Probability and Random Process (SWE3026)

Statistical Inference

JinYeong Bak
jy.bak@skku.edu
College of Computing, SKKU

H. Pishro-Nik, "Introduction to probability, statistics, and random processes", available at https://www.probabilitycourse.com, Kappa Research LLC, 2014.

Objectives

Instead of estimating a parameter pointwise such as

$$\widehat{\boldsymbol{\theta}} = \mathbf{34.25}$$

We might report the interval

$$\left[\widehat{\boldsymbol{\theta}}_{\boldsymbol{l}}, \widehat{\boldsymbol{\theta}_{\boldsymbol{h}}}\right] = [\mathbf{30.69}, \mathbf{37.81}]$$

with a confidence level that shows how confident we are about the interval

Interval Estimation

Definition

An interval estimator with confidence level $1 - \alpha$ consists of two estimators $\widehat{\Theta}_l(X_i)$ and $\widehat{\Theta}_h(X_i)$ such that

$$P\left(\widehat{\Theta}_l \le \theta \text{ and } \widehat{\Theta}_h \ge \theta\right) \ge 1 - \alpha$$

for every possible value of θ.

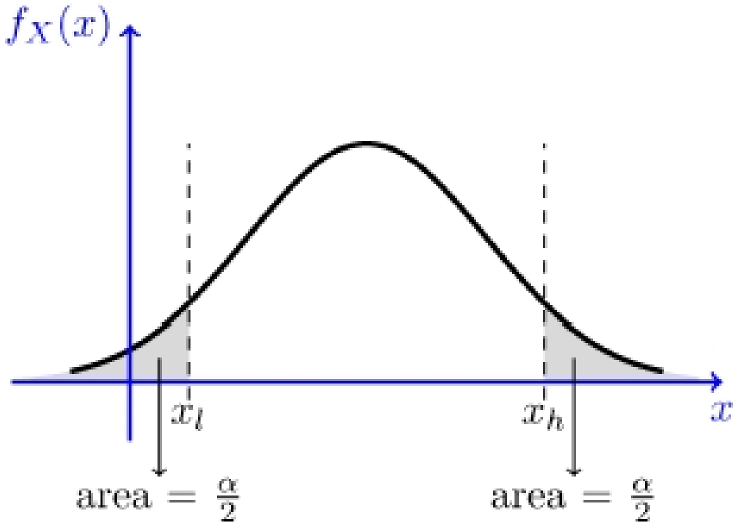
Interval Estimation

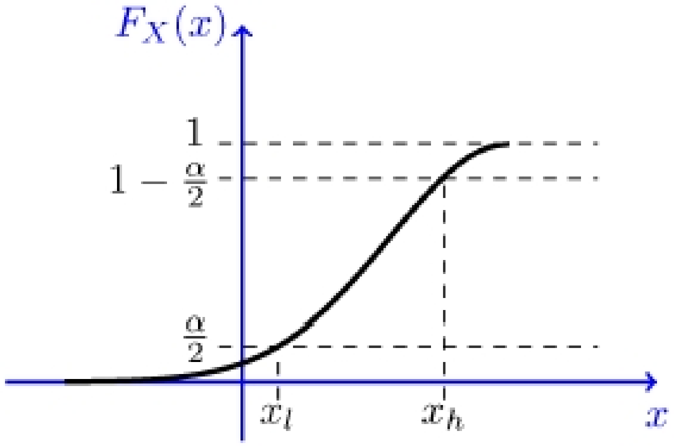
$$P(x_l \le X \le x_h) = 1 - \alpha$$

$$\Rightarrow P(X \le x_l) = \frac{\alpha}{2}, \text{ and } P(X \ge x_h) = \frac{\alpha}{2}$$

$$\Rightarrow F_X(x_l) = \frac{\alpha}{2}, \text{ and } F_X(x_h) = 1 - \frac{\alpha}{2}$$

$$\Rightarrow x_l = F_X^{-1}\left(\frac{\alpha}{2}\right), \text{ and } x_h = F_X^{-1}\left(1 - \frac{\alpha}{2}\right)$$





Example

Let $Z \sim N(0, 1)$, find x_l and x_h such that $P(x_l \leq Z \leq x_h) = 0.95$

z_p

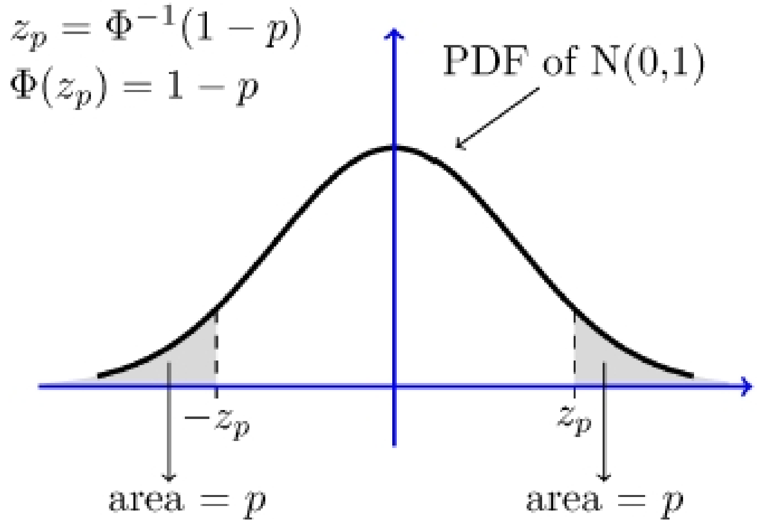
Definition

Let $Z \sim N(0,1)$. For any $p \in [0,1]$,

$$P\left(Z > z_p\right) = p$$

$$\Phi\left(\mathrm{z_p}\right) = 1 - \mathrm{p}, \mathrm{z_p} = \Phi^{-1}(1 - \mathrm{p})$$

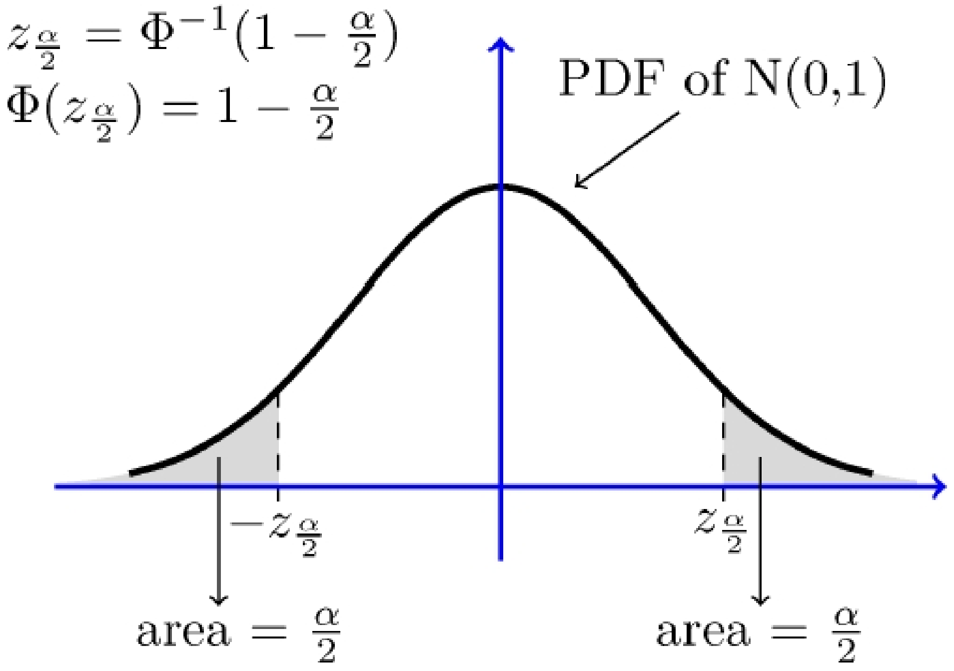
$$\mathrm{z}_{1-\mathrm{p}} = -\mathrm{z}_p$$



Interval Estimation

$(1-\alpha)$ **interval for the standard normal random variable** Z

$$P\left(-z_{\frac{\alpha}{2}} \le Z \le z_{\frac{\alpha}{2}}\right) = 1-\alpha$$



Pivotal Quantity

Definition

The random variable Q

1. **It is a function of the observed data and the unknown parameter θ**
2. **It does not depend on any other unknown parameters**
3. **The probability distribution of Q does not depend on θ or any other unknown parameters**

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with known variance $Var(X_i) = \sigma^2$, and unknown mean $E[X_i] = \theta$. Find a $(1 - \alpha)$ confidence interval for θ. Assume that n is large.

Example

We would like to estimate the portion of people who plan to vote for Candidate A in an upcoming election. It is assumed that the number of voters is large, and θ is the portion of voters who plan to vote for Candidate A. We define the random variable X as follows. A voter is chosen uniformly at random among all voters and we ask her/him: "Do you plan to vote for Candidate A?" If she/he says "yes," then $X = 1$, otherwise $X = 0$. Then, $X \sim Bernoulli(\theta)$.

Let $X_1, X_2, \ldots, X_n$ be a random sample from this distribution, which means that the X_i's are i.i.d. and $X_i \sim Bernoulli(\theta)$. In other words, we randomly select n voters (with replacement) and we ask each of them if they plan to vote for Candidate A. Find a $(1 - \alpha)$ confidence interval for θ based on $X_1, X_2, \ldots, X_n$

We would like to estimate the portion of people who plan to vote for Candidate A in an upcoming election. It is assumed that the number of voters is large, and θ is the portion of voters who plan to vote for Candidate A. We define the random variable X as follows. A voter is chosen uniformly at random among all voters and we ask her/him: "Do you plan to vote for Candidate A?" If she/he says "yes," then $X = 1$, otherwise $X = 0$. Then, $X \sim Bernoulli(\theta)$.

Let $X_1, X_2, \ldots, X_n$ be a random sample from this distribution, which means that

the X_i's are i.i.d. and $X_i \sim Bernoulli(\theta)$. In other words, we randomly select n voters (with replacement) and we ask each of them if they plan to vote for Candidate A. Find a $(1-\alpha)$ confidence interval for θ based on $X_1, X_2, \ldots, X_n$

Example

There are two candidates in a presidential election: Candidate A and Candidate B. Let θ be the portion of people who plan to vote for Candidate A. Our goal is to find a confidence interval for θ. Specifically, we choose a random sample (with replacement) of n voters and ask them if they plan to vote for Candidate A. Our goal is to estimate the θ such that the margin of error is 3 percentage points. Assume a 95% confidence level. That is, we would like to choose n such that $P(\bar{X} - 0.03 \leq \theta \leq \bar{X} + 0.03) \geq 0.95$ where $\bar{X}$ is the portion of people in our random sample that say they plan to vote for Candidate A. How large does n need to be?

There are two candidates in a presidential election: Candidate A and Candidate B. Let θ be the portion of people who plan to vote for Candidate A. Our goal is to find a confidence interval for θ. Specifically, we choose a random sample (with replacement) of n voters and ask them if they plan to vote for Candidate A. Our goal is to estimate the θ such that the margin of error is 3 percentage points. Assume a 95% confidence level. That is, we would like to choose n such that $P(\overline{X} - 0.03 \leq \theta \leq \overline{X} + 0.03) \geq 0.95$ where $\overline{X}$ is the portion of people in our random sample that say they plan to vote for Candidate A. How large does n need to be?

Chi-Squared Distribution

Definition

If $Z_1, Z_2, \ldots Z_n$ are independent standard normal R.V, the R.V Y defined as

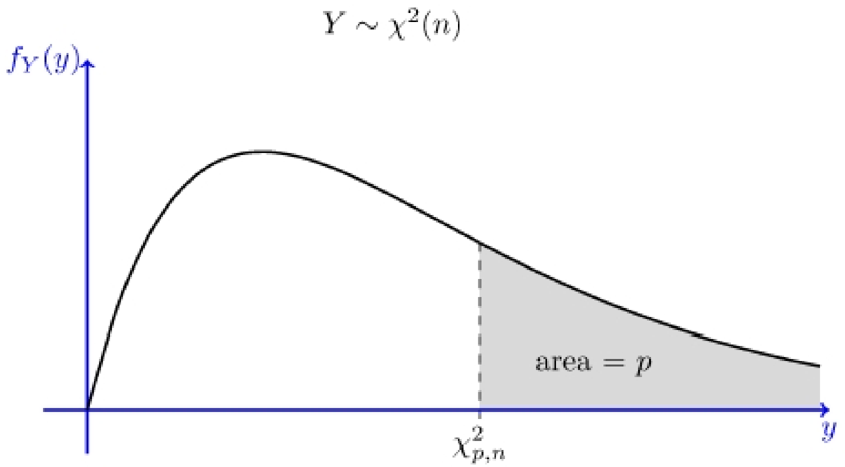
$$Y = Z_1^2 + Z_2^2 + \cdots + Z_n^2, Y \sim \chi^2(n)$$

Properties

- **The chi-squared distribution is a special case of the gamma distribution.**

$$Y \sim Gamma(\frac{n}{2}, \frac{1}{2})$$

- $E[Y] = n, Var(Y) = 2n$
- $P(Y > \chi^2_{p,n}) = p$



Chi-Squared Distribution

Let $X_1, X_2, \dots, X_n$ be i.i.d. $N(\mu, \sigma)$ random variables.

Let S^2 be the standard variance for this random sample.

Then, the random variable Y defined as

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

$$Y \sim \chi^2(n-1)$$

t-Distribution

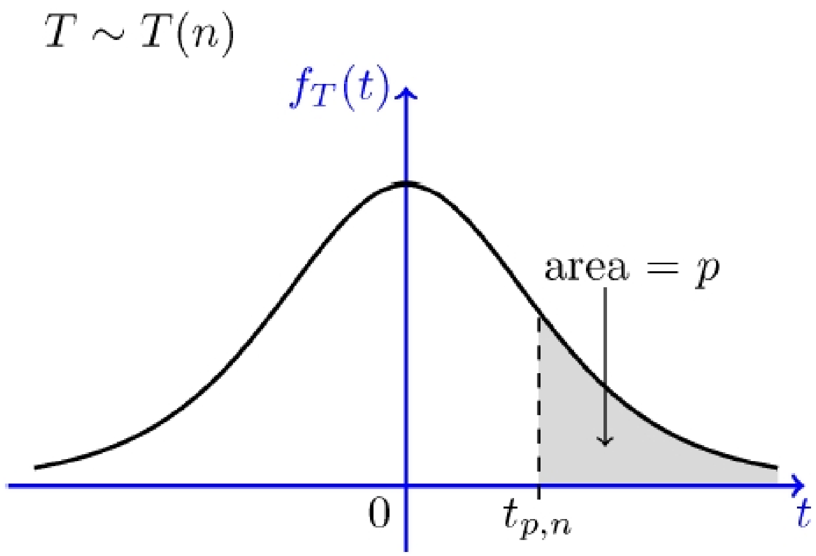
Definition

Let $Z \sim N(0,1)$ and $Y \sim \chi^2(n)$. Z and Y are independent. The R.V T defined as

$$T = \frac{Z}{\sqrt{Y/n}}, T \sim T(n)$$

Properties

- $E[T] = 0(n > 0, n \neq 1), Var(T) = \frac{n}{n-2}(n > 2)$
- $T(n) \to N(0,1)$ when n becomes large
- $P(T > t_{p,n}) = p$



t-Distribution

Let $X_1, X_2, \dots, X_n$ be i.i.d. $N(\mu, \sigma)$ random variables.

Let S^2 be the standard variance for this random sample.

Then, the random variable T defined as

$$\boldsymbol{T = \frac{\bar{X} - \mu}{S/\sqrt{n}}}$$

$$\boldsymbol{T \sim T(n-1)}$$

Confidence Intervals for the Mean of Normal R.V

Let $X_1, X_2, \dots, X_n$ be i.i.d. $N(\mu, \sigma)$ random variables

Let's find an interval estimator for μ

- **When we know the value of σ^2**
- **When we do not know the value of σ^2**

When we know the value of σ^2

Define Q

$$Q = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$Q \sim N(0, 1)$

Q is a pivotal quantity

$$[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$$

is a $(1 - \alpha)$ confidence interval for μ

When we do not know the value of σ^2

Define T

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

$T \sim T(n-1)$

T is a pivotal quantity

$$P\left(-t_{\frac{\alpha}{2},n-1} \leq T \leq t_{\frac{\alpha}{2},n-1}\right) = 1 - \alpha$$

$$[\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{S}{\sqrt{n}}]$$

is a $(1-\alpha)$ confidence interval for μ

Example

A farmer weighs 10 randomly chosen watermelons from his farm and he obtains the following values (in lbs):

7.72 9.58 12.38 7.77 11.27 8.80 11.10 7.80 10.17 6.00

Assuming that the weight is normally distributed with mean μ and variance σ^2, find a 95% confidence interval for μ.

Confidence Intervals for the Variance of Normal R.V

Let $X_1, X_2, \dots, X_n$ be i.i.d. $N(\mu, \sigma)$ random variables
Let's find an interval estimator for σ, We assume that μ is also unknown

Define Y

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

$Y \sim \chi^2(n-1)$
Y is a pivotal quantity

$$P\left(\chi^2_{1-\frac{\alpha}{2},n-1} \le Y \le \chi^2_{\frac{\alpha}{2},n-1}\right) = 1 - \alpha$$

$$\left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right]$$

is a $(1-\alpha)$ confidence interval for σ^2

Example

A farmer weighs 10 randomly chosen watermelons from his farm and he obtains the following values (in lbs):

7.72 9.58 12.38 7.77 11.27 8.80 11.10 7.80 10.17 6.00

Assuming that the weight is normally distributed with mean μ and variance σ^2, find a 95% confidence interval for σ^2 where μ and σ^2 are unknown.