SDS → Unit 2

→ Descreptive stat → summary stat that quantify describes or summarises features from collection of info.

→ Inferential stat → Methods that make decisions/pred on pop based on sampled data.

→ Probability → quantitative desc of chances/ likelihoods ~~[illegible]~~ associated with various outcomes.

→ Conditional prob $P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A\cap B)}{P(B)}$

For ind events → $P(A \,\&\, B) = P(A) \times P(B)$

For dep events → $P(A \,\&\, B) = P(A) \times P(B|A)$

$$P(A\cup B) = P(A) + P(B) - P(A\cap B)$$

$P(A) = P(A\cap B) + P(A\cap B^c)$ → Total Prob Rule

Baye's Rule → $P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$

→ Random variable → var whose value is det by chance.
→ continuous
→ Discrete

→ Discrete random variable → Op on discrete rv doesn't effect shape of prob histogram

→ Adding constant/subtracting constant →

$$\mu_{x\pm b} = \mu_x \pm b$$

$$\sigma^2_{x\pm b} = \sigma_x^2 \rightarrow \text{unchanged}$$

$\mu(ax+b) = a\mu_x + b$

$\sigma^2(ax+b) = a^2\sigma_x^2$

$\sigma(ax+b) = |a|\sigma_x$

→ Multiplying/dividing constant →

$$\mu_{ax} = a\mu_x$$

$$\sigma^2_{ax} = a^2\sigma_x^2$$

$$\sigma_{ax} = |a|\sigma_x$$

when $T = X+Y$, $D = X-Y$

$\mu_T = \mu_X + \mu_Y$, $\mu_D = \mu_X - \mu_Y$

$\sigma_T^2 = \sigma_X^2 + \sigma_Y^2$, $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 = \sigma_T^2$

→ If $x_1, x_2, \dots x_n$ → simple random sample then $x_1, x_2, \dots x_n$ are treated as indp random var, all with same distr

↓

independent → Outcome of one obs/ doesn't effect outcome of other obs/

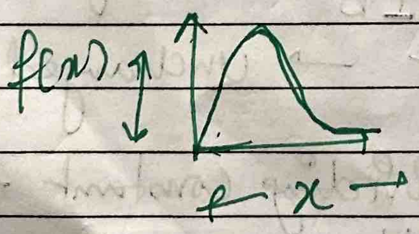
$$E(x) = \int_a^b x f(x)\, dx$$

$$E(x) = \Sigma x f(x)$$

→ Continuous random var →

→ Can assume infinitely many values corresponding to a point in a particular fine interval

→ Prob distribution → describes how prob/ are distr over all possible values. Denoted by $f(x)$ → density func. Graph of $f(x)$ → smooth curve.



for percentile → inv normal of value $z = \frac{x-\mu}{\sigma}$

$x = \mu - z(\sigma)$

→ Normal dist → Bell curve

→ Mean = Median = Mode
→ Symmetry about centre
→ 50% values left of mean (lesser than) & 50% to the right (greater than)

μ → change → shifts dist left/right
σ → change → changes spread.

PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Normal prob func: $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$

$\mu = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

$\sigma^2 = \int_{-\infty}^{\infty} x^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx - \mu^2$

→ Prop. → total area = 1,
→ 50% area → left of mean, 50% → right
→ no matter μ & σ:

68-95-99.7% Rule
- → area b/w $\mu-\sigma$ & $\mu+\sigma$ → 68%
- → area b/w $\mu-2\sigma$ & $\mu+2\sigma$ → 95%
- → area b/w $\mu-3\sigma$ & $\mu+3\sigma$ → 99.7%

$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = .68$ and so on.

(*) → Solve Q in slide 163, 167

→ Standard normal dist, $Z = \frac{X - \mu}{\sigma}$

→ Linear combination of normal random variables →

→ If X & Y are two ind, normal var, such that:

$X \sim N(\mu_1, \sigma_1^2)$ & $Y \sim N(\mu_2, \sigma_2^2)$

then $(X - Y) \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

→ If X & Y are two ind, normal var such that:

$X \sim N(\mu_1, \sigma_1^2)$ & $Y \sim N(\mu_2, \sigma_2^2)$

then,

$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$aX \sim N(a\mu_1, a^2\sigma_1^2)$

$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

$aX + b \sim N(a\mu + b, a^2\sigma^2)$

→ Let $X_1, X_2, \dots X_n$ be ind, & normally dist, with mean μ & variance σ^2, then,

$\bar{X} \sim N(\mu, \sigma^2/n)$

→ Random no,s → no, chosen by chance from specified ~~seq~~ dist, such that selecting large set of these no,s reproduces the dist.
Chosen by methods → give equal prob to nos occurring in dist

→ Prereq, → ind, & doesn't have correlation with successive nos,

→ freq, of occurence of random nos → approx same

True ↑ Pseudo ↗

→ Random no. gen. → O/P → seq. of ind. & identically dist. random no.s b/w 0 & 1.
Random nos. transformed into req. random dist.
→ It's hardware dev. / software algo. that gen. no. that is taken from limited/unlimited dist. & outputs it.

→ True Random No. Gen. → Measures some phy. phenomenon that is exp to be random & compensates for possible biases in the measurement process.

→ Pseudo Random No. Gen. →
→ Uses comp. algo. → prod. long seq. of random results → comp. det. by seed value/key.
→ Entire seq. → reprod. if seed value known.
→ Prop. →
→ Fast & not memory intensive
→ Able to reprod. given stream of random nos.
→ Provision for prod. sep. diff. indep. streams of random nos.

→ Random No. Seed → no./vector used to initialise Pseudo ". Set the starting point for random no. seq. Doesn't need to be random. Initialised with comp. real time clock → to have ext. noise.

uniformly or non-uniformly dist.
→ Random variate → var. gen. from uniformly dist. pseudorandom nos. Particular outcome of random var.

→ Random variate gen/ →

→ obj/ → prod/ obj/ that have stochastic prop/ of given random var/.

→ Relies on uniformly dist/ random no/ in the closed interval [0,1].

→ Distribution func/ $F(x) = Pr(x=x)$ where $-\infty < x < \infty$

→ Assumptions for random variate gen/ →

→ There exists perfect uniform random no/ gen/ that can prod/ seq/ of ind/ random var/ uniformly dist/ on [0,1]. U(0,1),

→ Computers can store & manipulate real no/.

→ Factors affecting random variate gen/ →

→ Exactness → how well gen/ prod/ random variates acc/ to char/ of desired dist/. Refers to theoretical exactness of random variate gen/ as well as errors induced by gen/ & comp/.

→ Speed → refers to computational set up & execution time req/ to gen/ random variates.

→ Space → refers to comp/ memory req/ for gen/.

→ Simplicity → simplicity of both algo/ & imp/. Includes lines of code, support routines, maths ops/, etc, req/.

→ Acceptance - Rejection Method →

→ Used when closed form cumulative dist, func doesn't exist / diff, to calculate.

→ Variates gen, from one dist, & are either accepted / rej, in such a way that accepted values have the desired dist,

→ Algo, →

→ for random var, X let $f(x)$ → density func,

→ $t(x)$ → majorising func such that $t(x) \geq f(x)$ for all values of x.

→ let $g(x) = t(x)/c$ denote density func, prop, to c such that $c = \int_{-\infty}^{\infty} t(x)\,dx$

→ generate $x \sim g(x)$

→ generate $U \sim U(0,1)$

→ If $u > f(x)/t(x)$ → reject x & go to step 1.

→ return x.

→ execution time dep, on →

→ time to gen, x from $g(x)$

→ time to perform comparison of $g(x)$ & $t(x)/c$

→ no, of iter, req, to return x.

→ Drawback → Trials ratio (avg, no, pts, needed to prod, accepted X) should be close to 1, else, gen, not eff, enough.

→ For Poisson dist, →

→ Set $n = 0$, $P = 1$.

→ Gen, random no, R_{n+1}, replace P by $P \times R_{n+1}$

→ If $P < e^{-\alpha}$, accept $N = n$, otherwise reject current n, increase n by one & return to step 2.

→ For Normal dist^y →

$X \sim N(0,1)$ then density func^y of x:

$$f(x) = \frac{2e^{-x^2/2}}{\sqrt{2\pi}}, \; x > 0$$

$g(x) = \sqrt{2e/\pi}\, e^{-x}$ majorises f(x)

→ Algo, →

→ Gen, exponential Y with mean 1.

→ Gen, U from U(0,1) ind, of Y.

→ If $U \leq e^{-(y-1)(y-1)/2}$, accept Y. Else, reject Y & return to step 1.

→ Return X = Y or X = −Y — both with prob, 0.5

→ Adv, of accept - reject →

→ Doesn't req, inverse CDF.

→ Eff,.

→ Disadv, of accept - reject →

→ Sample a lot to get accept

→ lot of unwanted samples being taken if func^y being sampled conc, in certain reg,.

→ Curse of dimensionality → prob, becomes larger → $\left(\frac{\text{embedded volume}}{\text{embedding volume}}\right) \to 0$, so lot of rej, takes place before useful sample gen,.

$1 - 1/k^2 = 0.5$

$1/k^2 = 0.5 \qquad k = \pm\sqrt{2} = \pm 1.414$

→ Chebyshev's inequality → value of random variable differs from its mean by more than a give multiple of its SD.

→ $P(|x - \mu| \geq k\sigma) \leq 1/k^2$

$P(\mu - k\sigma < x < \mu + k\sigma) \geq 1 - (1/k^2)$

→ Proportion of obs, within k SD $\geq 1 - (1/k^2)$.

No, of SD	$1 - (1/k^2)$
1.5	0.56
2	0.75
3	0.89
4	0.94

→ Only used when dist, of random variable is unknown.

→ learn

Probability plots →

→ Construction →

→ Sort data

→ Assign evenly spaced values for data b/w 0 & 1.

→ For each x_i, $\frac{i - 0.5}{n}$, i → pos, of data item, n → size of dataset

→ Find theoretical quantiles → Q_i ($X = z*\sigma + \mu$)

→ Plot (x_i, Q_i)

→ Plot (x_i, x_i)

→ Check if it forms approx, straight line.

→ Point estimate → single numeric value specified for the data. Also called sample stat. Used to estimate unknown constant/param. Infer about pop. param.

$\hat{p} = \frac{x}{n}$ — $\hat{p}$ → point est.
x → no. of successes
n → sample size

→ Prop. of point estimators →
→ Bias →
→ exp. value of ~~est.~~ estimator diff. from value of param. being est.
→ If equal → unbiased.
→ Consistency →
→ how close point estimator is to the true value of param.
→ ↑ consistent → sample more towards true value of pop. param.
→ Efficiency →
→ should have → smallest variance
→ unbiased obs.
→ consistent

lower → values closer to ~~mean & better fit for estimator~~

→ Mean Squared Error (MSE) →

→ Good estimator → both accurate & precise → measured by SD/uncertainty
accurate → measured by bias

DO again

→ bias $\hat{\theta} = \mu_{\hat{\theta}} - \theta$ → diff. b/w mean of estimator & true value

→ uncertainty $\sigma_{\hat{\theta}}$ → SD → standard error of estimator

$$MSE_{\hat{\theta}} = \underbrace{(\mu_{\hat{\theta}} - \theta)^2}_{\text{bias}} + \underbrace{\sigma_{\hat{\theta}}^2}_{\text{variance}}$$

→ Derivation of MSE →

$MSE_{\hat{\theta}} = E(\hat{\theta} - \theta)^2$ $\quad \hat{\theta} - \theta$ → diff b/w est. & true value

$MSE_{\hat{\theta}} = E(\hat{\theta}^2 + \theta^2 - 2\hat{\theta}\theta)$

$MSE_{\hat{\theta}} = E(\hat{\theta}^2) + \theta^2 - E(2\hat{\theta}\theta)$

Adding & subtracting with $E(\hat{\theta})^2$,

$MSE_{\hat{\theta}} = E(\hat{\theta}^2) + \theta^2 - 2\theta E(\hat{\theta}) + E(\hat{\theta})^2 - E(\hat{\theta})^2$

$MSE_{\hat{\theta}} = \underbrace{E(\hat{\theta} - \theta)^2}_{\text{bias}} + \underbrace{E(\hat{\theta}^2) - E(\hat{\theta})^2}_{\text{variance}}$

→ MSE for Binomial : $X \sim Bin(n, p)$, find MSE for $\hat{p} = X/n$

$bias = \mu_{\hat{p}} - p$

$\hat{p} = X/n$

$\mu_{\hat{p}} = \frac{\mu_X}{n} = \frac{np}{n} = p$

$\mu_{\hat{p}} = p$, $\hat{p}$ → unbiased, bias = 0.

uncertainty $\sigma_X = \sqrt{np(1-p)}$

$\sigma_{\hat{p}} = \frac{\sigma_X}{n} = \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{\cancel{n}p(1-p)}{n^{\cancel{2}}}}$

$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$

$MSE_{\hat{\theta}} = bias + variance = 0 + \frac{p(1-p)}{n}$

$MSE_{\hat{\theta}} = \frac{p(1-p)}{n}$

→ MSE for Poisson: $\hat{\mu} = \bar{x}$

$$MSE(\hat{\mu}) = Var(\hat{\mu}) + bias^2(\hat{\mu})$$

$$bias(\hat{\mu}) = E(\hat{\mu}) - \mu$$

$x_i \sim Poi(\mu)$
$E(x_i) = \mu$
$var(x_i) = \mu$

$$bias(\hat{\mu}) = E(\bar{x}) - \mu$$

$$bias(\hat{\mu}) = E\left(\frac{\sum x_i}{n}\right) - \mu$$

$$bias(\hat{\mu}) = \frac{1}{n}E\left(\sum x_i\right) - \mu$$

$$\cancel{bias(\hat{\mu}) = \frac{1}{n}E(n x_i) - \mu}$$

$$bias(\hat{\mu}) = \frac{1}{n}E[n x_i] - \mu$$

$$bias(\hat{\mu}) = E(x_i) - \mu$$

$$bias(\hat{\mu}) = \cancel{\mu} - \cancel{\mu} = 0$$

$$Var(\hat{\mu}) = Var(\bar{x})$$

$$= Var\left(\frac{\sum x_i}{n}\right)$$

$$= \frac{1}{n^2}\sum Var(x_i)$$

$$= \frac{1}{n^2}\sum x_i$$

$$= \frac{1}{\cancel{n^2}_{\,n}}\cancel{n}\mu$$

$$= \frac{\mu}{n}$$

$$MSE_{\hat{\mu}} = 0 + \frac{\mu}{n}$$

MSE for Bernoulli : $x \sim \text{Bernoulli}(p)$

x	$p(x)$	
0	$1-p$	→ failure
1	p	→ success

Mean $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{0(1-p) + 1(p)}{1} = p.$

Variance $=$ ~~$\sum(x-\bar{x})$~~ $p(1-p)$

Bias $= \mu_{\hat{p}} - p$

$\mu_{\hat{p}} = \frac{\mu_x}{n} = \frac{\cancel{n}p}{\cancel{n}} = p.$

Bias $= \cancel{p} - \cancel{p} = 0$

$MSE_{\hat{p}} = 0 + p(1-p) = p(1-p)$

→ MSE for Normal :

Mean $\bar{x} = \frac{\sum x_i}{n} = \frac{\cancel{n}\mu}{\cancel{n}} = \mu.$

~~Bias~~ Variance $= \text{var}\left(\frac{\sum x_i}{n}\right)$

$= \frac{1}{n^2} \text{var}(\sum x_i)$

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

$= \sigma^2/n.$

Bias $= \mu - \bar{x} = \cancel{\mu} - \cancel{\mu} = 0$

$MSE = 0^2 + \sigma^2/n$

$MSE = \sigma^2/n$

→ Sampling dist. → Prob. dist. of large no. of samples of size n from a given population

→ Mean of sample mean,

$\cancel{\mu_{\bar{x}} = \mu}$

$$\bar{x} = \frac{x_1}{n} + \frac{x_2}{n} + \dots \frac{x_n}{n}$$

$$\mu_{\bar{x}} = \frac{\mu_{x_1}}{n} + \frac{\mu_{x_2}}{n} + \dots \frac{\mu_{x_n}}{n}$$

$$\mu_{\bar{x}} = \cancel{n}\left(\frac{1}{\cancel{n}}\mu\right) = \mu$$

→ Variance of sample mean,

$$\sigma_{\bar{x}}^2 = \frac{1}{n^2}\sigma_{x_1}^2 + \frac{1}{n^2}\sigma_{x_2}^2 + \dots \frac{1}{n^2}\sigma_{x_n}^2$$

$$\sigma_{\bar{x}}^2 = n\left(\frac{\sigma_x^2}{n^2}\right)$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$$

→ SD of sample mean,

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

used when $n \approx 30$ or greater

→ Central limit theorem →

Let $x_1, x_2, \dots x_n$ be simple random sample

then, $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ → sample mean.

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

$$S_n \sim N(n\mu, n\sigma^2) \text{ when } n \to \text{large}$$

~~Slide~~ 412

or n — large → approx. normal dist.
when sample pop. symmetric → normal dist. for relatively small n values
when sample pop. skewed → normal dist. for approx. n ≥ 30.

→ Central Limit Theorem for sample proportion →

$p \to$ success, $q = 1-p$, $n \to$ size

$\hat{p} = x/n$, mean $\to p$, SD $\to \sqrt{\frac{pq}{n}}$

$$x \sim N(np,\ np(1-p))$$

$$\hat{p} \sim N\left(p,\ \frac{p(1-p)}{n}\right)$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

~~Slide 416~~

$S_x \sim N(N\mu_x,\ N\sigma_x^2)$ = $S_x \sim N(26, 0.65)$

$S_y \sim N(N\mu_y,\ N\sigma_y^2)$ = $S_y \sim N(29.25, 1.4625)$

$T = S_x + S_y$

$\mu_T = 26 + 29.25 = 55.25$

$\sigma_T^2 = (0.65) + (1.4625) = 2.1125$

Area b/w 50 & 55 hrs:

~~Area before 50:~~ $z < 50 = -3.61$

$z < 55 = -0.172$

$P(z < 50) = 0.0000153$

$P(z < 55) = 0.43171$

Area = 0.4317

Page No.

→ MLE — used for estimating parameters when pop. param, mean & variance are unknown!

MLE → value of estimators which when put for parameters maximises likelihood func.

Do again

→ MLE for Bernoulli:

$$f(x_i, p) = P(x = x_i) = p^{x_i}(1-p)^{1-x_i}$$

$$P(x_1, x_2 \ldots x_n / p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$L(p) = p^{\Sigma x_i}(1-p)^{n - \Sigma x_i}$$

Natural log both sides,

$$\ln L = \Sigma x_i \ln p + (n - \Sigma x_i) \ln(1-p)$$

$$\frac{d(\ln L)}{dp} = 0$$

$$0 = \Sigma x_i \frac{d(\ln p)}{dp} + (n - \Sigma x_i) \frac{d(\ln(1-p))}{dp}$$

$$0 = \frac{\Sigma x_i}{p} + (n - \Sigma x_i)\left(\frac{-1}{1-p}\right)$$

$$\frac{\Sigma x_i (1-p) - (n - \Sigma x_i)p}{p(1-p)} = 0$$

$$\Sigma x_i - \cancel{p\Sigma x_i} - np + \cancel{p\Sigma x_i} = 0$$

$$\frac{\Sigma x_i}{n} = \hat{p}, \quad \hat{p} = \bar{x}.$$

→ MLE for Binomial :

$$f(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$L(p; n, x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$\ln[L(p; n, x)] = \ln[\text{RHS}]$$

$$= \ln\left(\frac{n!}{x!(n-x)!}\right) + x\ln p + (n-x)\ln(1-p)$$

$$\frac{d}{dp}[L(p; n, x)] = \frac{d}{dp}[\text{RHS}]$$

$$0 = \cancel{\frac{d}{dp}\left(\ln\left(\frac{n!}{x!(n-x)!}\right)\right)}^{0} + \frac{d}{dp}(x\ln p) + (n-x)\frac{d}{dp}(\ln(1-p))$$

$$0 = \frac{x}{p} + \frac{x}{1-p} - \frac{n}{1-p}$$

$$0 = \frac{(1-p)x + px - np}{p(1-p)}$$

$$0 = x - \cancel{xp} + \cancel{px} - np$$

$$\hat{p} = \frac{x}{n}$$

→ MLE for poisson (λ):

$$L(x_1, \ldots x_n \mid \lambda) = \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$L(x_1, \ldots x_n ; \lambda) = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

$$\ln[L(x_i ; \lambda)] = \sum x_i \ln(\lambda) - n\lambda \ln(e) - \ln(x_1! \ldots x_n!)$$

$$\frac{d}{d\lambda}(LHS) = 0$$

$$0 = \sum x_i \frac{d}{d\lambda}[\ln(\lambda)] - n\frac{d\lambda}{d\lambda} + \cancel{\frac{d \ln(x_i!)}{d\lambda}}$$

$$0 = \frac{\sum x_i}{\lambda} - n$$

$$\hat{\lambda} = \frac{\sum x_i}{n} = \bar{\lambda}$$

→ MLE for normal:

$$f(x_i ; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

$$L(x_i ; \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

$$L(\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\cancel{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}}$$

$$\ln(LHS) = \frac{-n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\cancel{\log_e}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\cancel{\frac{d(\ln(LHS))}{d\sigma} = 0}.$$

$$\cancel{0 = \frac{d}{d\sigma}\left(\frac{-n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n}(x_i-\mu)^2\right)}$$

set to 0

$$\frac{\partial(\ln(LHS))}{\partial\mu} = \frac{\partial}{\partial\mu}\left[\cancel{\frac{-n}{2}\ln(2\pi\sigma^2)}\right]^{\to 0} + \frac{\partial}{\partial\mu}\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$0 = \cancel{\frac{-1}{2\sigma^2}}(\cancel{2})(-1)\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\sum x_i - n\mu = 0$$

$$\hat{\mu} = \frac{\sum x_i}{n} = \bar{x}$$

set to 0

$$\frac{\partial(\ln(LHS))}{\partial\sigma^2} = \frac{\partial}{\partial\sigma^2}\left(\frac{-n}{2}\ln(2\pi\sigma^2)\right] + \frac{\partial}{\partial\sigma^2}\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

$$0 = \frac{\partial}{\partial\sigma^2}\left(\frac{-n}{2}\ln(\sigma^2)\right] + \sum_{i=1}^{n}(x_i-\mu)^2\frac{\partial}{\partial\sigma^2}\left(\frac{1}{2\sigma^2}\right)$$

$$0 = -n/2\sigma^2 + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{n}{\cancel{2}\sigma^2} = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\cancel{2}\sigma^4}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

$$\hat{\sigma} = \sqrt{\sum_{i=1}^{n}\left(\frac{\sum x_i - \mu}{n}\right)^2}$$

→ MLE desirable prop. →
- → $n \uparrow$ → bias converges to 0.
- → $n \uparrow$ → variance converges to min.

→ Drawbacks of point estimators →
- → vary from sample to sample
- → Diff. from pop. param.
- → Cannot be confidently claimed to be close to actual parameter.

→ Do

→ Continuity correction → use normal dist for discrete dist. like binomial/poisson. Continuity correction factor used to approximate discrete prob. dist.