Math 252
Statistical Theory

Yujia Ding

2020

1

PROBABILITY AND STATISTICS

In this chapter, several examples are discussed to show the difference between types of questions probability addresses and those statistics addresses.

Let $T_1, T_2, \ldots, T_{10}$ be the lifetimes (time to failure) of 10 identical electrical parts. Suppose $T_1, T_2, \ldots, T_{10}$ are i.i.d[1] random variables, each said to have an exponential distribution[2] with parameter $1/\theta$. Therefore each T_i for $i = 1, 2, \ldots, 10$ has a CDF of the form

$$\mathbb{P}\{T_i \leq t\} = \begin{cases} 1 - e^{-t/\theta} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

According to this model, one can address both probability and statistics questions. Here are some examples.

Example 1.1. (**Probability analysis**). Given the parameter $\theta = 1000$, one can find the probabilities of some attributes.

1. $\mathbb{E}(T_1) = 1000$
2. $\mathbb{P}\{T_1 > 250\} = 1 - \mathbb{P}\{T_1 \leq 250\} = 1 - \left(1 - e^{-250/1000}\right) = 0.7788$
3. Let $T^* = \max(T_1, T_2)$ and one can find that

$$\mathbb{P}\{T^* > t\} = \mathbb{P}\{\max(T_1, T_2) > t\}$$

$$= 1 - \mathbb{P}\{T_1 \leq t\}\,\mathbb{P}\{T_2 \leq t\} = 1 - \left(1 - e^{-t/1000}\right)^2$$

$$\mathbb{P}\{T^* > 250\} = 1 - \left(1 - e^{-250/1000}\right)^2 = 0.951$$

[1] Independent and identically distributed

[2] Recall that the CDF of an exponential distribution with parameter λ is $1 - e^{-\lambda t}$ and that the expected value is $1/\lambda$.

$\boxed{\textit{Example 1.2.}}$ (**Statistics analysis**). The parameter θ is unknown. It is assumed that the i.i.d. random variables $T_1, T_2, \ldots, T_{10}$ are observed, resulting in a set of values[1] $t_1, t_2, \ldots, t_{10}$. Now assume

$$t_1 = 2574 \quad t_2 = 1310 \quad t_3 = 282 \quad t_4 = 1233 \quad t_5 = 1925$$

$$t_6 = 135 \quad t_7 = 281 \quad t_8 = 2254 \quad t_9 = 671 \quad t_{10} = 495$$

Based on the seeing values, one can make some statements (inferences) about θ.

1. $\sum_{i=1}^{10} t_i = 11160$
2. $\bar{t} = \frac{\sum_{i=1}^{10} t_i}{10} = 1116$
3. Point estimate of θ: $\bar{t} = 1116$
4. Interval estimation of θ: 80% confidence interval is $(786, 1800)$
5. Hypothesis test:

$$H_0 : \theta = 700 \quad H_1 : \theta = 1000$$

$$\frac{2\sum_{i=1}^{10} t_i}{\theta_0} = 3.19 \quad p < 0.05 \quad \text{reject}$$

[1] Note that capital letters (e.g. X) will be used to denote random variables. The lower case letters (e.g. x) will be used to denote possible values that the corresponding random variables can attain.

2

NORMAL AND RELATED DISTRIBUTIONS

2.1 Normal Distribution

Definition 2.1.1. **(Normal distribution).**

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

- PDF[1]: $f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Mean: μ
- Standard deviation: σ
- MGF[2]: $e^{\mu t + \sigma^2 t^2/2}$

Proof. To find the MGF of a normal random variable X:

$$\mathbb{M}_X(t) = \mathbb{E}\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tx} f_X(x;\mu,\sigma^2)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{tx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= e^{\mu t + \sigma^2 t^2/2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\left(\mu+t\sigma^2\right)^2}{\sigma}\right)^2} dx}_{=\int_{-\infty}^{\infty} f_X(x;(\mu+t\sigma^2)^2,\sigma^2)\,dx=1} = e^{\mu t + \sigma^2 t^2/2}$$

□

[1] Probability density function

[2] Moment generating function: $\mathbb{M}_X(t) = \mathbb{E}\left(e^{tX}\right)$

Definition 2.1.2. **(Standard normal distribution)**.

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

- PDF: $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
- CDF[1]: $\Phi(z) = \int_{-\infty}^{z} \varphi(z)\, dz$
- Mean: $\mu = 0$
- Standard deviation: $\sigma = 1$
- MGF: $e^{t^2/2}$

Lemma 2.1.1.. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

Proof.

$$\mathbb{M}_Z(t) = \mathbb{E}\left(e^{tZ}\right) = \mathbb{E}\left(e^{t\frac{X-\mu}{\sigma}}\right) = e^{\frac{-\mu t}{\sigma}} \mathbb{E}\left(e^{\frac{t}{\sigma}X}\right) = e^{\frac{-\mu t}{\sigma}} \mathbb{M}_X\left(\frac{t}{\sigma}\right) \overset{2}{=} e^{\frac{t^2}{2}}$$

□

Lemma 2.1.2.. If $X_1, X_2, \ldots, X_n$ are i.i.d $\mathcal{N}(\mu, \sigma^2)$ and let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

Proof. From $X_i \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathbb{E}\left(e^{t\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}\right) = e^{-\frac{t\mu}{\sigma/\sqrt{n}}} \mathbb{E}\left(e^{\frac{t}{\sigma/\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n} X_i}\right) = e^{-\frac{t\mu}{\sigma/\sqrt{n}}} \mathbb{E}\left(\prod_{i=1}^{n} e^{\frac{t}{\sigma\sqrt{n}}X_i}\right)$$

[1]Cumulative distribution function
[2]Recall the MGF of normal variable

$$\overset{i.i.d.}{=} e^{-\frac{t\mu}{\sigma/\sqrt{n}}} \prod_{i=1}^{n} \mathbb{E}\left(e^{\frac{t}{\sigma\sqrt{n}}X_i}\right) = e^{-\frac{t\mu}{\sigma/\sqrt{n}}} \prod_{i=1}^{n} \mathbb{M}_{X_i}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

$$= e^{-\frac{t\mu}{\sigma/\sqrt{n}}} \prod_{i=1}^{n} e^{\mu\frac{t}{\sigma\sqrt{n}}+\sigma^2\frac{t^2}{\sigma^2 n}/2} = e^{t^2/2} = \mathbb{M}_Z(t)$$

which is the MGF of standard normal variable. $\square$

Theorem 2.1.1. **(Linear combinations of normal variables).** If $X_1, \ldots, X_n$ are independent normal variables with $X_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right)$, then

$$Y = \sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$$

Proof.

$$\mathbb{M}_Y(t) = \mathbb{E}\left(e^{tY}\right) = \mathbb{E}\left(e^{t\sum_{i=1}^{n} a_i X_i}\right) = \mathbb{E}\left(\prod_{i=1}^{n} e^{ta_i X_i}\right)$$

Because of the independence of the variables $X_1, \ldots, X_n$,

$$\mathbb{E}\left(\prod_{i=1}^{n} e^{ta_i X_i}\right) = \prod_{i=1}^{n} \mathbb{E}\left(e^{ta_i X_i}\right)$$

Therefore

$$\mathbb{M}_Y(t) = \prod_{i=1}^{n} \mathbb{E}\left(e^{ta_i X_i}\right) = \prod_{i=1}^{n} \mathbb{M}_{X_i}(ta_i) = \prod_{i=1}^{n} e^{a_i\mu_i t + a_i^2\sigma_i^2 t^2/2}$$

$$= \exp\left\{t\underbrace{\sum_{i=1}^{n} a_i\mu_i}_{\text{new mean}} + \frac{t^2}{2}\underbrace{\sum_{i=1}^{n} a_i^2\sigma_i^2}_{\text{new variance}}\right\}$$

Since MGF uniquely determines the distribution and $e^{t\sum_{i=1}^{n} a_i\mu_i + \frac{t^2}{2}\sum_{i=1}^{n} a_i^2\sigma_i^2}$ is the MGF of $\mathcal{N}\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$, the result follows. $\square$

2.2 Gamma Distribution

Definition 2.2.1. **(Gamma function).**

$$\Gamma(\kappa) = \int_0^\infty t^{\kappa-1} e^{-t}\, dt \qquad \text{for } \mathcal{R}(\kappa)^{1} > 0$$

Properties:

- $\Gamma(\kappa+1) = \kappa\Gamma(\kappa)$
- $\Gamma(\kappa) = (\kappa-1)!$ [2] if κ is an integer and $\kappa \geq 1$

Definition 2.2.2. **(Gamma distribution).**

$$X \sim \mathcal{G}am\,(\theta, \kappa) \qquad \text{for } \theta > 0, \kappa > 0$$

- PDF: $f_X(x; \theta, \kappa) = \begin{cases} \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\theta} & x > 0 \\ 0 & otherwise \end{cases}$
- Mean: $\kappa\theta$
- Variance: $\kappa\theta^2$
- MGF: $\left(\frac{1}{1-\theta t}\right)^k$ for $t < \frac{1}{\theta}$

Lemma 2.2.1..

$$\int_0^\infty x^{\kappa-1} e^{-x/\theta}\, dx = \theta^\kappa \Gamma(\kappa) \qquad \text{for } \theta > 0, \kappa > 0$$

Proof. Apply change-of-variable by taking $y = x/\theta$ and recall Gamma function $\Gamma(\kappa) = \int_0^\infty t^{\kappa-1} e^{-t}\, dt$ for $\kappa > 0$

$$\int_0^\infty x^{\kappa-1} e^{-x/\theta}\, dx = \theta^\kappa \int_0^\infty y^{\kappa-1} e^{-y}\, dy = \theta^\kappa \Gamma(\kappa)$$

[1] Real part of κ; if κ is real, $\mathcal{R}(\kappa) = \kappa$

[2] Factorial: $n! = n \times (n-1) \times \cdots \times 2 \times 1$

Or use the fact that $\int_{\mathbb{R}} f_X(x)\,dx = 1$ and recall the PDF of Gamma random variable $f_X(x;\theta,\kappa) = \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\theta}$ for $x > 0, \theta > 0, \kappa > 0$.

$$\int_0^\infty x^{\kappa-1} e^{-x/\theta}\,dx = \theta^\kappa \Gamma(\kappa) \underbrace{\int_0^\infty \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\theta}\,dx}_{=\int_0^\infty f_X(x;\theta,\kappa)\,dx=1} = \theta^\kappa \Gamma(\kappa)$$

□

Proof of Definition 2.2.2.

- MGF:
  From the facts that $t < 1/\theta$ and $\theta > 0$, one has $\frac{\theta}{1-\theta t} > 0$ and since $\kappa > 0$

$$\mathbb{M}_X(t) = \mathbb{E}\left(e^{tX}\right) = \int_0^\infty e^{tx} \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\theta}\,dx$$

$$= \int_0^\infty \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x\left(\frac{1}{\theta}-t\right)}\,dx = \int_0^\infty \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\left(\frac{\theta}{1-\theta t}\right)}\,dx$$

$$= \frac{\left(\frac{\theta}{1-\theta t}\right)^\kappa}{\theta^\kappa} \underbrace{\int_0^\infty \frac{1}{\left(\frac{\theta}{1-\theta t}\right)^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\left(\frac{\theta}{1-\theta t}\right)}\,dx}_{\int_0^\infty f_X(x;\frac{\theta}{1-\theta t},\kappa)\,dx=1} = \left(\frac{1}{1-\theta t}\right)^\kappa$$

- Mean:
  Since $\theta > 0, \kappa + 1 > 0$

$$\mathbb{E}(X) = \int_0^\infty x \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\theta}\,dx = \theta\kappa \underbrace{\int_0^\infty \frac{1}{\theta^{\kappa+1}\Gamma(\kappa+1)} x^\kappa e^{-x/\theta}\,dx}_{=\int_0^\infty f_X(x;\theta,\kappa+1)\,dx=1} = \theta\kappa$$

  Or make use of the fact that $\mathbb{E}(X^r) = \mathbb{M}_X^{(r)}(0)$[1] for $r = 1, 2, \ldots$.

$$\mathbb{E}(X) = \mathbb{M}_X^{(1)}(0) = \frac{d}{dt}\mathbb{E}\left(e^{tX}\right)\Big|_{t=0} = \frac{d}{dt}\left(\frac{1}{1-\theta t}\right)^\kappa\Big|_{t=0} = \theta\kappa\left(\frac{1}{1-\theta t}\right)^{\kappa+1}\Bigg|_{t=0} = \theta\kappa$$

[1] $\mathbb{M}_X^{(r)}(0) = \frac{d^r}{dt^r}\mathbb{E}\left(e^{tX}\right)\Big|_{t=0}$

- Variance:
  Since $\theta > 0, \kappa + 2 > 0$

$$\mathbb{E}(X^2) = \int_0^\infty x^2 \frac{1}{\theta^\kappa \Gamma(\kappa)} x^{\kappa-1} e^{-x/\theta}\, dx$$

$$= \theta^2 \kappa(\kappa+1) \underbrace{\int_0^\infty \frac{1}{\theta^{\kappa+2}\Gamma(\kappa+2)} x^{\kappa+1} e^{-x/\theta}\, dx}_{=\int_0^\infty f_X(x;\theta,\kappa+2)\,dx = 1} = \theta^2\kappa(\kappa+1)$$

Or make use of the fact that $\mathbb{E}(X^r) = \mathbb{M}_X^{(r)}(0)$ for $r = 1, 2, \ldots$.

$$\mathbb{E}(X^2) = \mathbb{M}_X^{(2)}(0) = \frac{d^2}{dt^2}\mathbb{E}\left(e^{tX}\right)\Big|_{t=0} = \frac{d}{dt}\frac{d}{dt}\left(\frac{1}{1-\theta t}\right)^\kappa\Big|_{t=0}$$

$$= \frac{d}{dt}\,\theta\kappa\left(\frac{1}{1-\theta t}\right)^{\kappa+1}\Big|_{t=0} = \theta^2\kappa(\kappa+1)\left(\frac{1}{1-\theta t}\right)^{\kappa+2}\Big|_{t=0} = \theta^2\kappa(\kappa+1)$$

Hence

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \theta^2\kappa(\kappa+1) - \theta^2\kappa^2 = \theta^2\kappa$$

□

Theorem 2.2.1. **(Linear combinations of gamma variables)**. If $X_1, X_2, \ldots, X_n$ are independent and $X_i \sim \mathcal{Gam}(\theta, \kappa_i)$, then

$$U = \sum_{i=1}^n X_i \sim \mathcal{Gam}(\theta, \sum_{i=1}^n \kappa_i)$$

Proof.

$$\mathbb{M}_U(t) = \mathbb{E}\left(e^{tU}\right) = \mathbb{E}\left(e^{t\sum_{i=1}^n X_i}\right) = \mathbb{E}\left(\prod_{i=1}^n e^{tX_i}\right)$$

Because of the independence of the variables $X_1, \ldots, X_n$,

$$\mathbb{E}\left(\prod_{i=1}^n e^{tX_i}\right) = \prod_{i=1}^n \mathbb{E}\left(e^{tX_i}\right)$$

Recall the MGF of gamma random variable X_i is $\left(\frac{1}{1-\theta t}\right)^{\kappa_i}$. Therefore

$$\mathbb{M}_U(t) = \prod_{i=1}^{n} \mathbb{E}\left(e^{tX_i}\right) = \prod_{i=1}^{n} \mathbb{M}_{X_i}(t) = \prod_{i=1}^{n} \left(\frac{1}{1-\theta t}\right)^{\kappa_i} = \left(\frac{1}{1-\theta t}\right)^{\sum_{i=1}^{n} \kappa_i}$$

Since MGF uniquely determines the distribution and $\left(\frac{1}{1-\theta t}\right)^{\sum_{i=1}^{n} \kappa_i}$ is the MGF of $\mathcal{G}am(\theta, \sum_{i=1}^{n} \kappa_i)$, the result follows. □

2.3 Chi-Square Distribution

Definition 2.3.1. **(Chi-square distribution)**. Chi-square distribution is a special gamma distribution with $\theta = 2$ and $\kappa = v/2$ for $v > 0$, i.e.

$$Y \sim \mathcal{G}am\left(2, v/2\right) \qquad \text{for } v > 0$$

Because there is only one parameter v involved in the Chi-square distribution, a special notation for it is

$$Y \sim \chi^2\left(v\right) \qquad \text{for } v > 0$$

where v is the degree of freedom.

- PDF: $f_Y(y; v) = \begin{cases} \frac{1}{2^{v/2}\Gamma(v/2)} y^{\frac{v}{2}-1} e^{-\frac{y}{2}} & y > 0 \\ 0 & otherwise \end{cases}$
- Mean: v
- Variance: $2v$
- MGF: $\left(\frac{1}{1-2t}\right)^{v/2}$ for $t < \frac{1}{2}$

Lemma 2.3.1.. If $X \sim \mathcal{G}am\left(\theta, \kappa\right)$, the

$$\frac{2X}{\theta} \sim \chi^2(2\kappa)$$

Proof.

$$\mathbb{M}_{2X/\theta}(t) = \mathbb{E}\left(e^{t2X/\theta}\right) = \mathbb{E}\left(e^{(t2/\theta)X}\right) = \mathbb{M}_X(t2/\theta) = \left(\frac{1}{1-2t}\right)^k = \mathbb{M}_{\chi^2(2\kappa)}(t)$$

which requires $t < 1/2$. □

Theorem 2.3.1. **(Linear combinations of chi-square variables).** If $Y_1, Y_2, \ldots, Y_n$ are independent and $Y_i \sim \chi^2(v_i)$, then

$$V = \sum_{i=1}^{n} Y_i \sim \chi^2\left(\sum_{i=1}^{n} v_i\right)$$

Proof. Take $\kappa_i = v_i/2$ and $\theta = 2$ in Theorem 2.2.1. □

2.4 Connection Between Normal and Chi-Square Variables

Theorem 2.4.1. If $Z \sim \mathcal{N}(0,1)$, then

$$Z^2 \sim \chi^2(1)$$

Proof.

$$\mathbb{M}_{Z^2}(t) = \mathbb{E}\left(e^{tZ^2}\right) = \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}(1-2t)}\, dz$$

$$= \frac{1}{\sqrt{1-2t}} \underbrace{\int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}\,(1/\sqrt{1-2t})} e^{-\frac{1}{2}\left(\frac{z}{1/\sqrt{1-2t}}\right)^2}}_{\text{PDF of } \mathcal{N}(0,\frac{1}{1-2t})}\, dz}_{\text{integral of PDF over entire space = 1}} = \left(\frac{1}{1-2t}\right)^{1/2}$$

which is the MGF of $\chi^2(1)$. □

Corollary 2.4.1. If $X_1, X_2, \ldots, X_n$ are i.i.d $\mathcal{N}(\mu, \sigma^2)$ and let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then

(i) $\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$

(ii) $n\frac{(\overline{X}-\mu)^2}{\sigma^2} \sim \chi^2(1)$

Proof.

(i) From $X_i \sim \mathcal{N}(\mu, \sigma^2)$, one has $\frac{X_i-\mu}{\sigma} \sim \mathcal{N}(0,1)$[1]. By Theorem 2.4.1, $\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(1)$. The result then follows from Theorem 2.3.1.

(ii) By Lemma 2.1.2, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$. The result follows from Theorem 2.4.1.

□

Theorem 2.4.2. (Fundamental theorem of statistical inference for the normal distribution). If $X_1, X_2, \ldots, X_n$ are i.i.d $\mathcal{N}(\mu, \sigma^2)$ and let $\overline{X} = \sum_{i=1}^{n} X_i$, then

(i) $\overline{X}$ is independent of $(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X})$

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

(iii) $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$

Proof.

(i) Suffice to compute the joint MGF of $(\overline{X}, X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X})$ and show that it equals the product of MGF of $\overline{X}$ and the joint MGF of $(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X})$[2]. So we just need to show

$$M_{\overline{X}, X_1-\overline{X}, X_2-\overline{X}, \ldots, X_n-\overline{X}}(t, t_1, \ldots, t_n) = M_{\overline{X}}(t) M_{X_1-\overline{X}, X_2-\overline{X}, \ldots, X_n-\overline{X}}(t_1, \ldots, t_n)$$

First compute the MGF of $(\overline{X}, X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X})$. Let $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$.

$$\mathbb{M}_{\overline{X}, X_1-\overline{X}, X_2-\overline{X}, \ldots, X_n-\overline{X}}(t, t_1, \ldots, t_n) = \mathbb{E}e^{t\overline{X}+t_1(X_1-\overline{X})+\cdots+t_n(X_n-\overline{X})}$$

[1] $\mathbb{E}\left(e^{t\frac{X_i-\mu}{\sigma}}\right) = e^{-\frac{t\mu}{\sigma}}\mathbb{E}\left(e^{\frac{t}{\sigma}X_i}\right) = e^{-\frac{t\mu}{\sigma}}\mathbb{M}_{X_i}\left(\frac{t}{\sigma}\right) = e^{-\frac{t\mu}{\sigma}}e^{\mu\frac{t}{\sigma}+\sigma^2\frac{t^2}{\sigma^2}/2} = e^{t^2/2} = \mathbb{M}_Z(t)$

[2] X_1 and X_2 are independent if and only if $\mathbb{M}_{X_1,X_2}(t_1, t_2) = \mathbb{M}_{X_1}(t_1)\mathbb{M}_{X_2}(t_2)$

$$= \mathbb{E}e^{t\overline{X}+\sum_{i=1}^{n} t_i\left(X_i-\overline{X}\right)} = \mathbb{E}e^{t\overline{X}+\sum_{i=1}^{n} t_i X_i-\overline{X}\sum_{i=1}^{n} t_i}$$

$$= \mathbb{E}e^{t\overline{X}+\sum_{i=1}^{n} t_i X_i-\left(\sum_{i=1}^{n} X_i\right)\frac{1}{n}\left(\sum_{i=1}^{n} t_i\right)}$$

$$= \mathbb{E}e^{t\frac{1}{n}\sum_{i=1}^{n} X_i+\sum_{i=1}^{n} t_i X_i-\bar{t}\sum_{i=1}^{n} X_i} = \mathbb{E}e^{\sum_{i=1}^{n}\left(\frac{t}{n}+t_i-\bar{t}\right)X_i} = \mathbb{E}\prod_{i=1}^{n} e^{\left(\frac{t}{n}+t_i-\bar{t}\right)X_i}$$

$$\overset{i.i.d.}{=} \prod_{i=1}^{n}\mathbb{E}e^{\left(\frac{t}{n}+t_i-\bar{t}\right)X_i} = \prod_{i=1}^{n}\mathbb{M}_{X_i}\left(\frac{t}{n}+t_i-\bar{t}\right)$$

$$\overset{1}{=} \prod_{i=1}^{n} e^{\mu\left(\frac{t}{n}+t_i-\bar{t}\right)+\frac{\sigma^2}{2}\left(\frac{t}{n}+t_i-\bar{t}\right)^2} = e^{\mu\sum_{i=1}^{n}\left(\frac{t}{n}+t_i-\bar{t}\right)+\frac{\sigma^2}{2}\sum_{i=1}^{n}\left(\frac{t}{n}+t_i-\bar{t}\right)^2}$$

$$= e^{\mu\sum_{i=1}^{n}\frac{t}{n}+\mu\sum_{i=1}^{n}\left(t_i-\bar{t}\right)+\frac{\sigma^2}{2}\sum_{i=1}^{n}\left(\left(\frac{t}{n}\right)^2+\left(t_i-\bar{t}\right)^2+2\frac{t}{n}\left(t_i-\bar{t}\right)\right)}$$

$$\overset{2}{=} e^{\mu t+\frac{\sigma^2 t^2}{2n}+\frac{\sigma^2}{2}\sum_{i=1}^{n}\left(t_i-\bar{t}\right)^2}$$

The MGF of $\overline{X}$ and $\left(X_1-\overline{X}, X_2-\overline{X}, \ldots, X_n-\overline{X}\right)$ can be either computed following the same approach or can be directly obtained by

$$\mathbb{M}_{\overline{X}}(t) = \mathbb{M}_{\overline{X},X_1-\overline{X},X_2-\overline{X},\ldots,X_n-\overline{X}}(t,0,\ldots,0) = e^{\mu t+\frac{\sigma^2 t^2}{2n}}$$

$$\mathbb{M}_{X_1-\overline{X},X_2-\overline{X},\ldots,X_n-\overline{X}}(t_1,\ldots,t_n) = \mathbb{E}e^{t_1\left(X_1-\overline{X}\right)+\cdots+t_n\left(X_n-\overline{X}\right)}$$

$$= \mathbb{M}_{\overline{X},X_1-\overline{X},X_2-\overline{X},\ldots,X_n-\overline{X}}(0,t_1,\ldots,t_n) = e^{\frac{\sigma^2}{2}\sum_{i=1}^{n}\left(t_i-\bar{t}\right)^2}$$

Therefore, because

$$\mathbb{M}_{\overline{X},X_1-\overline{X},X_2-\overline{X},\ldots,X_n-\overline{X}}(t,t_1,\ldots,t_n) = e^{\mu t+\frac{\sigma^2 t}{2n}+\frac{\sigma^2}{2}\sum_{i=1}^{n}\left(t_i-\bar{t}\right)^2}$$

$$= e^{\mu t+\frac{\sigma^2 t}{2n}}e^{\frac{\sigma^2}{2}\sum_{i=1}^{n}\left(t_i-\bar{t}\right)^2} = \mathbb{M}_{\overline{X}}(t)\mathbb{M}_{X_1-\overline{X},X_2-\overline{X},\ldots,X_n-\overline{X}}(t_1,\ldots,t_n)$$

the result follows.

(ii) The results follows from (i), since S^2 depends only on the $X_i-\overline{X}, 1\leq i\leq n$.

[1] Recall MGF of normal random variables

[2] $\sum_{i=1}^{n}\left(t_i-\bar{t}\right) = \sum_{i=1}^{n} t_i - \sum_{i=1}^{n}\bar{t} = \sum_{i=1}^{n} t_i - n\bar{t} = \sum_{i=1}^{n} t_i - n\frac{1}{n}\sum_{i=1}^{n} t_i = 0$

(iii)

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

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

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

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

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

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

So

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

and then

$$\mathbb{M}_{\sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{\sigma^2} + n\frac{\left(\overline{X} - \mu\right)^2}{\sigma^2}}(t) = \mathbb{M}_{\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2}}(t) \qquad (*)$$

Since it follows from (i) that $\overline{X}$ is independent of $\left(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X}\right)$, $\sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{\sigma^2}$ and $n\frac{\left(\overline{X} - \mu\right)^2}{\sigma^2}$ are independent and therefore

$$\mathbb{M}_{\sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{\sigma^2} + n\frac{\left(\overline{X} - \mu\right)^2}{\sigma^2}}(t) = \mathbb{E}\left(e^{t\left(\sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{\sigma^2} + n\frac{\left(\overline{X} - \mu\right)^2}{\sigma^2}\right)}\right)$$

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

$$\stackrel{1}{=} \mathbb{E}\left(e^{t\left(\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2}\right)}\right)\mathbb{E}\left(e^{t\left(n\frac{(\overline{X}-\mu)^2}{\sigma^2}\right)}\right) = \mathbb{M}_{\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2}}(t)\mathbb{M}_{n\frac{(\overline{X}-\mu)^2}{\sigma^2}}(t) \qquad (**)$$

Combine $(*)$ and $(**)$

$$\mathbb{M}_{\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{\sigma^2}}(t)\mathbb{M}_{n\frac{(\overline{X}-\mu)^2}{\sigma^2}}(t) = \mathbb{M}_{\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}}(t) \qquad (***)$$

By Corollary 2.4.1, one has $\sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$ and $n\frac{(\overline{X}-\mu)^2}{\sigma^2} \sim \chi^2(1)$. Plugging them in $(***)$ and recall the MGF of chi-square variables

$$\mathbb{M}_{\frac{(n-1)S^2}{\sigma^2}}(t)\mathbb{M}_{\chi^2(1)}(t) = \mathbb{M}_{\chi^2(n)}(t)$$

which implies

$$\mathbb{M}_{\frac{(n-1)S^2}{\sigma^2}}(t)\left(\frac{1}{1-2t}\right)^{1/2} = \left(\frac{1}{1-2t}\right)^{n/2}$$

and then

$$\mathbb{M}_{\frac{(n-1)S^2}{\sigma^2}}(t) = \left(\frac{1}{1-2t}\right)^{(n-1)/2}$$

which is the MGF of $\chi^2(n-1)$.

□

Remark 2.4.1.. $S^2 = \sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{n-1}$ is called the sample variance counted for bias[2]. Use the result in Theorem 2.4.2

$$\mathbb{E}(S^2) = \mathbb{E}\left(\frac{\sigma^2}{n-1}\chi^2(n-1)\right) = \frac{\sigma^2}{n-1}\mathbb{E}(\chi^2(n-1)) = \sigma^2$$

so S^2 is an unbiased estimator of σ^2. What if used $\frac{1}{n}$ instead of $\frac{1}{n-1}$?

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

[1] If X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

[2] An estimator $(\hat{\theta})$ of θ is said to be unbiased if $\mathbb{E}(\hat{\theta}) = \theta$

2.5 Student's t-Distribution

Definition 2.5.1. **(Student's t distribution)**. Let $Z \sim \mathcal{N}(0,1)$ and $V \sim \chi^2(v)$, and Z and V are independent. The the random variable

$$T = \frac{Z}{\sqrt{V/v}}$$

has distribution referred to as student's t-distribution with v degree-of-freedom, denoted by

$$T \sim t(v) \qquad \text{for } v > 0$$

Theorem 2.5.1.. The PDF of $T \sim t(v)$ is

$$f_T(t;v) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{\sqrt{v\pi}} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}$$

Corollary 2.5.1.. Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\mathcal{N}(\mu, \sigma^2)$ and let $\overline{X} = \sum_{i=1}^n X_i$ and $S^2 = \sum_{i=1}^n \frac{(X_i - \overline{X})^2}{n-1}$, then

$$\sqrt{n}\frac{\overline{X} - \mu}{S} \sim t(n-1)$$

Proof. By Lemma 2.1.2, $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$. By Theorem 2.4.2, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, and $\overline{X}$ and S^2 are independent (which implies that $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ and $\frac{(n-1)S^2}{\sigma^2}$ are independent). Recall the student's t distribution, one has

$$\frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} \sim t(n-1)$$

The result yields after simplification. □

2.6 F-Distribution

Definition 2.6.1. **(F-distribution)**. Let $X \sim \chi^2(v_1)$ and $Y \sim \chi^2(v_2)$ be independent. Then

$$F = \frac{X/v_1}{Y/v_2} \qquad \text{for } v_1, v_2 > 0$$

has the F-distribution with numerator degree-of-freedom v_1 and denominator degree-of-freedom v_2, and is denoted by $F \sim F(v_1, v_2)$.

Theorem 2.6.1. The density of $F \sim F(v_1, v_2)$ is

$$f(x; v_1, v_2) = \begin{cases} \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)} \left(\frac{v_1}{v_2}\right)^{v_1/2} x^{\frac{v_1}{2}-1} \left(1 + \frac{v_1}{v_2}x\right)^{-\frac{v_1+v_2}{2}} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Corollary 2.6.1. Let $X_1, X_2, \ldots, X_{n_1}$ be i.i.d. $\mathcal{N}\left(\mu_1, \sigma_1^2\right)$ and let $Y_1, Y_2, \ldots, Y_{n_2}$ be i.i.d. $\mathcal{N}\left(\mu_2, \sigma_2^2\right)$ and suppose X's and Y's are independent of one another. Set

$$S_1^2 = \sum_{i=1}^{n_1} \frac{\left(X_i - \overline{X}\right)^2}{n_1 - 1}, \quad \overline{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$$

$$S_2^2 = \sum_{i=1}^{n_2} \frac{\left(Y_i - \overline{Y}\right)^2}{n_2 - 1}, \quad \overline{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$$

Then

$$\frac{\sigma_2^2}{\sigma_1^2} \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$$

Proof. $\frac{\sigma_2^2}{\sigma_1^2} \frac{S_1^2}{S_2^2} = \frac{\left((n_1-1)S_1^2/\sigma_1^2\right)/(n_1-1)}{\left((n_2-1)S_2^2/\sigma_2^2\right)/(n_2-1)}$. Apply Theorem 2.4.2

$$\frac{\left((n_1 - 1)\, S_1^2/\sigma_1^2\right)/(n_1 - 1)}{\left((n_2 - 1)\, S_2^2/\sigma_2^2\right)/(n_2 - 1)} = \frac{\chi^2(n_1 - 1)/(n_1 - 1)}{\chi^2(n_2 - 1)/(n_2 - 1)}$$

Because $\left((n_1 - 1)\, S_1^2/\sigma_1^2\right)/(n_1 - 1)$ only depends on X's and $\left((n_2 - 1)\, S_2^2/\sigma_2^2\right)/(n_2 - 1)$ only depends on Y's, they are independent. Hence the result follows by the definition of F-distribution. □

2.7 Quantiles

Definition 2.7.1. **(Quantiles)**. Let X has CDF $F(x) = \mathbb{P}\{X \leq x\}$. If $F(x_\rho) = \rho$ for $0 < \rho < 1$, then x_ρ is the ρth quantile of X.

Example 2.7.1.. If $X \sim \mathcal{N}(\mu, \sigma^2)$ and given ρ, how to find the ρth quantile of X?
We can express the relationship between ρ and x_ρ as

$$\mathbb{P}\{X \leq x_\rho\} = \rho$$

which implies

$$\mathbb{P}\{X \leq x_\rho\} = \mathbb{P}\{\frac{X-\mu}{\sigma} \leq \frac{x_\rho - \mu}{\sigma}\} \stackrel{1}{=} \Phi(\frac{x_\rho - \mu}{\sigma}) = \rho$$

Take $z_\rho = \frac{x_\rho - \mu}{\sigma}$, then $\Phi(z_\rho) = \rho$ where the value of z_ρ can be find in the distribution table. Hence $x_\rho = \sigma z_\rho + \mu$.

Example 2.7.2.. If $X \sim \mathcal{GAM}(\theta, \kappa)$, and given ρ, how to find the ρth quantile of X?
We can express the relationship between ρ and x_ρ as

$$\mathbb{P}\{X \leq x_\rho\} = \rho$$

which implies

$$\mathbb{P}\{X \leq x_\rho\} = \mathbb{P}\{2X/\theta \leq 2x_\rho/\theta\} \stackrel{2}{=} F_{\chi^2(2\kappa)}(2x_\rho/\theta) = \rho$$

Take $\chi^2_\rho(2\kappa) = 2x_\rho/\theta$, then $F_{\chi^2(2\kappa)}(\chi^2_\rho(2\kappa)) = \rho$ where the value of $\chi^2_\rho(2\kappa)$ can be find in the distribution table. Hence $x_\rho = \theta\chi^2_\rho(2\kappa)/2$.

---

[1] Recall we usually use Φ to represent the CDF of standard normal distribution

[2] Recall Lemma 2.3.1 and the CDF of X is $F_X(x) = \mathbb{P}\{X \leq x\}$.