Math 3070, Applied Statistics

Section 1

October 16, 2019

Lecture Outline, 10/16

Section 5.3

- Random Samples and Statistics
- Sampling Distributions

Random Samples

$X_1, X_2, \ldots, X_n$ are a called a **simple random sample** of size n if

1. The X_i's are independent.
2. Every X_i has the same probability distribution.

Note: These two assumptions are frequently referred to as **independently and identically distributed** (i.i.d.).

For example, measurements of parts from an assembly line or outcomes of sufficently randomized polls could be considered simple random sample. Framework is extermely flexible, can describe most settings, and be used to compute statistical errors.

A **statistic** is an quantity whose value can be calculated from a random sample.

Note: Statistics are also random variables since X_i's are random. Parameters are not.

Statistics

A **statistic** is an quantity whose value can be calculated from a random sample.

Note: Statistics are also random variables since X_i's are random. Parameters are not.

Examples:

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

Statistics

A **statistic** is an quantity whose value can be calculated from a random sample.

Note: Statistics are also random variables since X_i's are random. Parameters are not.

Examples:

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

Both $\overline{X}$ and S^2 functions of X_i so they are random. But, they estimate $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$ which are numbers and not random.

Why $n-1$?

Check the expected value of S^2 assuming $\mu = E[X_i]$, $\sigma^2 = Var(X_i)$, and the X_i's are i.i.d.

Why $n-1$?

Check the expected value of S^2 assuming $\mu = E[X_i]$, $\sigma^2 = Var(X_i)$, and the X_i's are i.i.d.

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X_i] = \mu$$

Why $n-1$?

Check the expected value of S^2 assuming $\mu = E[X_i]$, $\sigma^2 = Var(X_i)$, and the X_i's are i.i.d.

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X_i] = \mu$$

auxiliary identity

Why $n-1$?

Check the expected value of S^2 assuming $\mu = E[X_i]$, $\sigma^2 = Var(X_i)$, and the X_i's are i.i.d.

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X_i] = \mu$$

auxiliary identity

$$\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2 = \sum_{i=1}^{n}\left(X_i^2 - \frac{2}{n}X_i\sum_{j=1}^{n} X_j + \left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)^2\right)$$

Why $n-1$?

Check the expected value of S^2 assuming $\mu = E[X_i]$, $\sigma^2 = Var(X_i)$, and the X_i's are i.i.d.

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X_i] = \mu$$

auxiliary identity

$$\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2 = \sum_{i=1}^{n}\left(X_i^2 - \frac{2}{n}X_i\sum_{j=1}^{n} X_j + \left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)^2\right)$$

$$= \left(\sum_{i=1}^{n} X_i^2\right) - \frac{2}{n}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right) + \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} X_j\right)^2$$

Why $n-1$?

Check the expected value of S^2 assuming $\mu = E[X_i]$, $\sigma^2 = Var(X_i)$, and the X_i's are i.i.d.

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = E[X_i] = \mu$$

auxiliary identity

$$\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2 = \sum_{i=1}^{n}\left(X_i^2 - \frac{2}{n}X_i\sum_{j=1}^{n} X_j + \left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)^2\right)$$

$$= \left(\sum_{i=1}^{n} X_i^2\right) - \frac{2}{n}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right) + \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} X_j\right)^2$$

$$= \left(\sum_{i=1}^{n} X_i^2\right) - \frac{2}{n}\left(\sum_{j=1}^{n} X_j\right)^2 + \frac{1}{n}\left(\sum_{j=1}^{n} X_j\right)^2$$

$$= \left(\sum_{i=1}^{n} X_i^2\right) - \frac{1}{n}\left(\sum_{j=1}^{n} X_j\right)^2$$

Why $n-1$?

use $Var(Z) = E[Z^2] - E[Z]^2 \rightarrow E[Z^2] = Var(Z) + E[Z]^2$

Why $n-1$?

use $Var(Z) = E[Z^2] - E[Z]^2 \rightarrow E[Z^2] = Var(Z) + E[Z]^2$

$$E\left[\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2\right] = E\left[\left(\sum_{i=1}^{n} X_i^2\right) - \frac{1}{n}\left(\sum_{j=1}^{n} X_j\right)^2\right]$$

$$= \sum_{i=1}^{n} E[X_i^2] - \frac{1}{n} E\left[\left(\sum_{j=1}^{n} X_j\right)^2\right]$$

Why $n-1$?

use $Var(Z) = E[Z^2] - E[Z]^2 \rightarrow E[Z^2] = Var(Z) + E[Z]^2$

$$E\left[\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2\right] = E\left[\left(\sum_{i=1}^{n} X_i^2\right) - \frac{1}{n}\left(\sum_{j=1}^{n} X_j\right)^2\right]$$

$$= \sum_{i=1}^{n} E[X_i^2] - \frac{1}{n} E\left[\left(\sum_{j=1}^{n} X_j\right)^2\right]$$

$$= \left[\sum_{i=1}^{n}(\sigma^2 + \mu^2)\right] - \frac{1}{n}\left[Var\left(\sum_{j=1}^{n} X_j\right) + \left(E\left[\sum_{j=1}^{n} X_j\right]\right)^2\right]$$

Why $n-1$?

use $Var(Z) = E[Z^2] - E[Z]^2 \rightarrow E[Z^2] = Var(Z) + E[Z]^2$

$$E\left[\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right] = E\left[\left(\sum_{i=1}^{n}X_i^2\right) - \frac{1}{n}\left(\sum_{j=1}^{n}X_j\right)^2\right]$$

$$= \sum_{i=1}^{n}E[X_i^2] - \frac{1}{n}E\left[\left(\sum_{j=1}^{n}X_j\right)^2\right]$$

$$= \left[\sum_{i=1}^{n}(\sigma^2+\mu^2)\right] - \frac{1}{n}\left[Var\left(\sum_{j=1}^{n}X_j\right) + \left(E\left[\sum_{j=1}^{n}X_j\right]\right)^2\right]$$

$$= n(\sigma^2+\mu^2) - \frac{1}{n}\left[\sum_{j=1}^{n}Var(X_j) + \left(\sum_{j=1}^{n}E[X_j]\right)^2\right]$$

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

Why $n-1$?

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

Why $n-1$?

$$
\begin{aligned}
&= n(\sigma^2+\mu^2) - \frac{1}{n}[n\sigma^2 + (n\mu)^2] \\
&= n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 = (n-1)\sigma^2
\end{aligned}
$$

Why $n-1$?

$$
\begin{aligned}
&= n(\sigma^2+\mu^2) - \frac{1}{n}[n\sigma^2 + (n\mu)^2] \\
&= n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 = (n-1)\sigma^2 \\
E[S^2] &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right] \\
&= \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right] \\
&= \frac{n-1}{n-1}\sigma^2 = \sigma^2
\end{aligned}
$$

Why $n-1$?

$$
\begin{aligned}
&= n(\sigma^2+\mu^2) - \frac{1}{n}[n\sigma^2 + (n\mu)^2] \\
&= n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 = (n-1)\sigma^2 \\
E[S^2] &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right] \\
&= \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right] \\
&= \frac{n-1}{n-1}\sigma^2 = \sigma^2
\end{aligned}
$$

As n becomes very large, dividing by n or $n-1$ is mostly irrelevant.

Sampling Distributions

Sampling distributions are the PMF/PDF of a statistic.

Sampling Distributions, PMF Example

Weights X of randomly selected seeds follow PMF f. Compute the sampling distribution of the sample mean of two seeds weights.

$$f(x) = \begin{cases} 0.25, & x = 1 \\ 0.50, & x = 2 \\ 0.25, & x = 3 \end{cases}$$

Sampling Distributions, PMF Example

Weights X of randomly selected seeds follow PMF f. Compute the sampling distribution of the sample mean of two seeds weights.

$$f(x) = \begin{cases} 0.25, & x = 1 \\ 0.50, & x = 2 \\ 0.25, & x = 3 \end{cases}$$

$$\overline{X} = \frac{X_1 + X_2}{2}$$

Sampling Distributions, PMF Example

Weights X of randomly selected seeds follow PMF f. Compute the sampling distribution of the sample mean of two seeds weights.

$$f(x) = \begin{cases} 0.25, & x = 1 \\ 0.50, & x = 2 \\ 0.25, & x = 3 \end{cases}$$

$$\overline{X} = \frac{X_1 + X_2}{2}$$

$\overline{X}$	1	1.5	2	2.5	3
$f_X(x)$	0.0625	0.25	0.375	0.25	0.0625

Sampling Distributions, PMF Example

Weights X of randomly selected seeds follow PMF f. Compute the sampling distribution of the sample mean of two seeds weights.

$$f(x) = \begin{cases} 0.25, & x = 1 \\ 0.50, & x = 2 \\ 0.25, & x = 3 \end{cases}$$

$$\overline{X} = \frac{X_1 + X_2}{2}$$

$\overline{X}$	1	1.5	2	2.5	3
$f_X(x)$	0.0625	0.25	0.375	0.25	0.0625

$$\begin{aligned} P(\overline{X} = 1) &= P(X_1 = 1 \cap X_2 = 1) \\ &= P(X_1 = 1)P(X_2 = 1) = 0.25^2 = 0.0625 \\ P(\overline{X} = 1.5) &= P(X_1 = 2 \cap X_2 = 1) + P(X_1 = 1 \cap X_2 = 2) \\ &= 0.5(0.25) + 0.25(0.5) = 0.25 \end{aligned}$$

Sampling Distributions, PMF Example (Slick Approach)

Consider U_i to be indepdent random variables, $Bern(0.5)$.

Consider U_i to be indepdent random variables, $Bern(0.5)$.
Define $Y_i = U_{i,1} + U_{i,2}$. Notice that

$$f_Y(y) = \begin{cases} 0.25, & y = 0 \\ 0.50, & y = 1 \\ 0.25, & y = 2 \end{cases}$$

Sampling Distributions, PMF Example (Slick Approach)

Consider U_i to be indepdent random variables, $Bern(0.5)$.
Define $Y_i = U_{i,1} + U_{i,2}$. Notice that

$$f_Y(y) = \begin{cases} 0.25, & y = 0 \\ 0.50, & y = 1 \\ 0.25, & y = 2 \end{cases}$$

Since they have the same distribution, $X_i = Y_i + 1$ can be used.

$$\overline{X} = \frac{X_1 + X_2}{2} = \frac{Y_1 + 1 + Y_2 + 1}{2} = \frac{Y_1 + Y_2}{2} + 1$$
$$= \frac{U_{1,1} + U_{1,2} + U_{2,1} + U_{2,2}}{2} + 1 = \frac{Z}{2} + 1$$

Sampling Distributions, PMF Example (Slick Approach)

Consider U_i to be indepdent random variables, $Bern(0.5)$.
Define $Y_i = U_{i,1} + U_{i,2}$. Notice that

$$f_Y(y) = \begin{cases} 0.25, & y = 0 \\ 0.50, & y = 1 \\ 0.25, & y = 2 \end{cases}$$

Since they have the same distribution, $X_i = Y_i + 1$ can be used.

$$\begin{aligned} \overline{X} &= \frac{X_1 + X_2}{2} = \frac{Y_1 + 1 + Y_2 + 1}{2} = \frac{Y_1 + Y_2}{2} + 1 \\ &= \frac{U_{1,1} + U_{1,2} + U_{2,1} + U_{2,2}}{2} + 1 = \frac{Z}{2} + 1 \end{aligned}$$

Since Z is the sum of four independent bernoulli random variables, Z is also a bionomial random variable with $n = 4$ and $p = 0.5$.

Sampling Distributions, PMF Example (Slick Approach)

Transform the PMF of $\overline{X}$ into the PMF of Z.

Transform the PMF of $\overline{X}$ into the PMF of Z.

$$P(\overline{X} = t) = P\left(\frac{Z}{2} + 1 = t\right) = P(Z = 2(t-1))$$

Sampling Distributions, PMF Example (Slick Approach)

Transform the PMF of $\overline{X}$ into the PMF of Z.

$$P(\overline{X} = t) = P\left(\frac{Z}{2} + 1 = t\right) = P(Z = 2(t-1))$$

$\overline{X} = t$	1	1.5	2	2.5	3
$f_X(x)$	0.0625	0.25	0.375	0.25	0.0625
$z = 2(t-1)$	0	1	2	3	4
$f_Z(z)$	0.0625	0.25	0.375	0.25	0.0625

In the previous problem, determine the mean and variance of $\overline{X}$, and the mean of $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Sampling Distributions, Mean and Variance Example

In the previous problem, determine the mean and variance of $\overline{X}$, and the mean of $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

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

Sampling Distributions, Mean and Variance Example

In the previous problem, determine the mean and variance of $\overline{X}$, and the mean of $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$.

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

$$\begin{aligned} E[X_i] &= E[U_{i,1} + U_{i,1} + 1] = E[U_{i,1}] + E[U_{i,2}] + 1 \\ &= p + p + 1 = 2 \\ Var(X_i) &= Var(U_{i,1} + U_{i,1} + 1] = Var(U_{i,1}) + Var(U_{i,2}) \\ &= p(1-p) + p(1-p) = 0.5 \end{aligned}$$

In the previous problem, determine the mean and variance of $\overline{X}$, and the mean of $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

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

$$\begin{aligned} E[X_i] &= E[U_{i,1} + U_{i,1} + 1] = E[U_{i,1}] + E[U_{i,2}] + 1 \\ &= p + p + 1 = 2 \\ Var(X_i) &= Var(U_{i,1} + U_{i,1} + 1] = Var(U_{i,1}) + Var(U_{i,2}) \\ &= p(1-p) + p(1-p) = 0.5 \\ E[\overline{X}] &= E\left[\frac{X_1 + X_2}{2}\right] = \frac{E[X_1] + E[X_2]}{2} \end{aligned}$$

In the previous problem, determine the mean and variance of $\overline{X}$, and the mean of $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

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

$$\begin{aligned}
E[X_i] &= E[U_{i,1} + U_{i,1} + 1] = E[U_{i,1}] + E[U_{i,2}] + 1 \\
&= p + p + 1 = 2 \\
Var(X_i) &= Var(U_{i,1} + U_{i,1} + 1] = Var(U_{i,1}) + Var(U_{i,2}) \\
&= p(1-p) + p(1-p) = 0.5 \\
E[\overline{X}] &= E\left[\frac{X_1 + X_2}{2}\right] = \frac{E[X_1] + E[X_2]}{2} \\
&= E[X_i] = 2
\end{aligned}$$

Sampling Distributions, Mean and Variance Example

In the previous problem, determine the mean and variance of $\overline{X}$, and the mean of $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

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

$$\begin{aligned}
E[X_i] &= E[U_{i,1} + U_{i,1} + 1] = E[U_{i,1}] + E[U_{i,2}] + 1 \\
&= p + p + 1 = 2 \\
Var(X_i) &= Var(U_{i,1} + U_{i,1} + 1] = Var(U_{i,1}) + Var(U_{i,2}) \\
&= p(1-p) + p(1-p) = 0.5 \\
E[\overline{X}] &= E\left[\frac{X_1 + X_2}{2}\right] = \frac{E[X_1] + E[X_2]}{2} \\
&= E[X_i] = 2 \\
Var(\overline{X}) &= Var\left(\frac{X_1 + X_2}{2}\right) = \frac{Var(X_1) + Var(X_2)}{4} =
\end{aligned}$$

Sampling Distributions, Mean and Variance Example

In the previous problem, determine the mean and variance of $\overline{X}$, and the mean of $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

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

$$\begin{aligned}
E[X_i] &= E[U_{i,1} + U_{i,1} + 1] = E[U_{i,1}] + E[U_{i,2}] + 1 \\
&= p + p + 1 = 2 \\
Var(X_i) &= Var(U_{i,1} + U_{i,1} + 1] = Var(U_{i,1}) + Var(U_{i,2}) \\
&= p(1-p) + p(1-p) = 0.5 \\
E[\overline{X}] &= E\left[\frac{X_1 + X_2}{2}\right] = \frac{E[X_1] + E[X_2]}{2} \\
&= E[X_i] = 2 \\
Var(\overline{X}) &= Var\left(\frac{X_1 + X_2}{2}\right) = \frac{Var(X_1) + Var(X_2)}{4} = \frac{Var(X_i)}{2} = 0.25
\end{aligned}$$

Note:

$$\hat{S}^2 = \frac{n-1}{n} S^2$$

Sampling Distributions, Mean and Variance Example

Note:

$$\hat{S}^2 = \frac{n-1}{n} S^2$$

$$E[\hat{S}^2] = E\left[\frac{n-1}{n} S^2\right] = \frac{n-1}{n} E[S^2] =$$

Sampling Distributions, Mean and Variance Example

Note:

$$\hat{S}^2 = \frac{n-1}{n} S^2$$

$$E[\hat{S}^2] = E\left[\frac{n-1}{n} S^2\right] = \frac{n-1}{n} E[S^2] = \frac{2-1}{2} 0.5 = 0.5$$

Sampling Distributions, Mean and Variance Example

Note:

$$\hat{S}^2 = \frac{n-1}{n} S^2$$

$$E[\hat{S}^2] = E\left[\frac{n-1}{n} S^2\right] = \frac{n-1}{n} E[S^2] = \frac{2-1}{2} 0.5 = 0.5$$

Fun Fact:

$$Var(S^2) = \frac{\mu_4}{n} - \frac{\sigma^2(n-3)}{n(n-1)}$$

where

$$\mu_4 = E[(X_i - \mu)^4]$$

https://math.stackexchange.com/questions/2476527/variance-of-sample-variance

Sampling Distributions, Sum of Normals Example

Suppose that the weight X_i of pumpkins is normally distributed with a mean of 9 pounds with a standard deviation of 4 pounds. Determine the distribution of the sample average weight of n pumpkins.

Recall, the sum of normal random variables is normally distributed. We only need to find the expected value and standard deviation to identify a normal distribution.

Sampling Distributions, Sum of Normals Example

Suppose that the weight X_i of pumpkins is normally distributed with a mean of 9 pounds with a standard deviation of 4 pounds. Determine the distribution of the sample average weight of n pumpkins.

Recall, the sum of normal random variables is normally distributed. We only need to find the expected value and standard deviation to identify a normal distribution.

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

Sampling Distributions, Sum of Normals Example

Suppose that the weight X_i of pumpkins is normally distributed with a mean of 9 pounds with a standard deviation of 4 pounds. Determine the distribution of the sample average weight of n pumpkins.

Recall, the sum of normal random variables is normally distributed. We only need to find the expected value and standard deviation to identify a normal distribution.

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

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] =$$

Sampling Distributions, Sum of Normals Example

Suppose that the weight X_i of pumpkins is normally distributed with a mean of 9 pounds with a standard deviation of 4 pounds. Determine the distribution of the sample average weight of n pumpkins.

Recall, the sum of normal random variables is normally distributed. We only need to find the expected value and standard deviation to identify a normal distribution.

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

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} 9 = 9$$

Sampling Distributions, Sum of Normals Example

Suppose that the weight X_i of pumpkins is normally distributed with a mean of 9 pounds with a standard deviation of 4 pounds. Determine the distribution of the sample average weight of n pumpkins.

Recall, the sum of normal random variables is normally distributed. We only need to find the expected value and standard deviation to identify a normal distribution.

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

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} 9 = 9$$

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

Sampling Distributions, Sum of Normals Example

Suppose that the weight X_i of pumpkins is normally distributed with a mean of 9 pounds with a standard deviation of 4 pounds. Determine the distribution of the sample average weight of n pumpkins.

Recall, the sum of normal random variables is normally distributed. We only need to find the expected value and standard deviation to identify a normal distribution.

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

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} 9 = 9$$

$$Var(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} 4^2 = \frac{16}{n}$$

Sampling Distributions, Sum of Normals Example

Suppose that the weight X_i of pumpkins is normally distributed with a mean of 9 pounds with a standard deviation of 4 pounds. Determine the distribution of the sample average weight of n pumpkins.

Recall, the sum of normal random variables is normally distributed. We only need to find the expected value and standard deviation to identify a normal distribution.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \overline{X} \sim N\left(9, \frac{4}{\sqrt{n}}\right)$$

$$E[\overline{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n} 9 = 9$$

$$Var(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} 4^2 = \frac{16}{n}$$

Sampling Distributions, Sum of Exponentials Example

Suppose that T_i, wait tines between redlights, follow an exponential distribution with mean 2. Compute the PDF of the sample average of two wait times.

Suppose that T_i, wait tines between redlights, follow an exponential distribution with mean 2. Compute the PDF of the sample average of two wait times.

$$\overline{T} = \frac{T_1 + T_2}{2}$$

Sampling Distributions, Sum of Exponentials Example

Suppose that T_i, wait tines between redlights, follow an exponential distribution with mean 2. Compute the PDF of the sample average of two wait times.

$$\overline{T} = \frac{T_1 + T_2}{2}$$

Find CDF and differentiate. Due to the memoryless property wait times are independent.

Sampling Distributions, Sum of Exponentials Example

Suppose that T_i, wait tines between redlights, follow an exponential distribution with mean 2. Compute the PDF of the sample average of two wait times.

$$\overline{T} = \frac{T_1 + T_2}{2}$$

Find CDF and differentiate. Due to the memoryless property wait times are independent.

$$P(\overline{T} < t) = P(T_1 + T_2 < 2t) = P(T_1 < 2t - T_2)$$

Sampling Distributions, Sum of Exponentials Example

Suppose that T_i, wait tines between redlights, follow an exponential distribution with mean 2. Compute the PDF of the sample average of two wait times.

$$\overline{T} = \frac{T_1 + T_2}{2}$$

Find CDF and differentiate. Due to the memoryless property wait times are independent.

$$\begin{aligned} P(\overline{T} < t) &= P(T_1 + T_2 < 2t) = P(T_1 < 2t - T_2) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{2t-y} f_{T_1}(x) f_{T_2}(y) dx dy \\ &= \int_{-\infty}^{\infty} f_{T_2}(y) \int_{-\infty}^{2t-y} f_{T_1}(x) dx dy \end{aligned}$$

Sampling Distributions, Sum of Exponentials Example

Suppose that T_i, wait tines between redlights, follow an exponential distribution with mean 2. Compute the PDF of the sample average of two wait times.

$$\overline{T} = \frac{T_1 + T_2}{2}$$

Find CDF and differentiate. Due to the memoryless property wait times are independent.

$$\begin{aligned}
P(\overline{T} < t) = P(T_1 + T_2 < 2t) &= P(T_1 < 2t - T_2) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{2t-y} f_{T_1}(x) f_{T_2}(y) dx dy \\
&= \int_{-\infty}^{\infty} f_{T_2}(y) \int_{-\infty}^{2t-y} f_{T_1}(x) dx dy \\
&= \int_{-\infty}^{\infty} f_{T_2}(y) F_{T_1}(2t - y) dy
\end{aligned}$$

Sampling Distributions, Sum of Exponentials Example

$$\frac{d}{dt}P(\overline{T} < t) = \frac{d}{dt}\int_{-\infty}^{\infty} f_{T_2}(y)F_{T_1}(2t - y)dy$$

$$\begin{aligned}\frac{d}{dt}P(\overline{T} < t) &= \frac{d}{dt}\int_{-\infty}^{\infty} f_{T_2}(y)F_{T_1}(2t-y)dy \\ &= \int_{-\infty}^{\infty} f_{T_2}(y)\frac{d}{dt}F_{T_1}(2t-y)dy\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt}P(\overline{T} < t) &= \frac{d}{dt}\int_{-\infty}^{\infty} f_{T_2}(y)F_{T_1}(2t-y)dy \\
&= \int_{-\infty}^{\infty} f_{T_2}(y)\frac{d}{dt}F_{T_1}(2t-y)dy \\
&= \int_{-\infty}^{\infty} 2f_{T_2}(y)f_{T_1}(2t-y)dy
\end{aligned}$$

$$
\begin{aligned}
\frac{d}{dt}P(\overline{T} < t) &= \frac{d}{dt}\int_{-\infty}^{\infty} f_{T_2}(y)F_{T_1}(2t-y)dy \\
&= \int_{-\infty}^{\infty} f_{T_2}(y)\frac{d}{dt}F_{T_1}(2t-y)dy \\
&= \int_{-\infty}^{\infty} 2f_{T_2}(y)f_{T_1}(2t-y)dy \\
\text{note: } & y > 2t \rightarrow f_{T_1}(2t-y) = 0 \\
&= \int_{0}^{2t} 2\frac{\exp(-y/2)}{2}\frac{\exp(-t+y/2)}{2}dy
\end{aligned}
$$

Sampling Distributions, Sum of Exponentials Example

$$
\begin{aligned}
\frac{d}{dt}P(\overline{T} < t) &= \frac{d}{dt}\int_{-\infty}^{\infty} f_{T_2}(y)F_{T_1}(2t-y)dy \\
&= \int_{-\infty}^{\infty} f_{T_2}(y)\frac{d}{dt}F_{T_1}(2t-y)dy \\
&= \int_{-\infty}^{\infty} 2f_{T_2}(y)f_{T_1}(2t-y)dy \\
&\text{note: } y > 2t \rightarrow f_{T_1}(2t-y) = 0 \\
&= \int_{0}^{2t} 2\frac{\exp(-y/2)}{2}\frac{\exp(-t+y/2)}{2}dy \\
&= \int_{0}^{2t} \frac{\exp(-t)}{2}dy = \frac{t}{2}\exp(-t)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}P(\overline{T} < t) &= \frac{d}{dt}\int_{-\infty}^{\infty} f_{T_2}(y)F_{T_1}(2t-y)dy \\
&= \int_{-\infty}^{\infty} f_{T_2}(y)\frac{d}{dt}F_{T_1}(2t-y)dy \\
&= \int_{-\infty}^{\infty} 2f_{T_2}(y)f_{T_1}(2t-y)dy \\
&\text{note: } y > 2t \rightarrow f_{T_1}(2t-y) = 0 \\
&= \int_{0}^{2t} 2\frac{\exp(-y/2)}{2}\frac{\exp(-t+y/2)}{2}dy \\
&= \int_{0}^{2t} \frac{\exp(-t)}{2}dy = \frac{t}{2}\exp(-t) \\
\overline{T} &\sim Gamma(2,1)
\end{aligned}
$$

- Simple random sample or independent identically distributed assumption applies in many cases, but not all.
- Sampling distribution can model the probablistic behavior of statistics, but are often hard to find.
- Sum of Normal Random Variables is a Normal Random Variable. It's mean and standard deviation must be identified.
- Assuming X_i's are i.i.d.

$$E[\overline{X}] = E[X_i] = \mu \qquad Var(\overline{X}) = \frac{Var(X_i)}{n} = \frac{\sigma}{n}$$

$$E[S^2] = Var(X_i) = \sigma^2$$