Nov 22 Skip Brownian Motion

Now class next week → Thanksgiving

Now Statistics : → applied probability
→ dealing with data

Previously : we know the distribution (uniform, Poisson, normal, etc)
↳ from this we studied μ and σ^2

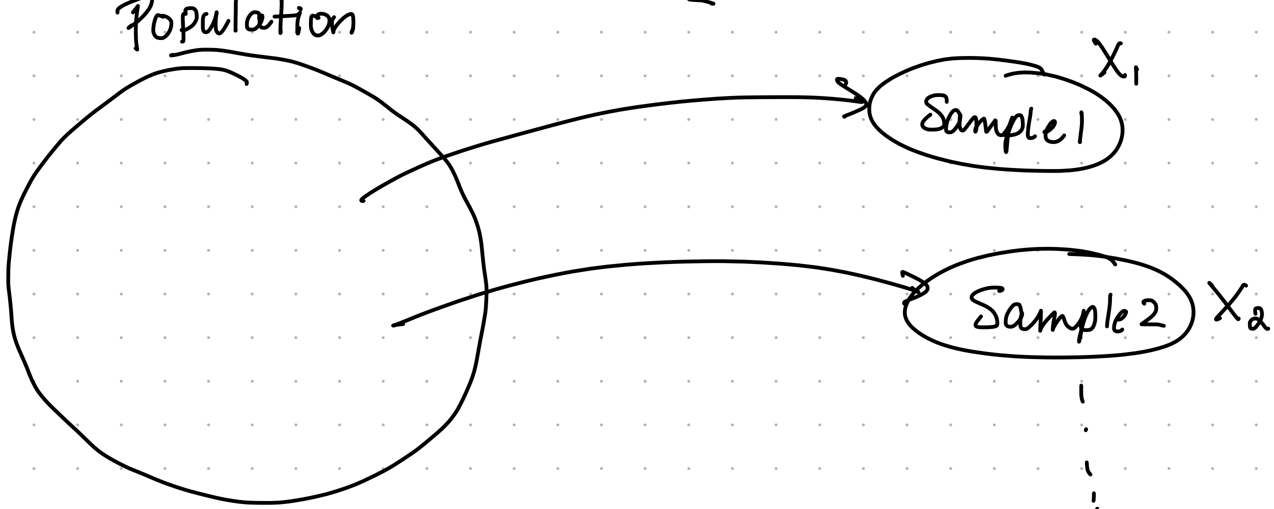
Now: try to guess μ & σ^2 & other properties w/ minimal knowledge of the actual distribution but know it is affected by randomness
↳ predict election
↳ guess people's height
↳ quality control → produce goods

2 main ways of thinking

Frequentist/Classical Methods
we discuss in this class
↓
repeat many times same process

Bayesian Inference
↓
update beliefs/methods w/ new information

Point Estimation



Quantitative features w/ population → parameters (mean, variance, proportion, etc)

Quantities calculated from a random sample
↓
statistic

inference

Def'n Suppose θ is a population parameter

- ↳ average
- ↳ variance
- ↳ proportion

parameter ↓ not subject to randomness ↓ constant

A statistic (from data) $\hat{\theta}$, constructed using a random sample $\{X_1, \dots, X_n\}$, is called a point estimator of θ

go back to this soon

ex Say the population parameter of interest is the mean μ.

Let $\{X_1, \ldots, X_n\}$ be a random sample.

What are possible estimators for μ?

① $\hat{\theta}_1 = \overline{X}_n$ = sample mean = $\dfrac{X_1 + X_2 + \ldots + X_n}{n}$

② $\hat{\theta}_2 = X_1$

③ $\hat{\theta}_3 = \text{median}(X_1, \ldots, X_n)$

④ $\hat{\theta}_4 = X_1^2 + X_2^2 + \ldots + X_n^2$

⑤ $\hat{\theta}_5 = \frac{1}{2}(X_1 + X_2)$

there are ∞ many ways to define a point estimator. What makes it a "good" estimator?

In general: $\hat{\theta}$ = some function of X_i's

$= h(X_1, X_2, \ldots X_n)$

First: Random Sampling : it is unrealistic to collect info on full population

people, goods, etc

Draw samples: R.V's : $X_1, \ldots, X_n$

↳ drawing w/ replacement → to make sure R.V. are independent

↳ $X_i \quad i = 1, \ldots, n$ are i.i.d

↳ technically one person can be polled twice. However, for a large population, this would be extremely low probability.

What makes $\hat{\theta}$, point estimator, for θ "good"?

↳ Need techniques :
- bias/unbiased
- mean square error
- consistent

<u>Given</u> $\hat{\theta} = h(X_1, \ldots, X_n)$ be a point estimator for θ. The <u>bias</u> of $\hat{\theta}$ is defined by

$$B(\hat{\theta}) = E[\hat{\theta}] - \theta$$

Want close to 0

↳ if $B(\hat{\theta}) = 0$ for all possible θ, then $\hat{\theta}$ is unbiased.

<u>ex</u> $\hat{\theta} = \overline{X} = \dfrac{X_1 + \ldots + X_n}{n}$ is this unbiased for $\theta = E[X_i]$?

$$\begin{aligned}
B(\hat{\theta}) &\overset{\text{def'n}}{=} E[\hat{\theta}] - \theta \\
&= E[\overline{X}] - \theta \\
&= E\left[\frac{X_1 + \ldots + X_n}{n}\right] - \theta \\
&= \frac{1}{n}\left(E[X_1] + \ldots + E[X_n]\right) - \theta
\end{aligned}$$

by linearity of $E[X]$

$$= \frac{1}{n}\left(E[X_1] + \dots + E[X_1]\right) - \theta$$

by i.i.d.

$$= \frac{\cancel{n}\, E[X_1]}{\cancel{n}} - \theta = \theta - \theta = 0$$

$\Rightarrow$ Yes! Unbiased! $\bar{X} = \frac{X_1 + \dots + X_n}{n}$ is an unbiased point estimator of μ, the population mean.

ex

What about $\hat{\theta}_6 = \frac{X_1 + \dots + X_n}{n+1}$ point estimator

is it biased?

Work: $B(\hat{\theta}) = E[\hat{\theta}] - \theta$

$$= E\left[\frac{X_1 + \dots + X_n}{n+1}\right] - \theta$$

Same work as above

$$= \frac{n\, E[X_1]}{n+1} - \theta$$

$$= \frac{n\,\theta}{n+1} - \theta = \frac{n\theta - \theta n - \theta}{n+1}$$

$$= \frac{-\theta}{n+1} \neq 0$$

biased since $B[\hat{\theta}] \neq 0$

but as $n \to \infty$, the bias $\downarrow$, gets small

However, not all unbiased estimators are necessarily "good" estimators.

Another tool $\rightarrow$ mean squared error

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

↳ measures the distance btwn $\hat{\theta}$ & θ

$\theta \downarrow$ true parameter

↳ smaller MSE makes a better estimator

* slightly different than <u>variance</u> of $\hat{\theta}$

$$Var(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2]$$

$E[\hat{\theta}] \downarrow$ mean of sample

<u>ex</u> Let $X_1, \ldots, X_n$ be a random sample from distribution with mean $E[X_i] = \theta$ and $Var(X_i) = \sigma^2$.

Let $\hat{\theta}_1 = X_1$ $\qquad$ $\hat{\theta}_2 = \bar{X} = \dfrac{X_1 + \dots + X_n}{n}$

Find $MSE(\hat{\theta}_1)$ and $MSE(\hat{\theta}_2)$.

$$MSE(\hat{\theta}_1) \underset{\text{by def'n}}{=} E[(\hat{\theta}_1 - \theta)^2] = E[(X_1 - E[X_1])^2]$$

$$= Var(X_1) = \sigma^2$$

Rewrite:

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

$\Rightarrow E[Y^2] = Var(Y) + (E[Y])^2$

$$MSE(\hat{\theta}_2) = MSE(\bar{X}) = E[(\hat{\theta}_2 - \theta)^2]$$

$$= E[(\bar{X} - \theta)^2]$$

θ = constant

$Var(X-a) = Var(X)$

$$= Var(\bar{X} - \theta) + (E[\bar{X} - \theta])^2$$

$$= Var(\bar{X}) + (E[\bar{X}] - E[\theta])^2$$

$$= Var(\bar{X}) + (E[\bar{X}] - \theta)^2$$

$$= Var(\bar{X}) + \left(E\left[\frac{X_1 + \dots + X_n}{n}\right] - \theta\right)^2$$

$$= Var(\bar{X}) + \left(\frac{\cancel{n} E[X_1]}{\cancel{n}} - \theta\right)^2$$

$$= Var(\bar{X}) + (\theta - \theta)^2$$

$$= Var\left(\frac{X_1 + \dots + X_n}{n}\right) + 0$$

independent

$$= \frac{1}{n^2}\left(Var(X_1) + \dots + Var(X_n)\right)$$

$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$\Rightarrow \hat{\theta}_1 = X_1 \quad MSE(\hat{\theta}_1) = \sigma^2$

$\hat{\theta}_2 = \overline{X} = \frac{X_1 + \dots + X_n}{n} \quad MSE(\hat{\theta}_2) = \frac{\sigma^2}{n}$

$n \to \infty$, $MSE(\hat{\theta}_2) \to 0$

$$MSE(\hat{\theta}_1) > MSE(\hat{\theta}_2) = MSE(\overline{X})$$

$\therefore$ $\overline{X}$ is a better point estimator

Summary: If $\hat{\theta}$ is a pt estimator for θ:

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

$$= Var(\hat{\theta} - \theta) + \left(E[\hat{\theta} - \theta]\right)^2$$

$$= Var(\hat{\theta}) + B(\hat{\theta})^2$$

by def'n of bias

So far: showed $\overline{X}$ is a good point estimator for the mean,

what about variance? Want to estimate σ^2.

$$\text{Variance} = \sigma^2 = E[(X-\mu)^2] \rightarrow \text{distance from } \mu$$

before $\overline{X} = \dfrac{X_1 + \dots + X_n}{n}$ is a good pt est for μ

try: $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{K=1}^{n}(X_K - \mu)^2 = \dfrac{(X_1-\mu)^2 + \dots + (X_n-\mu)^2}{n}$

Not always known

↳ parameter of population

↳ replace w/ $\overline{X}$

$$\approx \overline{S}^2 = \frac{1}{n}\sum_{K=1}^{n}(X_K - \overline{X})^2$$

is this "good"? Is it biased?

First let's rewrite $\overline{S}^2$ before computing bias:

$$\overline{S}^2 = \frac{1}{n}\sum_{K=1}^{n}(X_K - \overline{X})^2 = \frac{1}{n}\sum_{K=1}^{n}(X_K^2 - 2\overline{X}X_K + \overline{X}^2)$$

does not depend on k

$$= \frac{1}{n}\left[\sum_{K=1}^{n}X_K^2 - 2\overline{X}\sum_{K=1}^{n}X_K + \sum_{K=1}^{n}\overline{X}^2\right]$$

$$= \frac{1}{n}\left[\sum_{K=1}^{n}(X_K^2) - 2\bar{X}\,n\bar{X} + n\bar{X}^2\right] \qquad \bar{X} = \frac{1}{n}\sum_{K=1}^{n} X_K$$

$$= \frac{1}{n}\left[\sum_{K=1}^{n}(X_K^2) - \underbrace{n\bar{X}^2}_{\text{outside of sum}}\right]$$

Is it biased?

$E[X_K^2] = Var(X_K^2) + (E[X_K])^2$ by rewriting variance def'n

$$E[\bar{S}^2] = \frac{1}{n}\left[\sum_{K=1}^{n}(E[X_K^2]) - n\,E[\bar{X}^2]\right]$$

by Linearity of $E[X]$

$$= \frac{1}{n}\left[\sum_{K}^{n}(\sigma^2 + \mu^2) - n\,E[\bar{X}^2]\right]$$

$$= \frac{1}{n}\left[n(\sigma^2+\mu^2) - n\left(\underbrace{var(\bar{X})}_{\frac{\sigma}{n}} + \underbrace{(E[\bar{X}])^2}_{\mu^2}\right)\right]$$

$$= \sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{n} - \cancel{\mu^2}$$

$$= \frac{n\sigma^2 - \sigma^2}{n} = \frac{(n-1)}{n}\sigma^2$$

$$\Rightarrow E[\bar{S}^2] = \frac{(n-1)\sigma^2}{n}$$

$B(\hat{\theta}) = E[\hat{\theta}] - \theta$

$$\therefore B(\bar{S}^2) = E[\bar{S}^2] - \sigma^2$$

$$= \frac{(n-1)\sigma^2}{n} - \sigma^2 = \frac{\cancel{n\sigma^2} - \sigma^2 - \cancel{n\sigma^2}}{n}$$

$$= \frac{-\sigma^2}{n} \neq 0$$

$\therefore$ it is biased, but as $n \to \infty$, the bias gets small.

Can we adjust $\bar{S}^2$ to not have the bias?

Go back $E[\bar{S}^2] = \frac{(n-1)}{n}\sigma^2$

Multiply by $\frac{n}{n-1}$ to get rid of coefficient in front!

$\hookrightarrow B(\bar{S}^2) = \sigma^2 - \sigma^2 = 0$ unbiased

Adjust $\bar{S}$:

$$S^2 = \frac{1}{n}\sum_{K=1}^{\infty}(X_K - \bar{X})^2 \cdot \frac{n}{n-1}$$

$$= \frac{1}{\cancel{n}} \cdot \frac{\cancel{n}}{n-1} \sum_{K=1}^{\infty}(X_K - \bar{X})^2 = \underbrace{\frac{1}{n-1}\sum_{K=1}^{\infty}(X_K - \bar{X})^2}_{\text{best choice}}$$

Define Sample Variance (standard) unbiased point estimator for $\theta = \sigma^2$

$$S^2 = \frac{1}{n-1}\sum_{K=1}^{\infty}(X_K - \bar{X})^2$$

ex sample standard deviation is defined as

$$S = \sqrt{S^2}$$

* however, this is biased estimator for σ

<u>ex</u> heights randomly chosen people to estimate population mean, variance, standard deviation of heights

data: $\overset{X_1}{166.8}, \overset{X_2}{171.4}, \overset{X_3}{169.1}, \overset{X_4}{178.5}, \overset{X_5}{168.0}, \overset{X_6}{157.9}, \overset{X_7}{170.1}$

find sample mean, variance, s.d.,

$$\tilde{X} = \frac{X_1 + \dots + X_7}{7} = 168.8$$

$$S^2 = \frac{(X_1 - 168.8)^2 + \dots + (X_7 - 168.8)^2}{7-1} = 37.7$$

$$S = \sqrt{S^2} = 6.1$$

Beyond mean, variance:

ex Bag contains 3 colored balls

↳ red
↳ yellow } 2 colors

No other info on color or proportion but allowed to draw one at a time and record color & replace → 4 times

We want to know total yellow balls → θ

$\theta = 0, 1, 2, \text{or } 3$

We define R.V. $X_i = \begin{cases} 1 & \text{if } i\text{th ball is yellow} \\ 0 & \text{if } \text{" " " red} \end{cases}$

data: $X_1 = 1$ $X_2 = 0$ $X_3 = 1$ $X_4 = 1$

yellow red yellow yellow

and $X_i \sim \text{Bern}(\underbrace{\tfrac{\theta}{3}}_{p})$ i.i.d.

$$P_{X_i} = \begin{cases} \frac{\theta}{3} \; p & X = 1 \\ 1 - \frac{\theta}{3} & X = 0 \end{cases}$$

$1 - p$

⇒ joint prob

$$P_{X_1 X_2 X_3 X_4}(1,0,1,1) = P_{X_1}(1) P_{X_2}(0) P_{X_3}(1) P_{X_4}(1)$$

by independence

$$= \frac{\theta}{3}\left(1-\frac{\theta}{3}\right)\frac{\theta}{3}\,\frac{\theta}{3}$$

$$= \underbrace{\left(\frac{\theta}{3}\right)^3\left(1-\frac{\theta}{3}\right)}_{\text{function on } \theta}$$

function on θ → depends on θ

likelihood function

$$P_{X_1X_2X_3X_4}(x_1,x_2,x_3,x_4;\theta)$$

Plug in values of θ to maximize likelihood

θ	$P_{X_1X_2X_3X_4}(\theta)$
0	0
1	0.0247
2	0.0988
3	0

← the highest probability → more likely to occur

$\hat{\theta}=2$ is an estimate of θ, total # of yellow balls "maximum likelihood estimate (MLE) of θ"

Generalized from previous example

Maximum Likelihood Estimate (MLE)

Def'n the likelihood function for a given sample data $\{X_1, \ldots, X_n\}$ from a distribution X is defined as

$$L(x_1, x_2, \ldots, x_n; \theta) = \begin{cases} \underbrace{P_{X_1 X_2 \cdots X_n}(x_1, \ldots, x_n; \theta)}_{} & \text{discrete} \\ \underbrace{f_{X_1 X_2, \cdots X_n}(x_1, \ldots, x_n; \theta)}_{\text{then use independence to get a product}} & \text{cont} \end{cases}$$

then use independence to get a product

* want to maximize $\longrightarrow$ take derivative & set $= 0$

* in some cases, it is easier to work w/ "log likelihood function" $\rightarrow \ln L(x_1, \ldots, x_n; \theta)$

ex 3 repeated coin tosses from unfair coin w/ prob p. We count # heads & after 4 experiments we counted 1, 3, 2, 2

Want to find MLE of $\theta = p$, unfair coin parameter

$X_i \sim Bin(3, p) = Bin(3, \theta)$ i.i.d

$$P_{X_i}(x) = \binom{3}{x} p^x (1-p)^{3-x}$$

$$P_{X_i}(x;\theta) = \binom{3}{x} \theta^x (1-\theta)^{3-x}$$

joint distribution

$$L(1,3,2,2;\theta) = P_{X_1 X_2 X_3 X_4}(x_1, x_2, x_3, x_4; \theta)$$ independence

$$= P_{X_1}(1;\theta)\, P_{X_2}(3;\theta)\, P_{X_3}(2;\theta)\, P_{X_4}(2;\theta)$$

$$= \binom{3}{1}\theta(1-\theta)^2 \binom{3}{3}\theta^3(1-\theta)^0 \binom{3}{2}\theta^2(1-\theta)^1 \binom{3}{2}\theta^2(1-\theta)^1$$

#s

$$L(1,3,2,2;\theta) = \underbrace{27\,\theta^8(1-\theta)^4}_{\text{function of } \theta = p,\ 0 \le \theta \le 1} \longleftarrow$$ want to maximize

$\hookrightarrow$ der = 0

$$\frac{d}{d\theta} L(1,3,2,2;\theta) = 27\left(8\theta^7(1-\theta)^4 - 4(1-\theta)^3\theta^8\right) = 0$$ Product Rule

$$8\theta^7(1-\theta)^4 - 4(1-\theta)^3\theta^8 = 0$$

$$8\theta^7(1-\theta)^4 = 4\theta^8(1-\theta)^3$$

$$2(1-\theta) = \theta$$

$$2 - 2\theta = \theta$$

$$2 = 3\theta$$

$$\theta = \frac{2}{3}$$

So, $\hat{\theta}_{ML} = \frac{2}{3}$ → estimate unfair coin to have $p = \frac{2}{3}$ of heads occured

→ *Note θ could be a vector of parameters

$$L(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_K)$$

Extra example: Find $\hat{\theta}_{ML}$ for $X_i \sim Exp(\theta)$ when the observed data is

$$(x_1, x_2, x_3, x_4) = (1.23, 3.32, 1.98, 2.12)$$

$X_i \sim Exp(\theta)$ so $f_{X_i}(x;\theta) = \theta e^{-\theta x}$

and the joint prob can be simplified using independence:

$$L(X_1, X_2, X_3, X_4; \theta) = f_{X_1 X_2 X_3 X_4}(x_1, x_2, x_3, x_4; \theta)$$

$$= f_{X_1}(x_1; \theta) f_{X_2}(x_2; \theta) f_{X_3}(x_3; \theta) f_{X_4}(x_4; \theta)$$

$$= \theta e^{-1.23\theta} \theta e^{-3.32\theta} \theta e^{-1.98\theta} \theta e^{-2.12\theta}$$

$$= \theta^4 e^{-(1.23+3.32+1.98+2.12)\theta}$$

$$L = \theta^4 e^{-8.65\theta}$$

try ln to make computation easier, since ln is a monotone function, the max will occur at the same location. Although without the ln, you can take the derivative as is and still get the correct answer.

$\ln(ab) = \ln a + \ln b$

$$\ln L = \ln(\theta^4 e^{-8.65\theta})$$

$\ln(x^a) = a \ln x$

$$= \ln(\theta^4) + \ln(e^{-8.65\theta})$$

$\ln e = 1$

$$= 4 \ln \theta - 8.65\theta \ln e$$

$$\ln L = 4 \ln \theta - 8.65\theta$$

take derivative: $\frac{4}{\theta} - 8.65$

Set $=0$

$$\frac{4}{\theta} - 8.65 = 0$$

$$4 = 8.65\,\theta$$

$$\theta = \frac{4}{8.65} = .4624$$

$$\hat{\theta}_{ML} = 0.4624$$

W/o ln trick:

$$L = \theta^4 e^{-8.65\theta}$$

$$L' = 4\theta^3 e^{-8.65\theta} - 8.65\, e^{-8.65\theta}\,\theta^4 = 0$$

$$4\theta^3 \cancel{e^{-8.65\theta}} = 8.65\,\theta^4 \cancel{e^{-8.65\theta}}$$

$$\theta = \frac{4}{8.65}$$ <u>same!</u>