Ch8 Sampling distributions of estimators

Let $X_1, \dots, X_n | \theta$ be an iid random sample from $f(x|\theta)$ for unknown $\theta \in \mathbb{R}$. We consider an estimator $\hat{\theta}$ for θ.

1. We aim to quantify how likely it is that the observed value of $\hat{\theta}$ will be close to the true parameter θ, prior to sampling.

Compute $P(|\hat{\theta} - \theta| < .01)$, and see whether it is close to 1.

2. We aim to construct an interval around our estimate that is likely to contain the true value of θ.

point estimate $\quad$ interval estimate

$\hat{\theta} = 2.3 \longrightarrow 2.2 < \theta < 2.4$.

To resolve these goals, the dist. of $\hat{\theta}$ is required!

"sampling dist. of $\hat{\theta}$" Ch8.1

Particularly, we study sampling dist. of two estimators:

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\sigma'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ Ch 8.2 – 8.4

We study how to construct a confidence interval Ch8.5

"Unbiased" Ch8.7

"C-R lower bound" Ch 8.8.

evaluating the performance of $\hat{\theta}$.

Ch8.1 Sampling distribution of a statistic

a function of $X_1, \dots, X_n$

Definition. Let T be a statistic, $T(X_1, \dots, X_n)$. The distribution of T is called the sampling dist. of T.

$\leftarrow$ (conditional on θ in general)

Notation: $E_\theta[T] = E[T|\theta]$: mean of sampling dist. of T.

EG $X_1, \dots, X_n|\theta \sim$ iid normal θ, σ^2, where θ is unknown and σ^2 is known.

Consider $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ to estimate θ.

Recall from Ch5.6, a linear combination of $X_1, \dots, X_n$ (normally distributed) is also normally distributed.

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

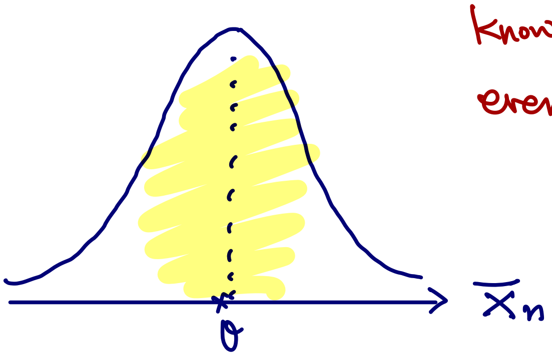
$$\overset{id}{=} \frac{1}{n}\, n\, E[X_1] = E[X_1] = \theta.$$

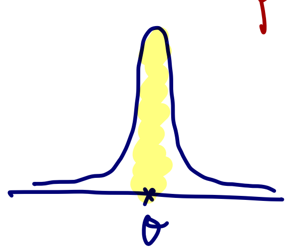
$$Var[\bar{X}_n] = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right] \overset{indep.}{=} \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] \quad + \cancel{cov(\quad)} \; 0$$

$$\overset{id}{=} \frac{1}{n^2}\, n \cdot Var[X_1] = \frac{Var[X_1]}{n} = \frac{\sigma^2}{n}.$$

$\Rightarrow$ Sampling dist. of $\bar{X}_n \sim$ normal $\theta, \frac{\sigma^2}{n}$

Know the performance of $\bar{X}_n$ even prior to sampling!





EG [Lifetimes of Electronic Components] $X_1, X_2, X_3 | \theta \sim$ iid exponential θ

MLE of $\theta = \hat{\theta} = \dfrac{1}{\overline{X}_3} = \dfrac{3}{\sum_{i=1}^{3} X_i}$

Ch5

$\sum_{i=1}^{3} X_i \sim \text{gamma } 3, \theta \Rightarrow \theta \sum_{i=1}^{3} X_i \sim \text{gamma } 3, 1$

$E[\sum X_i] = \dfrac{3}{\theta}$. $\qquad E[\theta \sum X_i] = \theta \times \dfrac{3}{\theta} = 3 = \dfrac{3}{1}$

$P(\hat{\theta}$ is within 10% of $\theta | \theta)$

$= P\left(\left| \dfrac{\hat{\theta}}{\theta} - 1 \right| < .1 \,\middle|\, \theta \right) = P\left(.9 < \dfrac{\hat{\theta}}{\theta} < 1.1 \,\middle|\, \theta \right)$ $\quad \hat{\theta} = \dfrac{3}{\sum X_i}$

$= P\left(.9 < \dfrac{3}{\theta \sum X_i} < 1.1 \,\middle|\, \theta \right) = P\left(\dfrac{3}{1.1} < \underbrace{\theta \sum X_i}_{\sim \text{gamma } 3,1 \ (\theta \text{ free})} < \dfrac{3}{.9} \,\middle|\, \theta \right)$

$\approx P\left(2.73 < \underbrace{\theta \sum X_i}_{\sim \text{gamma } 3,1} < 3.33 \right)$

$= P(\theta \sum X_i < 3.33) - P(\theta \sum X_i \leq 2.73)$

$=$ "pgamma(3.33, 3, 1)" $-$ "pgamma(2.73, 3, 1)"

$= .6465628 - .5136956$

$= .134$ (no matter what θ is)

low because of small sample size ($n = 3$).

Computing Probabilities: By computer or table

The purpose of this handout is to lay out how you can double-check a calculation with R.

The Comprehensive R Archive Network
https://cran.r-project.org › bin › windows › base

Download R-4.3.2 for Windows - CRAN

A table of R's special distribution functions is below, e.g., use the R command "?pnorm" to learn more about the pnorm function or look online.

Ch5

Special Distribution	CDF $F(x) = p$	Inverse CDF $F^{-1}(p) = x$
binomial n, π	pbinom(x, n, π)	qbinom(p, n, π)
Poisson λ	ppois(x, λ)	qpois(p, λ)
exponential β	pexp(x, β)	qexp(p, β)
gamma α, β	pgamma(x, α, β)	qgamma(p, α, β)
beta α, β	pbeta(x, α, β)	qbeta(p, α, β)
normal μ, σ^2	pnorm(x, μ, σ)	qnorm(p, μ, σ)
t with v df	pt(x, v)	qt(x, v)
χ^2 with v df	pchisq(x, v)	qchisq(x, v)

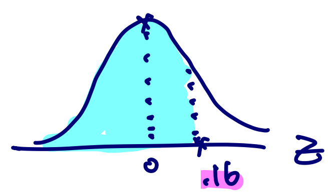
Ch8

Example 1: Let Z be standard normal. Compute $P(Z < 0.16) \approx 0.5636$.

$\mu=0, \sigma^2=1$

```
> pnorm(.16)
[1] 0.5635595
> qnorm(.5635595)
[1] 0.1600001
```

= pnorm(.16, $\mu=0$, $\sigma^2=1$)



Example 2: Let X be beta 2,3. Find x such that $P(X < x) = 0.3$, i.e., the 0.3 quantile $x_{0.3}$ or 30th percentile of X.

```
> qbeta(.3,2,3)
[1] 0.2723839
```

Ch8.2 Chi-square distributions

$X_1, \ldots, X_n \sim$ iid normal μ, σ^2

The sample mean $\bar{X}_n \sim$ normal $\mu, \frac{\sigma^2}{n}$. Ch8.1

The sample variance $\sigma'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim ?$

The sampling distribution of σ'^2 is related to Chi-square distributions.

Definition. The gamma $\alpha = m/2, \beta = 1/2$ distribution is called the χ^2 dist. w/ m df $\forall m > 0$.

gamma α, β

$$f(x) = \frac{\left(\frac{1}{2}\right)^{m/2}}{\Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{1}{2}x}, \quad x > 0. \qquad \Rightarrow f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \cdot e^{-\beta x}, \quad x > 0.$$

Note.

χ^2 dist. w/ 2 df $\overset{d}{=}$ gamma $1, \frac{1}{2}$ $\overset{d}{=}$ exponential $\frac{1}{2}$.

Theorem. If X is χ^2 with m df, then

χ^2 w/ m $\overset{d}{=}$ gamma $\frac{m}{2}, \frac{1}{2}$

$$E[X] = \frac{m/2}{1/2} = m \qquad \mathrm{Var}(X) = \frac{m/2}{(1/2)^2} = 2m.$$

$$\psi(t) = \left(\frac{\frac{1}{2}}{\frac{1}{2}-t}\right)^{\frac{m}{2}} = \left(\frac{1}{1-2t}\right)^{\frac{m}{2}}, \quad t < \frac{1}{2}.$$

Proof. Use the fact that if $Y \sim$ gamma α, β, then $E[Y] = \frac{\alpha}{\beta}$, $\mathrm{Var}[Y] = \frac{\alpha}{\beta^2}$,

and $\psi(t) = \left(\frac{\beta}{\beta-t}\right)^{\alpha}, t < \beta$.

χ^2 w/ m df $\overset{d}{=}$ gamma $\frac{m}{2}, \frac{1}{2}$

Theorem. If $X_1, \dots, X_k$ are independent χ^2 with df $m_1, \dots, m_k$, then $X_1 + \cdots + X_k$ is χ^2 w/ $m_1 + \cdots + m_k$ df.

Proof. Using the fact that if $Y_i \sim$ gamma α_i, β and $Y_1, \dots, Y_k$ are independent, then $Y_1 + \cdots + Y_k \sim$ gamma $\alpha_1 + \cdots + \alpha_k, \beta$. Ch5.

$$X_i \sim \chi^2 \text{ w/ } m_i \text{ df} \overset{d}{=} \text{gamma } \frac{m_i}{2}, \frac{1}{2}.$$

Since $X_1, \dots, X_k$ are independent,

$$X_1 + \cdots + X_k \sim \text{gamma } \frac{m_1}{2} + \cdots + \frac{m_k}{2}, \frac{1}{2} \overset{d}{=} \chi^2 \text{ w/ } m_1 + \cdots + m_k.$$

pdf

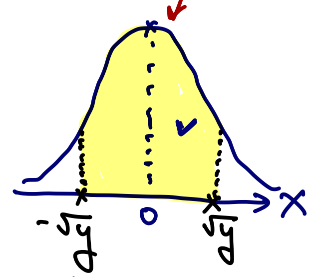
$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \Phi_X'(x), \quad \Phi_X(x): \text{CDF of } X.$$

Theorem. If X is standard normal, then $Y = X^2$ is χ^2 w/ 1 df

Proof. CDF of Y

$$P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y})$$

$$= 2\left(\Phi_X(\sqrt{y}) - \frac{1}{2}\right) = 2\Phi_X(\sqrt{y}) - 1.$$



PDF of Y

$$f_Y(y) = \frac{dP(Y \le y)}{dy} = \cancel{2}\, \Phi_X'(\sqrt{y}) \times \frac{1}{\cancel{2}\sqrt{y}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \times \frac{1}{\sqrt{y}} = \frac{1/\sqrt{2}}{\sqrt{\pi}} \times y^{-\frac{1}{2}} e^{-\frac{y}{2}} = \frac{\left(\frac{1}{2}\right)^{1/2}}{\Gamma\left(\frac{1}{2}\right)} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y}$$

$$\Rightarrow Y \sim \text{gamma } \frac{1}{2}, \frac{1}{2} \overset{d}{=} \chi^2 \text{ w/ } 1 \text{ df}.$$

Corollary. If $X_1, \dots, X_k$ are iid standard normal, then $X_1^2 + \cdots + X_k^2$ is χ^2 w/ k df

Proof. $X_i^2 \sim \chi^2$ w/ 1 df, and $X_1^2, \dots X_k^2$ are independent,

$$\Rightarrow X_1^2 + \cdots + X_k^2 \sim \chi^2 \text{ w/ } k \text{ df}.$$

unknown.

EG $X_1, \ldots, X_n \sim$ iid normal μ, σ^2, where μ is known.

$X_i \sim$ normal μ, σ^2.

$\Rightarrow \dfrac{X_i - \mu}{\sigma} \sim$ standard normal.

$\Rightarrow \dfrac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2$ w/ 1 df.

Since X_i's are independent,

$\sum_{i=1}^{n} \dfrac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2$ w/ n df. (when μ is known)

$\sum_{i=1}^{n} \dfrac{(X_i-\bar{X}_n)^2}{\sigma^2} \sim \chi^2$ w/ $n-1$ df (when μ is unknown)

Ch8-3

EG $X_1, X_2, X_3 \sim$ iid normal $0, 1$. What is the distribution of $\frac{1}{2}(X_1+X_2)^2 + X_3^2$?

$\sim \chi^2$ w/ 1 df

$\frac{1}{2}(X_1+X_2)^2 = \left\{ \frac{1}{\sqrt{2}}(X_1+X_2) \right\}^2 \sim \chi^2$ w/ 1 df.

dist. of $\frac{1}{\sqrt{2}}(X_1+X_2) \sim$ normal $0, 1$

- $\frac{1}{\sqrt{2}}(X_1+X_2)$ is a linear combination of two independent normal X_1 and X_2. $\Rightarrow$ it is also normally distributed!
- $E\left[\frac{1}{\sqrt{2}}(X_1+X_2)\right] = \frac{1}{\sqrt{2}} E[X_1+X_2] = \frac{1}{\sqrt{2}}\left(\underbrace{E[X_1]}_{0} + \underbrace{E[X_2]}_{0}\right) = 0$
- $Var\left[\frac{1}{\sqrt{2}}(X_1+X_2)\right] = \frac{1}{2} Var[X_1+X_2] = \frac{1}{2}\left\{\underbrace{Var[X_1]}_{1} + \underbrace{Var[X_2]}_{1} + 2\underbrace{cov(X_1,X_2)}_{0}\right\}$
$= \frac{1}{2}(1+1+0) = 1.$

Since $\frac{1}{2}(X_1+X_2)^2$ and X_3^2 are independent,

$\frac{1}{2}(X_1+X_2)^2 + X_3^2 \sim \chi^2$ w/ 2 df

Ch8.3 Joint distribution of the sample mean and sample variance

Theorem. Suppose that $X_1, \dots, X_n \sim$ iid normal μ, σ^2 where μ and σ^2 are unknown. Then,

← sample mean sample variance

1. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim$ normal $\mu, \frac{\sigma^2}{n}$, and $\sigma'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim \frac{\sigma^2 \chi^2 \text{ w/ } n-1 \text{ df}}{n-1}$

2. $\bar{X}_n$ and σ'^2 are independent.

$\bar{X}_n$	1	5
σ'^2	$.5$	$.4$

We make use of orthogonal matrices for the proof!

Definition. An $n \times n$ matrix $\boldsymbol{A}$ is orthogonal if $\boldsymbol{A}^{\mathbf{T}} = \boldsymbol{A}^{-1}$, where $\boldsymbol{A}^{\mathbf{T}}$ is the transpose of $\boldsymbol{A}$.

Consider orthogonal transformation $\boldsymbol{Y} = \boldsymbol{AX}$, where $\boldsymbol{X} = [X_1, \cdots, X_n]^T$ and $\boldsymbol{Y} = [Y_1, \cdots, Y_n]^T$ are n-dimensional random vectors, where $\boldsymbol{A}$ is an orthogonal matrix.

Fact 1. Orthogonal transformation preserves the squared length.

Fact 2. If $X_1, \dots, X_n \sim$ iid standard normal, then $Y_1, \dots, Y_n \sim$ iid standard normal.

Proof.

EG $X_1, \dots, X_n | \mu, \sigma^2 \sim$ iid normal μ, σ^2. If the sample size n is 10, determine the value of the following probability

MLE of σ^2

$$P\left(\frac{\widehat{\sigma^2}}{\sigma^2} \leq 1.5\right) = P\left(\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{\sigma^2} \leq 1.5\right)$$

$\times n$ $\times n$

$$= P\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{\sigma^2} \leq 1.5n\right)$$

$$\boxed{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim \frac{\sigma^2 \chi^2 \text{ w/ } n-1 \text{ df}}{n-1}} \Rightarrow \frac{\sum(X_i - \bar{X}_n)^2}{\sigma^2} \sim \chi^2 \text{ w/ } n-1 \text{ df.}$$

$$= P(Y \leq 15), \quad Y \sim \chi^2 \text{ w/ } 9 \text{ df.}$$

$$= \text{"pchisq}(15, 9)\text{" in R}$$

$$\approx .909.$$

Ch8-pg11

$P(Y \le 15) = ?$

χ^2 w/ 9

Table of the χ^2 Distribution

If X has a χ^2 distribution with m degrees of freedom, this table gives the value of x such that $\Pr(X \le x) = p$, the p quantile of X.

	p								
m	.005	.01	.025	.05	.10	.20	.25	.30	.40
1	.0000	.0002	.0010	.0039	.0158	.0642	.1015	.1484	.2750
2	.0100	.0201	.0506	.1026	.2107	.4463	.5754	.7133	1.022
3	.0717	.1148	.2158	.3518	.5844	1.005	1.213	1.424	1.869
4	.2070	.2971	.4844	.7107	1.064	1.649	1.923	2.195	2.753
5	.4117	.5543	.8312	1.145	1.610	2.343	2.675	3.000	3.655
6	.6757	.8721	1.237	1.635	2.204	3.070	3.455	3.828	4.570
7	.9893	1.239	1.690	2.167	2.833	3.822	4.255	4.671	5.493
8	1.344	1.647	2.180	2.732	3.490	4.594	5.071	5.527	6.423
9	1.735	2.088	2.700	3.325	4.168	5.380	5.899	6.393	7.357
10	2.156	2.558	3.247	3.940	4.865	6.179	6.737	7.267	8.295
11	2.603	3.053	3.816	4.575	5.578	6.989	7.584	8.148	9.237
12	3.074	3.571	4.404	5.226	6.304	7.807	8.438	9.034	10.18
13	3.565	4.107	5.009	5.892	7.042	8.634	9.299	9.926	11.13
14	4.075	4.660	5.629	6.571	7.790	9.467	10.17	10.82	12.08
15	4.601	5.229	6.262	7.261	8.547	10.31	11.04	11.72	13.03
16	5.142	5.812	6.908	7.962	9.312	11.15	11.91	12.62	13.98
17	5.697	6.408	7.564	8.672	10.09	12.00	12.79	13.53	14.94
18	6.265	7.015	8.231	9.390	10.86	12.86	13.68	14.43	15.89
19	6.844	7.633	8.907	10.12	11.65	13.72	14.56	15.35	16.85
20	7.434	8.260	9.591	10.85	12.44	14.58	15.45	16.27	17.81
21	8.034	8.897	10.28	11.59	13.24	15.44	16.34	17.18	18.77
22	8.643	9.542	10.98	12.34	14.04	16.31	17.24	18.10	19.73
23	9.260	10.20	11.69	13.09	14.85	17.19	18.14	19.02	20.69
24	9.886	10.86	12.40	13.85	15.66	18.06	19.04	19.94	21.65
25	10.52	11.52	13.12	14.61	16.47	18.94	19.94	20.87	22.62
30	13.79	14.95	16.79	18.49	20.60	23.36	24.48	25.51	27.44
40	20.71	22.16	24.43	26.51	29.05	32.34	33.66	34.87	36.16
50	27.99	29.71	32.36	34.76	37.69	41.45	42.94	44.31	46.86
60	35.53	37.48	40.48	43.19	46.46	50.64	52.29	53.81	56.62
70	43.27	45.44	48.76	51.74	55.33	59.90	61.70	63.35	66.40
80	51.17	53.54	57.15	60.39	64.28	69.21	71.14	72.92	76.19
90	59.20	61.75	65.65	69.13	73.29	78.56	80.62	82.51	85.99
100	67.33	70.06	74.22	77.93	82.86	87.95	90.13	92.13	95.81

"Table of the X2 Distribution" adapted in part from "A new table of percentage points of the chi-square distribution" by H. Leon Harter. From BIOMETRIKA, vol 51(1964), pp. 231–239.
"Table of the X2 Distribution" adapted in part from the BIOMETRIKA TABLES FOR STATISTICIANS, Vol. 1, 3rd ed., Cambridge University Press, © 1966, edited by E.S. Pearson and H.O. Hartley.

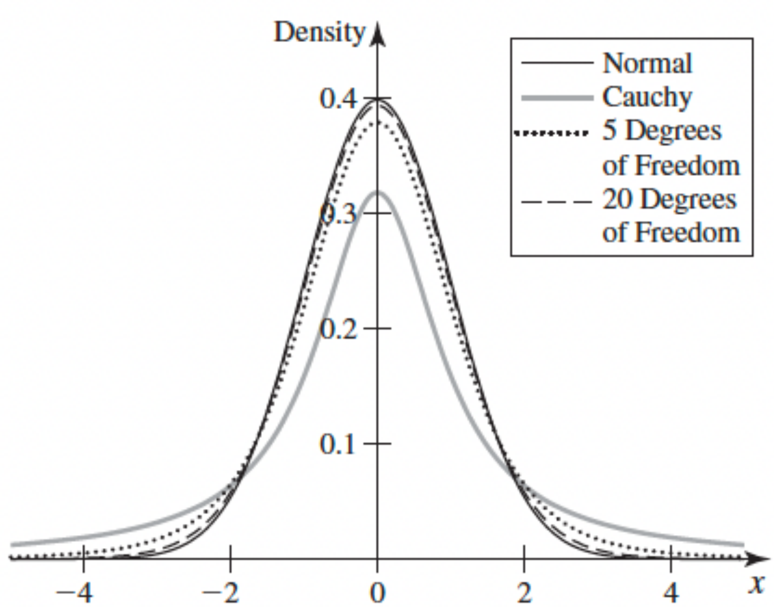
Table of the χ^2 Distribution *(continued)*

p									
.50	.60	.70	.75	.80	.90	.95	.975	.99	.995
.4549	.7083	1.074	1.323	1.642	2.706	3.841	5.024	6.635	7.879
1.386	1.833	2.408	2.773	3.219	4.605	5.991	7.378	9.210	10.60
2.366	2.946	3.665	4.108	4.642	6.251	7.815	9.348	11.34	12.84
3.357	4.045	4.878	5.385	5.989	7.779	9.488	11.14	13.28	14.86
4.351	5.132	6.064	6.626	7.289	9.236	11.07	12.83	15.09	16.75
5.348	6.211	7.231	7.841	8.558	10.64	12.59	14.45	16.81	18.55
6.346	7.283	8.383	9.037	9.803	12.02	14.07	16.01	18.48	20.28
7.344	8.351	9.524	10.22	11.03	13.36	15.51	17.53	20.09	21.95
8.343	9.414	10.66	11.39	12.24	14.68	16.92	19.02	21.67	23.59
9.342	10.47	11.78	12.55	13.44	15.99	18.31	20.48	23.21	25.19
10.34	11.53	12.90	13.70	14.63	17.27	19.68	21.92	24.72	26.76
11.34	12.58	14.01	14.85	15.81	18.55	21.03	23.34	26.22	28.30
12.34	13.64	15.12	15.98	16.98	19.81	22.36	24.74	27.69	29.82
13.34	14.69	16.22	17.12	18.15	21.06	23.68	26.12	29.14	31.32
14.34	15.73	17.32	18.25	19.31	22.31	25.00	27.49	30.58	32.80
15.34	16.78	18.42	19.37	20.47	23.54	26.30	28.85	32.00	34.27
16.34	17.82	19.51	20.49	21.61	24.77	27.59	30.19	33.41	35.72
17.34	18.87	20.60	21.60	22.76	25.99	28.87	31.53	34.81	37.16
18.34	19.91	21.69	22.72	23.90	27.20	30.14	32.85	36.19	38.58
19.34	20.95	22.77	23.83	25.04	28.41	31.41	34.17	37.57	40.00
20.34	21.99	23.86	24.93	26.17	29.62	32.67	35.48	38.93	41.40
21.34	23.03	24.94	26.04	27.30	30.81	33.92	36.78	40.29	42.80
22.34	24.07	26.02	27.14	28.43	32.01	35.17	38.08	41.64	44.18
23.34	25.11	27.10	28.24	29.55	33.20	36.42	39.36	42.98	45.56
24.34	26.14	28.17	29.34	30.68	34.38	37.65	40.65	44.31	46.93
29.34	31.32	33.53	34.80	36.25	40.26	43.77	46.98	50.89	53.67
39.34	41.62	44.16	45.62	47.27	51.81	55.76	59.34	63.69	66.77
49.33	51.89	54.72	56.33	58.16	63.17	67.51	71.42	76.15	79.49
59.33	62.13	65.23	66.98	68.97	74.40	79.08	83.30	88.38	91.95
69.33	72.36	75.69	77.58	79.71	85.53	90.53	95.02	100.4	104.2
79.33	82.57	86.12	88.13	90.41	96.58	101.9	106.6	112.3	116.3
89.33	92.76	96.52	98.65	101.1	107.6	113.1	118.1	124.1	128.3
99.33	102.9	106.9	109.1	111.7	118.5	124.3	129.6	135.8	140.2

<u>Ch8.4</u> The t-distributions

<u>Definition.</u> Consider two independent random variables Z, Y such that

The distribution of $X =$ is called the ___________________________.



<u>Facts:</u>

$*$ t with 1 df is __________________.

* The mean of t with m df exists when $m > 1$. $E[X] =$

* The variance of t with m df exists when $m > 2$. $Var[X] = m/(m-2)$

* t with m df is approximately standard normal if m is _______________.

* t with m df only have moments up to $k < m$. So the mgf does not exist.

Theorem. Suppose that $X_1, \dots, X_n \sim$ iid normal μ, σ^2. Then,

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $\sigma' = \left[\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n-1}\right]^{1/2}$.

Proof.

EG. $X_1, \dots, X_n \sim$ iid normal μ, σ^2, and $n = 16$. Find a number c such that $P(\bar{X}_n \le \mu + c\sigma') = .95$.

Table of the *t* Distribution

If X has a t distribution with m degrees of freedom, the table gives the value of x such that $\Pr(X \leq x) = p$.

m	$p = .55$	.60	.65	.70	.75	.80	.85	.90	.95	.975	.99	.995
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	.137	.277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032
6	.131	.265	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707
7	.130	.263	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106
12	.128	.259	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055
13	.128	.259	.394	.538	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012
14	.128	.258	.393	.537	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977
15	.128	.258	.393	.536	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947
16	.128	.258	.392	.535	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921
17	.128	.257	.392	.534	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898
18	.127	.257	.392	.534	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878
19	.127	.257	.391	.533	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861
20	.127	.257	.391	.533	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845
21	.127	.257	.391	.532	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831
22	.127	.256	.390	.532	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819
23	.127	.256	.390	.532	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807
24	.127	.256	.390	.531	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797
25	.127	.256	.390	.531	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787
26	.127	.256	.390	.531	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779
27	.127	.256	.389	.531	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771
28	.127	.256	.389	.530	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763
29	.127	.256	.389	.530	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756
30	.127	.256	.389	.530	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750
40	.126	.255	.388	.529	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704
60	.126	.254	.387	.527	.679	.848	1.046	1.296	1.671	2.000	2.390	2.660
120	.126	.254	.386	.526	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	.126	.253	.385	.524	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576

Ch8.5 Confidence intervals (CIs)

Let $X_1, \dots, X_n | \theta$ be an iid random sample from $f(x|\theta)$ for unknown $\theta \in \mathbb{R}$.

We consider an estimator $\hat{\theta}$ for θ, and based on $\hat{\theta}$, we aim to construct an interval which is expected to contain the true value of θ.

Definition. Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a random sample from $f(x|\theta)$ for unknown θ. Let $A(\boldsymbol{X}) \leq B(\boldsymbol{X})$ be two statistics such that

Then $(A(\boldsymbol{X}), B(\boldsymbol{X}))$ is called a ____.

EX7. We are interested in estimating the average calorie context of all beef hot dogs. A random sample was taken, and data on the calorie content of $n = 20$ hot dog brands are given in the textbook. From the data, $\bar{x}_{20} = 156.85, \sigma' = 22.64$.

We assume that $X_1, \dots, X_{20} | \mu, \sigma^2 \sim$ iid normal μ, σ^2, both unknown.

Theorem. Suppose that $X_1, \dots, X_n | \mu, \sigma^2 \sim$ iid normal μ, σ^2, both unknown. Then, an (exact) $100\gamma\%$ confidence interval for μ is

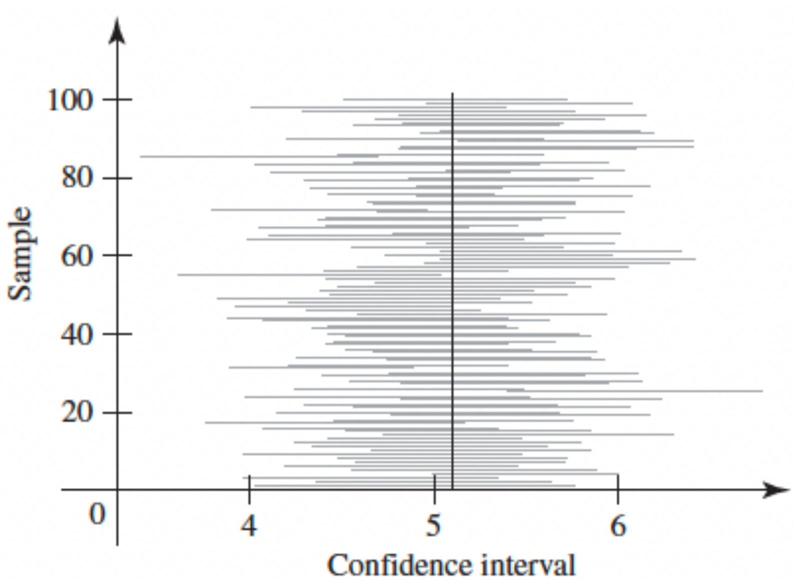
where T_ν is the CDF of a t distribution with $n-1$ df.

EX7. Construct a 95% confidence interval for μ.

Interpretation (EX7)

We are ____________ confident that the population mean content μ across all brands is between _________ and _________.

Confidence level $100\gamma\%$ means the long-term success rate of the method.



We expect that intervals constructed this way capture the true value of the parameter $100\gamma\%$ of the time.

<u>EG.</u> Suppose that $X_1, \ldots, X_n | \mu, \sigma^2 \sim$ iid normal μ, σ^2, both unknown. Find an (exact) $100\gamma\%$ confidence interval for σ^2.

<u>EG</u> [Lifetimes of Electronic Components] $X_1, X_2, X_3 | \theta \sim$ iid exponential θ. Find an upper bound on θ.

<u>Ch8.7</u> Unbiased estimators

The quality of an estimator can be discussed in many ways.

<u>Definition.</u> Suppose that $X_1, \ldots, X_n|\theta \sim$ iid $f(x|\theta)$. An estimator $\delta(\boldsymbol{X}) = \delta(X_1, \ldots, X_n)$ is ____________________________________ for $g(\theta)$ if

<u>EG</u> $X_1, \ldots, X_n|\theta \sim$ iid such that $E_{\boldsymbol{\theta}}[X_1] = \mu$.

Bias $= E_\theta[\delta(\boldsymbol{X})] - g(\theta)$

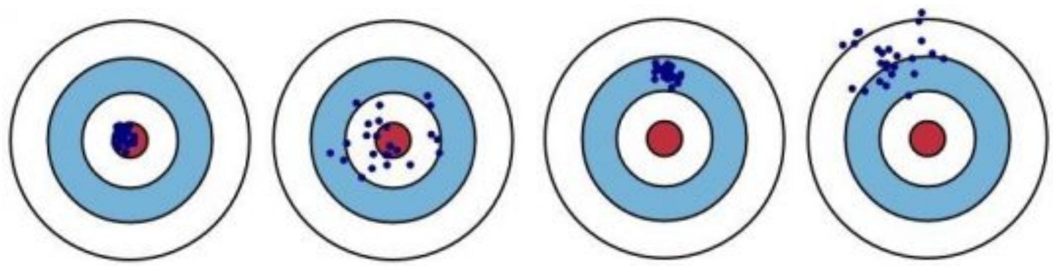
Bias $\neq 0 \Rightarrow \delta(\boldsymbol{X})$ is ________________ for $g(\theta)$.

<u>EG</u>[Lifetimes of Electronic Components] $X_1, X_2, X_3|\theta \sim$ iid exponential θ.

<u>EG</u> $X_1, \ldots, X_n | \boldsymbol{\theta} = (\mu, \sigma^2) \sim$ iid such that $E_{\boldsymbol{\theta}}[X_1] = \mu$ and $\mathrm{Var}_{\boldsymbol{\theta}}[X_1] = \sigma^2$

Consider $\sigma'^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$. Which one is unbiased?

We want our unbiased estimator to have a small variance!



Corollary.

In many problems, there exist biased estimators that have smaller MSE than every unbiased estimator.

EG. Suppose that $X_1, \dots, X_n | \mu \sim$iid normal $\mu, 1$. Consider the form of $\delta_c(\boldsymbol{X}) = c\bar{X}_n$, where $0 < c < 1$. Find c that yields the smallest MSE.

<u>Ch8.8</u> Fisher information

A method for measuring the amount of information a random variable or random sample carries about an unknown parameter.

<u>Definition [Fisher Information in a Random Variable]</u>

Let X be a random variable whose distribution is $f(x|\theta)$ such that $(*)$

- $\Omega \subset \mathbb{R}$ is an open interval.
- the support $X = \{x \in \mathbb{R} \mid f(x|\theta) > 0\}$ is the same $\forall\, \theta \in \Omega$
- $\lambda(x|\theta) := \log f(x|\theta)$ is twice differentiable w.r.t. θ.

The Fisher information $I(\theta)$ in X is defined as

Check the support

normal μ, σ^2 | Bernoulli p | uniform $[0, \theta]$

Theorem. Under the conditions $(*)$ (that typically apply with say Ch5 special distribution),

EG. $X|p \sim$ Bernoulli p, $0 < p < 1$

EG. $X|\mu \sim$ normal μ, σ^2, μ unknown and σ^2 known.

Definition [Fisher Information in a Random Sample]

Let $X_1, \dots, X_n|\theta$ be a random sample from $f(x|\theta)$, and let $f_n(x_1, \dots, x_n|\theta)$ be the joint distribution of $X_1, \dots, X_n|\theta$. Assume $(*)$, and let

The Fisher information $I_n(\theta)$ in the random sample $X_1, \dots, X_n$ is defined as

There is a simple relation between the Fisher information $I_n(\theta)$ in the entire sample and the Fisher information $I(\theta)$ in a single observation X_i.

Theorem.

Proof.

EG $X_1, \dots, X_n|p$ ~iid Bernoulli p, $0 < p < 1 \Rightarrow I_n(p) =$

EG $X_1, \dots, X_n|\mu$ ~ normal μ, σ^2, μ unknown and σ^2 known $\Rightarrow I_n(\mu) =$

Fisher information can be used to determine a lower bound for the variance of an arbitrary estimator of the parameter θ.

<u>Cramer-Rao Lower Bound.</u> Let $T = r(X_1, \dots, X_n)$ be a statistic with finite variance and define $m(\theta) = E_\theta[T]$. Assume that $m(\theta)$ is a differentiable function of θ. Then

$$\mathrm{Var}_\theta(T) \geq$$

<u>Corollary.</u> If T is unbiased for θ, then $\mathrm{Var}_\theta(T) \geq$

<u>Proof.</u>

<u>Theorem [Asymptotic distribution of an MLE]</u> Suppose the MLE $\hat{\theta}_n$ solves $\lambda_n'(\boldsymbol{x}|\theta) = 0$ and λ_n'' and λ_n''' exist. Then, the asymptotic distribution of $\sqrt{nI(\theta)}\left(\hat{\theta}_n - \theta\right)$ is the standard normal distribution.

<u>Note</u>

Definition. An estimator T is called __________________________ for its expected value $m(\theta)$ if its variance achieves the Cramer-Rao lower bound.

EG $X_1, \ldots, X_n|\theta \sim$ iid exponential with rate θ

EG $X_1, \ldots, X_n|\theta \sim$ iid Poisson with mean θ