STAT2006 Basic Concepts in Statistics and Probability II

Review of Normal Distribution

Chan Chun Ho, Benjamin*†

January 21, 2018

Abstract

It aims to review normal distribution, which can be considered to be the most important probability distribution in Statistics. It serves as optional reading materials. It would not be covered in tutorials but some of the results will be used directly in subsequent tutorials. See you starting from tutorial 5! Some materials are extracted from STAT2001 lecture note and do credit to Dr Ho Kwok Wah while some are extracted from textbooks "Probability and Statistical Inference" (Hogg and Tanis) used in STAT2001/2006 and "Statistical Inference" (Casella and Berger) used in STAT4003.

Suggestion for future study: If students are interested in statistical theory, they should consider to take STAT4003. There are other popular Statistics courses, including but not limited to STAT3008 Regression and 4005 Time Series. Statistics and Risk Management Science are closely related subjects. RMSC2001 is a good starting point.

Notations and Definitions

- Set membership: $x \in A$ means "x is an element of the set A".
- Set of natural numbers: $\mathbb{N} = \{1, 2, \dots\}$; Interval: $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$
- Set of real numbers: $\mathbb{R} = (-\infty, \infty)$; Absolute value: $|x| = x$ if $x \geq 0$, $|x| = -x$ if $x < 0$
- Differentiation: $\frac{d}{dx} f(x)$; Integration: $\int f(x)dx$
- Expectation: $E(\cdot)$; Variance: $Var(\cdot)$; Standard deviation: $SD(\cdot)$
- Moment-generating function (mgf): $M_X(t)$ / $M_Z(t)$
- Probability distribution: A tilde ($\sim$) means "has the probability distribution of".
- Parameter(s): θ denotes population characteristic(s) that can be set to different values to produce different probability distributions.
- Parameter space: Ω denotes the set of all possible values for all the different parameters in a distribution.

*For enquiry, please email to 1155049861@link.cuhk.edu.hk.
†Personal profile: www.linkedin.com/in/benjamin-chan-chun-ho

1 Introduction

1.1 Continuous Random Variables

- A **random variable** is a function that maps each element in a sample space into a real number $x \in \mathbb{R}$.
- A random variable X is said to be **continuous** if the **cumulative distribution function (cdf)** $F(x) = P(X \leq x)$ is a continuous function for all $x \in \mathbb{R}$.

1.2 Continuous Probability Distributions

- For a continuous random variable X, the **probability density function (pdf)** $f(x)$ has to satisfy the following conditions:
  1. $f(x) \geq 0$
  2. $P(X = x) = 0$ for any x
  3. $P(a \leq X \leq b) = \int_a^b f(x)dx$
  4. $\int_{-\infty}^{\infty} f(x)dx = 1$

  The pdf is a non-negative function, which has no direct probability interpretation, that is $f(x) \neq P(X = x)$. From (1) and (2), $f(x)$ can be greater than 1 and the probability of X at any point x is 0. From (3), the probability of X falling in $[a, b]$ is given by integrating $f(x)$ over $[a, b]$. Intuitively, this probability is given by the area bounded by the function, the x-axis and the vertical lines $x = a$ and $x = b$. In fact, $F(X) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$. By Fundamental Theorem of Calculus, $f(x) = \frac{d}{dx}F(x)$. Informally, "density" means the rate of change in cumulative probability.

1.3 Expectation and Variance of Continuous Random Variables

- The **mean**, or **expectation**, or **expected value** of X, $g(X)$ and e^{tX} are

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \quad \text{and} \quad M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx}f(x)dx$$

- The **variance** of X is

$$\sigma^2 = Var(X) = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 = E(X^2) - \mu^2$$

- The **standard deviation** of X is simply $\sigma = SD(X) = \sqrt{Var(X)}$.

2 Normal Distribution

A **normal random variable** X with parameters (μ, σ^2), denoted by $X \sim N(\mu, \sigma^2)$, where $\Omega = \{(\mu, \sigma^2) : \mu \in \mathbb{R},\ \sigma > 0\}$, has a pdf as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}, \quad (\mu, \sigma^2) \in \Omega.$$

- $E(X) = \mu$ (Proof in Theorem 3.5)
- $Var(X) = \sigma^2$
- $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ (Proof in Theorem 3.8)

The normal distribution (sometimes called the Gaussian distribution) is generally considered to be the most important continuous distribution in Statistics.

Since $f(x)$ is (1) maximized at $x = \mu$, (2) symmetric about $x = \mu$ and (3) bell-shaped, the mean, mode and median are equal to μ. Intuitively, μ and σ^2 control the location and scale of $f(x)$. If μ is greater, the center of $f(x)$ is located more to the right. If σ^2 is greater, the spread is wider while the peak is lower since the total area under $f(x)$ is always 1 (Proof in Theorem 3.6 and 3.7).

2.1 Standard Normal Distribution

A **standard normal random variable** Z, denoted by $Z \sim N(0, 1)$ has a pdf as

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}.$$

- $E(Z) = 0$ (Proof in Theorem 3.4)
- $Var(Z) = 1$
- $M_Z(t) = e^{\frac{1}{2}t^2}$ (See Corollary 3.9)

Standard normal distribution is a special case of normal distribution with $\mu = 0$ and $\sigma^2 = 1$. The numerical values, $P(Z \leq z)$, can be obtained from many computer packages or from table.

Theorem 3.1. If $X \sim N(\mu, \sigma^2)$, define $Z = \frac{X-\mu}{\sigma}$. Since X is a random variable, Z is also a random variable. Then, $Z \sim N(0, 1)$. (Proof in Section 3)

It therefore follows that all normal probabilities can be calculated in terms of the standard normal. The cdf of $N(\mu, \sigma^2)$ cannot be obtained in closed form. We have to transform it into $N(0, 1)$ in order to find the cumulative probabilities using a standard normal table. In fact, we do not need a table for each value of μ and σ^2 because of it.

Theorem 3.2. If $Z \sim N(0, 1)$, define $X = \mu + \sigma Z$. Since Z is a random variable, X is also a random variable. Then, $X \sim N(\mu, \sigma^2)$. (Proof in Section 3)

It therefore follows that all normal theorems can be built for the standard normal. The result can then be generalized to $N(\mu, \sigma^2)$ by using a transformation. In fact, it is true for any location-scale family. Here μ is the location parameter and σ^2 is the scale parameter.

Theorem 3.3.
$P(|X - \mu| \leq \sigma) = P(|Z| \leq 1) \approx 0.68$,
$P(|X - \mu| \leq 2\sigma) = P(|Z| \leq 2) \approx 0.95$,
$P(|X - \mu| \leq 3\sigma) = P(|Z| \leq 3) \approx 0.997$. (Proof in Section 3)

It therefore follows that normal quantiles and coverage probabilities of normal random variables can be found by similar arguments.

2.2 Useful Formulas

Assume that $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$. Here $f(x)$, $f(z)$ and $F(x)$ are defined as before.

(a) $P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2)$

(b) $P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$

(c) $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$

(d) $P(Z \leq -z) = P(Z \geq z)$

Sketch of proofs:

For formula (a), since X is a continuous random variable, from Section 1, it is known that $P(X = x_1) = P(X = x_2) = 0$.

For formula (b), it is by definition of $F(x) = P(X \leq x)$ and the use of formula (a).

For formula (c), note that $P(-\infty < X < \infty) = 1$, hence $P(X \leq x) + P(X > x) = 1$.

For formula (d), note that $f(z)$ is symmetric about $z = 0$.

3 Theorems and Proofs

3.1 Normal and Standard Normal Distribution

Theorem 3.1. If $X \sim N(\mu, \sigma^2)$, define $Z = \frac{X-\mu}{\sigma}$. Since X is a random variable, Z is also a random variable. Then, $Z \sim N(0, 1)$.[1]

Proof.

$$\begin{aligned} P(Z \leq z) &= P\Big(\frac{X-\mu}{\sigma} \leq z\Big) \\ &= P(X \leq z\sigma + \mu) \\ &= \int_{-\infty}^{z\sigma+\mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \ , \ (\text{substitute } t = \frac{x-\mu}{\sigma}) \end{aligned}$$

showing that $P(Z \leq z)$ is the standard normal cdf. □

Theorem 3.2. If $Z \sim N(0, 1)$, define $X = \mu + \sigma Z$. Since Z is a random variable, X is also a random variable. Then, $X \sim N(\mu, \sigma^2)$.[2]

Proof.

$$\begin{aligned} P(X \leq x) &= P(\mu + \sigma Z \leq x) \\ &= P(Z \leq \frac{x-\mu}{\sigma}) \\ &= \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \ , \ (\text{substitute } t = \mu + \sigma z) \end{aligned}$$

showing that $P(X \leq x)$ is $N(\mu, \sigma^2)$ cdf. □

[1] "Statistical Inference" 2nd ed. (Casella and Berger) Ch3 p.102
[2] "Statistical Inference" 2nd ed. (Casella and Berger) Ch3 p.103

Theorem 3.3.
$P(|X-\mu| \leq \sigma) = P(|Z| \leq 1) \approx 0.68$,
$P(|X-\mu| \leq 2\sigma) = P(|Z| \leq 2) \approx 0.95$,
$P(|X-\mu| \leq 3\sigma) = P(|Z| \leq 3) \approx 0.997$.

Proof.

$$\begin{aligned} P(|X-\mu| \leq \sigma) &= P(|Z| \leq 1) \quad \text{, by Theorem 3.1} \\ &= P(-1 \leq Z \leq 1) \\ &= P(Z \leq 1) - P(Z \leq -1) \\ &= 0.8413 - 0.1587 \\ &= 0.6826 \\ &\approx 0.68 \end{aligned}$$

$$\begin{aligned} P(|X-\mu| \leq 2\sigma) &= P(|Z| \leq 2) \quad \text{, by Theorem 3.1} \\ &= P(-2 \leq Z \leq 2) \\ &= P(Z \leq 2) - P(Z \leq -2) \\ &= 0.9772 - 0.0228 \\ &= 0.9544 \\ &\approx 0.95 \end{aligned}$$

$$\begin{aligned} P(|X-\mu| \leq 3\sigma) &= P(|Z| \leq 3) \quad \text{, by Theorem 3.1} \\ &= P(-3 \leq Z \leq 3) \\ &= P(Z \leq 3) - P(Z \leq -3) \\ &= 0.9987 - 0.0013 \\ &= 0.9974 \\ &\approx 0.997 \end{aligned}$$

□

Theorem 3.4. If $Z \sim N(0,1)$, then $E(Z) = 0$ and $Var(Z) = 1$.

Proof.

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]_{-\infty}^{\infty} \\ &= 0 \end{aligned}$$

$$\begin{aligned} E(Z^2) &= \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= \left[-z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad \text{(integration by parts)} \\ &= 1, \end{aligned}$$

since the first integrand is 0 and the second integrand is 1 ($\int_{-\infty}^{\infty} f(z)dz = 1$ by Theorem 3.6).

$$\begin{aligned} Var(Z) &= E(Z^2) - [E(Z)]^2 \\ &= 1 - 0^2 \\ &= 1 \end{aligned}$$

Alternatively you may use the mgf approach. □

Theorem 3.5. If $X \sim N(\mu, \sigma^2)$, then $E(X) = \mu$ and $Var(X) = \sigma^2$.

Proof. Using transformation of random variable,

$$\begin{aligned} Z &= \frac{X - \mu}{\sigma} \sim N(0,1) \quad \text{by Theorem 3.1} \\ \Rightarrow X &= \mu + \sigma Z \end{aligned}$$

$$\begin{aligned} E(X) &= \mu + \sigma E(Z) \\ &= \mu + \sigma \cdot 0 \qquad \text{by Theorem 3.4} \\ &= \mu \end{aligned}$$

$$\begin{aligned} Var(X) &= Var(\mu + \sigma Z) \\ &= E(\mu + \sigma Z - \mu)^2 && \text{by definition} \\ &= \sigma^2 E(Z^2) && \text{pull out non-random } \sigma^2 \\ &= \sigma^2 Var(Z) && \text{by } Var(Z) = E(Z^2) - [E(Z)]^2 = E(Z^2) \\ &= \sigma^2 \cdot 1 && \text{by Theorem 3.4} \\ &= \sigma^2 \end{aligned}$$

Alternatively you may use the mgf approach. □

Theorem 3.6. If $Z \sim N(0,1)$ and $f(z)$ is the pdf of Z, then $\int_{-\infty}^{\infty} f(z)dz = 1$.[3]

Proof. The target is to show that

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 1.$$

Notice that the integrand is symmetric around 0, implying that the integral over $(-\infty, 0)$ is equal to the integral over $(0, \infty)$. We need only to show that

$$\int_0^{\infty} e^{-\frac{z^2}{2}} dz = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}.$$

The function $e^{-\frac{z^2}{2}}$ does not have an antiderivative that can be written in closed form, so we cannot perform the integration directly. Since both sides are positive, the equality will hold if we establish that the squares are equal.

$$\begin{aligned}
\Big(\int_0^{\infty} e^{-\frac{z^2}{2}} dz\Big)^2 &= \Big(\int_0^{\infty} e^{-\frac{t^2}{2}} dt\Big)\Big(\int_0^{\infty} e^{-\frac{u^2}{2}} du\Big) \\
&= \int_0^{\infty} \int_0^{\infty} e^{-\frac{t^2+u^2}{2}} dtdu.
\end{aligned}$$

To convert to polar coordinates. Define

$$t = r\text{cos}\theta \quad \text{and} \quad u = r\text{sin}\theta.$$

Then $t^2 + u^2 = r^2$ and $dtdu = rd\theta dr$. The limits of integration becomes $0 < r < \infty$ and $0 < \theta < \frac{\pi}{2}$ (the upper limit on θ is $\frac{\pi}{2}$ because t and u are restricted to be positive). So,

$$\begin{aligned}
\int_0^{\infty} \int_0^{\infty} e^{-\frac{t^2+u^2}{2}} dtdu &= \int_0^{\infty} \int_0^{\frac{\pi}{2}} re^{-\frac{r^2}{2}} d\theta dr \\
&= \frac{\pi}{2} \int_0^{\infty} re^{-\frac{r^2}{2}} dr \\
&= \frac{\pi}{2} \Big[-e^{-\frac{r^2}{2}}\Big]_0^{\infty} \\
&= \frac{\pi}{2},
\end{aligned}$$

which establishes the result. □

[3] "Statistical Inference" 2nd ed. (Casella and Berger) Ch3 p.103-104

Theorem 3.7. If $X \sim N(\mu, \sigma^2)$ and $f(x)$ is the pdf of X, then $\int_{-\infty}^{\infty} f(x)dx = 1$.

Proof. Using transformation of random variable,

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1) \quad \text{by Theorem 3.1}$$

Let $f(z)$ be the pdf of transformed Z.

$$\int_{-\infty}^{\infty} f(z)dz = 1 \quad \text{by Theorem 3.6}$$

It follows that

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

□

Theorem 3.8. If $X \sim N(\mu, \sigma^2)$, then $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.[4]

Proof.

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} \frac{e^{tx}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left[x^2 - 2(\mu + \sigma^2 t)x + \mu^2\right]\right\} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left[[x - (\mu + \sigma^2 t)]^2 - 2\mu\sigma^2 t - \sigma^4 t^2\right]\right\} dx \quad \text{(complete the square)} \\
&= \exp\left(\frac{2\mu\sigma^2 t + \sigma^4 t^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left[[x - (\mu + \sigma^2 t)]^2\right]\right\} dx \\
&= \exp\left(\frac{2\mu\sigma^2 t + \sigma^4 t^2}{2\sigma^2}\right) \quad \text{(The normal pdf integrates to 1 with } \mu \text{ replaced by } \mu + \sigma^2 t.) \\
&= \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)
\end{aligned}$$

□

Corollary 3.9. If $Z \sim N(0, 1)$, then $M_Z(t) = e^{\frac{1}{2}t^2}$.

Proof. Please refer to Theorem 3.8. □

[4] "Probability and Statistical Inference" 8th ed. (Hogg and Tanis) Ch3 p.162

3.2 Review of Chi-squared Distribution

- The pdf of $X \sim \text{Gamma}(\alpha, \beta)$ is[5]

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty, \quad \alpha > 0, \quad \beta > 0,$$

  in which $\Gamma(\cdot)$ denotes the gamma function

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

- The chi-squared distribution with p degrees of freedom (χ_p^2), where p is an integer, is a special case of the gamma distribution, with $\alpha = \frac{p}{2}$ and $\beta = 2$.

- The pdf of $X \sim \chi_p^2$ is

$$f(x|p) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad 0 < x < \infty, \quad p \in \mathbb{N},$$

  with mgf

$$M_X(t) = (1-2t)^{-p/2}, \quad t < \frac{1}{2}.$$

3.3 Sampling from Normal Distribution

Theorem 3.10. If $Z \sim N(0,1)$, then $Z^2 \sim \chi_1^2$.[6]

Proof. Consider $Y = Z^2$. For $y > 0$, the cdf of $Y = Z^2$ is

$$F_Y(y) = P(Y \le y) = P(Z^2 \le y) = P(-\sqrt{y} \le Z \le \sqrt{y}) = F_Z(\sqrt{y}) - F_Z(-\sqrt{y}).$$

Then,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy}[F_Z(\sqrt{y}) - F_Z(-\sqrt{y})] \\ &= \frac{1}{2\sqrt{y}} f_Z(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_Z(-\sqrt{y}). \end{aligned}$$

[5] "Statistical Inference" 2nd ed. (Casella and Berger) Ch3 p.99-101
[6] "Statistical Inference" 2nd ed. (Casella and Berger) Ch2 p.52-53

Since $Z \sim N(0,1)$, we know that

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.$$

So,

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}} \Big[\frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} + \frac{1}{\sqrt{2\pi}} e^{-(-\sqrt{y})^2/2} \Big] \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2}, \quad 0 < y < \infty. \end{aligned}$$

Therefore, Z^2 is a chi-squared random variable with 1 degree of freedom, i.e. $Z^2 \sim \chi^2_1$. □

Theorem 3.11. If $X_1, \ldots, X_n$ are independent and $X_i \sim \chi^2_{p_i}$, then $X_1 + \cdots + X_n \sim \chi^2_{p_1 + \cdots + p_n}$.[7]

Proof. The mgf of X_i is

$$M_{X_i}(t) = (1-2t)^{-p_i/2}, \quad t < \frac{1}{2}.$$

Consider $Y = X_1 + \cdots + X_n$. By independence, the mgf of Y is

$$M_Y(t) = M_{X_1}(t) \times \cdots \times M_{X_n}(t) = (1-2t)^{-p_1/2} \times \cdots \times (1-2t)^{-p_n/2} = (1-2t)^{-(p_1+\cdots+p_n)/2}, \ t < \frac{1}{2}.$$

This is the mgf of chi-squared distribution with $p_1 + \cdots + p_n$ degrees of freedom, i.e. $X_1 + \cdots + X_n \sim \chi^2_{p_1+\cdots+p_n}$. □

Corollary 3.12. If $Z_1, \ldots, Z_n \sim N(0,1)$ independently, then $Z_1^2 + \cdots + Z_n^2 \sim \chi^2_n$.

Proof. If $Z_i \sim N(0,1)$ for $i = 1, \ldots, n$, then $Z_i^2 \sim \chi^2_1$ by Theorem 3.10. So, $Z_1^2 + \cdots + Z_n^2 \sim \chi^2_n$ by Theorem 3.11 if we set $p_i = 1$ for $i = 1, \ldots, n$. □

Corollary 3.13. If $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ independently, then $(\frac{X_1 - \mu}{\sigma})^2 + \cdots + (\frac{X_n - \mu}{\sigma})^2 \sim \chi^2_n$.

Proof. If $X_i \sim N(\mu, \sigma^2)$ for $i = 1, \ldots, n$, then $Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$. Then, $Z_1^2 + \cdots + Z_n^2 \sim \chi^2_n$ by Corollary 3.12. □

Corollary 3.14. If $\bar{X}_1, \ldots, \bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ independently, then $(\frac{\bar{X}_1 - \mu}{\sigma/\sqrt{n}})^2 + \cdots + (\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}})^2 \sim \chi^2_n$.

Proof. If $\bar{X}_i \sim N(\mu, \frac{\sigma^2}{n})$ for $i = 1, \ldots, n$, then $Z_i = \frac{\bar{X}_i - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$. Then, $Z_1^2 + \cdots + Z_n^2 \sim \chi^2_n$ by Corollary 3.12. □

[7] "Statistical Inference" 2nd ed. (Casella and Berger) Ch4 p.183

Theorem 3.15. (The Sample Mean and Variance in Normal Distribution). Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution and let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Proof. Consider[8]

$$
\begin{aligned}
W &= \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \\
&= \sum_{i=1}^n \left[\frac{(X_i - \bar{X}) + (\bar{X} - \mu)}{\sigma}\right]^2 \\
&= \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 + n\left(\frac{\bar{X} - \mu}{\sigma}\right)^2 \\
&= \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2
\end{aligned}
$$

because the cross-product term is

$$2\sum_{i=1}^n \frac{(X_i - \bar{X})(\bar{X} - \mu)}{\sigma^2} = 2\frac{\bar{X} - \mu}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X}) = 0.$$

Here, $W \sim \chi^2_n$ by Corollary 3.13. Moreover, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ and $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. Consider $Y = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$. Hence, $Y \sim \chi^2_1$. Rewrite

$$W = \frac{(n-1)S^2}{\sigma^2} + Y.$$

We know that $\bar{X}$ and S^2 are independent. Interested students please refer to "Statistical Inference" 2nd ed. (Casella and Berger) Ch5 p.218-219 for the proof. So Y and S^2 are also independent. Hence,

$$
\begin{aligned}
E[e^{tW}] &= E[e^{t(n-1)S^2/\sigma^2}]E[e^{tY}] \\
(1-2t)^{-n/2} &= E[e^{t(n-1)S^2/\sigma^2}](1-2t)^{-1/2}
\end{aligned}
$$

So,

$$E[e^{t(n-1)S^2/\sigma^2}] = (1-2t)^{-(n-1)/2}, \quad t < \frac{1}{2}.$$

This is the mgf of chi-squared distribution with $n-1$ degrees of freedom, i.e. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. $\square$

[8] "Probability and Statistical Inference" 8th ed. (Hogg and Tanis) Ch5 p.258-260