Vinay Maturi
SID: 3031783444

STAT135: Homework 1

1] $x_1, x_2, \ldots, x_N$ are list of numbers with mean μ and standard deviation σ. $a_1, \ldots, a_M$ are distinct values of list with frequencies $n_1, \ldots, n_M$.

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

Let X = list of numbers containing $x_1, x_2, \ldots, x_N$

$E[X] = \mu$

definition 2: (discrete case) $E[X] = \frac{1}{N}\sum_{i=1}^{N} x_i = \sum_{j=1}^{M} \frac{n_j}{N} \cdot a_j$

where x_i are the elements of the list, a_j are the distinct values of the list, and n_j are the frequencies of the unique values.

N = # of elements in list.

M = ~~# of unique elements~~ in list.

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

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 i.i.d. 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

(due to locational differences)

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.

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 $\left[\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2\right]$ and the definition of sample variance $\left[\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \text{ or } \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right]$ requires knowledge of the population size or sample size to compute properly.

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

$\mu = 80$, $SD = 10$ (section 1); $\mu = 87$, $SD = 10$ (section 2)

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.

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

$$Var(X) = E[X^2] - [E(X)]^2 \quad (1)$$

μ_X = mean of combined group

$$Var(X) = \frac{1}{50} \sum_{i=1}^{50} x_i^2 - (\mu_X)^2$$

$$Var(X) = \frac{1}{50} \left[\sum_{i=1}^{30} x_i^2 + \sum_{i=31}^{50} x_i^2 \right] - (\mu_X)^2$$

re-arranging (1): $E(X^2) = Var(X) + [E(X)]^2$

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2 = \sigma^2 + \mu^2$$

$$\sum_{i=1}^{n} x_i^2 = n\sigma^2 + n\mu^2$$

sub in for both summations

$$Var(X) = \frac{1}{50} \left[30\sigma_{s_1}^2 + 30\mu_{s_1}^2 + 20\sigma_{s_2}^2 + 20\mu_{s_2}^2 \right] - (\mu_X)^2$$

$\sigma_{s_1}^2 = 100$, $\sigma_{s_2}^2 = 100$, $\mu_{s_1}^2 = 6400$, $\mu_{s_2}^2 = 7569$

$$\mu_X = \frac{30(80) + 20(87)}{50} = 82.8$$

plug in:

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

$$Var(X) = 111.76$$

$$SD(X) = \sqrt{Var(X)} = 10.5716602$$

$$SD(X) \approx 10.572$$

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 X_i is just a value drawn w/ some probability from the distribution.]

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

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

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

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

Hence, $2\frac{1}{6} < E(X) < 4$ and thus $\bar{X}_n \neq E(X)$.

($2\frac{1}{6}$ at $\theta = \frac{1}{6}$; 4 at $\theta = 0$)

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

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

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

Hence, the statement is true.

(c) False.

Since $X_1, X_2, \ldots, X_n$ are iid samples from distribution X, 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 $\varepsilon > 0$,

$E(X_i) = \mu$ & $Var(X_i) = \sigma^2$

$E(X_i) = E(X)$

$Var(X_i) = Var(X)$

$$P(|\bar{X}_n - \mu| > \varepsilon) \to 0 \text{ as } n \to \infty.$$

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 $n \to \infty$, $\bar{X}_n$ converges in probability to μ for $X_1, \ldots, X_n$ 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:

$$f(x_1, x_2 | \theta) = \prod_{i=1}^{2} f(x_i|\theta) = f(x_1|\theta) \cdot f(x_2|\theta)$$

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

$f(x_1, x_2 \mid \theta) =$

	$x_1 = 1$	$x_1 = 2$	$x_1 = 3$	$x_1 = 4$
$x_2 = 1$	$4\theta^2$	$2\theta^2$	$6\theta^2$	$2\theta - 12\theta^2$
$x_2 = 2$	$2\theta^2$	θ^2	$3\theta^2$	$\theta - 6\theta^2$
$x_2 = 3$	$6\theta^2$	$3\theta^2$	$9\theta^2$	$3\theta - 18\theta^2$
$x_2 = 4$	$2\theta - 12\theta^2$	$\theta - 6\theta^2$	$3\theta - 18\theta^2$	$1 - 12\theta + 36\theta^2$

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

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

$$= [\theta - 6\theta^2] + [\theta - 6\theta^2] + [9\theta^2] + [3\theta - 18\theta^2] + [9\theta^2] + [1 - 12\theta + 36\theta^2] = 1 - 7\theta + 24\theta^2$$

$P\left(\frac{X_1 + X_2}{2} \geq 3\right) = 1 - 7\theta + 24\theta^2$

(4) ~~Yes, I can. I construct method of moments estimator;~~

~~$E(X) =$~~ ~~Yes, I can. The maximum likelihood estimator for θ would be unbiased.~~

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

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

$\frac{1}{N} \cdot Nc^2 = c^2$

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

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

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

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

Proof: X is the random variable with distribution in question

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

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

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

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

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

← by law of iterated expectation

$$E(\theta) = \frac{4 - \frac{1}{n}\sum_{i=1}^{n} X_i}{11} = \frac{4 - E(X)}{11}$$

(from definition in question.)

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

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

$$E(\hat{\theta}_n) = \frac{4 - \bar{X}_n}{11}$$

sample ~~first moment~~ of estimator

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:

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

set equal to zero:

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

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

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

this is the mean μ

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

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

~~this is the definition of population variance.~~

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

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

$$= E[x^2] + \mu^2 - 2\mu^2 = E[x^2] - \mu^2 =$$

$$E[x^2] - (E(x))^2 = \sigma^2$$

by definition.

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

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

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

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

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

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

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

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

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

$E(\bar{X}_n^2) = Var(\bar{X}_n) + [E(\bar{X}_n)]^2 = \frac{\sigma^2}{n} + \mu^2$

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

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

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

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

Alternate proof for Problem 6, part (2)

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

$$\hat{\sigma}_n = \sqrt{E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right) - \left[E\left(\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}\right)\right]^2}$$

~~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 $E[f(x)] < f(E(X))$.

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

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

Cauchy-Schwarz potentially,

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

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

---

7) (1) WTS $E[\hat{\mu}_{m+n}] = \mu$

$E[\hat{\mu}_{m+n}] = E[\alpha\bar{X}_n + \beta\bar{Y}_m] = \alpha E[\bar{X}_n] + \beta E[\bar{Y}_m]$

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

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

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

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

(2) $Var(\hat{\mu}_{m+n}) = Var(\alpha \bar{X}_n + \beta \bar{Y}_m)$ because $\bar{X}_n$ & $\bar{Y}_m$ are independent

$= \alpha^2 Var(\bar{X}_n) + \beta^2 Var(\bar{Y}_m)$

$= \alpha^2 Var(\frac{1}{n}\sum_{i=1}^{n} X_i) + \beta^2 Var(\frac{1}{m}\sum_{i=1}^{m} Y_i)$

$Var(\hat{\mu}_{m+n}) = \frac{\alpha^2}{n^2} n\sigma^2 + \frac{\beta^2}{m^2} m\sigma^2$ ← because iid $= \frac{\alpha^2}{n}\sigma^2 + \frac{\beta^2}{m}\sigma^2$

s.t. $\alpha = 1-\beta \Leftrightarrow \beta = 1-\alpha$

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

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

$$\frac{\partial}{\partial\alpha}\left(\frac{\sigma^2}{n}\alpha^2 + \frac{\sigma^2}{m} - 2\alpha\frac{\sigma^2}{m} + \alpha^2\frac{\sigma^2}{m}\right) = 0$$

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

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

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

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

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

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

and

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

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

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

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

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

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

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

$$2 - 2\Phi\left(\frac{\sqrt{200}\,\delta}{\frac{1}{\sqrt{800}}}\right) = 0.025$$

$$-2\Phi\left(\frac{\sqrt{200}\,\delta}{\frac{1}{\sqrt{800}}}\right) = -1.975$$

$$\Phi\left(\frac{\sqrt{200}\,\delta}{\frac{1}{\sqrt{800}}}\right) = 0.9875$$

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

$$2.24 = \frac{\sqrt{200}\,\delta}{\frac{1}{\sqrt{800}}}$$

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

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

As we found in the last question,

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

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, then 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, tossed 400 times.

(1) 280 heads observed.

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

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

(because squaring doesn't change values here.)

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

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

bootstrap interval:

$n = 400$

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

$$= [0.655116, 0.744884]$$

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)

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

$$= (1-p)p^2 + p(1-p)^2$$

$$\sigma^2 = p(1-p)$$

Assuming a fair coin: $p = 0.5$

$$\sigma^2 = 0.25$$

$$\sigma = 0.5$$

conservative interval:

$$\bar{X}_n \pm \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \Rightarrow 0.7 \pm \frac{1.96(0.5)}{20}$$

$$= [0.651, 0.749]$$

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 $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i - \bar{X}_n^2$.

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

$= [0.327568, 0.422432]$ (bootstrap estimate)

conservative estimate:

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

conservative interval:

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

$$= [0.326, 0.424].$$

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

[redacted]

10]

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

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

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

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

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

$$= -2 \int_0^\infty v^2 e^{-v^2}\, dv \qquad v = \sqrt{-u}$$

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

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

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

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

$$= \frac{-v e^{-v^2}}{2} - \left(\frac{\sqrt{\pi}}{4} \int -\frac{2 e^{-v^2}}{\sqrt{\pi}}\, dv \right.$$

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

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

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

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

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

$E(T) = 1$

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

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

$= \int_0^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{t}{2}} (t^2-2t+1)\,dt$ ← I used an integral calculator because this would be tedious to write out.

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

$Var(T) = 2.$

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

$E(T_n) = n$

$Var(T_n) = 2n$

← from the probability review 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$\left(\frac{1}{2}, \frac{1}{2}\right)$ RV.

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

Thus, $T_n \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$ 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, $z \sim Gamma(\frac{n}{2}, \frac{1}{2})$
- if $Y_1 \sim Gamma(\alpha_1, \beta)$, $Y_2 \sim Gamma(\alpha_2, \beta)$ and Y_1 & Y_2 are independent, $Y_1 + Y_2 \sim Gamma(\alpha_1 + \alpha_2, \beta)$.

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

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

$T_n + S_m \sim Gamma(\frac{n+m}{2}, \frac{1}{2})$ since T_n & S_m are independent

which is the same distribution as χ_{n+m}^2. 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.

$T_n = Z_1^2 + Z_2^2 + \dots + Z_n^2$ where $E(Z_i^2) = 1$ from (1).
$Var(Z_i^2) = 2 < \infty$

as $n \to \infty$, $P\left(\frac{T_n - n}{\sqrt{2n}/\sqrt{n}} \leq z\right) \xrightarrow{d} \Phi(z)$ where $\Phi(\cdot)$ is the CDF for a standard normal distribution.

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

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