$ based on the Markov inequality,

(b) the upper bound to $P[R \geq 250]$ based on the Chebyshev inequality,

(c) the exact value of $P[R \geq 250]$.

$P[R \geq 250]$

$= P[R - 108 \geq 142]$

$P[|R - \mu_R| \geq c] \leq \frac{Var[R]}{c^2}$

$P[|R - 3(36)| \geq c] \leq \frac{Var[R]}{c^2}$

$\{|R-108| \geq 142\} = \{-(R-108) \geq 142\} \cup \{R-108 \geq 142\}$

Section 7.3

Point Estimates of Model Parameters

$\hat{R}_1 = \frac{X_1}{1}$ $\hat{R}_2 = \frac{X_1 + X_2}{2}$ $\hat{R}_3 = \frac{X_1 + X_2 + X_3}{3}$ [if estimator is sample mean]

$r = E[X]$

Definition 7.2 Consistent Estimator

The sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *of the parameter* r *is* consistent *if for any* $\epsilon > 0$,

$$\lim_{n \to \infty} P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

$X \leftarrow \mathcal{N}$ $\qquad E\left[\frac{X_1 + X_2 + X_3}{3}\right] = E[X]$

Definition 7.3 ***Unbiased Estimator***

An estimate, $\hat{R}$*, of parameter* r *is* unbiased *if* $E[\hat{R}] = r$*; otherwise,* $\hat{R}$ *is* biased*.*

Asymptotically Unbiased

Definition 7.4 Estimator

The sequence of estimators $\hat{R}_n$ *of parameter* r *is* asymptotically unbiased *if*

$$\lim_{n\to\infty} E[\hat{R}_n] = r.$$

Definition 7.5 ***Mean Square Error***

The mean square error *of estimator* $\hat{R}$ *of parameter* r *is*

$$e = E\left[(\hat{R} - r)^2\right].$$

$e_n = E[(\hat{R}_n - r)^2]$

For unbiased $\hat{R}_n$

$e_n = E[(\hat{R}_n - E[\hat{R}_n])^2]$

$= Var[\hat{R}_n]$

Theorem 7.4

Fact given that estimates are unbiased

If a sequence of unbiased estimates $\hat{R}_1, \hat{R}_2, \ldots$ of parameter r has mean square error $e_n = \mathrm{Var}[\hat{R}_n]$ satisfying $\lim_{n\to\infty} e_n = 0$, then the sequence $\hat{R}_n$ is consistent.

Given

$\lim_{n\to\infty} P[|\hat{R}_n - r| \geq \epsilon] = 0$ To prove

$P[|\hat{R}_n - r| \geq \epsilon] = P[|\hat{R}_n - E[\hat{R}_n]| \geq \epsilon] \leq \frac{Var(\hat{R}_n)}{\epsilon^2}$

$$E[\hat{R}_k] = E\left[\frac{N_k}{k}\right] = \frac{1}{k} E[N_k] = r$$

Example 7.4 Problem

In any interval of k seconds, the number N_k of packets passing through an Internet router is a Poisson random variable with expected value $E[N_k] = kr$ packets. Let $\hat{R}_k = N_k/k$ denote an estimate of r. Is each estimate $\hat{R}_k$ an unbiased estimate of r? What is the mean square error e_k of the estimate $\hat{R}_k$? Is the sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ consistent?

$$e_k = E\left[(\hat{R}_k - r)^2\right] = E\left[(\hat{R}_k - E[\hat{R}_k])^2\right] = Var[\hat{R}_k]$$

$$= \frac{Var[N_k]}{k^2}$$

Theorem 7.5

The sample mean $M_n(X)$ is an unbiased estimate of $E[X]$.

Theorem 7.6

The sample mean estimator $M_n(X)$ has mean square error

$$e_n = E\left[(M_n(X) - E[X])^2\right] = \mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}.$$

This is also the variance of the sample mean estimator. Square-root of this is called the standard error.
Given the mean and variance of the estimator, what can we say about the sample mean that results from n IID trials?

Markov's Inequality

Let X be a nonnegative RV.

We want to find an upper bound for the tail probability.

$$E[X] = \int_0^\infty x f_X(x)\,dx$$
$$= \int_0^c x f_X(x)\,dx + \int_c^\infty x f_X(x)\,dx$$
$$\geq \int_c^\infty x f_X(x)\,dx$$
$$\geq c \int_c^\infty f_X(x)\,dx$$
$$\Rightarrow \boxed{P[X \geq c] \leq \frac{E[X]}{c}}$$

Let $X = (Y-\mu_Y)^2$ where $E[Y] = \mu_Y$.

$$\therefore P[(Y-\mu_Y)^2 \geq c^2] \leq \frac{E[(Y-\mu_Y)^2]}{c^2}$$
$$= \frac{Var[Y]}{c^2} \quad \text{Var}[Y] = \sigma_Y^2$$

Or yet $\boxed{P[|Y-\mu_Y| \geq c] \leq \frac{\sigma_Y^2}{c^2}}$

This is called the Chebyshev Ineq.

(NB) The smaller the $Var[Y]$, the more probably that a value drawn from Y is far from the $E[Y]$.

(NB) Chebyshev is true for any RV. It does not need to be nonnegative.

Problem 7-31

$P[\text{Occurrence of a 3 in a roll of a fair die}] = \left(\frac{1}{6}\right)^k$ for k.

Let N be the number of dice rolls needed to observe the kth occurrence of a roll of a 3.

$$P[N=n] = p \cdot \binom{n-1}{k-1} p^{k-1} \cdot (1-p)^{n-k}$$
$$= \frac{k \binom{n-1}{k-1}}{6^k} \cdot 5^{n-k}$$

$E[N] = E[X_1 + X_2 + \cdots X_k]$

where $X_i = \text{Geom}(p)$. Let X be the total number of rolls.

$$\therefore E[N] = \frac{k}{p} \quad \text{——— (1)}$$

$$Var[N] = E[N^2] - (E[N])^2$$
$$= E[(X_1 + X_2 + \cdots X_k)^2] - \frac{k^2}{p^2}$$
$$= E[X_1^2 + X_2^2 + \cdots X_k^2 + 2X_1X_2 + 2X_1X_3 + \cdots 2X_{k-1}X_k] - \frac{k^2}{p^2}$$

$$= \sum E[X_i^2] + 2E[X_1X_2] - \frac{k^2}{p^2}$$
$$= \sum E[X_i^2] + 2(E[X_1])^2 - \frac{k^2}{p^2}$$
$$E[X_i^2] = \frac{2-p}{p^2} + \frac{2}{p^2} - \frac{k^2}{p^2}$$
$$\therefore Var[N] = \frac{k(1-p)}{p^2} + \frac{k}{p^2} \cdot \frac{2-p+2-p}{p^2}$$
$$= \frac{k(1-p)}{p^2}$$

(1) and (2) $\Rightarrow$ Mean and Variance of Number of Chebyshev

$$P[N \geq 200] \leq \frac{E[N]}{200} + \frac{Var}{200}$$
$$= \left(\frac{3}{200}\right) k$$
$$= \frac{18}{200} = \frac{9}{100}$$

From Chebyshev

$$P[N \geq 200] = P[N - \mu_N \geq 200 - \mu_N]$$
$$= P[N - 108 \geq 18]$$
$$= P[|N - 108| \geq 18] \quad \because 200$$
$$\leq \frac{Var[N]}{(18)^2} = \frac{k(1-p)}{(18)^2 p^2} = 0.0987 < 9\%$$

$|X-\mu| \geq k\sigma$.
If $X-\mu \geq 0$,
$X-\mu \geq k\sigma$.
$X-\mu \geq -k\sigma$.
$X \geq \mu + k\sigma$
If $X-\mu < 0$
$X-\mu < -k\sigma$
$X < \mu - k\sigma$
$\therefore P[|X-\mu| \geq k\sigma]$
$= P[X \geq \mu + k\sigma]$
$+ P[X \leq \mu - k\sigma]$
$= P[X - \mu \geq k\sigma]$

Theorem 7.1

The sequence $\hat{\Theta}_n$ of estimators is consistent if for any $\epsilon > 0$

$$\lim_{n\to\infty} P[|\hat{\Theta}_n - \alpha| \geq \epsilon] = 0.$$

Mean-squared error, $a_n = E[(\hat{\Theta}_n - \alpha)^2]$

Given that $\hat{\Theta}_n$ are unbiased estimators

$\Rightarrow E[\hat{\Theta}_n] = \alpha$

$\therefore a_n = E[(\hat{\Theta}_n - \alpha)^2] = Var[\hat{\Theta}_n]$.

Consider

$$P[|\hat{\Theta}_n - \alpha| > \epsilon] = P[|\hat{\Theta}_n - E[\hat{\Theta}_n]| > \epsilon]$$
$$\leq \frac{Var[\hat{\Theta}_n]}{\epsilon^2} = \frac{a_n}{\epsilon^2}$$

Now given $\lim_{n\to\infty} a_n = 0$.

$$\therefore \lim_{n\to\infty} P[|\hat{\Theta}_n - \alpha| > \epsilon] = 0 \quad \text{// Proved}$$

Example 7.4

$\hat{\Theta}_n = \frac{X_1 + \cdots X_n}{n} \Rightarrow E[\hat{\Theta}_n] = \frac{E[X_1]}{n} = \frac{n\alpha}{n} = \alpha$

$\therefore \hat{\Theta}_n$ is an unbiased estimator of parameter α.

$$a_n = E[(\hat{\Theta}_n - \alpha)^2]$$
$$= E\left[\left(\frac{X_1 + \cdots X_n}{n} - \alpha\right)^2\right] = \frac{E[X_1^2]}{n^2} + \frac{n^2\alpha^2}{n^2} - \frac{2\alpha \cdot n\alpha}{n}$$
$$= \frac{E[X_1^2]}{n^2} - \alpha^2$$
$$= \frac{(\text{Var}+\alpha^2)}{n^2} - \alpha^2 = \frac{\text{Var}}{n^2} + \frac{\alpha^2}{n^2}$$

If $\lim_{n\to\infty} a_n = 0$. $\therefore$ From Thm 7.1, the sequence of estimators $\hat{\Theta}_n$ is consistent.

We could show an upper bound for a_n $P[|\hat{\Theta}_n - \alpha| > \epsilon]$, that is

$$P[|\hat{\Theta}_n - \alpha| > \epsilon] = P[|\hat{\Theta}_n - E[\hat{\Theta}_n]| > \epsilon]$$
$$\leq \frac{Var[\hat{\Theta}_n]}{\epsilon^2} = \frac{\text{Var}}{n\epsilon^2}$$

$\therefore \lim_{n\to\infty} P[|\hat{\Theta}_n - \alpha| > \epsilon] = 0$ //

Example 7.5

$\hat{\Theta}_n(X)$ is the solution for my estimate of $p[X]$.

Standard error = ?

$$a_n = E[(\hat{\Theta}_n(X) - p[X])^2]$$
$$\hat{\Theta}_n(X) = \frac{X_1 + \cdots + X_n}{n}$$
$$E[\hat{\Theta}_n(X)] = p[X]$$

$\therefore E[\hat{\Theta}_n(X)] = p[X]$, that is $\hat{\Theta}_n(X)$ is an unbiased estimator of $p[X]$.

$\therefore a_n = Var[\hat{\Theta}_n(X)]$

For any $p \geq 0$

50 heads = 0.1,

$a_n < 0.01$.

$$Var[\hat{\Theta}_n(X)] = E[(\hat{\Theta}_n(X))^2] - (p[X])^2$$
$$= \frac{1}{n} Var[\hat{X}] \quad \because \text{Unbiased}$$

$\boxed{\hat{X} = 1 \text{ w.p. } p[X] \quad 0 \text{ otherwise.}}$

$E[\hat{X}] = p[X] \quad E[\hat{X}^2] = p[X]$

$Var[\hat{X}] = p[X] - (p[X])^2$

$= p[X](1 - p[X])$.

$\Rightarrow Var[\hat{\Theta}_n(X)] = \frac{1}{n} p[X](1 - p[X])$

For $Var[\hat{\Theta}_n(X)] < 0.01$,

we need

$$n > \boxed{\frac{p[X](1-p[X])}{0.01}}$$

Note that since $\hat{\Theta}_n(X)$ is unbiased

$$P[|\hat{\Theta}_n(X) - p[X]| > \epsilon] \leq \frac{Var[\hat{\Theta}_n(X)]}{\epsilon^2}$$

So for any given ϵ picking a smaller $Var[\hat{\Theta}_n(X)]$ $\Rightarrow$ Value drawn from $\hat{\Theta}_n(X)$ is even closer to the unknown $p[X]$ of $p[X]$.

(NB) $Var[\hat{\Theta}_n(X)] = \frac{1}{n} Var[\hat{X}]$

If $Var[\hat{X}]$ is finite for $n \to \infty$,

$Var[\hat{\Theta}_n(X)] \to 0$.

The sequence $\hat{\Theta}_n(X)$ is a consistent estimator of $p[X]$.

Theorem 7.2

Sample mean $M_n(X) = \frac{X_1 + \cdots + X_n}{n}$

Note, it is an unbiased estimator of $E[X]$.

From Chebyshev (for any $\epsilon > 0$)

$$P[|M_n(X) - E[M_n(X)]| > \epsilon] \leq \frac{Var[M_n(X)]}{\epsilon^2}$$

$$\therefore P[|M_n(X) - E[X]| > \epsilon] \leq \frac{Var[M_n(X)]}{\epsilon^2}$$

$$Var[M_n(X)] = \frac{1}{n} Var[X].$$

$\therefore$ As $n \to \infty$

$$P[|M_n(X) - E[X]| > \epsilon] = 0 \quad \left[\lim_{n\to\infty} \frac{Var[X]}{n\epsilon^2} = 0\right]$$

Theorem Proved //

Problem 7.31

Given X is Gaussian.

$M_n(X) \Rightarrow$ the sample mean.

$$P[|M_n(X) - \mu_X| \geq \sqrt{\sigma_X^2}] = ?$$

Since $M_n(X)$ is an unbiased estimator

the mean squared error $E[a] = Var[M_n(X)]$

$\therefore$ We want to calculate

$$P[|M_n(X) - \mu_X| \geq \sigma_X]$$
$$= P\left[\frac{|M_n(X) - \mu_X|}{\sigma_X} \geq 1\right]$$
$$= 1 - P[\ldots \leq 1]$$
$$= 1 - 0.68$$

The problem asks for $P[|M_n(X) - \mu_X| \geq \sigma_X]$

$\therefore$ We get $P[|M_n(X) - \mu_X| \geq \sigma_X] = 0.32$

If n is large enough for the CLT approx, $M_n(X)$ can be approximated to a Gaussian with mean μ_X and variance $\frac{\sigma_X^2}{n}$. Therefore the above evaluation may hold as long as CLT can be used to approximate. Note, $\sigma_X = \frac{Var[X]}{n}$

Point Estimator of the Variance

Two cases:
(a) $\mu_X = E[X]$ is known.
(b) μ_X is not known.

If known:

We want to estimate $Var[X]$

$$= E[(X-\mu_X)^2]$$

Let $Y = (X-\mu_X)^2$

$\therefore E[Y] = Var[X]$.

[$X_1, X_2, \ldots, X_n$ sub-sequence of $X_1, X_2, \ldots, X_n$]

The sample mean $M_n(Y)$ of Y is therefore an unbiased and consistent estimator of $Var[X]$.

If Unknown:

The mean μ_X also needs to be estimated.

Define sample variance.

$$V_n(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n(X))^2$$

What kind of an estimator is sample variance?

1) Is it unbiased?

$$E[V_n(X)] = \frac{1}{n}\sum_{i=1}^{n} E[(X_i - M_n(X))^2]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left[\left(X_i - \frac{(X_1+X_2+\cdots+X_n)}{n}\right)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left[X_i^2 + \frac{1}{n^2}(X_1+\cdots+X_n)^2 - \frac{2}{n}X_i(X_1+\cdots+X_n)\right]$$

$$= \frac{1}{n^2}E[(X_1+\cdots+X_n)^2] + \frac{1}{n}E[X_1^2+\cdots+X_n^2] - \frac{2}{n}E[(X_1+\cdots+X_n)^2]$$

$$= \frac{1}{n}E[X_1^2+\cdots+X_n^2] - \frac{1}{n^2}E[(X_1+X_2+\cdots+X_n)^2]$$

$$= \frac{1}{n}nE[X^2] - \frac{1}{n^2}E[X_1^2+\cdots+X_n^2+\underbrace{2X_1X_2+\cdots}_{2\binom{n}{2}\text{ terms}}]$$

$$= E[X^2] - \frac{1}{n^2}\left[nE[X^2] - \frac{n(n-1)}{2}E[X_1X_2]\right]$$

$$= E[X^2] - \frac{1}{n}E[X^2] - \frac{n(n-1)}{n^2}E^2[X]$$

$$= \left(\frac{n-1}{n}\right)E[X^2] - \left(\frac{n-1}{n}\right)E^2[X]$$

$$= \left(\frac{n-1}{n}\right)Var[X].$$

$\therefore$ The Sample Var is biased

This leads us to an unbiased estimate given by $\left(\frac{n}{n-1}\right)V_n(X)$.

Problem 7.23

The parameter we want to estimate is $\eta = E[X_1X_2]$.

The estimate used is

$$\hat{\eta}_n = \frac{1}{n}\sum_{i=1}^{n} X_1(i)X_2(i).$$

Given $Var[X_1X_2]$ is finite.

To show: $\hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_n, \ldots$ is an unbiased consistent sequence of estimators of η.

Proof:

$$E[\hat{\eta}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_1(i)X_2(i)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_1(i)X_2(i)] = \frac{1}{n}(n\eta) = \eta$$

(Unbiased)

From Chebyshev:

$$P[|\hat{\eta}_n - E[\hat{\eta}_n]| > \epsilon] \le \frac{Var[\hat{\eta}_n]}{\epsilon^2}$$

To show consistency, we need to show that

$$\lim_{n\to\infty} P[|\hat{\eta}_n - \eta| > \epsilon] = 0$$

same as (because $\hat{\eta}_n$ is an unbiased est)

$$\lim_{n\to\infty} P[|\hat{\eta}_n - E[\hat{\eta}_n]| > \epsilon] = 0$$

If we set $Y_i = X_1(i)X_2(i)$, and note that Y_i and Y_j are independent and identically distributed, $\hat{\eta}_n$ is the sample mean of Y_i.

$$\therefore Var[\hat{\eta}_n] = Var\left[\frac{Y_1+\cdots+Y_n}{n}\right]$$

$$= \frac{n}{n^2}Var[Y]$$

$$= \frac{1}{n}Var[X_1X_2]$$

$$= \left[\frac{Var[X_1X_2]}{n}\right].$$

~~If $Var[X_1X_2] \to 0$,~~

$$\lim_{n\to\infty} Var[\hat{\eta}_n] = 0$$

$$\Rightarrow \lim_{n\to\infty} P[|\hat{\eta}_n - E[X_1X_2]| > \epsilon] = 0$$

for any $\epsilon > 0$.

$\therefore \hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_n$ are consistent.

Problem 7.34

(a) Straightforward.

(b) Toshow:

$$\lim_{n\to\infty} P\left[\max_{j=1,\ldots,k} |M_j(X) - \mu_j| \ge c\right] = 0$$

$$P[\text{Atleast one of } |M_j(X)-\mu_j| \ge c, 1 \le j \le k]$$

$$= P[|M_1(X)-\mu_1| \ge c \cup \cdots \cup |M_k(X)-\mu_k| \ge c]$$

$$\le P[|M_1(X)-\mu_1| \ge c] + \cdots + P[|M_k(X)-\mu_k| \ge c]$$

$$= P[|M_1(X)-E[M_1(X)]| \ge c] + \cdots + P[|M_k(X)-E[M_k(X)]| \ge c]$$

($\because M_j(X)$ is an unbiased estimate of μ_j)

Taking the limit $\lim_{n\to\infty}$ we get the desired result.

Problem 7.37

$$Var[\hat{\eta}_{ij}(n)] = Var\left[\frac{1}{n}\sum_{k=1}^{n} X_i(k)X_j(k)\right]$$

$$= \frac{1}{n^2}\, n\, Var[X_iX_j]$$

[$X_i(k)X_j(k)$, $k=1,\ldots,n$ are independent]

$$Var[X_iX_j] = E[X_i^2X_j^2] - E^2[X_iX_j]$$

$$E[X_i^2X_j^2] = E[E[X_i^2X_j^2 \mid X_i]]$$

$$= E[X_i^2\, E[X_j^2 \mid X_i]]$$

(Check teacher's notes) Exp 1 class

Conditional pdf of Gaussian

$$E[X_j^2 \mid X_i] = Cov[X_i, X_j] + E[X_i]E[X_j]$$

$$= \rho_{X_i,X_j}\sigma_{X_i}\sigma_{X_j} + \mu_{X_i}\mu_{X_j}$$

And so on... to show that $Var[X_iX_j]$ is finite.

Probability and Random Processes

Sanjit Kaul

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Parameter Estimation

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Sampling From a Finite Population

$\underline{n}$ $\quad \underbrace{k=2}$

$X_1, X_2, \ldots, X_n$

$X_i \sim X$

$X^{(1)}$

$X^{(2)}$

Are $X^{(1)}$ and $X^{(2)}$ independent.

Sampling w/o replacement

with

Definition 7.1 ***Sample Mean***

For iid random variables $X_1, \ldots, X_n$ *with PDF* $f_X(x)$*, the* sample mean *of* X *is the random variable*

$$M_n(X) = \frac{X_1 + \cdots + X_n}{n}.$$

Theorem 7.1

The sample mean $M_n(X)$ has expected value and variance

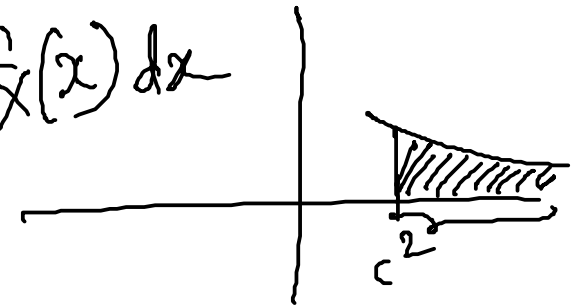
$$E\left[M_n(X)\right] = E\left[X\right], \qquad \mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}.$$

Quiz 7.1

Let X be an exponential random variable with expected value 1. Let $M_n(X)$ denote the sample mean of n independent samples of X. How many samples n are needed to guarantee that the variance of the sample mean $M_n(X)$ is no more than 0.01?

$$E[X] = \int_0^\infty x f_X(x)\,dx = \int_0^{c^2} x f_X(x)\,dx + \int_{c^2}^\infty x f_X(x)\,dx$$

$$\geq \int_{c^2}^\infty x f_X(x)\,dx \geq \int_{c^2}^\infty c^2 f_X(x)\,dx$$



Theorem 7.2 Markov Inequality

For a random variable X such that $P[X < 0] = 0$ and a constant c,

$$\int_{c^2}^\infty f_X(x)\,dx$$

$$P\left[X \geq c^2\right] \leq \frac{E[X]}{c^2}.$$

$$Z = (Y - E[Y])^2$$

$$P\left[|Y - E[Y]| \leq c\right] \geq \frac{Var[Y]}{c^2}$$

$$P\left[Z \geq c^2\right] \leq \frac{Var[Y]}{c^2}$$

$$P\left[(Y - E[Y])^2 \geq c^2\right] \leq \frac{Var[Y]}{c^2}$$

Theorem 7.3 Chebyshev Inequality

For an arbitrary random variable Y and constant $c > 0$,

$$P\left[|Y - \mu_Y| \geq c\right] \leq \frac{\mathrm{Var}[Y]}{c^2}.$$

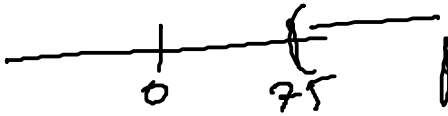
$$|Y-\mu_Y| \geq c \iff (Y-\mu_Y)^2 \geq c^2$$

$$-(Y-\mu_Y) \geq c$$

Quiz 7.2

Elevators arrive randomly at the ground floor of an office building. Because of a large crowd, a person will wait for time W in order to board the third arriving elevator. Let X_1 denote the time (in seconds) until the first elevator arrives and let X_i denote the time between the arrival of elevator $i-1$ and i. Suppose X_1, X_2, X_3 are independent uniform $(0, 30)$ random variables. Find upper bounds to the probability W exceeds 75 seconds using

$E[W] = 45$



$P[W > 75] \le \frac{E[W]}{75}$

(1) the Markov inequality,

$|W-45| > 30$

$-(W-45) > 30$

$W-45 < -30$

$W < -75$

$P[W - E[W] > 75 - E[W]]$

$\boxed{P[|W-\mu_W| \ge c] \le \frac{Var[W]}{c^2}}$

(2) the Chebyshev inequality.

$P[|W-45| > 30]$

$W-45 > 30$

$p = \frac{1}{36}$ $\quad P[R=r] = {}_{r-1}C_2\, p^2 (1-p)^{r-3}\, p \quad P[R \geq r] = \sum$

Problem 7.2.4

$R = X_1 + X_2 + X_3 \qquad E[R] = 3E[X] = \frac{3}{p}$

In a game with two dice, the event *snake eyes* refers to both dice showing one spot. Let R denote the number of dice rolls needed to observe the third occurrence of *snake eyes*. Find

$P[R \geq (\sqrt{250})^2] \leq \frac{3/p}{250}$

(a) the upper bound to $P[R \geq 250]$ based on the Markov inequality,

(b) the upper bound to $P[R \geq 250]$ based on the Chebyshev inequality,

(c) the exact value of $P[R \geq 250]$.

$P[R \geq 250]$

$= P[R - 108 \geq 142]$

$P[|R - \mu_R| \geq c] \leq \frac{Var[R]}{c^2}$

$P[|R - 3(36)| \geq c] \leq \frac{Var[R]}{c^2}$

$\{|R-108| \geq 142\} = \{-(R-108) \geq 142\} \cup \{R-108 \geq 142\}$

Section 7.3

Point Estimates of Model Parameters

$\hat{R}_1 = \frac{X_1}{1}$ $\hat{R}_2 = \frac{X_1 + X_2}{2}$ $\hat{R}_3 = \frac{X_1 + X_2 + X_3}{3}$ [if estimator is sample mean]

$r = E[X]$

Definition 7.2 Consistent Estimator

The sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *of the parameter* r *is* consistent *if for any* $\epsilon > 0$,

$$\lim_{n \to \infty} P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

$X \longleftarrow \pi$

$$E\left[\frac{X_1 + X_2 + X_3}{3}\right] = E[X]$$

Definition 7.3 ***Unbiased Estimator***

An estimate, $\hat{R}$*, of parameter* r *is* unbiased *if* $E[\hat{R}] = r$*; otherwise,* $\hat{R}$ *is* biased*.*

Asymptotically Unbiased

Definition 7.4 Estimator

The sequence of estimators $\hat{R}_n$ *of parameter* r *is* asymptotically unbiased *if*

$$\lim_{n \to \infty} E[\hat{R}_n] = r.$$

Definition 7.5 ***Mean Square Error***

The mean square error *of estimator* $\hat{R}$ *of parameter* r *is*

$$e = E\left[(\hat{R} - r)^2\right].$$

$e_n = E[(\hat{R}_n - r)^2]$

For unbiased $\hat{R}_n$

$e_n = E[(\hat{R}_n - E[\hat{R}_n])^2]$

$= Var[\hat{R}_n]$

Theorem 7.4

Fact given that estimates are unbiased

If a sequence of unbiased estimates $\hat{R}_1, \hat{R}_2, \ldots$ of parameter r has mean square error $e_n = \mathrm{Var}[\hat{R}_n]$ satisfying $\lim_{n\to\infty} e_n = 0$, then the sequence $\hat{R}_n$ is consistent.

Given

$\lim_{n\to\infty} P[|\hat{R}_n - r| \geq \epsilon] = 0$ To prove

$P[|\hat{R}_n - r| \geq \epsilon] = P[|\hat{R}_n - E[\hat{R}_n]| \geq \epsilon] \leq \frac{Var(\hat{R}_n)}{\epsilon^2}$

$$E[\hat{R}_k] = E\left[\frac{N_k}{k}\right] = \frac{1}{k} E[N_k] = r$$

Example 7.4 Problem

In any interval of k seconds, the number N_k of packets passing through an Internet router is a Poisson random variable with expected value $E[N_k] = kr$ packets. Let $\hat{R}_k = N_k/k$ denote an estimate of r. Is each estimate $\hat{R}_k$ an unbiased estimate of r? What is the mean square error e_k of the estimate $\hat{R}_k$? Is the sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ consistent?

$$e_k = E\left[(\hat{R}_k - r)^2\right] = E\left[(\hat{R}_k - E[\hat{R}_k])^2\right] = Var[\hat{R}_k]$$

$$= \frac{Var[N_k]}{k^2}$$

Theorem 7.5

The sample mean $M_n(X)$ is an unbiased estimate of $E[X]$.

Theorem 7.6

The sample mean estimator $M_n(X)$ has mean square error

$$e_n = E\left[(M_n(X) - E[X])^2\right] = \mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}.$$

This is also the variance of the sample mean estimator. Square-root of this is called the standard error.
Given the mean and variance of the estimator, what can we say about the sample mean that results from n IID trials?

$\text{Bern}(P[A]) \sim X$

$$\frac{X_1 + X_2 \cdots + X_n}{n}$$

Example 7.5 Problem

How many independent trials n are needed to guarantee that $\hat{P}_n(A)$, the relative frequency estimate of $P[A]$, has standard error less than 0.1?

$$E\left[\left(\hat{P}_n(A) - P[A]\right)^2\right] = \underline{\underline{\text{Var}\left[\hat{P}_n(A)\right]}} \leq 0.01$$

$$\text{Var}[X] = P[A]\left(1 - P[A]\right)$$

$$\frac{\text{Var}[X]}{n} \leq 0.01$$

$$n \geq \frac{\text{Var}[X]}{0.01} = \frac{P[A]\left(1 - P[A]\right)}{0.01}$$

$\hat{R}_1 \quad \hat{R}_2 \quad \ldots \quad \hat{R}_n$

$\hat{R}_1 = X_1$

$\hat{R}_2 = \frac{X_1 + X_2}{2} \ldots$

Theorem 7.7

If X has finite variance, then the sample mean $M_n(X)$ is a sequence of consistent estimates of $E[X]$.

$\lim_{n \to \infty} Var[\hat{R}_n] = 0$

Theorem 7.8 Weak Law of Large Numbers

If X has finite variance, then for any constant $c > 0$,

(a) $\lim_{n\to\infty} P[|M_n(X) - \mu_X| \geq c] = 0,$

(b) $\lim_{n\to\infty} P[|M_n(X) - \mu_X| < c] = 1.$

Theorem 7.9

As $n \to \infty$, the relative frequency $\hat{P}_n(A)$ converges to $P[A]$; for any constant $c > 0$,

$$\lim_{n \to \infty} P\left[\left|\hat{P}_n(A) - P[A]\right| \geq c\right] = 0.$$

Definition 7.6 Convergence in Probability

The random sequence Y_n *converges in probability to a constant* y *if for any* $\epsilon > 0$,

$$\lim_{n \to \infty} P\left[|Y_n - y| \geq \epsilon\right] = 0.$$

$$E\left[\left(M_n(X) - E(X)\right)^2\right]$$

$$= \mathrm{Var}\left[M_n(X)\right]$$

Problem 7.3.1

When X is Gaussian, verify the claim of Equation (7.16) that the sample mean is within one standard error of the expected value with probability 0.68.

Estimating the Variance

- Mean is known
- Mean Unknown

$X \quad \mu_X$ ✓

$$\mathrm{Var}[X]$$

$$= E\left[(X-\mu_X)^2\right]$$

$$= E[Y]$$

$$Y = (X-\mu_X)^2$$

Definition 7.7 Sample Variance

The sample variance of a set of n independent observations of random variable X is

$$V_n(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n(X))^2 .$$

Is

$$E[V_n(X)] = Var[X] \; ?$$

$$E[(X_i - M_n(X))^2]$$

$$= E[X$$

$$E[V_n(X)]$$

$$= \frac{n-1}{n} Var[X]$$

Theorem 7.10

$$E\left[V_n(X)\right] = \frac{n-1}{n}\,\mathrm{Var}[X].$$

Theorem 7.11

The estimate

$$V_n'(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M_n(X))^2$$

is an unbiased estimate of $\mathrm{Var}[X]$.

$$\mathrm{Var}[\hat{R}_n] = \frac{\mathrm{Var}[Y]}{n} = \frac{\mathrm{Var}[X_1 X_2]}{n}$$

$$Y_i = X_1(i) X_2(i)$$

Problem 7.3.3

An experimental trial produces random variables X_1 and X_2 with correlation $r = E[X_1 X_2]$. To estimate r, we perform n independent trials and form the estimate

$$\hat{R}_n = \frac{1}{n} \sum_{i=1}^{n} X_1(i) X_2(i) = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

where $X_1(i)$ and $X_2(i)$ are samples of X_1 and X_2 on trial i. Show that if $\mathrm{Var}[X_1 X_2]$ is finite, then $\hat{R}_1, \hat{R}_2, \ldots$ is an unbiased, consistent sequence of estimates of r.

$$Y \sim Y_i \sim \underset{\downarrow \atop X_1}{X_1(i)} \; \underset{\downarrow \atop X_2}{X_2(i)}$$

Problem 7.3.4

An experiment produces random vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_k \end{bmatrix}'$ with expected value $\boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix} \mu_1 & \cdots & \mu_k \end{bmatrix}'$. The ith component of $\mathbf{X}$ has variance $\mathrm{Var}[X_i] = \sigma_i^2$. To estimate $\boldsymbol{\mu}_{\mathbf{X}}$, we perform n independent trials such that $\mathbf{X}(i)$ is the sample of $\mathbf{X}$ on trial i, and we form the vector mean

$$\mathbf{M}(n) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}(i).$$

(a) Show $\mathbf{M}(n)$ is unbiased by showing $E[\mathbf{M}(n)] = \boldsymbol{\mu}_{\mathbf{X}}$.

(b) Show that the sequence of estimates $\mathbf{M}_n$ is consistent by showing that for any constant $c > 0$,

$$\lim_{n\to\infty} P\left[\max_{j=1,\ldots,k} \left|M_j(n) - \mu_j\right| \geq c\right] = 0.$$

Hint: Let $A_i = \{|M_i(n) - \mu_i| \geq c\}$ and apply the union bound (see Problem 1.4.5) to upper bound $P[A_1 \cup A_2 \cup \cdots \cup A_k]$. Then apply the Chebyshev inequality.

Problem 7.3.7

An experiment produces a zero mean Gaussian random vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_k \end{bmatrix}'$ with correlation matrix $\mathbf{R} = E[\mathbf{X}\mathbf{X}']$. To estimate $\mathbf{R}$, we perform n independent trials, yielding the iid sample vectors $\mathbf{X}(1), \mathbf{X}(2), \ldots, \mathbf{X}(n)$, and form the sample correlation matrix

$$\hat{\mathbf{R}}(n) = \frac{1}{n} \sum_{m=1}^{n} \mathbf{X}(m)\mathbf{X}'(m).$$

(a) Show $\hat{\mathbf{R}}(n)$ is unbiased by showing $E[\hat{\mathbf{R}}(n)] = \mathbf{R}$.

(b) Show that the sequence of estimates $\hat{\mathbf{R}}(n)$ is consistent by showing that every element $\hat{R}_{ij}(n)$ of the matrix $\hat{\mathbf{R}}$ converges to R_{ij}. That is, show that for any $c > 0$,

$$\lim_{n\to\infty} P\left[\max_{i,j} \left|\hat{R}_{ij} - R_{ij}\right| \geq c\right] = 0.$$

Hint: Extend the technique used in Problem 7.3.4. You will need to use the result of Problem 4.11.8 to show that $\mathrm{Var}[X_i X_j]$ is finite.

Section 7.4

μ

$(x-c, x+c)$

Estimator $\triangleq f(X_1, X_2, \ldots, X_n)$

$M_n(X)$ $X_i \sim X$

Confidence Intervals

Suppose $M_n(X) = 10$

$$P\left[\,|M_n(X) - \mu| \le c\,\right] = P\left[M_n(X) \in (x-c, x+c)\right] \ge 1-\alpha$$

$2c$

Theorem 7.12

For any constant $c > 0$,

(a) $P[|M_n(X) - \mu_X| \geq c] \leq \dfrac{\mathrm{Var}[X]}{nc^2} = \alpha,$

(b) $P[|M_n(X) - \mu_X| < c] \geq 1 - \dfrac{\mathrm{Var}[X]}{nc^2} = 1 - \alpha.$

Confidence *Coefficient* of your estimate being in an interval of length 2c

Example 7.6 Problem

Suppose we perform n independent trials of an experiment and we use the relative frequency $\hat{P}_n(A)$ to estimate $P[A]$. Use the Chebyshev inequality to calculate the smallest n such that $\hat{P}_n(A)$ is in a confidence interval of length 0.02 with confidence 0.999.

Example 7.7 Problem

$X_1 \; X_2 \cdots X_n \quad X_i \sim X$

Suppose we perform n independent trials of an experiment. For an event A of the experiment, use the Chebyshev inequality to calculate the number of trials needed to guarantee that the probability the relative frequency of A differs from $P[A]$ by more than 10% is less than 0.001.

$x = P[A]$ $\quad 2c \quad c = 0.1\,P[A]$

$P\left[M_n(X) \in (x-c, x+c)\right] \geq 1 - 0.001$

$0.1x$

$\rightarrow P\left[|M_n(X) - x| \leq 0.1x\right] \geq 1 - 0.001$

$P\left[|M_n(X) - E[M_n(X)]| \leq 0.1E[M_n(X)]\right] \geq 1 - \frac{Var[X]}{n(0.1)^2 x^2}$

$n > \frac{10^5}{P[A]}$

Note $E[M_n(X)] = x$; $\quad P[A](1-P[A])$

Example 7.8 Problem

Based on a sample of 1103 potential voters, the percentage of people supporting Candidate Jones is 58% with an accuracy of plus or minus 3 percentage points.

$0.58 = \hat{P}_n(A)$ $n = 1103$

$c = 0.03$

With what minimum probability does your method for estimating support for Jones satisfy the accuracy requirement?

$$P\left[\left|\hat{P}_n(A) - P(A)\right| \geq 0.03\right] \leq \frac{Var[X]}{(1103)(0.03)^2}$$

0.25

In other words, what is the confidence coefficient?

Interval Estimates

- The interval is random
- A probability is associated with the parameter lying within the random interval
  - You estimate a random interval
  - With some probability your parameter of interest lies within your estimated interval

$P[A \le \mu \le B] \ge 1-\alpha$ $\quad \mu = E[X]$

(a, b)

$(\hat{\mu}-c, \hat{\mu}+c)$

$\ge 1 - \frac{Var(X)}{nc^2}$

$P[M_n(X) - c \le \mu \le M_n(X) + c]$

$= P[-c \le M_n(X) - \mu \le c]$

$= P[-c \le M_n(X) - E[X] \le c]$

Example 7.9 Problem

Suppose X_i is the ith independent measurement of the length (in cm) of a board whose actual length is b cm. Each measurement X_i has the form

$$X_i = b + Z_i, \qquad E(X_i) = b \tag{7.43}$$

where the measurement error Z_i is a random variable with expected value zero and standard deviation $\sigma_Z = 1$ cm. Since each measurement is fairly inaccurate, we would like to use $M_n(X)$ to get an accurate confidence interval estimate of the exact board length. How many measurements are needed for a confidence interval estimate of b of length $2c = 0.2$ cm to have confidence coefficient $1 - \alpha = 0.99$?

$$P[M_n(X) - c \le E[X] \le M_n(X) + c] \ge 1 - \frac{\mathrm{Var}(X)}{nc^2} = 1 - \frac{1}{n(0.1)^2}$$

$$n \ge 10^4$$

$$P[-c \le M_n(X) - E[X] \le c]$$

$$= P[-c \le M_n(X) - E[M_n(X)] \le c]$$

$$= P\left[\frac{-c}{\sigma_{M_n(X)}} \le \frac{M_n(X) - E[M_n(X)]}{\sigma_{M_n(X)}} \le \frac{c}{\sigma_{M_n(X)}}\right] = 1 - 2Q\left(\frac{c\sqrt{n}}{\sigma_X}\right) \ge 0.99$$

$$Q(0.1\sqrt{n}) \le 0.05 \Rightarrow n \ge 666$$

Example 7.10 Problem

In Example 7.9, suppose we know that the measurement errors Z_i are iid Gaussian random variables. How many measurements are needed to guarantee that our confidence interval estimate of length $2c = 0.2$ has confidence coefficient $1 - \alpha \geq 0.99$?

Theorem 7.13

$\sigma_{M_n(X)} = \sigma/\sqrt{n}$

$\mu = E[M_n(X)]$

Let X be a Gaussian (μ, σ) random variable. A confidence interval estimate of μ of the form

$$P\left[M_n(X) - c \le \mu \le M_n(X) + c\right] =$$

$$= P\left[-c \le \mu - M_n(X) \le c\right]$$

$$= P\left[c \ge M_n(X) - \mu \ge -c\right]$$

has confidence coefficient $1 - \alpha$ where

$$\alpha/2 = Q\left(\frac{c\sqrt{n}}{\sigma}\right) = 1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right).$$

Z

$$= P\left[\frac{-c}{(\sigma/\sqrt{n})} \le \boxed{\frac{M_n(X) - \mu}{(\sigma/\sqrt{n})}} \le \frac{c}{(\sigma/\sqrt{n})}\right] = \underbrace{P\left[-\frac{c}{\sigma/\sqrt{n}} \le Z \le \frac{c}{\sigma/\sqrt{n}}\right]}_{= 1 - \alpha}$$

$$P\left[Z \le \frac{c}{\sigma/\sqrt{n}}\right] - P\left[Z \le \frac{-c}{\sigma/\sqrt{n}}\right]$$

$$\hookrightarrow 1 - P\left[Z \le -\frac{c}{\sigma/\sqrt{n}}\right] - P\left[Z \ge \frac{c}{\sigma/\sqrt{n}}\right]$$

$$= 1 - 2P\left[Z \ge \frac{c}{\sigma/\sqrt{n}}\right]$$

Example 7.11 Problem

$E(Y)$: truth of interest.

Y is a Gaussian random variable with unknown expected value μ but known variance σ_Y^2. Use $M_n(Y)$ to find a confidence interval estimate of μ_Y with confidence 0.99. If $\sigma_Y^2 = 10$ and $M_{100}(Y) = 33.2$, what is our interval estimate of μ formed from 100 independent samples?

$n = 100$ ✓

$1-\alpha = 0.99$

$c = 10$

$(-10, 10)$

$(-10\sqrt{10}, 10\sqrt{10})$

$(-10+33.2, 10+33.2)$

$1-\alpha$

We want to compare the mean family income in two states. For state 1, we had a random sample of $n_1 = 100$ families with a sample mean of $\bar{x}_1 = 35000$. For state 2, we had a random sample of $n_2 = 144$ families with a sample mean of $\bar{x}_2 = 36000$. Past studies have shown that for both states $\sigma = 4000$. Estimate $\mu_1 - \mu_2$ and place a two-standard-error bound on the error of estimation. How much confidence do we have in this interval?

X_1 X_2

$c = 2e$

http://old.stat.duke.edu/courses/Spring02/sta103/lec/ch8b_4.pdf

$M_n(X_1) - M_n(X_2)$

$$P\left[M_n(X_1) - M_n(X_2) - c \le \mu_1 - \mu_2 \le M_n(X_1) - M_n(X_2) + c\right]$$

$$e^2 = E\left[\left(M_n(X_1) - M_n(X_2) - (\mu_1 - \mu_2)\right)^2\right] = Var\left[M_n(X_1) - M_n(X_2)\right]$$

$$\mathrm{Var}\left[M_n(X_1) - M_n(X_2)\right] = \underbrace{\mathrm{Var}\left[M_n(X_1)\right]} + \mathrm{Var}\left[M_n(X_2)\right]$$

$$\geq 1 - \frac{\mathrm{Var}(X_1 - X_2)}{n 4 e^2}$$

$$\left(-1000 - c, 1000 + c\right)$$

$2e$

EXAMPLE 7.3a Suppose that when a signal having value μ is transmitted from location A the value received at location B is normally distributed with mean μ and variance 4. That is, if μ is sent, then the value received is $\mu + N$ where N, representing noise, is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, let us construct a 95 percent confidence interval for μ.

$$\hookrightarrow M_n(X) = \hat{\mu} \quad \frac{1}{8}\left((5-\hat{\mu})^2 + (8.5-\hat{\mu})^2 + \cdots\right)$$

α α

$$\rightarrow P\left[\,|M_n(X) - \mu| \le c\,\right] \ge 1-\alpha$$

$$P\left[\frac{|M_n(X) - \mu|}{\sigma_{M_n(X)}} \le \frac{c}{\sigma_{M_n(X)}}\right]$$

$\sigma_{M_n(X)} \hookrightarrow \sigma/\sqrt{n}$

$\sigma_{M_n(X)} \hookrightarrow \sigma/\sqrt{n}$

S^2

Unbiased estimate of σ^2

$$P\left[\frac{|M_n(X)-\mu|}{(S/\sqrt{n})} \le \frac{c}{(S/\sqrt{n})}\right]$$

Student-t distribution

Symmetric about the mean

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - M_n(X)\right)^2$$

$$P\left[\frac{-c}{S/\sqrt{n}} \le \frac{M_n(X)-\mu}{S/\sqrt{n}} \le \frac{c}{S/\sqrt{n}}\right] \ge \underbrace{1-\alpha}_{0.95}$$

$$P\left[-t_{\alpha/2,(n-1)} \le \frac{M_n(X)-\mu}{S/\sqrt{n}} \le t_{\alpha/2,(n-1)}\right] = 1-\alpha$$

$$\left(M_n(X) - \frac{S}{\sqrt{n}}t_{\alpha/2,(n-1)},\ M_n(X) + \frac{S}{\sqrt{n}}t_{\alpha/2,(n-1)}\right)$$

EXAMPLE 7.3e Let us again consider Example 7.3a but let us now suppose that when the value μ is transmitted at location A then the value received at location B is normal with mean μ and variance σ^2 but with σ^2 being unknown. If 9 successive values are, as in Example 7.3a, 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, compute a 95 percent confidence interval for μ.

$t_{0.025,8} = 2.306$

$1-\alpha = 0.95$

$\alpha = 0.05$

$\alpha/2 = 0.025$

Hypothesis Testing

Sanjit Kaul
(and Slides from RY)

Hypothesis Testing

- Perform Experiment
- Observe Outcome
- Make a conclusion/inference with high accuracy

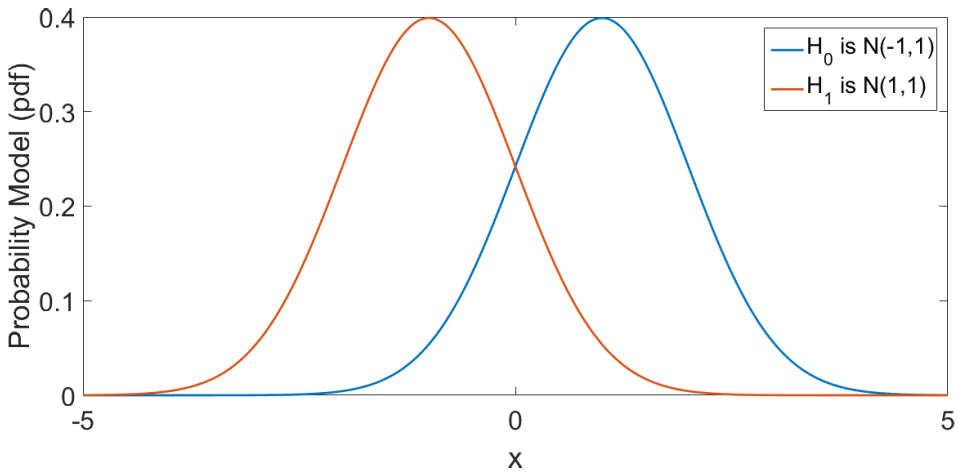
Two Categories of Statistical Inference

- Goal is to assign a probability model to the outcome of your experiment as accurately as possible
- Start with one or more possible models (one or more hypotheses) a priori
- Significance Testing
  - Start with a probability model as your hypothesis
    - Often named H_0 and called the null hypothesis
  - Conclusion: Accept or reject the hypothesis that outcome is from the probability model
  - Accuracy Measure: Probability of rejecting a true hypothesis

Two Categories of Statistical Inference

- Hypothesis Testing
  - Start with multiple probability models.
    - They are your hypotheses H_0, H_1, H_2, …, H_{M-1}
  - Conclusion: Observation from the experiment is from one of the starting Hypotheses
  - Accuracy Measure: Probability you conclude that the observation is from H_i when it is truly from H_k

The Communications Question



- The receiver makes an observation resulting from a transmitter sending a bit (either +1V or -1V)
- Receiver knows that two probability models N(-1,1) and N(1,1) are possible
- Receiver receives a voltage 0.00001 V.
- The receiver must conclude either N(-1,1) or N(1,1)
    - Equivalent to picking -1 or +1

Significance Testing

- Split sample space into two mutually exclusive and collectively exhaustive sets A (accept) and R (reject).
- If outcome belongs to A, then say H_0 occurred, null hypothesis is accepted
  - Else, the null hypothesis is rejected
- Accuracy is a function of how A (and R) are created

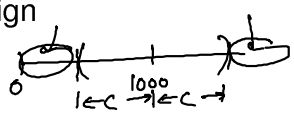
Significance Level

- $P[s \in R|H_0]$
  - Probability that the hypothesis is rejected when the observation is from it
- You may start with this value as a design parameter
- Then decide on sets R and A
- The error is called a Type I error or False Rejection
  - Reject H0 when H0 is true
- Another possible error is Type II or False Acceptance
  - Accept H0 when H0 is false
  - Can't calculate as we don't know the *other* probability model

Example 8.2 Problem

Suppose that on Thursdays between 9:00 and 9:30 at night, the number of call attempts N at a telephone switching office is a Poisson random variable with expected value 1000. Next Thursday, the President will deliver a speech at 9:00 that will be broadcast by all radio and television networks. The null hypothesis, H_0, is that the speech does not affect the probability model of telephone calls. In other words, H_0 states that on the night of the speech, N is a Poisson random variable with expected value 1000. Design a significance test for hypothesis H_0 at a significance level of $\alpha = 0.05$.

$P[S \in R | H_0] = 0.05$



$P[S \in R | H_0] = P[|N - 1000| \geq c] = 0.05$

$P[R \mid H_0] = 0.01$

Type-I error

$P[M_n(X) < 190 - c \mid H_0]$

$P[M_n(X) - 190 < -c \mid H_0]$

$\boxed{Q\left(\frac{c}{3}\right) = 0.01}$

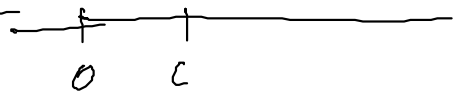
$c > 0$

Example 8.3 Problem

Before releasing a diet pill to the public, a drug company runs a test on a group of 64 people. Before testing the pill, the probability model for the weight of the people measured in pounds, is a Gaussian $(190, 24)$ random variable W. Design a test based on the sample mean of the weight of the population to determine whether the pill has a significant effect. The significance level is $\alpha = 0.01$.

$$P[X \le x \mid H_0] = \frac{P[X \le x, H_0]}{P[H_0]}$$

$$P[X \le x]$$



Quiz 8.1

$$P[X \in (0,c) \mid H_0] = 0.01$$

$$(1-e^{-c})^{15} = 0.01$$

$$c = 1.33$$

Under hypothesis H_0, the interarrival times between phone calls are independent and identically distributed exponential (1) random variables. Given X, the maximum among 15 independent interarrival time samples $X_1, \ldots, X_{15}$, design a significance test for hypothesis H_0 at a level of $\alpha = 0.01$.

$$P[X \le x] = P[\max(X_1, X_2, \ldots, X_{15}) \le x] = P[X_1 \le x, X_2 \le x, \ldots, X_{15} \le x]$$

$$= (1-e^{-x})^{15}$$

$E[L|H] = 2$

$P(L > c|H) = 0.05$

$(0.5)^c = 0.05$ $\quad c = 4.32$

Problem 8.1.1

Let L equal the number of flips of a coin up to and including the first flip of heads. Devise a significance test for L at level $\alpha = 0.05$ to test the hypothesis H that the coin is fair. What are the limitations of the test?

Problem 8.1.3

When a chip fabrication facility is operating normally, the lifetime of a microchip operated at temperature T, measured in degrees Celsius, is given by an exponential (λ) random variable X with expected value $E[X] = 1/\lambda = (200/T)^2$ years. Occasionally, the chip fabrication plant has contamination problems and the chips tend to fail much more rapidly. To test for contamination problems, each day m chips are subjected to a one-day test at $T = 100°C$. Based on the number N of chips that fail in one day, design a significance test for the null hypothesis test H_0 that the plant is operating normally.

(a) Suppose the rejection set of the test is $R = \{N > 0\}$. Find the significance level of the test as a function of m, the number of chips tested.

(b) How many chips must be tested so that the significance level is $\alpha = 0.01$.

(c) If we raise the temperature of the test, does the number of chips we need to test increase or decrease?

$P[N>0]$

$= 1 - P[N=0]$

$= 1 - (1-e^{(\ )})^m$

$m = \left(\frac{200}{T}\right)^2 \times 365 \log\left(\frac{1}{1-\alpha}\right)$

$P[A \text{ chip passes} \mid H_0]$

$= P\left(X \geq \frac{1}{365}\right)$

$= e^{-\left(\frac{T}{200}\right)^2 \frac{1}{365}}$

Problem 8.1.5

When a pacemaker factory is operating normally (the null hypothesis H_0), a randomly selected pacemaker fails a test with probability $q_0 = 10^{-4}$. Each day, an inspector randomly tests pacemakers. Design a significance test for the null hypothesis with significance level $\alpha = 0.01$. Note that testing pacemakers is expensive because the pacemakers that are tested must be discarded. Thus the significance test should try to minimize the number of pacemakers tested.

$$\frac{\ln(1-0.01)}{\ln(1-10^{-4})} = M$$

$R = \{N > 0\}$

M

$P[N>0] = 0.01$

$1 - P[N=0]$

$P[N=0] = (1-10^{-4})^M$

Problem 8.1.5

When a pacemaker factory is operating normally (the null hypothesis H_0), a randomly selected pacemaker fails a test with probability $q_0 = 10^{-4}$. Each day, an inspector randomly tests pacemakers. Design a significance test for the null hypothesis with significance level $\alpha = 0.01$. Note that testing pacemakers is expensive because the pacemakers that are tested must be discarded. Thus the significance test should try to minimize the number of pacemakers tested.

$$\frac{\ln(1-0.01)}{\ln(1-10^{-4})} = M$$

$R = \{N > 0\}$

M

$P[N>0] = 0.01$

$1 - P[N=0]$

$P[N=0] = (1-10^{-4})^M$

Binary Hypothesis Testing

$P[R|H_0]$

A_0, A_1 $\quad P[A_0|H_1]$

$\uparrow \quad \uparrow$

$H_0 \quad H_1$ $\quad P[A_1|H_0]$

- Start with two probability models.
  - They are your hypotheses H_0, H_1
    - What are the PDF(s)/CDF(s)?
      - Also called likelihood functions
- Conclusion: Observation from the experiment is from one of the starting Hypotheses
- Outcome is governed by H_0 when there is no target
- Else by H_1
- Accuracy Measure:
  - Miss Detection (Type II): Probability you conclude that the observation is from H_0 when it is truly from H_1
    - Target missed by RADAR when it is present
  - False Alarm (Type I): Probability you conclude that the observation is from H_1 when it is truly from H_0
    - Target detected by RADAR when none exists

Acceptance Sets

- A_0, A_1
- What is False alarm probability? → $P[A_0 | H_1]$
- What is Miss Detection probability? → $P[A_1 | H_0]$

Receiver Operating → ROC

- Plot a graph (Pfa (x-axis) vs. Pmiss)

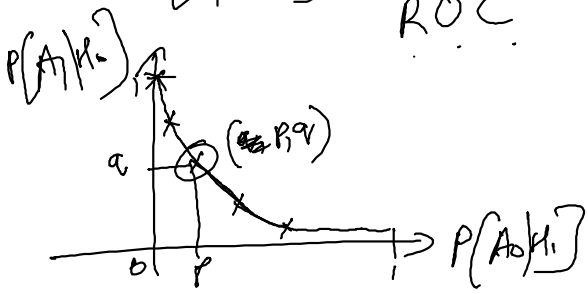
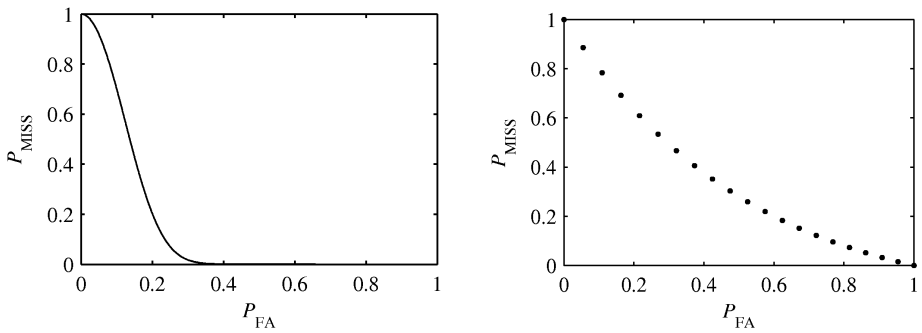


Figure 8.1



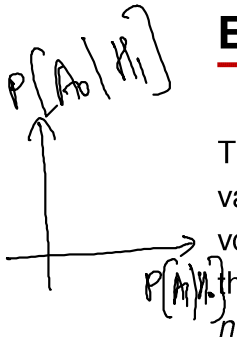
ROC for continuous X **ROC for discrete X**

Continuous and discrete examples of a receiver operating curve (ROC).

Example 8.4 Problem

$H_1 \sim N(v,1)$

$H_0 \sim N(0,1)$

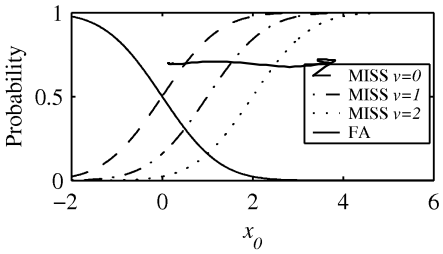


The noise voltage in a radar detection system is a Gaussian $(0, 1)$ random variable, N. When a target is present, the received signal is $X = v + N$ volts with $v \geq 0$. Otherwise the received signal is $X = N$ volts. Periodically, the detector performs a binary hypothesis test with H_0 as the hypothesis *no target* and H_1 as the hypothesis *target present.* The acceptance sets for the test are $A_0 = \{X \leq x_0\}$ and $A_1 = \{X > x_0\}$. Draw the receiver operating curves of the radar system for the three target voltages $v = 0, 1, 2$ volts.

$\rightarrow P[A_0|H_1] = P[X \leq x_0 | H_1] =$

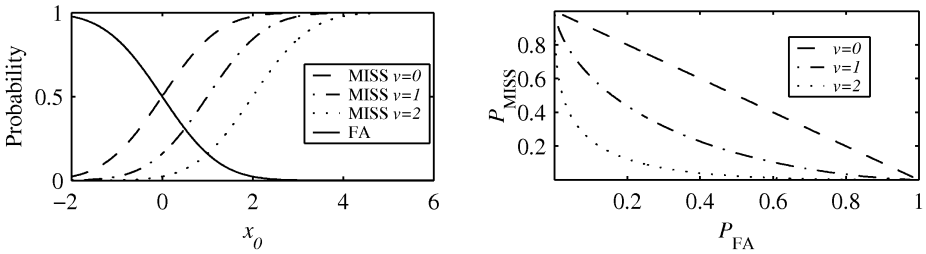
$\rightarrow P(A_1|H_0) = P[X > x_0 | H_0]$

Figure 8.2



(a)

Figure 8.2


Probability
1
0.5
0
−2
0
2
4
6
x_0
MISS ν=0
MISS ν=1
MISS ν=2
FA
P_{MISS}
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1
P_{FA}
ν=0
ν=1
ν=2

(a) **(b)**

How Do We Select the Set A_0

$$P[H_0|s] = \frac{P[s|H_0]\,P[H_0]}{P[s]} \rightarrow P[s|H_0]P[H_0] + P[s|H_1]P[H_1]$$

$P[H_0]$

$P[H_1]$

Theorem 8.1 Maximum A posteriori Probability (MAP) Binary Hypothesis Test

Assign $s \in A_0$ $\rightarrow P[s|H_1]P[H_1]$

$\leq P[s|H_0]\,P[H_0]$

$P[\text{Error}]$

$= P[A_0|H_1]P[H_1]$

$+ P[A_1|H_0]P[H_0]$

Given a binary hypothesis testing experiment with outcome s, the following rule leads to the lowest possible value of P_{ERR}:

$$s \in A_0 \text{ if } P[H_0|s] \geq P[H_1|s]; \qquad s \in A_1 \text{ otherwise.}$$

$$\frac{P[s|H_0]\,P[H_0]}{P[s]} \geq \frac{P[s|H_1]\,P[H_1]}{P[s]}$$

$$\frac{1}{P[H_1]}P[H_1|s]\,\cancel{P[s]}\,P[H_1] \leq P[H_0|s]\,\cancel{P[s]}\,P[H_0]\frac{1}{P[H_0]}$$

$$P[H_1|s] \leq P[H_0|s]$$

MAP Rule: Accept the hypothesis with the larger a posteriori probability

$$\frac{P[s|H_0]}{P[s|H_1]} \geq \frac{P[H_1]}{P[H_0]}$$

Theorem 8.2

For an experiment that produces a random vector $\mathbf{X}$, the MAP hypothesis test is

Discrete: $\mathbf{x} \in A_0$ if $\dfrac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]}{P[H_0]}$; $\mathbf{x} \in A_1$ otherwise

Continuous: $\mathbf{x} \in A_0$ if $\dfrac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]}{P[H_0]}$; $\mathbf{x} \in A_1$ otherwise.

Evidence in favor of H_1 prior to observation

The likelihood ratio. Evidence in favor of H_0 after an observation

$$\rightarrow \frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} \geq \frac{P[H_1]}{P[H_0]}$$

Note that $f_{X|H_0}(x) \approx \frac{P[x \leq X < x + \Delta x]}{\Delta x}$

$H_0: N(-v, \sigma)$ $\quad P[H_0] = p$

$H_1: N(v, \sigma)$ $\quad P[H_1] = (1-p)$

Assign to A_0:

$$\frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x+v)^2/(2\sigma^2)]}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-v)^2/2\sigma^2}} \geq \frac{(1-p)}{p}$$

Example 8.6 Problem

With probability p, a digital communications system transmits a 0. It transmits a 1 with probability $1-p$. The received signal is either $X = -v + N$ volts, if the transmitted bit is 0; or $v + N$ volts, if the transmitted bit is 1. The voltage $\pm v$ is the information component of the received signal, and N, a Gaussian $(0, \sigma)$ random variable, is the noise component. Given the received signal X, what is the minimum probability of error rule for deciding whether 0 or 1 was sent?

Is the threshold is a function of the prior probabilities? How? What else matters?

Select $x \in A_0$ if x $\boxed{}$ (Expr.)

Ex

$$x \leq \frac{\sigma^2}{2v} \ln\left(\frac{p}{1-p}\right) \quad x^*$$

$A_0 = \{X \le x^*\}$

$A_1 = \{X > x^*\}$

Example 8.7 Problem

Find the error probability of the communications system of Example 8.6.

$$P[E] = P[A_0 | H_1] P[H_1] + P[A_1 | H_0] P[H_0]$$

$$P[X \le x^* | H_1] P[H_1] + P[X > x^* | H_0] P[H_0]$$

Example 8.8 Problem

$H_0 : \text{Geom}(q_0)$

$H_1 : \text{Geom}(q_1)$

Assign n to A_0

$$\frac{P(N=n \mid H_0)}{P(N=n \mid H_1)} \geq P[H_1]/P[H_0]$$

$$\frac{(1-q_0)^{n-1} q_0}{(1-q_1)^{n-1} q_1} \geq \frac{1}{9}$$

Assign n to A_0:

$$n \geq 1 + \frac{\ln(\quad)}{\ln(\quad)}$$

At a computer disk drive factory, the manufacturing failure rate is the probability that a randomly chosen new drive fails the first time it is powered up. Normally the production of drives is very reliable, with a failure rate $q_0 = 10^{-4}$. However, from time to time there is a production problem that causes the failure rate to jump to $q_1 = 10^{-1}$. Let H_i denote the hypothesis that the failure rate is q_i.

Every morning, an inspector chooses drives at random from the previous day's production and tests them. If a failure occurs too soon, the company stops production and checks the critical part of the process. Production problems occur at random once every ten days, so that $P[H_1] = 0.1 = 1 - P[H_0]$. Based on N, the number of drives tested up to and including the first failure, design a MAP hypothesis test. Calculate the conditional error probabilities P_{FA} and P_{MISS} and the total error probability P_{ERR}.

Theorem 8.3 Minimum Cost Binary Hypothesis Test

For an experiment that produces a random vector $\mathbf{X}$, the minimum cost hypothesis test is

Discrete: $\mathbf{x} \in A_0$ if $\dfrac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]C_{01}}{P[H_0]C_{10}}$; $\mathbf{x} \in A_1$ otherwise

Continuous: $\mathbf{x} \in A_0$ if $\dfrac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]C_{01}}{P[H_0]C_{10}}$; $\mathbf{x} \in A_1$ otherwise.

Example 8.9 Problem

Continuing the disk drive test of Example 8.8, the factory produces 1,000 disk drives per hour and 10,000 disk drives per day. The manufacturer sells each drive for \$100. However, each defective drive is returned to the factory and replaced by a new drive. The cost of replacing a drive is \$200, consisting of \$100 for the replacement drive and an additional \$100 for shipping, customer support, and claims processing. Further note that remedying a production problem results in 30 minutes of lost production. Based on the decision statistic N, the number of drives tested up to and including the first failure, what is the minimum cost test?

Maximum Likelihood Decision

Definition 8.1 Rule

For a binary hypothesis test based on the experimental outcome $s \in S$, the maximum likelihood (ML) decision rule is

$$s \in A_0 \text{ if } P\left[s|H_0\right] \geq P\left[s|H_1\right]; \qquad s \in A_1 \text{ otherwise.}$$

Theorem 8.6

If an experiment produces a random vector $\mathbf{X}$, the ML decision rule states

Discrete: $\mathbf{x} \in A_0$ if $\dfrac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq 1;$ $\mathbf{x} \in A_1$ otherwise

Continuous: $\mathbf{x} \in A_0$ if $\dfrac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq 1;$ $\mathbf{x} \in A_1$ otherwise.