

Vijay Masuri
SID: 3031783444

STAT 135: Homework 1

- 1) x_1, x_2, \dots, x_N are a list of numbers with mean μ and standard deviation σ .
 a_1, \dots, a_M are distinct values of list with frequencies n_1, \dots, n_M .

Lecture Slides → Definition 1: $\sigma^2 = E[X^2] - (E[X])^2$

Let X = list of numbers containing x_1, x_2, \dots, x_N .

$$E[X] = \mu$$

definition 2: $E[X] = \frac{1}{N} \sum_{i=1}^N x_i = \frac{M}{N} \sum_{j=1}^{n_j} a_j$ where x_i are the elements of the list, a_j are the distinct values.

$$E[X^2] = \sum_{j=1}^M n_j \cdot a_j^2$$

frequencies don't

change, only values do.

Hence,
(Substitute
into definition)

$$\sigma^2 = \sum_{i=1}^M \frac{n_i a_i^2}{N} - \mu^2$$

1.)

- 2) (1) The samples from Design A are not iid. The view hours of different sessions from the same user might be correlated with each other because the user will likely exhibit similar behavior when using the YouTube app. This means that they will likely watch similar content every time they come on, and thus are likely to have similar watch times. Hence, watch times for a user across multiple sessions are likely to be linearly related, and thus dependent events. Hence, the samples from design A are not iid.

- (2) No, it is not. Because the sampling of users was not fully randomized, we are told that Design A was shown to mostly rural users, while Design B was shown mostly to Bay Area users. This will lead to problems in analysis because of the biased sampling procedure. From this data, we cannot

INDEPENDENT
SAMPLES TEST
(due to locational
differences)

WHAT'S THE DATA?

conclude that the mobile app design is the reason that explains session duration differences between the 2 groups, because there is a significant difference in internet speed between the two groups that explains some of the difference in session duration. For example, a user with a better internet connection might be inclined to stay online longer on YouTube due to higher video quality from more bandwidth, while a user with bad internet could be inclined to use YouTube less due to excessive buffering times and very low video quality due to low bandwidth. Hence, because this sample was not fully randomized and biased in terms of internet speed and location, it's not appropriate to analyze this data set to decide which design is better.

~~ANSWER~~
~~QUESTION~~
~~ANSWER~~

3] (1) It's not possible to make the choice with the information given. we need to know the class sizes for each of the sections in order to compute the standard deviation of the combined group of students. This is because both the definition of population variance $\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$ and the definition

of sample variance $\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ or $\sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ requires knowledge of the population size or sample size to compute properly.

(2) section 1 = 30 students, section 2 = 20 students

$$\begin{array}{ccc} \downarrow & \searrow & \\ M = 80 & SD = 10 & \end{array} \quad \begin{array}{ccc} \downarrow & \nearrow & \\ M = 87 & & SD = 10 \end{array}$$

Let X_1, X_2, \dots, X_{30} represent section 1.

Let $X_{31}, X_{32}, \dots, X_{50}$ represent section 2.

Let X be the combined group.

$$X = [X_1, X_2, \dots, X_{30}, X_{31}, X_{32}, \dots, X_{50}]$$

$$\text{Var}(x) = E[x^2] - [E(x)]^2 \quad (1)$$

$$\text{Var}(x) = \frac{1}{50} \sum_{i=1}^{50} x_i^2 - (\bar{M}_x)^2$$

\bar{M}_x = mean of combined group

$$\text{Var}(x) = \frac{1}{50} \left[\sum_{i=1}^{30} x_i^2 + \sum_{i=31}^{50} x_i^2 \right] - (\bar{M}_x)^2$$

re-arranging (1): $E(x^2) = \text{Var}(x) + [E(x)]^2$

$$\frac{1}{n} \sum_{i=1}^n x_i^2 = \sigma^2 + \bar{M}_x^2$$

$$\sum_{i=1}^n x_i^2 = n\sigma^2 + n\bar{M}_x^2$$

sub in for both summations

$$\text{Var}(x) = \frac{1}{50} \left[30\sigma_{S_1}^2 + 30\bar{M}_{S_1}^2 + 20\sigma_{S_2}^2 + 20\bar{M}_{S_2}^2 \right] - (\bar{M}_x)^2$$

$$\sigma_{S_1}^2 = 100, \sigma_{S_2}^2 = 100, \bar{M}_{S_1}^2 = 6400, \bar{M}_{S_2}^2 = 7569$$

$$\bar{M}_x = \frac{30(80) + 20(87)}{50} = 82.8$$

plug in:

$$\text{Var}(x) = \frac{1}{50} \left[30(100) + 30(6400) + 20(100) + 20(7569) \right] - (82.8)^2$$

$$\text{Var}(x) = 111,76$$

$$SD(x) = \sqrt{\text{Var}(x)} = 10.5716602$$

$$SD(x) \approx 10.572$$

4] (1) (a) X is a random variable.

(b) $E(X)$ is a real number.

(c) X_1, X_n is a random variable.

(d) \bar{X}_n is a random variable.

(2) (a) False, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ where X_1, \dots, X_n are iid samples from distribution X .

Counter-example: Let our sample size be $n=1$.

Let X_1 be 4. [Since it is just a value drawn w/ some probability from the distribution.]

$$\text{Then, } \bar{X}_1 = \frac{1}{1} \sum_{i=1}^1 4 = 4.$$

$$E(X) = 1(2\theta) + 2(6) + 3(7\theta) + 4(1-6\theta)$$

$$E(X) = 2\theta + 2\theta + 9\theta + 4 - 24\theta$$

$$E(X) = 4 - 11\theta \quad \text{where } \theta \in (0, \frac{1}{6})$$

Hence, $2\frac{1}{6} < E(X) < 4$ and thus
 $\uparrow \quad \uparrow$
 $\theta = \frac{1}{6} \quad \theta = 0 \quad \bar{X}_1 \neq E(X).$

Hence, the counter-example is proven and statement is false.

(b) True,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

take

expectation: $E(\bar{X}_n) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$

$$= \frac{1}{n} \sum_{i=1}^n E(X_i)$$

since all X_i are iid samples from distribution X ,

$$E(X_i) = E(X).$$

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i)$$

$$E(\bar{X}_n) = \frac{1}{n} \cdot n E(X) = E(X).$$

Hence, the statement is true.

(c) False.

Since X_1, X_2, \dots, X_n are iid samples from distribution X ,

$E(X_i) = \mu$ & $\text{Var}(X_i) = \sigma^2$ and $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, we can apply the Law of

Large Numbers to state that for any $\epsilon > 0$,

$$\text{Var}(\bar{X}_n) = \text{Var}(X) \quad P(|\bar{X}_n - \mu| > \epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

This theorem does not state that $\bar{X}_n = \mu$, but rather that

\bar{X}_n converges in probability to μ , meaning that it will be close to μ , with the difference in magnitudes between \bar{X}_n & μ approaching probability zero for a non-zero positive difference.

Since $E(X) = \mu$, this statement is false.

(d) True. This is the correct interpretation of the Law of Large Numbers, which states that as $n \rightarrow \infty$, \bar{X}_n converges in probability to μ for X_1, \dots, X_n sequence of iid random variables,

and $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. \bar{X}_n will be very close to $E(X)$ for a

sufficiently large n .

(3) Since X_1 & X_2 are iid samples from population X , the joint distribution can be written as:

$$f(X_1, X_2 | \theta) = \prod_{i=1}^2 f(X_i | \theta) = f(X_1 | \theta) \cdot f(X_2 | \theta)$$

whole $f(x_i | \theta)$ represents the probability distribution in the question.
this can be summarized in the following table:

$$f(x_1, x_2 | \theta) = \begin{array}{cccc} x_1 = 1 & x_1 = 2 & x_1 = 3 & x_1 = 4 \end{array}$$

$$\begin{array}{ccccc} x_2 = 1 & 4\theta^2 & 2\theta^2 & 6\theta^2 & 2\theta - 12\theta^2 \\ x_2 = 2 & 2\theta^2 & \theta^2 & 3\theta^2 & \theta - 6\theta^2 \\ x_2 = 3 & 6\theta^2 & 3\theta^2 & 9\theta^2 & 3\theta - 18\theta^2 \end{array}$$

$$x_2 = 4 \quad 2\theta - 12\theta^2 \quad \theta - 6\theta^2 \quad 3\theta - 18\theta^2 \quad 1 - 12\theta + 36\theta^2$$

$$\frac{x_1 + x_2}{2} \geq 3 \Rightarrow x_1 + x_2 \geq 6$$

$$P(x_1 = 2, x_2 = 4) + P(x_1 = 4, x_2 = 2) + \\ P(x_1 + x_2 \geq 6) = P(x_1 = 3, x_2 = 3) + P(x_1 = 3, x_2 = 4)$$

$$+ P(x_1 = 4, x_2 = 3) + P(x_1 = 4, x_2 = 4)$$

$$= [\theta - 6\theta^2] + [\theta - 6\theta^2] + [9\theta^2] + [3\theta - 18\theta^2] + [9\theta^2] +$$

$$[1 - 12\theta + 36\theta^2] = 1 - 7\theta + 24\theta^2$$

~~$\theta - 6\theta^2 + \theta - 6\theta^2 + 9\theta^2 + 3\theta - 18\theta^2 + 9\theta^2 + 1 - 12\theta + 36\theta^2$~~

$$P\left(\frac{x_1 + x_2}{2} \geq 3\right) = 1 - 7\theta + 24\theta^2$$

see page

after for ~~for~~ yes, I can. I construct method of moments estimator:

4.4

~~$E(x) =$ Yes, I can. The maximum likelihood~~

~~MLE~~

~~estimator for θ would be unbiased.~~

~~(If this is out of scope, then no because the method of moments estimator is biased.)~~

$$5) \text{ mse}_c = \frac{1}{N} \sum_{i=1}^N (x_i^2 - 2Cx_i + C^2) \quad \text{as } \frac{1}{N} Nc^2 = C^2$$

proof for 5 continues on page 8

$$\text{mse}_c = \frac{1}{N} \sum_{i=1}^N x_i^2 - \frac{1}{N} \sum_{i=1}^N 2Cx_i + \frac{1}{N} \sum_{i=1}^N C^2$$

Problem 4-(4) Let $\hat{\theta}_n$ be the estimator for θ .

By definition, bias($\hat{\theta}_n$) = $E[\hat{\theta}_n] - \theta$

$$\text{WTS } E[\hat{\theta}_n] - \theta = 0$$

Proof: X is the random variable with distribution in question

$$E(X) = 1(2\theta) + 2(\theta) + 3(3\theta) + 4(1-6\theta)$$

$$E(X) = 4 - 11\theta$$

$$E(X) - 4 = -11\theta$$

$$4 - E(X) = 11\theta \quad \theta = \frac{4 - E(X)}{11}$$

$$E(\theta) = E\left(\frac{4 - E(X)}{11}\right) = \frac{4}{11} - \frac{1}{11}E(E(X))$$

$$E(\theta) = \underbrace{4 - \frac{1}{11} \sum_{i=1}^n x_i}_{\text{from definition}} = \frac{4 - E(X)}{11}$$

$$E(\theta) = \frac{4 - \bar{x}_n}{11} \quad (\text{in question.})$$

$$\theta = \frac{4 - \bar{x}_n}{11}$$

$$E(\hat{\theta}_n) = \underbrace{4 - \bar{x}_n}_{\text{first moment sample of estimator}} = \frac{4 - \bar{x}_n}{11}$$

first moment
sample
of estimator

$$\text{Hence, } E(\hat{\theta}_n) - \theta = \frac{4 - \bar{x}_n}{11} - \left(\frac{4 - \bar{x}_n}{11}\right) = 0$$

Thus, we can construct an unbiased estimator of θ
in $\hat{\theta}_n$.

differentiate w.r.t. c :

$$mse_c(c) = -\frac{2}{N} \sum_{i=1}^N x_i + 2c$$

set equal to zero:

$$0 = -\frac{2}{N} \sum_{i=1}^N x_i + 2c$$

$$\frac{2}{N} \sum_{i=1}^N x_i = 2c$$

$$\frac{1}{N} \sum_{i=1}^N x_i = c \Rightarrow m = c.$$

\uparrow
this is the mean m

Hence the value of c that minimizes mse_c is $c = m$.

$$mse_m = \frac{1}{N} \sum_{i=1}^N (x_i - m)^2 = \sigma^2$$

\uparrow
this is the definition of population variance.

$$\text{more formally: } \frac{1}{N} \sum_{i=1}^N x_i^2 + m^2 - 2m x_i =$$

$$\frac{1}{N} \sum_{i=1}^N x_i^2 + \frac{1}{N} N m^2 - 2m \cdot \frac{1}{N} \sum_{i=1}^N x_i$$

$$= E(x^2) + m^2 - 2m^2 = E(x^2) - m^2 = \\ E(x^2) - (E(x))^2 = \sigma^2$$

\uparrow
by definition.

6] (i) want to show: $E(\hat{\sigma}_n^2) = \sigma^2$.

$$E(\hat{\sigma}_n^2) = E\left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right)$$

$$\begin{aligned}
 E(\hat{\sigma}_n^2) &= \frac{1}{n-1} E \left(\sum_{i=1}^n (x_i - \bar{x}_n)^2 \right) = \frac{1}{n-1} \left[\sum_{i=1}^n E(x_i^2) - nE(x_i) + n\bar{x}_n^2 \right] \\
 &= \frac{1}{n-1} \left[\sum_{i=1}^n (x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2) \right] = \frac{1}{n-1} \left[\sum_{i=1}^n (x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2) \right] \\
 &= \frac{1}{n-1} \left[\sum_{i=1}^n (x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2) \right] = \frac{1}{n-1} \left[\sum_{i=1}^n (x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2) \right]
 \end{aligned}$$

$$E(\hat{\sigma}_n^2) = \frac{1}{n-1} E \left(\sum_{i=1}^n (x_i - \bar{x}_n)^2 \right) = \frac{1}{n-1} E \left(\sum_{i=1}^n (x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2) \right)$$

$$= \frac{1}{n-1} E \left(\sum_{i=1}^n x_i^2 - 2\bar{x}_n \sum_{i=1}^n x_i + n(\bar{x}_n)^2 \right)$$

using $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$

$$= \frac{1}{n-1} E \left(\sum_{i=1}^n x_i^2 - 2\bar{x}_n(n\bar{x}_n) + n\bar{x}_n^2 \right)$$

$$= \frac{1}{n-1} E \left(\sum_{i=1}^n x_i^2 - n\bar{x}_n^2 \right)$$

$$\text{Var}(x_i) = E(x_i^2) - [E(x_i)]^2$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^n E(x_i^2) - n E(\bar{x}_n^2) \right]$$

$$E(\bar{x}_n^2) = \text{Var}(\bar{x}_n) +$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^n [\text{Var}(x_i) + (E(x_i))^2] - n E(\bar{x}_n^2) \right]$$

$$[E(\bar{x}_n)]^2$$

$$= \frac{1}{n-1} \left[n[\sigma^2 + \mu^2] - n \left[\frac{\sigma^2}{n} + \mu^2 \right] \right]$$

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

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

Hence, $\hat{\sigma}_n^2$ is an unbiased estimator of σ^2 .

Alternate proof for problem 6, part (2)

$$\text{Var}(\hat{\sigma}_n) = E(\hat{\sigma}_n^2) - [E(\hat{\sigma}_n)]^2$$

$$\hat{\sigma}_n = \sqrt{E\left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right)} = \sqrt{E\left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right)}$$

since this term will always
be non-zero and positive

because variance is non zero
and positive, this will make

$\hat{\sigma}_n$ smaller than σ .
 $\hat{\sigma}_n$ as constructed here is a
biased estimator of σ and
tends to underestimate σ .

$$(2) \hat{\sigma}_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2} = \sqrt{\hat{\sigma}_n^2}$$

No, $\hat{\sigma}_n$ is not an unbiased estimator of σ , and tends to underestimate σ .

PROOF :

Jensen's Inequality states that for a concave function $f(x)$ and for $E[f(x)]$ and $f(E[x])$ being finite,

$$E[f(x)] \leq f(E(x))$$

Because the square root function is strictly concave, this becomes $E[\sqrt{x}] < \sqrt{E[x]}$,

Let $f(x)$ be the square root function, and let x be $\hat{\sigma}_n^2$.

Then, $E[\sqrt{\hat{\sigma}_n^2}] < \sqrt{E[\hat{\sigma}_n^2]} = \sigma$.

$$\Rightarrow E[\hat{\sigma}_n] < \sigma.$$

Thus, $\hat{\sigma}_n$ is a biased estimator of σ , and tends to underestimate.

$$7)(i) \text{ WTS } E[\hat{M}_{m+n}] = M$$

$$E[\hat{M}_{m+n}] = E[\alpha \bar{x}_n + \beta \bar{y}_m] = \alpha E[\bar{x}_n] + \beta E[\bar{y}_m]$$

$$= \alpha \frac{1}{n} \sum_{i=1}^n E[x_i] + \beta \frac{1}{m} \sum_{i=1}^m E[y_i]$$

$$= \alpha \frac{1}{n} \cdot nM + \beta \frac{1}{m} \cdot mM$$

$$E[\hat{M}_{m+n}] = \alpha M + \beta M$$

$$E[M] = M / (\alpha + \beta)$$

condition for unbiased estimate: $\alpha + \beta = 1$.

$$(2) \text{Var}(\hat{\mu}_{m+n}) = \text{Var}(\alpha \bar{x}_n + \beta \bar{y}_m)$$

because \bar{x}_n & \bar{y}_m
are independent

$$= \alpha^2 \text{Var}(\bar{x}_n) + \beta^2 \text{Var}(\bar{y}_m)$$

$$= \alpha^2 \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) + \beta^2 \text{Var}\left(\frac{1}{m} \sum_{i=1}^m y_i\right)$$

$$\text{Var}(\hat{\mu}_{m+n}) = \frac{\alpha^2}{n^2} \text{Var}x^2 + \frac{\beta^2}{m^2} \text{Var}y^2 \quad \leftarrow \text{because iid}$$

$$= \frac{\alpha^2}{n} \text{Var}x^2 + \frac{\beta^2}{m} \text{Var}y^2$$

$$\text{s.t. } \alpha = 1 - \beta \quad (\Rightarrow \beta = 1 - \alpha)$$

take derivative wrt α & set to 0. [$\beta = 1 - \alpha$]

$$\frac{\partial}{\partial \alpha} \left(\frac{\alpha^2}{n} \text{Var}x^2 + \frac{(1-\alpha)^2}{m} \text{Var}y^2 \right) = 0$$

$$\frac{\partial}{\partial \alpha} \left(\frac{\alpha^2}{n} \text{Var}x^2 + \frac{\alpha^2}{m} \text{Var}y^2 - 2\alpha \frac{\text{Var}x^2}{n} + \alpha^2 \frac{\text{Var}y^2}{m} \right) = 0$$

$$2\alpha \frac{\text{Var}x^2}{n} - 2\alpha \frac{\text{Var}y^2}{m} + 2\alpha \frac{\text{Var}y^2}{m} = 0$$

$$\alpha \frac{1}{n} - 2 \frac{1}{m} + 2\alpha \frac{1}{m} = 0$$

$$\alpha \left(\frac{1}{n} + 2 \frac{1}{m} \right) = \frac{2}{m}$$

$$\frac{\alpha - \frac{2}{m}}{\frac{1}{n} + \frac{2}{m}} \quad \beta = 1 - \frac{\frac{2}{m}}{\frac{1}{n} + \frac{2}{m}}$$

8] (1) I invoke Central Limit Theorem, as in lecture 2. It's stated that regardless of true $f(x)$, so long as X_1, X_2, \dots, X_n are iid,

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) \rightarrow \Phi(z) \text{ as } n \rightarrow \infty$$

and

$$P(|\bar{X}_n - \mu| \leq \delta) \approx 2\Phi\left(\frac{\sqrt{n}\delta}{\sigma}\right) = 1$$

Since this is one sample from same population, iid assumption holds.

$$\text{Var}(\bar{p}_n) = \sigma^2 = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot p \cdot (1-p) \cdot n \quad n = 200$$

$$\sigma^2 = \frac{1}{200} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{800}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{800}}$$

$$P(|\bar{X}_n - \mu| \leq \delta) = 1 - P(|\bar{X}_n - \mu| \geq \delta)$$

$$P(|\bar{X}_n - \mu| \geq \delta) = 1 - \left[2\Phi\left(\frac{\sqrt{200}\delta}{\sqrt{800}}\right) - 1 \right] = 0.025$$

$$2 - 2\Phi\left(\frac{\sqrt{200}\delta}{\sqrt{800}}\right) = 0.025$$

$$-2\Phi\left(\frac{\sqrt{200}\delta}{\sqrt{800}}\right) = -0.975$$

$$\Phi\left(\frac{\sqrt{200}\delta}{\sqrt{800}}\right) = 0.9875$$

$$\Phi^{-1}(0.9875) = \frac{\sqrt{200}\delta}{\sqrt{800}}$$

$$2.24 = \frac{\sqrt{200}\delta}{\sqrt{800}}$$

$$\delta = 75.6 \times 10^{-3}$$

$$(2) \hat{P}_n = 0.25.$$

As we found in the last question,

$$P(|\hat{P}_n - P| \geq \delta) = 0.025$$

$$\text{when } \delta = 5.6 \times 10^{-3},$$

This is equivalent to a 97.5% confidence interval

With an upper bound of $\hat{P}_n + (5.6 \times 10^{-3})$ and a lower bound of

$$\hat{P}_n - (5.6 \times 10^{-3}).$$

If $\hat{P}_n = 0.25$, and we use a 95% confidence interval, which by definition is a tighter interval than a 97.5% confidence interval, than the 95% confidence interval for p will not contain the true value of p .

[The 95% interval would be $[0.25 - \delta, 0.25 + \delta]$,

$$\delta = 5.6 \times 10^{-3}].$$

9] 95% bootstrap confidence interval for P , to 500 bootstraps.

(1) 280 heads observed.

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (\bar{x}_i - \bar{\bar{x}}_n)^2 = \frac{1}{n} \sum_{i=1}^n \bar{x}_i^2 - 2\bar{x}_i \bar{\bar{x}}_n + \bar{\bar{x}}_n^2 = \frac{1}{n} \sum_{i=1}^n \bar{x}_i^2 - \frac{2\bar{\bar{x}}_n}{n} \sum_{i=1}^n \bar{x}_i$$

$$+ \bar{\bar{x}}_n^2 = \frac{1}{n} \sum_{i=1}^n \bar{x}_i^2 - \bar{\bar{x}}_n^2 = \frac{1}{n} \sum_{i=1}^n \bar{x}_i - \bar{\bar{x}}_n^2 = 0.7 - (0.7)^2$$

$$\hat{\sigma}_n^2 = 0.21$$

(because squaring doesn't change values here.)

$$\hat{P}_n \approx 0.458$$

bootstrap interval :

$$n=400 \quad \bar{x}_n \pm \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}} \Rightarrow 0.7 \pm \frac{1.96(0.458)}{20}$$

$$= [0.655116, 0.744884]$$

Conservative estimate uses population variance:

(instead of sample

variance.)

By lecture 1, population variance for Bernoulli random variables
(which flipping coins in this problem is)

$$\begin{aligned}\sigma^2 &= \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2 = \frac{N-n}{N} (0-p)^2 + \frac{n}{N} (1-p)^2 \\ &= (1-p)p^2 + p(1-p)^2 \\ \sigma^2 &= p(1-p)\end{aligned}$$

Assuming a fair coin, $p=0.5$ $\sigma^2 = 0.25$
 $\sigma = 0.5$

conservative interval:

$$\begin{aligned}\bar{x}_n \pm \frac{z_{\alpha/2} \sigma}{\sqrt{n}} &\Rightarrow 0.7 \pm \frac{1.96(0.5)}{\sqrt{20}} \\ &= [0.651, 0.749]\end{aligned}$$

The bootstrap interval is tighter than the conservative interval for p in this case. This is because the bootstrap sample variance was smaller than the population variance.

(2) From (1), I proved that $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \hat{x}_i^2 - \bar{\hat{x}}_n^2$.

Now, 150 heads are observed.

$$\hat{\sigma}_n^2 = 0.375 - (0.375)^2$$

$$\hat{\sigma}_n^2 = 0.234375$$

$$\hat{\sigma}_n \approx 0.484$$

bootstrap interval: $\bar{x}_n \pm \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}} \Rightarrow 0.375 \pm \frac{1.96(0.484)}{\sqrt{20}}$

$$= [0.327568, 0.422432] \quad (\text{bootstrap estimate})$$

conservative estimate:

population variance is still $\sigma^2 = 0.25$, hence $\sigma = 0.5$

conservative interval:

$$\bar{x}_n \pm \frac{z_{\alpha/2} \sigma}{\sqrt{n}} \Rightarrow 0.375 \pm \frac{1.96(0.5)}{\sqrt{20}}$$

$$= [0.326, 0.424],$$

The bootstrap interval is tighter than the conservative interval for P , again due to the fact that sample variance was smaller than population variance.

10]

$$(1) E(T) = \int_t^\infty t f(t) dt$$

$$= \int_0^\infty t \frac{1}{\sqrt{2\pi t}} e^{-\frac{t}{2}} dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^\infty \sqrt{t} e^{-\frac{t}{2}} dt \quad u = -\frac{t}{2}$$

$$du = -\frac{1}{2} dt$$

$$= -2^{\frac{3}{2}} \int_0^\infty e^u \sqrt{-u} du \quad dt = -2du$$

$$= -2 \int_0^\infty v^2 e^{-v^2} dv \quad V = \sqrt{-u}$$

$$dv = \frac{1}{2\sqrt{-u}} du$$

$$f = v \quad g' = ve^{-v^2}$$

$$f' = 1 \quad g = -\frac{e^{-v^2}}{2} \quad (\text{by parts})$$

$$= -\frac{ve^{-v^2}}{2} - \int -\frac{e^{-v^2}}{2} dv$$

$$= -\frac{ve^{-v^2}}{2} - \frac{\sqrt{n}}{4} \int \frac{2e^{-v^2}}{\sqrt{n}} dv$$

$$= -\frac{ve^{-v^2}}{2} + \frac{\sqrt{n}}{4} \operatorname{erf}(v)$$

$$= \overline{\int_u^\infty e^u - \frac{\sqrt{\pi}}{2} \operatorname{erf}(\sqrt{-u})}$$

$$= \sqrt{2} \sqrt{\pi} \operatorname{erf}(\sqrt{-u}) - 2^{\frac{3}{2}} \sqrt{-u} e^u$$

$$= \sqrt{2} \sqrt{\pi} \operatorname{erf}\left(\frac{\sqrt{t}}{\sqrt{2}}\right) - 2\sqrt{t} e^{-\frac{t}{2}}$$

$$= \frac{1}{\sqrt{\pi}} \left(\sqrt{\pi} \operatorname{erf}\left(\frac{\sqrt{t}}{\sqrt{2}}\right) - 2\sqrt{t} e^{-\frac{t}{2}} \right) = \operatorname{erf}\left(\frac{\sqrt{t}}{\sqrt{2}}\right) - \frac{\sqrt{t} e^{-\frac{t}{2}}}{\sqrt{\pi}}$$

$$E(T) = 1$$

$$\text{Var}(T) = \int_0^{\infty} (t-1)^2 f(t) dt$$

$$= \int_0^{\infty} (t-1)^2 f(t) dt = \int_0^{\infty} f(t) (t^2 - 2t + 1) dt$$

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{t}{2}} (t^2 - 2t + 1) dt \quad \leftarrow \text{Used an integral calculator because this would be too hard to do by hand.}\right.$$

$$= \left[2\text{erf}\left(\frac{\sqrt{t}}{\sqrt{2}}\right) - \frac{\sqrt{2}e^{-(t+1)}e^{-\frac{t}{2}}}{\sqrt{\pi}} \right]_0^{\infty} = 2$$

$$\text{Var}(T) = 2.$$

(2) By definition, since T_n is a chi-square random variable, with n degrees of freedom,

$$E(T_n) = n \quad \leftarrow \text{from the probability review}$$

$$\text{Var}(T_n) = 2n \quad \leftarrow \text{document; these are properties of chi-square random variables.}$$

$$(3) T_n = z_1^2 + z_2^2 + \dots + z_n^2$$

where z_1, z_2, \dots, z_n are iid

standard normal RVs.

Let $Y_i = z_i^2$. By definition given

in problem statement, Y is a χ^2 RV which is a Gamma($\frac{1}{2}, \frac{1}{2}$) RV.

Hence, $T_n = \sum_{i=1}^n Y_i$ where $Y_i \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$

Thus, $T_n \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$ since there are n Y_i random variables that are iid and by the hint in question that adding independent gammas sums up the α 's and keeps β constant.

(4) From previous (2) & (3), we know that

- for a χ_n^2 random variable $Z_i \sim \text{Gamma}(\frac{n}{2}, \frac{1}{2})$
- if $Y_1 \sim \text{Gamma}(\alpha_1, \beta)$, $Y_2 \sim \text{Gamma}(\alpha_2, \beta)$ and $Y_1 + Y_2$ are independent, $Y_1 + Y_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$.

$$\text{Hence, } T_n \sim \chi_n^2 \Rightarrow T_n \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$S_m \sim \chi_m^2 \Rightarrow S_m \sim \text{Gamma}\left(\frac{m}{2}, \frac{1}{2}\right)$$

$T_n + S_m \sim \text{Gamma}\left(\frac{n+m}{2}, \frac{1}{2}\right)$ since T_n & S_m are independent
which is the same distribution as

$$\chi_{n+m}^2. \quad \text{Hence, } T_n + S_m \sim \chi_{n+m}^2.$$

(5) since T_n is a sum of squared iid standard normal random variables, we can apply CLT.

$$T_n = Z_1^2 + Z_2^2 + \dots + Z_n^2 \quad \text{where } E(Z_i^2) = 1 \quad \text{from (1),}$$
$$\text{Var}(Z_i^2) = 2 \quad \text{as } n \rightarrow \infty$$

$$\text{as } n \rightarrow \infty, \quad P\left(\frac{\bar{T}_n - n}{\sqrt{n}/\sqrt{n}} \leq z\right) \xrightarrow{d} \Phi(z) \quad \text{where } \Phi(\cdot) \text{ is the CDF for a standard normal distribution.}$$

$$\text{CLT: } P\left(\frac{\bar{T}_n - n}{\sigma/\sqrt{n}} \leq z\right) \xrightarrow{d} \Phi(z) \text{ as } n \rightarrow \infty$$

Hence, a good normal approximation distribution for T_n is $N(0, 1)$.