OCT 16

Review: homework 03

Review: central limit theorem

What is the distribution of $\bar{X}$?

What is $\mathbb{E}(\bar{X})$?

What is $\mathrm{Var}(\bar{X})$

$X_1, \dots, X_n$ IID

$\mathbb{E}X_i = \mu$

$\mathrm{Var}(X_i) = \sigma^2$

Today: Estimating parameters

$X_1, X_2, \dots$ IID $N(\mu, \sigma)$ How do we find μ?

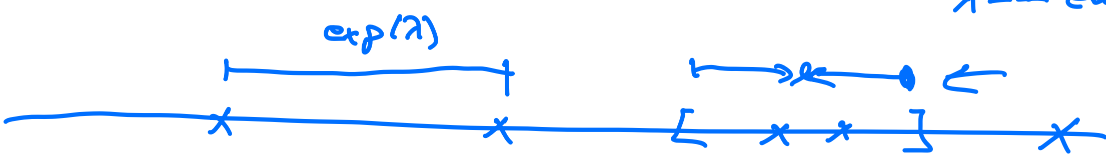
→ Poisson (λ) DISTRIBUTION <u>HOW</u> & <u>WHERE</u>

→ Normal (μ, σ) DISTRIBUTION (CLT)

Normal approx to $\bar{X}$

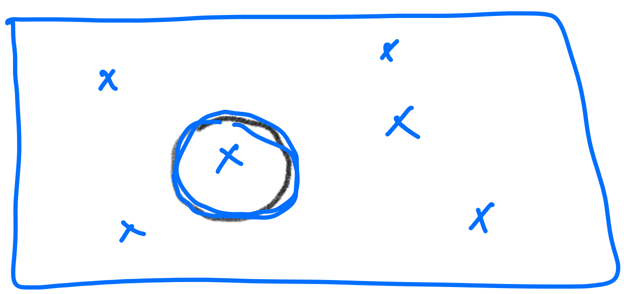
to $\sum_{i=1}^n X_i = S_n$

λ = # events/unit time



N = # interval of length l

$N \sim \text{Poisson}(l \times \lambda \quad)$



falling / unit time in table

$N \simeq \text{Poisson}(\lambda t)$

$5 = \mathbb{E}N = \lambda t$

Last time: $X_1, X_2, \ldots\ldots X_n$ IID

$\mathbb{E}X_i = \mu$ $\quad Var(X_i) = \sigma^2$
$\quad\quad\quad\quad SD = \sigma$

$\bar{X} = (X_1 + X_2 + \ldots + X_n)/n$

$\bar{X} \simeq \text{Normal}(\mu, \sigma/\sqrt{n})$

$\mathbb{E}(\bar{X}) = \mathbb{E}[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i = \frac{1}{n}\sum_{i=1}^n \mu = \mu$

$Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = n\sigma^2/n^2$

$= \sigma^2/n$

$SD(\bar{X}) = \sigma/\sqrt{n}$

HW03 PR 3

$S_{100} = \sum_{i=1}^{100} X_i$ $\qquad X_i \sim$ $\mathbb{E}(X_i) = 49$ cm
$\qquad\qquad SD(X_i) = 30$ cm

(a)

What is Dist'n of S_{100}?

$\bar{X} \simeq N(\mu, \sigma/\sqrt{n})$ $\qquad S_n = \sum_{i=1}^n X_i$

$S_n \simeq N(n\mu, \sqrt{n}\,\sigma)$ $\qquad \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$

$\mathbb{E}S_n = n\,\mathbb{E}\bar{X} \overset{?}{=} n\mu$ $\qquad n = 100$

$SD(S_n) = n\,SD(\bar{X}) = n\,\sigma/\sqrt{n} = \sqrt{n}\,\sigma$

$\mathbb{P}(S_n > 55)$

Estimating parameters

Toss a coin w/ $p = \mathbb{P}(\text{"heads"})$ → fixed unknown

toss 100 times. $X = 62$ # heads

If p is unknown, want to guess its values based on data

DATA = 62 HEADS

Guess at $p \Leftrightarrow \hat{p} = 62/100$

$\hat{p}$ is the outcome of experiment so it has a prob. distribution

$\hat{p}$ is a <u>statistic</u>: depends on data

$\hat{p}$ is an estimator of p

$X_1, X_2, ---, X_{100}$ IID $\mathbb{E}X_i = \mu$ $Var(X_i) = \sigma^2$

unknown fixed constants

Guess μ:

Goal: guess μ is $\overline{X}$

$\hookrightarrow$ statistic

General for finding estimates of parameters given data

Likelihood estimation

Coin example: $X = \#$ heads in 100 tosses

$$\mathbb{P}(X = k \mid p) = \binom{100}{k} p^k (1-p)^{100-k}$$

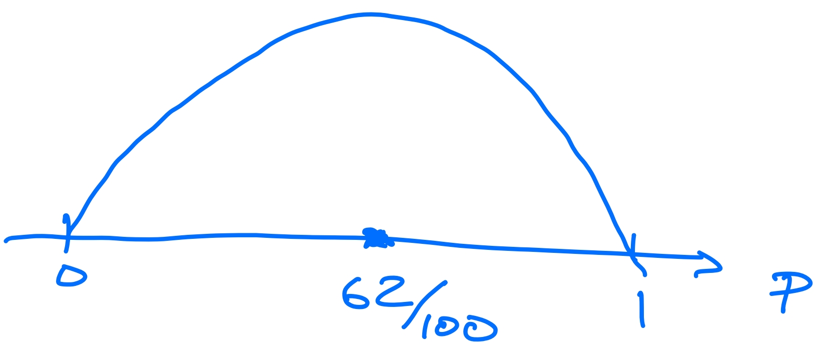
$\hookrightarrow$ Binomial $(100, p)$

Observe $X = 62$ want to know value of p

Likelihood function

$$L(p \mid X = 62) = \binom{100}{62} p^{62} (1-p)^{38}$$

Likelihood principle: pick value of p making this as large as possible



$$l(p) = \log L(p) = \log\binom{100}{62} + 62\log p + 38\log(1-p)$$

$$l'(p) = 0 + 62/p - 38/(1-p)$$

$$l'(p) = 0 \iff 62/p - 38/(1-p) = 0 \qquad p(1-p)$$

$$62(1-p) - 38p = 0$$

$$62 - 62p - 38p = 0$$

$$62 - 100p = 0$$

$$62 = 100p$$

$$p = 62/100$$

Max. Likelihood estimate of unknown parameter $p \iff$

$$\hat{p} = 62/100$$

In general: if p is unknown $X \sim \mathrm{Bin}(n,p)$

$$\hat{p} = X/n$$

$\lambda = \mathbb{E}X_i$

<u>Example</u>: $X_1, \ldots, X_n$ IID Poisson (λ)

Don't know λ but want to estimate from data

$P(X_1 = k_1, \ldots, X_n = k_n \mid \lambda)$

Suppose 5 data points $X_1 = 2$, $X_2 = 0$, $X_3 = 3$, $X_4 = 3$, $X_5 = 1$

$$\mathbb{P}(X_1 = 2, X_2 = 0, X_3 = 3, X_4 = 3, X_5 = 1 \mid \lambda)$$

$$= \mathbb{P}(X_1 = 2 \mid \lambda)\,\mathbb{P}(X_2 = 0 \mid \lambda)\,\mathbb{P}(X_3 = 3 \mid \lambda)\,\mathbb{P}(X_4 = 3 \mid \lambda)\,\mathbb{P}(X_5 = 1 \mid \lambda)$$

$$= \frac{e^{-\lambda}\lambda^2}{2!}\,\frac{e^{-\lambda}\lambda^0}{0!}\,\frac{e^{-\lambda}\lambda^3}{3!}\,\frac{e^{-\lambda}\lambda^3}{3!}\,e^{-\lambda}\lambda^1/1!$$

$$L(\lambda) = \frac{e^{-5\lambda}\lambda^{2+0+3+3+1}}{2!\,0!\,3!\,3!\,1!} = \frac{e^{-5\lambda}\lambda^{9}}{2!\,0!\,3!\,3!\,1!}$$

$$\ell(\lambda) = \log L(\lambda) = -5\lambda + 9\log\lambda - \sum_{i=1}^{5}\log k_i$$

$$\ell'(\lambda) = -5 + 9/\lambda$$

$$= 0 \iff -5 + 9/\lambda = 0$$

$$9/\lambda = 5$$

$$\lambda = 9/5$$

Value maximizing $L(\lambda) \iff \hat{\lambda} = 9/5$

In general $\hat{\lambda} = \sum_{i=1}^{n} k_i / n$ maximum likelihood estimator

$\hat{\lambda} = \overline{X}$

<u>Example</u>

Note: If $X_1, \ldots, X_n$ IID r.v.s

$\mathbb{E}X_i = \mu, \quad Var(X_i) = \sigma^2$

Want to estimate μ always $\overline{X}$ is a reasonable estimator

$$\left.\begin{array}{l} \mathbb{E}(\overline{X}) = \mu \\ Var(\overline{X}) = \dfrac{\sigma^2}{n} \qquad SD(\overline{X}) = \sigma/\sqrt{n} \end{array}\right\}$$ Fundamental Theorem of Statistic

---

$X_1, \ldots, X_n$ Poisson (λ)

$\mathbb{E}X_i = \lambda$

Good estimator of λ should be: $\overline{X}$

Maximum Likelihood estimator is $\overline{X}$

$X_1, X_2, \ldots, X_n$ IID $N(\mu, \sigma)$

estimator of μ is $\bar{X}$ $\qquad \mu = \mathbb{E}X_1$ = pop. mean

$\bar{X}$ as best guess at μ

How certain are we of this estimate?

In this example: estimate μ by $\hat{\mu} = 118$.

How likely is $\hat{\mu}$ to be more than 5 gm from truth?

$\mathbb{P}(|\hat{\mu} - \mu| > 5)$? $\qquad$ Suppose know $\sigma = 3$

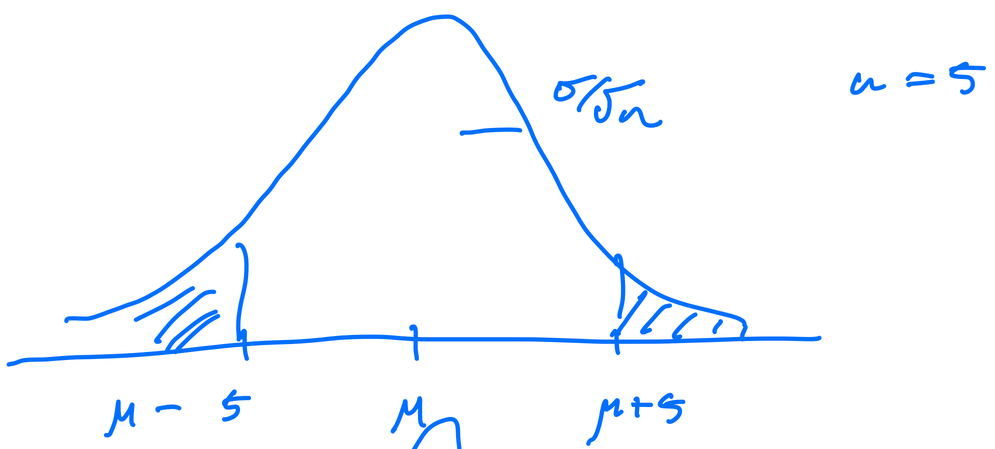
$\mathbb{P}(|\bar{X} - \mu| > 5)$

$\hookrightarrow \bar{X} \sim N(\mu, \overset{3}{\sigma}/\sqrt{5})$

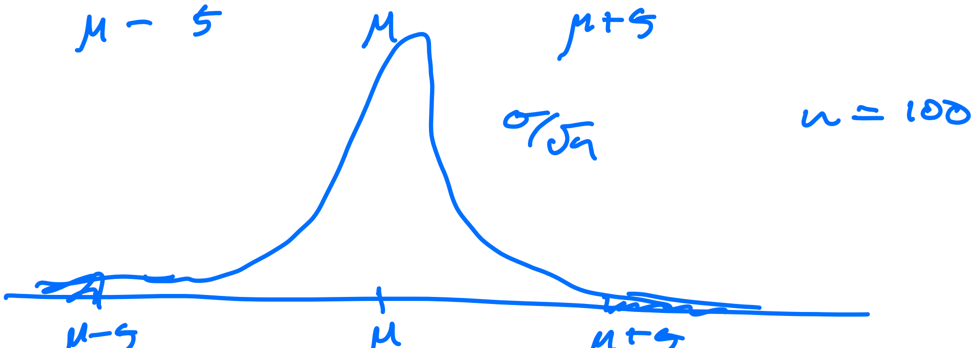
$\bar{X} - \mu \sim N(0, 3/\sqrt{5})$

$\mathbb{P}(\bar{X} - \mu < -5) + \mathbb{P}(\bar{X} - \mu > 5)$

What is advantage of taking sample bigger than 5?



$SD(\bar{x}) = \sigma/\sqrt{n}$



$$\mathbb{P}\left(|\bar{x} - \mu| > \square \right) \longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty$$