

Probability & Statistics

Directions: There are 4 questions
For credit, show the details of your work and/or send me your R code as well as

You are allowed to use R and/or RStudio

Problem 1:

Let $w = \langle w_1, w_2, \dots, w_n \rangle$ be a vector of fixed numbers (weights). For a sample $X = \langle X_1, X_2, \dots, X_n \rangle$, let the weighted sum be defined by

$$\bar{X}_w = \sum_{i=1}^n w_i X_i,$$

and assume that X comes from iid sampling.

- (1) Determine conditions on w that make \bar{X}_w an unbiased estimator of μ , where μ is the population mean. Your answer will be in terms of the weights w_i . (Note that under these conditions \bar{X}_w is called a **weighted mean** or **weighted average**.)

- (2) What is the variance of a weighted average in terms of the weights w ? Your answer will involve the weights w_i .

Problem 2:

Under iid random sampling, the sample variance is given by

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

We can show that the sample variance as defined above is *biased*; i.e., that $E(\hat{\sigma}^2) \neq \sigma^2$. Specifically, we can show that

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

(1) An estimator $\hat{\theta}$ is **asymptotically unbiased** if

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta.$$

Show that $\hat{\sigma}^2$ (from above) is an asymptotically unbiased estimator for σ^2 .

(2) Our usual version of the sample variance is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

which uses $n-1$ in the denominator instead of n . Use the result from the first part to show that S^2 is an unbiased estimator of σ^2 . (Note that this gives one justification for using $n-1$ and not n .)

Problem 3:

[Note: This problem will continue on the next page.]

We say a random variable X has a **chi-square distribution** with n degrees of freedom if its pdf is

$$f_X(x; n) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} x^{n/2-1} e^{-x/2}, \quad x \geq 0$$

and 0 for $x < 0$. We write $X \sim \text{Chisq}(n)$ or $X \sim \chi_n^2$.

We can write $X \sim \chi_n^2$ as the sum of the squares of iid standard normal RVs; i.e., for $Z_i \stackrel{iid}{\sim} \text{Norm}(0, 1)$,

$$X = Z_1^2 + Z_2^2 + \dots + Z_n^2.$$

(This is very useful when modeling the sampling distribution of the variance, which can never be negative and has a right-skewed distribution.)

For $X \sim \chi_n^2$, the MGF, for $t < \frac{1}{2}$, is

$$M_X(t) = (1 - 2t)^{-n/2}.$$

(1) Here we will prove that if $Z \sim \text{Norm}(0, 1)$, then $V = Z^2 \sim \chi_1^2$, and, subsequently, the MGF result from above. Observe,

$$\begin{aligned} M_V(t) = M_{Z^2}(t) &= \boxed{} \\ &= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2(1/2-t)} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2(1-2t)/2} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2 a^2/2} dz \quad (a = \sqrt{1-2t}) \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(az)^2/2} dz \quad (u = az, du = a dz) \\ &= \frac{1}{a} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\ &= \frac{1}{a} \\ &= \frac{1}{\sqrt{1-2t}}. \end{aligned}$$

Then, by independence, for $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$,

$$M_X(t) = \left(\frac{1}{\sqrt{1-2t}} \right)^n = [(1-2t)^{-1/2}]^n = (1-2t)^{-n/2}. \quad \square$$

Your work is to complete/justify two of the steps in the proof above:

(a) What should go in the **red** box at the top?

(b) Justify the penultimate inequality; that is, justify how we get from the **blue** expression to the **green** expression.

[#3 continued]

(2) Find $E(X)$ for $X \sim \chi_n^2$ using the MGF above, which is $M_X(t) = (1-2t)^{-n/2}$.

(3) Find $\text{Var}(X)$ for $X \sim \chi_n^2$ using the MGF above.

(4) The chi-square distribution with n degrees of freedom (χ_n^2) is also a special case of the gamma distribution (when $\alpha = n/2$ and $\lambda = 1/2$). Show that $\text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right) = \chi_n^2$.
(Hint: There is more than one approach to show this.)

Problem 4: [10 points]

Two players—Player A, who goes first; and Player B, who goes second—are playing a game of *Memory*. In *Memory*, all cards have a specific color, but every card begins “face down,” so that only the back of the card is seen (all backs look the same). On a player’s turn, she selects one card and turns it over; then she selects another card and turns this one over. If the colors match, she keeps those cards (called a **set**) and gets another turn. If the colors do not match, she turns them back over (so they are face down again) and the other player goes. Assume that both players have perfect memories, meaning that they always remember which cards were previously turned up and where they are. Also assume that both players are wise, choosing a match if possible.

(So assume that when a player reveals her first card, if its match has already been turned over on a previous turn, she will choose it and get a set. Also assume that if two matching cards are known but face down, the next player will choose them and get a set. Also assume that if the known face-down cards do not contain a match, the next player will choose a different card first.)

Remember: *Get a match — go again!*

This game of *Memory* is a bit different. There are only 5 cards: 2 cards are **blue**, and 2 cards are **green**, and 1 card is **red**. Thus the **red** card has no match. If you are playing this game, is it better to go first or second? Answer by finding the following probabilities:

- the first player (Player A) gets two sets and wins;
- the second player (Player B) gets two sets and wins;
- each player gets one set, which results in a tie.