The Chi-Square distribution

If $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables, then

$$X = \sum_{i=1}^{n} Z_i^2$$

$$\left[\; Z_i \sim N(0,1), \quad i = 1, 2, \ldots, n \; \right]$$

is said to have a chi-square distribution with n-degrees of freedom. We write

$$X \sim \chi_n^2$$

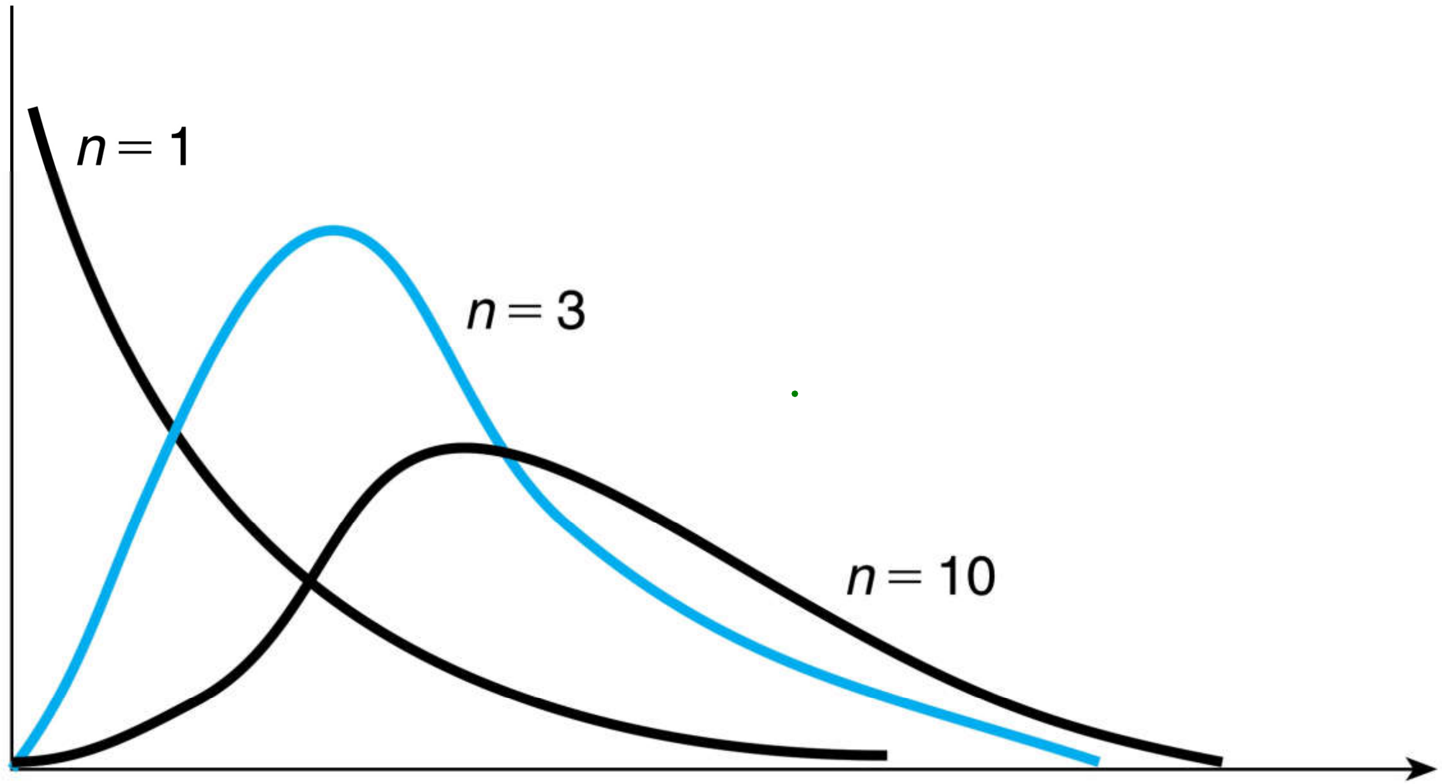


FIGURE 5.13 *The chi-square density function with n degrees of freedom.*

Ex: Let $X \sim \chi^2_n$. Find its m.g.f. Hence, find its mean and variance.

X & Y are independent r.v., then so is g(X) & h(Y).

Sol: $\chi^2_n = Z_1^2 + Z_2^2 + \cdots + Z_n^2$

$$\rightarrow M_{\chi^2_n}(t) = E\left[e^{t\chi^2_n}\right] = \prod_{i=1}^{n} E\left[e^{tZ_i^2}\right]$$

$$= \left[[1-2t]^{-\frac{1}{2}}\right]^n$$

$\boxed{Z \sim N(0,1)}$

$$E\left[e^{tZ^2}\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz^2} e^{-z^2/2}\, dz$$

$$\therefore = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2\left[\frac{1-2t}{2}\right]} dz$$

let $\sigma^2 = \frac{1}{1-2t}$

$$\boxed{E[e^{tZ^2}] = \frac{\sigma}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{-z^2/2\sigma^2} dz = \sigma = (1-2t)^{-1/2}}$$

$N(0, \sigma^2)$

$$\therefore \quad M_{\chi^2_n}(t) = (1-2t)^{-n/2}$$

$$= \frac{1}{(1-2t)^{n/2}}$$

$$= \frac{\left(\frac{1}{2}\right)^{n/2}}{\left(\frac{1}{2}-t\right)^{n/2}}$$

$$\Rightarrow \quad \chi^2_n \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right).$$

$$X \sim G(\alpha, \lambda)$$

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}$$

$$E[X] = \frac{\alpha}{\lambda}$$

$$Var(X) = \frac{\alpha}{\lambda^2}$$

$$\Rightarrow E[\chi^2_n] = n$$

$$\& \; Var(\chi^2_n) = 2n.$$

Def: $\left(\chi^2_{\alpha,n}\right)$

If $X \sim \chi^2_n$, then for any $\alpha \in (0,1)$, the quantity $\chi^2_{\alpha,n}$ is defined to be such that

$$P\left[X \geq \chi^2_{\alpha,n}\right] = \alpha.$$

Or $$P\left[X \leq \chi^2_{\alpha,n}\right] = 1-\alpha.$$

Ex:

Find (a) $P\left[\chi^2_{26} \leq 30\right]$

(b) $\chi^2_{.05,15}$

(a) $P\left[\chi^2_{26} \leq 30\right] = 1 - P\left[\chi^2_{26} \geq 30\right]$

$= 0.1$

(b) $\chi^2_{.05,15} = 24.996.$

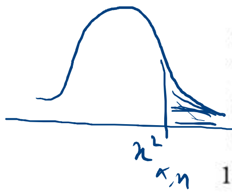
α, n

$P[X \geq x^2_{\alpha,n}]$

$P[X \geq 15.379] = 0.95$

$P[X \geq 38.885] = 0.05$

chi2cdf(30,26) = 0.9



$0.05 \leq P[X \geq 30] \leq 0.95$

$15.379 < 30 < 38.885$

TABLE A2 *Values of* $x^2_{\alpha,n}$

n	$\alpha = .995$	$\alpha = .99$	$\alpha = .975$	$\alpha = .95$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	.0000393	.000157	.000982	.00393	3.841	5.024	6.635	7.879
2	.0100	.0201	.0506	.103	5.991	7.378	9.210	10.597
3	.0717	.115	.216	.352	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	11.070	12.832	13.086	16.750
6	.676	.872	1.237	1.635	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.484	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	42.557	45.772	49.588	52.336
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672

Other chi-square probabilities:
$x^2_{.9,9} = 4.2 \quad P\{x^2_{16} < 14.3\} = .425 \quad P\{x^2_{11} < 17.1875\} = .8976.$

Ex: Suppose that we are attempting to locate a target in 3-D space, and that the three coordinate errors (in meters) of the point chosen are independent normal random variables with mean 0 and standard deviation 2.

Find the prob. that the distance between the point chosen and the target exceeds 3 meters.

Sol: Let X_i denotes the error in the i^{th} coordinate. $(i=1,2,3)$

$$D^2 = X_1^2 + X_2^2 + X_3^2 , \qquad \frac{X_i - 0}{2} \sim N(0,1).$$

Given that $X_i \sim N(0,4)$

$$\Rightarrow \quad \frac{D^2}{4} = \frac{X_1^2}{4} + \frac{X_2^2}{4} + \frac{X_3^2}{4}$$

$$= Z_1^2 + Z_2^2 + Z_3^2 , \quad Z_i \sim N(0,1)$$

$$= \chi_3^2 \qquad Z_i = \frac{X_i}{2} , \; i = 1,2,3.$$

$$P(D^2 > 9) = P\left(\frac{D^2}{4} > \frac{9}{4}\right) = P\left[\chi_3^2 > \frac{9}{4}\right)$$

$$= P\left[\chi_3^2 > 2.25\right] \in (.05, .95).$$

$$= 1 - \text{Chi2cdf}(2.25, 3).$$

In the case: Attempt to locate a target in **2D space:**

$$D^2 = X_1^2 + X_2^2 \Rightarrow \frac{D^2}{4} = \frac{X_1^2}{4} + \frac{X_2^2}{4} \sim \chi_2^2$$

$$\sim \text{Gamma}\left(\frac{2}{2}, \frac{1}{2}\right)$$

$$= \text{Gamma}\left(1, \frac{1}{2}\right)$$

$$F_X(x) = 1 - e^{-\lambda x}$$

$$P\left[D^2 > 9\right] = P\left[\frac{D^2}{4} > 2.25\right] = e^{-9/8} \approx 0.3247.$$

The t-distribution

The random variable

$$\boxed{T_n = \frac{Z}{\sqrt{\chi^2_n / n}}},$$

where $Z \sim N(0,1)$ & χ^2_n is chi square distribu-tion, is said to have t-distribution with n degrees of freedom.

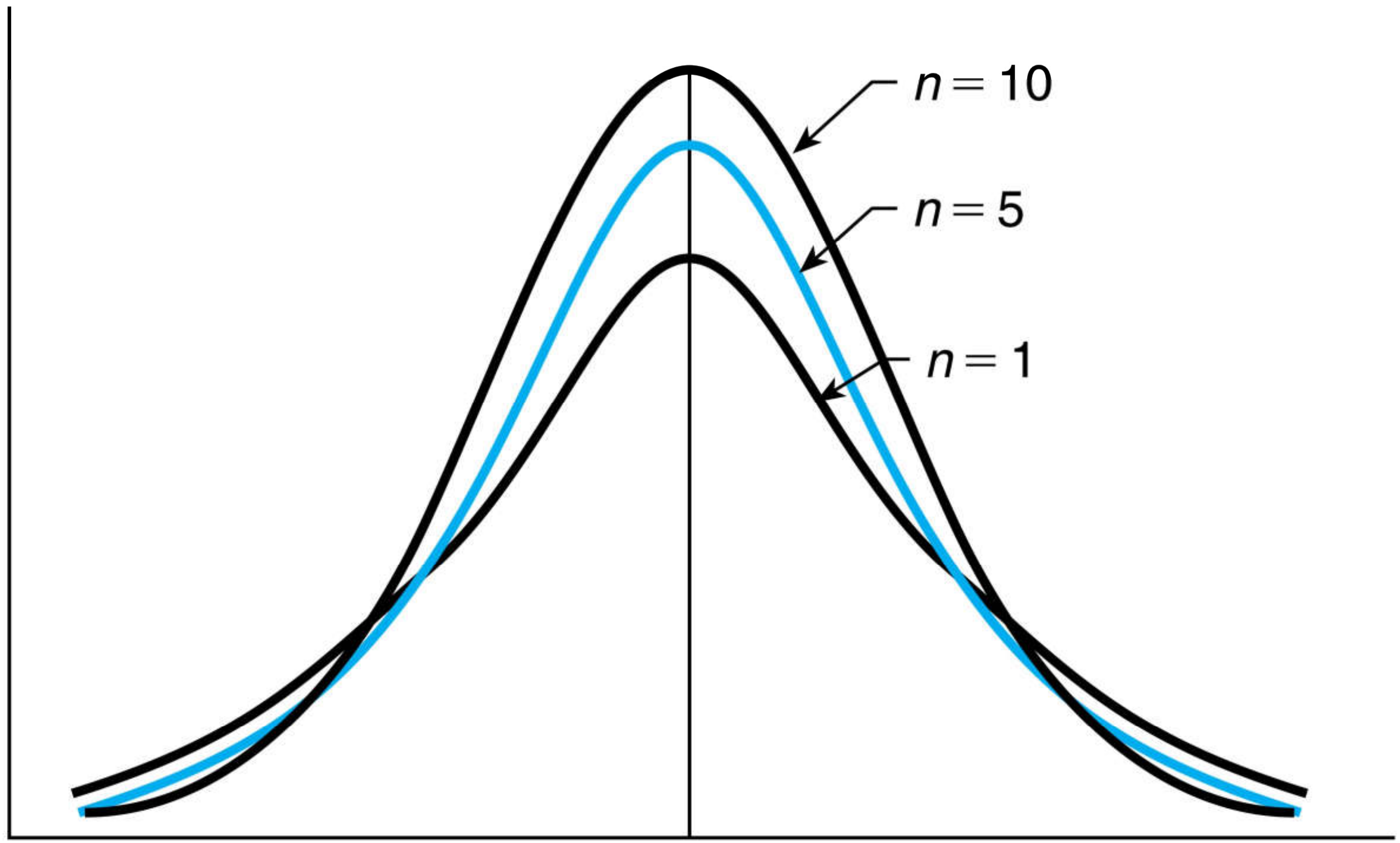


FIGURE 5.14 *Density function of T_n.*

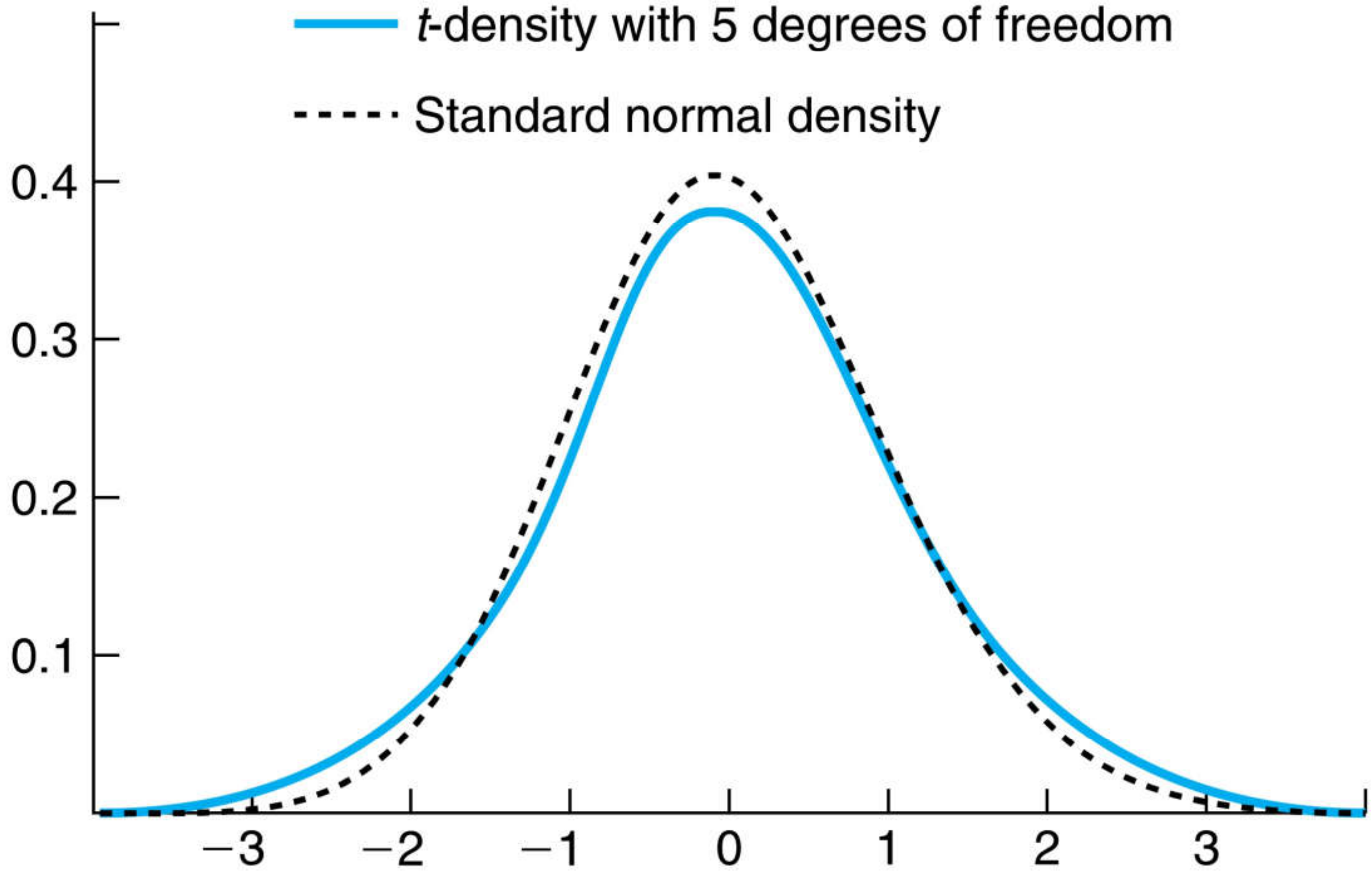


FIGURE 5.15 *Comparing standard normal density with the density of T_5.*

$$E\left[\frac{\chi_n^2}{n}\right] = \frac{1}{n} E\left[\chi_n^2\right] = \frac{1}{n} \times n = \underline{1}.$$

$$\chi_n^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

Weak law of large number implies that

$$P\left[\left|\frac{\chi_n^2}{n} - \mu\right| > \varepsilon\right] \longrightarrow 0 \quad \text{as } n \to \infty.$$

$$\Rightarrow \quad \frac{\chi_n^2}{n} \approx \underline{1} \quad \text{as } n \to \infty. \quad \Rightarrow \quad T_n = \frac{Z}{\sqrt{\chi_n^2/n}} \simeq Z$$

$E[T_n] = E\left[\dfrac{Z}{\sqrt{\chi_n^2/n}}\right]$

$= 0 \qquad$ for $n>1$

$$Var(T_n) = \frac{n}{n-2}, \quad \text{for } n>2$$

$$= \frac{n-2}{n-2} + \frac{2}{n-2}$$

$$= 1 + \frac{2}{n-2}$$

$$\longrightarrow 1.$$

Ex:

Find the m.g.f. of t-distribution.

Def: $(t_{\alpha,n})$

For $\alpha \in (0,1)$, the quantity $t_{\alpha,n}$ be such that

$$P[T_n \geq t_{\alpha,n}] = \alpha.$$

Since t-density funⁿ is symmetric about zero, therefore $-T_n$ has the same distribution as T_n, and so.

$$\Rightarrow P\left[-T_n > t_{\alpha,n}\right] = \alpha$$

$$\Rightarrow P\left[T_n < -t_{\alpha,n}\right] = \alpha$$

$$\Rightarrow P\left[T_n > -t_{\alpha,n}\right] = 1-\alpha$$

$$\Rightarrow -t_{\alpha,n} = t_{1-\alpha,n}$$

Ex : (a) $P[T_{12} < 1.4]$

TABLE A3 *Values of* $t_{\alpha,n}$

$P[T_{12} > 1.4]$

$.05 < P[T_{12} > 1.4] < .1$

$\underset{\alpha}{t_{0.025,\,\underset{n}{9}}} = 2.262$

n	$\alpha = .10$	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$	$\alpha = .005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.474	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
∞	1.282	1.645	1.960	2.326	2.576

$P[T_{12} < 1.4] = tcdf(1.4, 12)$

Other t probabilities:
$P\{T_8 < 2.541\} = .9825$ $P\{T_8 < 2.7\} = .9864$ $P\{T_{11} < .7635\} = .77$ $P\{T_{11} < .934\} = .81$ $P\{T_{11} < 1.66\} = .94$ $P\{T_{12} < 2.8\} = .984$.

The Sample mean ($\overline{X}$).

Let $X_1, X_2, \ldots, X_n$ be a sample of values from a certain population. The sample mean, denoted by $\overline{X}$, is defined as

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

where $X_1, \ldots, X_n$ are i.i.d. r.vs with mean μ and variance σ^2.

Ex: Show that

$$E[\bar{X}] = \mu$$

and $Var(\bar{X}) = \dfrac{\sigma^2}{n}$.

Sol:

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

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

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

The Central Limit Theorem

Let $X_1, X_2, \dots, X_n, \dots$ be a sequence of i.i.d. random variables each having mean μ and variance σ^2. Then, for n large

$\sum_{i=1}^{n} X_i$ follows $N(n\mu, n\sigma^2)$ approximately.

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

$Var(\bar{X}) = \frac{\sigma^2}{n}$

OR

$$P\left[\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}<x\right]$$

$$\simeq P[Z<x],$$

where $Z\sim N(0,1)$.

$X\sim N(\mu,\sigma^2)$

$\frac{X-\mu}{\sigma}\sim Z$

$\frac{\sum_{i=1}^{n}X_i-n\mu}{\sigma\sqrt{n}}\sim Z$

Ex: An insurance company has 25,000 automobile policy holders. If the yearly claim of a policy holder is a random variable with mean 320 and a standard deviation 540, approximate the prob. that the total yearly claim exceeds 8.3 million.

Sol: Let X = Total yearly claim

X_i = yearly claim of a policy holder i.

$i = 1, 2, 3, \ldots, 25000$.

$$X = \sum_{i=1}^{25000} X_i$$

$\mu = E[X_i] = 320$, & $\sigma = 540$

CLT $\Rightarrow$ $X \sim N(n\mu, n\sigma^2)$

$$E[X] = n\mu = (25000)(320) = 8 \times 10^{6}$$

$$\text{S.D. of } X = \sigma\sqrt{n} = 540\sqrt{25000} = 8.5381 \times 10^{4}$$

$$P[X > 8.3 \times 10^{6}] = P\left[\frac{X - 8\times 10^{6}}{8.5381\times 10^{4}} > \frac{8.3\times 10^{6} - 8\times 10^{6}}{8.5381 \times 10^{4}}\right]$$

$$\approx P\left[Z \geq \frac{0.3 \times 10^{6}}{8.5381 \times 10^{4}}\right]$$

$$\approx P[Z \geq 3.51] \approx 0.00023.$$

Ex: Civil engineers believe that W, the amount of weight (in units of 1000 pounds) that a certain span of bridge can withstand without structural damage resulting, is normally distributed with mean 400 and standard deviation 40.

Suppose that the weight (in units of 1000 pounds) of a car is a random variable with mean 3 and S.D. 0.3.

How many cars would have to be on the bridge span for the prob. of structural damage to exceed 0.1?

Sol: Let X_i = the weight of the ith car.

$i = 1, 2, \ldots, n$

$$P\left[\sum_{i=1}^{n} X_i \geq W\right] \simeq 0.1$$

CLT $\Rightarrow$ $\displaystyle\sum_{i=1}^{n} X_i \simeq N(n\mu, n\sigma^2)$

$$= N(3n, 0.09n)$$

Also, $W \sim N(400, (40)^2)$, W is independent of X_i

$$\therefore \sum_{i=1}^{n} X_i - W \sim N\left(3n-400,\ 0.09n+(40)^2\right)$$

$$\Rightarrow \frac{\sum_{i=1}^{n} X_i - W - (3n-400)}{\sqrt{0.09n+1600}} \sim N(0,1)$$

$$\Rightarrow P\left[\sum_{i=1}^{n} X_i - W \geq 0\right] \approx 0.1$$

$$\Rightarrow P\left[Z \geq \frac{-(3n-400)}{\sqrt{0.09n+1600}}\right] \approx 0.1$$

$$\boxed{P\left[Z \geq 1.28\right] \approx 0.1,}$$

$$\frac{400-3n}{\sqrt{0.09n+1600}} \leq 1.28 \iff n \geq 117.$$

Approximate Distribution of $\bar{X}$

CLT $\quad \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$

$$\Rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

$$\Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Ex: The weights of a population of workers have mean 167 and S.D. 27.

(a) If a sample of 36 workers is chosen, find $P[163 < \overline{X} < 170]$.

(b) Repeat the part (a) when the sample is of size 144.

The Sample Variance

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Let $\overline{X}$ be the sample mean. The statistic S^2, defined by

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

is called the sample variance.

$E[(X-\mu)^2]$

$\boxed{E[S^2] = \sigma^2}$

$S = \sqrt{S^2}$ is called the sample S.D.

Ex: Show that $E[S^2] = \sigma^2$.

Sol:

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

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

$$= \left(\frac{1}{n-1}\right)\left[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \overline{X}^2 - 2\sum_{i=1}^{n} \overline{X} X_i\right]$$

$$= \left(\frac{1}{n-1}\right)\left[\sum_{i=1}^{n} X_i^2 + n\overline{X}^2 - 2n\overline{X}^2\right]$$

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

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

$$\Rightarrow E\left[(n-1)S^2\right] = \sum_{i=1}^{n} E[X_i^2] - n\,E[\bar{X}^2]$$

$$= \sum_{i=1}^{n}\left[Var(X_i) + \left(E[X_i]\right)^2\right]$$

$$- n\left[Var(\bar{X}) + \left(E[\bar{X}]\right)^2\right]$$

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

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

$$\Rightarrow \quad E[S^2] = \sigma^2 .$$