50.034 - Introduction to Probability and Statistics

Week 10 – Lecture 17

January–May Term, 2019

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Lecture 17 Assigned Readings: 8.1, 8.2, 8.3

Outline of Lecture

- Posterior mean
- Sampling distribution
- Chi-squared distribution
- Computations involving sample variance

Recall: An estimator is a statistic

Let $X_1, \ldots, X_n$ be observable R.V.'s whose joint distribution is parametrized by a parameter θ.

- An estimator of θ is a real-valued function $\delta(X_1, \ldots, X_n)$.
- Given δ and a vector $\mathbf{x} = (x_1, \ldots, x_n)$ of observed values, the real number $\delta(\mathbf{x})$ is called an estimate of θ.

Note: An estimator is a statistic.

- **Recall:** A statistic is a function of observable R.V.'s.

Examples of estimators:

- (Lecture 15) Bayes estimator $\delta^*(X_1, \ldots, X_n)$
  - δ^* minimizes Bayes risk over all possible estimates.
  - Given an estimate $a = \delta^*(\mathbf{x})$ and a loss function $L(x, y)$, the Bayes risk of δ^* is the expected loss $\mathbf{E}[L(\theta, a)|\mathbf{x}]$.
- (Lecture 16) Maximum likelihood estimator $\hat{\theta}(X_1, \ldots, X_n)$
  - $\hat{\theta}$ maximizes likelihood function over all possible estimates.
  - The likelihood function of θ is defined using the exact same expression for the joint condition pmf/pdf (either $p_n(\mathbf{x}|\theta)$ or $f_n(\mathbf{x}|\theta)$), but treated as a function only in terms of θ.

Posterior mean as an estimator

Let $X_1, \dots, X_n$ be observable R.V.'s whose joint distribution is parametrized by a parameter θ.

Theorem: (Lecture 15) The **Bayes estimator** of θ with respect to the **squared error loss function** $L(x, y) = (x - y)^2$ is the estimator

$$\delta^*(X_1, \dots, X_n) = \mathbf{E}[\theta | X_1, \dots, X_n].$$

- **Definition:** $\mathbf{E}[\theta | X_1, \dots, X_n]$, treated as a estimator, is called the posterior mean of θ.
- **Note:** $\mathbf{E}[\theta | X_1, \dots, X_n]$ is a function of $X_1, \dots, X_n$, similar to how we saw in Lecture 9 that $\mathbf{E}[X|Y]$ is a function of Y.

Technicality: Frequently, the posterior mean is treated as an estimator (i.e. a function of $X_1, \dots, X_n$). However, the posterior mean is also sometimes treated as an estimate, i.e. a real number $\mathbf{E}[\theta | x_1, \dots, x_n]$ given some actual observed values $x_1, \dots, x_n$. Whether the posterior mean is a function or a real number will depend on the context. For example, if the observed values are not given, then it is implicitly assumed that the posterior mean is treated as an estimator.

Example 1

Light Bulb Example Revisited: A lighting company has a new light bulb design. Before mass production, they would first like to know how long each light bulb manufactured by them would last.

Consider a statistical model consisting of observable exponential R.V.'s $X_1, \ldots, X_n$ that are conditionally iid given the parameter θ. Each X_i represents the lifespan (in hours) of the i-th light bulb.

- Suppose the prior distribution of θ is gamma with prior hyperparameters 10 and 4500.

Question: What is the posterior mean of θ?

Example 1 - Solution

Since $X_1, \ldots, X_n$ are sampled from the exponential distribution, and since the prior distribution is gamma, we know that the posterior distribution must also be gamma.

- Prior hyperparameters: $\alpha = 10$ and $\beta = 4500$.
- If the observed values are $X_1 = x_1, \ldots, X_n = x_n$, then the posterior hyperparameters are:

$$\alpha' = \alpha + \begin{pmatrix} \text{number of} \\ \text{experiments} \end{pmatrix} = 10 + n;$$

$$\beta' = \beta + \begin{pmatrix} \text{total time} \\ \text{elapsed} \end{pmatrix} = 4500 + (x_1 + \cdots + x_n).$$

Fact: The mean of a gamma R.V. with parameters α and β is $\frac{\alpha}{\beta}$.

- Hence, the posterior mean of θ (treated as an estimator) is

$$\mathbf{E}[\theta | X_1, \ldots, X_n] = \frac{10 + n}{4500 + (X_1 + \cdots + X_n)}.$$

Sampling distribution

Definition: Given a random sample $\{X_1, \ldots, X_n\}$ and any statistic $Y = \delta(X_1, \ldots, X_n)$, the distribution of Y is called the sampling distribution of Y.

- **Recall:** A random sample is a collection of **iid** R.V.'s.
- **Recall:** A statistic is a function of observable R.V.'s.
- **Note:** A statistic is a R.V. (since it is a function of R.V.'s).
- Any estimator (e.g. posterior mean) is a statistic, hence the "sampling distribution of an estimator" makes sense.

Example: Let $\{X_1, \ldots, X_n\}$ be a random sample of observable **normal** R.V.'s with mean μ and variance σ^2.

- (Lecture 16) The M.L.E. of μ is exactly the sample mean $\overline{X}_n$.
  - This estimator is usually denoted by $\hat{\mu}$.
- (Lecture 10) Since each X_i is normal, the distribution of $\overline{X}_n$ is exactly the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.
- **Consequence:** The sampling distribution of the M.L.E. $\hat{\mu}$ is the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

Example 2

Michael has a steel company. He wants to determine the tensile strength of the steel his factory manufactures, so he considers the following statistical model:

- $X_1, \ldots, X_n$ are iid observable normal R.V.'s, each with unknown mean μ and known standard deviation 0.5.
- Each X_i represents the tensile strength (in MPa or megapascals) of a randomly cut steel sample piece.

Michael decides to use the M.L.E. of μ as an estimator of μ.

Questions:

1. What is the sampling distribution of the M.L.E. of μ?
2. What is the minimum size of a random sample $\{X_1, \ldots, X_n\}$ that he should consider, so that any estimate given by the M.L.E. would be within 0.1 of the "true" mean μ with probability at least 95%?

Example 2 - Solution

1. Since the X_i's are normal, the M.L.E. of μ is precisely $\overline{X}_n$. Any linear function of normal R.V.'s is normal (Lecture 10), so the sampling distribution of $\overline{X}_n$ is normal with mean μ and variance $\frac{\sigma^2}{n} = \frac{0.25}{n} = \frac{1}{4n}$.
2. We want to find the smallest possible n such that

$$\Pr(|\overline{X}_n - \mu| \leq 0.1) \geq 0.95,$$

or equivalently, such that $\Pr(-0.1 \leq \overline{X}_n - \mu \leq 0.1) \geq 0.95$.
Let $Z = 2\sqrt{n}(\overline{X}_n - \mu)$, and note that $Z \sim N(0, 1)$. Thus

$$\begin{aligned} \Pr(-0.1 \leq \overline{X}_n - \mu \leq 0.1) &= \Pr(-0.2\sqrt{n} \leq Z \leq 0.2\sqrt{n}) \\ &= \Pr(Z \leq 0.2\sqrt{n}) - \Pr(Z \leq -0.2\sqrt{n}) \\ &= \Phi(0.2\sqrt{n}) - (1 - \Phi(0.2\sqrt{n})) \\ &= 2 \cdot \Phi(0.2\sqrt{n}) - 1, \end{aligned}$$

where $\Phi(z)$ denotes the standard normal cdf.

Example 2 - Solution (continued)

We want $2 \cdot \Phi(0.2\sqrt{n}) - 1 \geq 0.95$, i.e. that $\Phi(0.2\sqrt{n}) \geq 0.975$. Checking the standard normal table, the closest value we can find for z satisfying $\Phi(z) = 0.975$ is $z = 1.96$.
Hence, $0.2\sqrt{n} \geq 1.96$, which implies that $n \geq 96.04$. Therefore, Michael should consider a sample size of at least 97.

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}u^2\right) du$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0.00	0.5000	0.60	0.7257	1.20	0.8849	1.80	0.9641	2.40	0.9918
0.01	0.5040	0.61	0.7291	1.21	0.8869	1.81	0.9649	2.41	0.9920
0.02	0.5080	0.62	0.7324	1.22	0.8888	1.82	0.9656	2.42	0.9922
0.03	0.5120	0.63	0.7357	1.23	0.8907	1.83	0.9664	2.43	0.9925
0.04	0.5160	0.64	0.7389	1.24	0.8925	1.84	0.9671	2.44	0.9927
0.05	0.5199	0.65	0.7422	1.25	0.8944	1.85	0.9678	2.45	0.9929
0.06	0.5239	0.66	0.7454	1.26	0.8962	1.86	0.9686	2.46	0.9931
0.07	0.5279	0.67	0.7486	1.27	0.8980	1.87	0.9693	2.47	0.9932
0.08	0.5319	0.68	0.7517	1.28	0.8997	1.88	0.9699	2.48	0.9934
0.09	0.5359	0.69	0.7549	1.29	0.9015	1.89	0.9706	2.49	0.9936
0.10	0.5398	0.70	0.7580	1.30	0.9032	1.90	0.9713	2.50	0.9938
0.11	0.5438	0.71	0.7611	1.31	0.9049	1.91	0.9719	2.52	0.9941
0.12	0.5478	0.72	0.7642	1.32	0.9066	1.92	0.9726	2.54	0.9945
0.13	0.5517	0.73	0.7673	1.33	0.9082	1.93	0.9732	2.56	0.9948
0.14	0.5557	0.74	0.7704	1.34	0.9099	1.94	0.9738	2.58	0.9951
0.15	0.5596	0.75	0.7734	1.35	0.9115	1.95	0.9744	2.60	0.9953
0.16	0.5636	0.76	0.7764	1.36	0.9131	1.96	0.9750	2.62	0.9956
0.17	0.5675	0.77	0.7794	1.37	0.9147	1.97	0.9756	2.64	0.9959
0.18	0.5714	0.78	0.7823	1.38	0.9162	1.98	0.9761	2.66	0.9961
0.19	0.5753	0.79	0.7852	1.39	0.9177	1.99	0.9767	2.68	0.9963

Why the sampling distribution is important

As Example 2 shows, if we know the sampling distribution of an estimator $\hat{\theta}$ (of a random sample $\{X_1, \ldots, X_n\}$), then we can use this sampling distribution to determine a suitable sample size:

- e.g. minimum sample size so that any estimate given by $\hat{\theta}$ is "close" to the "true" value of θ with "high probability".
  - Note: We need to specify thresholds for what we mean by "close" and "high probability".

Practical Considerations: In many real-world statistical models, estimators are used to estimate a certain value of interest (e.g. tensile strength of steel, lifespan of light bulb).

- The larger the random sample, the better the estimate.
- The larger the random sample, the higher the cost incurred.

Trade-off: cost-effectiveness versus accuracy of estimate.

Thus, by knowing the sampling distribution of an estimator **before** carrying out experiments on a random sample, we can find the smallest sample size required for any given threshold for accuracy.

Sampling distribution of sample variance

Let $\{X_1, \ldots, X_n\}$ be a random sample of observable **normal** R.V.'s with mean μ and variance σ^2.

Theorem: (Lecture 16) The maximum likelihood estimator of the variance σ^2 is the sample variance, which is defined by

$$\hat{\sigma}^2(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Notation: The sample variance is usually denoted by $\hat{\sigma}^2$ or $\widehat{\sigma^2}$.

What we know so far:

- The M.L.E. of the mean μ is the sample mean $\overline{X}_n$.
  - The sampling distribution of $\overline{X}_n$ is normal.
- The M.L.E. of the variance σ^2 is the sample variance $\hat{\sigma}^2$.
  - **Question:** What is the sampling distribution of $\hat{\sigma}^2$?

To find the sampling distribution of sample variance, we need a new class of distributions called χ^2 **distributions**.

Chi-squared distribution

The χ^2 distribution is a special kind of gamma distribution.

- χ is a greek letter spelled in English as "chi".
- When pronounced, χ sounds like "kai".

Definition: A continuous R.V. X is called chi-squared (or χ^2) if its pdf is given by:

$$f(x) = \begin{cases} \dfrac{1}{2^{m/2}\Gamma(\frac{m}{2})} x^{(m/2)-1} e^{-0.5x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0; \end{cases}$$

for some positive integer m, which is called the degree of freedom.

- We say that X is the χ^2 R.V. with m degrees of freedom.
- Its distribution is called χ^2 distribution.
- This distribution is **exactly the same** as the gamma distribution with parameters $\alpha = \frac{m}{2}$ and $\beta = \frac{1}{2}$.

Common Notation: We write $X \sim \chi^2(m)$ to mean that X is a χ^2 R.V. with m degrees of freedom.

Special cases of chi-squared distribution

Special Case: The following distributions are all exactly the same:

- The χ^2 distribution with two degrees of freedom.
- The gamma distribution with parameters $\alpha = 1$ and $\beta = \frac{1}{2}$.
- The exponential distribution with parameter $\frac{1}{2}$ (i.e. mean $= 2$).

Another Special Case: χ^2 distribution with one degree of freedom.

- The pdf of a χ^2 R.V. X with one degree of freedom is

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} x^{-0.5} e^{-0.5x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0; \end{cases}$$

- **Useful Fact:** $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
  - This is an amazing fact! $\Gamma(n) = (n-1)!$ for positive integers n, so this function that generalizes factorials is also related to π.

More about chi-squared distributions

Useful properties:

- $\mathbf{E}[X] = m$ and $\mathrm{var}(X) = 2m$.
  - Recall that if Y is a gamma R.V. with parameters α and β, then $\mathbf{E}[Y] = \frac{\alpha}{\beta}$ and $\mathrm{var}(Y) = \frac{\alpha}{\beta^2}$.
  - Recall that a χ^2 R.V. with m degrees of freedom is precisely a gamma R.V. with parameters $\frac{m}{2}$ and $\frac{1}{2}$.
- Moment generating function: $\psi(t) = \left(\frac{1}{1-2t}\right)^{m/2}$, for $t < \frac{1}{2}$.

Common Uses of χ^2 distribution:

- To model the sample variances of random samples of normal R.V.'s, to be used as estimates for the actual variances, e.g. of the following:
  - Heights of individuals in a homogeneous population of people.
  - Weights of Fuji apples harvested from Fujisaki, Japan.
  - Tensile strength of pieces of steel produced in a factory.
- To model how well the observed data fits a given distribution
  - This is called "goodness of fit" (covered later in Lecture 22).

χ^2 distribution versus standard normal distribution

Theorem: If $Z \sim N(0,1)$ is the standard normal R.V., then the R.V. $Y = Z^2$ has the χ^2 distribution with one degree of freedom.

Proof: Let $f(y)$, $F(y)$ be the pdf and cdf of Y respectively. Also, let $\phi(z)$ and $\Phi(z)$ be the pdf and cdf of Z respectively. For $y > 0$,

$$F(y) = \Pr(Z^2 \le y) = \Pr(-\sqrt{y} \le Z \le \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(\sqrt{y}).$$

Since $f(y) = F'(y)$ and $\phi(z) = \Phi'(z)$, the chain rule yields:

$$f(y) = \phi(\sqrt{y})(\tfrac{1}{2\sqrt{y}}) - \phi(\sqrt{y})(-\tfrac{1}{2\sqrt{y}}) \quad \text{(for } y > 0\text{)}.$$

By the definition of the pdf of Z,

$$\phi(\sqrt{y}) = \phi(-\sqrt{y}) = \tfrac{1}{\sqrt{2\pi}} e^{-0.5y},$$

therefore

$$f(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-0.5} e^{-0.5y}, & \text{if } y > 0; \\ 0, & \text{otherwise;} \end{cases}$$

which is exactly the pdf of the χ^2 R.V. with one degree of freedom.

List of results involving χ^2 distribution

Theorem: (previous slide) If $Z \sim N(0,1)$ is the standard normal R.V., then $Z^2 \sim \chi^2(1)$, i.e. Z^2 has the χ^2 distribution with one degree of freedom.

Theorem: Let $Y_1, \dots, Y_n$ be **independent** R.V.'s, such that $Y_i \sim \chi^2(m_i)$ for each $1 \leq i \leq m$. Then the sum $Y_1 + \cdots + Y_n$ has the χ^2 distribution with $m_1 + \cdots + m_n$ degrees of freedom.

Corollary: Let $Z_1, \dots, Z_n$ be iid **standard normal** R.V.'s. Then $(Z_1^2 + \cdots + Z_n^2) \sim \chi^2(n)$, i.e. $Z_1^2 + \cdots + Z_n^2$ has the χ^2 distribution with n degrees of freedom.

Corollary: Let $X_1, \dots, X_n$ be iid **normal** R.V.'s with mean μ and variance σ^2. Then $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n)$, i.e. it has the χ^2 distribution with n degrees of freedom.

Computing the cdf of a χ^2 R.V.

If $X \sim \chi^2(m)$, then by definition, its cdf is

$$F_X(x) = \begin{cases} \int_{-\infty}^{x} \frac{1}{2^{m/2}\Gamma(\frac{m}{2})} u^{(m/2)-1} e^{-0.5u}\, du, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

- **Problem:** This integral (for the case $x > 0$) has no closed form formula!
- Hence, the cdf $F_X(x)$ can only be computed approximately, using integral approximation methods (e.g. trapezoidal rule).

How your calculator/software package "computes" $\mathbf{F_X(x)}$ today:

- Approximations to $F_X(x)$ for various possible x and different degrees of freedom m are calculated beforehand, then stored in the calculator/software package in lookup tables.
- e.g. when you query "For $X \sim \chi^2(10)$, what value of x satisfies $F_X(x) = 0.95$?", the stored value for the closest value of x found would be retrieved.
  - Queries are possible only for those stored values of x and m.

Table for values of the χ^2 distribution

(Before there were computers:) Computed values are stored as tables of values. For example, there is a table at the back of the course textbook called the "Table of the χ^2 Distribution".

- For example, if $X \sim \chi^2(11)$, then the closest value found for x satisfying $F_X(x) = 0.025$ is $x \approx 3.816$.

	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	$\boxed{3.816}$	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

Example 3

Suppose $Y \sim \chi^2(10)$, i.e. Y has the χ^2 distribution with 10 degrees of freedom. Determine an approximate value of x that satisfies $\Pr(Y \leq x) = 0.4$.

Solution: From the table, the closest value is $x \approx 8.295$.

	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	$\boxed{8.295}$
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

Sample mean and sample variance

Let $\{X_1, \ldots, X_n\}$ be a random sample of observable **normal** R.V.'s with mean μ and variance σ^2.

- Let $\hat{\mu}$ and $\hat{\sigma}^2$ denote the sample mean and sample variance respectively, of the random sample $\{X_1, \ldots, X_n\}$.

Theorem: $\hat{\mu}$ and $\hat{\sigma}^2$ are **independent** R.V.'s.

- **Important Note:** This theorem is false if the random sample consists of non-normal R.V.'s.
- **Fact:** The sample mean and sample variance are independent **only** when the R.V.'s in the random sample are normal!

Theorem: $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \hat{\mu})^2 = \frac{n\hat{\sigma}^2}{\sigma^2}$ has the χ^2 distribution with $n-1$ degrees of freedom.

- **Note:** $\frac{n\hat{\sigma}^2}{\sigma^2}$ equals the sample variance divided by the variance of the sample mean.

Underestimation of sample variance

Let $\{X_1, \ldots, X_n\}$ be a random sample of observable **normal** R.V.'s with mean μ and variance σ^2. Also, let $\hat{\mu}$ and $\hat{\sigma}^2$ be the sample mean and sample variance respectively, of the random sample $\{X_1, \ldots, X_n\}$.

Recall: In Lecture 16, we mentioned that $\hat{\sigma}^2$ consistently underestimates σ^2 for finite samples, although its deviation from σ^2 approaches 0 as $n \to \infty$.

- For this reason, we say that $\hat{\sigma}^2$ is the biased sample variance or the uncorrected sample variance.

Question: Given some sample size n, to what extent will the sample variance $\hat{\sigma}^2$ underestimate the "true" variance σ^2?

- In other words, we want to understand the R.V. $\frac{\hat{\sigma}^2}{\sigma^2}$.
  - If $\frac{\hat{\sigma}^2}{\sigma^2} \approx 1$, then the sample variance is approximately σ^2.
  - If $\frac{\hat{\sigma}^2}{\sigma^2} < 1$ by a significant margin, then the sample variance underestimates the "true" variance by a significant margin.
  - If $\frac{\hat{\sigma}^2}{\sigma^2} > 1$ by a significant margin, then the sample variance overestimates the "true" variance by a significant margin.

Example 4

Let $\{X_1, \ldots, X_{10}\}$ be a random sample of 10 observable **normal** R.V.'s with mean μ and variance σ^2. Let $\hat{\sigma}^2$ denote the sample variance of this random sample $\{X_1, \ldots, X_{10}\}$.

What is the probability that the sample variance $\hat{\sigma}^2$ will underestimate the variance σ^2 by at least 26.4%?

Hint: How is $\frac{\hat{\sigma}^2}{\sigma^2}$ related to the χ^2 distribution?

Example 4 - Solution

Key Idea: Since our random sample has sample size 10, the R.V. $\frac{10\hat{\sigma}^2}{\sigma^2}$ has the χ^2 distribution with 9 degrees of freedom.
Note that $1 - 0.264 = 0.736$. We want to find some value p such that $\Pr(\frac{\hat{\sigma}^2}{\sigma^2} \leq 0.736) \approx p$, or equivalently, such that

$$\Pr\left(\frac{10\hat{\sigma}^2}{\sigma^2} \leq 7.36\right) \approx p.$$

From the table, $\Pr\left(\frac{10\hat{\sigma}^2}{\sigma^2} \leq 7.36\right) \approx 0.40$. Therefore, there is 40% probability that $\hat{\sigma}^2$ will underestimate σ^2 by at least 26.4%.

	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	$\boxed{7.357}$
10	2.156	2.558	3.247	3.940	4.865	6.179	6.737	7.267	8.295

Example 5

Let $\{X_1, \ldots, X_n\}$ be a random sample of n observable **normal** R.V.'s with mean μ and variance σ^2. Let $\hat{\sigma}^2$ denote the sample variance of this random sample $\{X_1, \ldots, X_n\}$.

Determine the smallest values for the sample size n for which the following relations are satisfied:

1. $\Pr\left(\frac{\hat{\sigma}^2}{\sigma^2} \leq 1.7\right) \geq 0.95$.
2. $\Pr\left(|\hat{\sigma}^2 - 1.35\sigma^2| \leq \frac{1}{2}\sigma^2\right) \geq 0.45$.

Example 5 - Solution

1. The R.V. $\frac{n\hat{\sigma}^2}{\sigma^2}$ has a χ^2 distribution with $(n-1)$ degrees of freedom, i.e. $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$.

Note: $\Pr\left(\frac{\hat{\sigma}^2}{\sigma^2} \leq 1.7\right) \geq 0.95 \Longleftrightarrow \Pr\left(\frac{n\hat{\sigma}^2}{\sigma^2} \leq 1.7n\right) \geq 0.95.$

Hence, what we need to do is to find a value for n (from the table of values of the χ^2 distribution) satisfying the following:

- $\Pr\left(\frac{n\hat{\sigma}^2}{\sigma^2} \leq 1.7n\right) \geq 0.95.$
- $\Pr\left(\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \leq 1.7(n-1)\right) < 0.95.$

Example 5 - Solution

1. (continued) Looking through the values in the table, we find:
   - $\Pr\left(\frac{10\hat{\sigma}^2}{\sigma^2} \leq 1.7 \times (10)\right) > \Pr\left(\frac{10\hat{\sigma}^2}{\sigma^2} \leq 16.92\right) = 0.95.$
     - **Note:** $1.7 \times (10) = 17$; and $\frac{10\hat{\sigma}^2}{\sigma^2} \sim \chi^2(9)$.
   - $\Pr\left(\frac{9\hat{\sigma}^2}{\sigma^2} \leq 1.7 \times (9)\right) < \Pr\left(\frac{9\hat{\sigma}^2}{\sigma^2} \leq 15.51\right) = 0.95.$
     - **Note:** $1.7 \times (9) = 15.3$; and $\frac{9\hat{\sigma}^2}{\sigma^2} \sim \chi^2(8)$.

Therefore, the smallest sample size is $n = 10$.

	p							
.50	.60	.70	.75	.80	.90	.95	.975	.9
.4549	.7083	1.074	1.323	1.642	2.706	3.841	5.024	6.6
1.386	1.833	2.408	2.773	3.219	4.605	5.991	7.378	9.2
2.366	2.946	3.665	4.108	4.642	6.251	7.815	9.348	11.3
3.357	4.045	4.878	5.385	5.989	7.779	9.488	11.14	13.2
4.351	5.132	6.064	6.626	7.289	9.236	11.07	12.83	15.0
5.348	6.211	7.231	7.841	8.558	10.64	12.59	14.45	16.8
6.346	7.283	8.383	9.037	9.803	12.02	14.07	16.01	18.4
7.344	8.351	9.524	10.22	11.03	13.36	$\boxed{15.51}$	17.53	20.0
8.343	9.414	10.66	11.39	12.24	14.68	$\boxed{16.92}$	19.02	21.6
9.342	10.47	11.78	12.55	13.44	15.99	18.31	20.48	23.2

Example 5 - Solution

2. Same as in the previous part, we know that $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$. Notice that

$$\Pr\left(|\hat{\sigma}^2 - 1.35\sigma^2| \leq \tfrac{1}{2}\sigma^2\right) \geq 0.45$$

$$\iff \Pr\left(|\tfrac{\hat{\sigma}^2}{\sigma^2} - 1.35| \leq \tfrac{1}{2}\right) \geq 0.45$$

$$\iff \Pr\left(-0.5 \leq \tfrac{\hat{\sigma}^2}{\sigma^2} - 1.35 \leq 0.5\right) \geq 0.45$$

$$\iff \Pr\left(0.85 \leq \tfrac{\hat{\sigma}^2}{\sigma^2} \leq 1.85\right) \geq 0.45$$

$$\iff \Pr\left(0.85n \leq \tfrac{n\hat{\sigma}^2}{\sigma^2} \leq 1.85n\right) \geq 0.45$$

Hence, what we need to do is to find a value for n (from the table of values of the χ^2 distribution) satisfying the following:

$$\Pr\left(\tfrac{n\hat{\sigma}^2}{\sigma^2} \leq 1.85n\right) - \Pr\left(\tfrac{n\hat{\sigma}^2}{\sigma^2} \leq 0.85n\right) \geq 0.45.$$

$$\Pr\left(\tfrac{(n-1)\hat{\sigma}^2}{\sigma^2} \leq 1.85(n-1)\right) - \Pr\left(\tfrac{(n-1)\hat{\sigma}^2}{\sigma^2} \leq 0.85(n-1)\right) < 0.45.$$

Example 5 - Solution

2. (continued) Looking through the values in the table, we find:

$$\Pr\left(\frac{6\hat{\sigma}^2}{\sigma^2} \leq 1.85 \times (6) = 11.1\right) - \Pr\left(\frac{6\hat{\sigma}^2}{\sigma^2} \leq 0.85 \times (6) = 5.1\right)$$
$$> \Pr\left(\frac{6\hat{\sigma}^2}{\sigma^2} \leq 11.07\right) - \Pr\left(\frac{6\hat{\sigma}^2}{\sigma^2} \leq 5.132\right) = 0.95 - 0.5 = 0.45.$$

$$\Pr\left(\frac{5\hat{\sigma}^2}{\sigma^2} \leq 1.85 \times (5) = 9.25\right) - \Pr\left(\frac{5\hat{\sigma}^2}{\sigma^2} \leq 0.85 \times (5) = 4.25\right)$$
$$< \Pr\left(\frac{5\hat{\sigma}^2}{\sigma^2} \leq 9.488\right) - \Pr\left(\frac{5\hat{\sigma}^2}{\sigma^2} \leq 4.045\right) = 0.95 - 0.5 = 0.45.$$

Therefore, the smallest sample size is $n = 6$.

				p				
.50	.60	.70	.75	.80	.90	.95	.975	.9
.4549	.7083	1.074	1.323	1.642	2.706	3.841	5.024	6.6
1.386	1.833	2.408	2.773	3.219	4.605	5.991	7.378	9.2
2.366	2.946	3.665	4.108	4.642	6.251	7.815	9.348	11.3
3.357	$\boxed{4.045}$	4.878	5.385	5.989	7.779	$\boxed{9.488}$	11.14	13.2
4.351	$\boxed{5.132}$	6.064	6.626	7.289	9.236	$\boxed{11.07}$	12.83	15.0
5.348	6.211	7.231	7.841	8.558	10.64	12.59	14.45	16.8
6.346	7.283	8.383	9.037	9.803	12.02	14.07	16.01	18.4

Summary

- Posterior mean
- Sampling distribution
- Chi-squared distribution
- Computations involving sample variance