# Chapter 2. Sampling Distributions

BY YUEJIAN MO

May 26, 2018

## 1 Distribution of the Function of Random Varables

In statistics, a **sample distribution** or **finite-sample distribution** is the probability distribution of a given statistic based on a random sample. Sampling disribution provide a major simplification enroute to statistical inference.

Given a set of r.v.'s $X_1, \cdots, X_n$ with the cdf $F(x_1, \cdots, x_n)$ or the pdf $f(x_1, \cdots, x_n)$, we can seek the distribution of $Y = h(X_1, \cdots, X_n)$ for some functions $h(\cdot)$. We can use following techniques to solve it: cdf technique, transformation technique and moment generating function technique.

### 1.1 Cumulative distribution function technique

The distribution of $Y$ can be determined by the transformation $h(\cdot)$ together with the joint distribution of $X_1, \ldots, X_n$. If $X_1, \ldots, X_n$ are continuous r.v.'s, then the cdf of $Y$ can determined by integrating $f(x_1, \ldots, x_n)$ over the domain

$$\mathbb{D} = \{(x_1, \ldots, x_n) : h(x_1, \ldots, x_n) \leqslant y\};$$

that is

$$\begin{aligned} G(y) &= \Pr(Y \leqslant y) \\ &= \Pr(h(X_1, \ldots, X_n) \leqslant y) \\ &= \int_{\mathbb{D}} f(x_1, \ldots, x_n) dx_1 \cdots dx_n. \end{aligned}$$

Then by differentiating it with respect to $y$, we obtain the density of $Y$ as $g(y) = G'(y)$.

If $X$ is a discrete r.v. taking values $\{x_i\}$ with probabilities $\{p_i\}$, then the distribution of $Y = h(X)$ is determined directly the laws of probality. It may be that several values of $X$ give rise to the same value of $Y$. The probability that $Y$ takes on a given value, say $y_i$,is

$$\Pr(Y = y_j) = \sum_{\{i : h(x_i) = y_j\}} p_i$$

**Example 2.4** (Joint discrete distribution). Let $(X_1, X_2, X_3)$ have a joint discrete distribution given by

| $(X_1, X_2, X_3)$ | $(0,0,0)$ | $(0,0,1)$ | $(0,1,1)$ | $(1,0,1)$ | $(1,1,0)$ | $(1,1,1)$ |
|---|---|---|---|---|---|---|
| Probability | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ |

Find the pmf of $Y = h(X_1, X_2, X_3) = X_1 + X_2 + X_3$.

Solution. We note that $Y$ can take on values 0, 1, 2 and 3; then

$$\begin{aligned} \Pr(Y = 0) &= \frac{1}{8}, \\ \Pr(Y = 1) &= \frac{3}{8}, \\ \Pr(Y = 2) &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}, \\ \Pr(Y = 3) &= \frac{1}{8}. \end{aligned} \qquad \|$$

## 1.2 Transformation technique

### 1.2.1 Monotone transformation

Let $f(x)$ and $F(x)$ denote the corresponding pdf and cdf of an r.v. X. If $y = h(x)$ is a differentiable and monotone function and the inverse function is $x = h^{-1}(y)$, then the pdf of $Y = h(y)$ is given by

$$g(y) = f(x) \times |dx/dy| = f(h^{-1}(y)) \times \left|\frac{dh^{-1}(y)}{dy}\right|.$$

### 1.2.2 Piecewise monotone transformation(分段)

Let $\mathbb{A}_1, ..., \mathbb{A}_n$ be a partition of the real line $\mathbb{R} = (-\infty, +\infty)$, i.e., they are mutually exclusive and $\cup_{i=1}^n \mathbb{A}_i = \mathbb{R}$. If $y = h(x)$ is monotone on monotone on each $\mathbb{A}_i$, then $h_i(x) \hat{=} h(x)\boldsymbol{I}_{\mathbb{A}_i}(x)$ has a unique inverse $h_i^{-1}$ on $\mathbb{A}_i$, and the pdf of $Y$ is given by

$$g(y) = \sum_{i=1}^n f(h_i^{-1}(y)) \times \left|\frac{dh_i^{-1}(y)}{dy}\right|.$$

### 1.2.3 Bivariate transormation

Let $(X_1, X_2) \sim f(x_1, x_2)$ and the functions $y_i = h_i(x_1, x_2)$ for $i = 1, 2$ are differentiable and their inverse functions

$$x_i = h_i^{-1}(y_1, y_2) \, \text{for} \, i = 1, 2$$

exist. Then, the joint pdf of $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$ is

$$\begin{aligned} g(y_1, y_2) &= f(x_1, x_2) \times |J(x_1, x_2 \to y_1, y_2) \\ &= f(x_1, x_2) \times \left|\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right| \\ &= f(h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2)) \times |J(x_1, x_2 \to y_1, y_2)| \end{aligned}$$

where

$$J(x_1, x_2 \to y_1, y_2) = \left|\begin{array}{c} \partial(x_1, x_2) \\ \partial(y_1, y_2) \end{array}\right| = \det\begin{pmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{pmatrix}$$

denotes the Jacobian determinant of the transformation from $(x_1, x_2)$ to $(y_1, y_2)$.

### 1.2.4 Multivariate transformation

Let $(X_1, ..., X_n)^T \sim f(x_1, ..., x_n)$. If the functions $y_i = h_i(x_1, ..., x_n)$ for $i = 1, ..., n$ are differentiable, then the joint pdf of $Y_i = h_i(X_1, ..., X_n)$ for $i = 1, ..., n$ is given by

$$\begin{aligned} g(y_1, ..., y_n) &= f(x_1, ..., x_n) \times |J(x_1, ..., x_n \to y_1, ..., y_n)| \\ & \quad f(x_1, ..., x_n) \times \left|\frac{\partial(x_1, ..., x_n)}{\partial(y_1, ..., y_n)}\right| \end{aligned}$$

## 1.3 Moment generating function technique

Let $Y = \sum_{i=1}^n X_i$. If $\{X_i\}_{i=1}^n$ are independent r.v.'s, then

$$M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$$

We can find that binomial r.v., poisson r.v., chi-squared r.v.are additive.

$\{X_i\} \overset{\text{iid}}{\sim} N(0,1), Y = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^n X_i$

$\Rightarrow Y \sim \chi^2(n)$

$\Leftrightarrow Y \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$

Possion $\rightarrow$ Exponention $\longrightarrow$ Gamma

# 2 Statistics, Sample Mean and Sample Variance

Let $F(x)$ be the cdf of an r.v. $X$. If $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} F(x)$, then $\{X_i\}_{i=1}^n$ is said to be a *random sample* of $X$, or $\{X_i\}_{i=1}^n$ is a random sample from $F(x)$. An arbitrary function $T(X_1, ..., X_n)$ of $\{X_i\}_{i=1}^n$ is called a *statistic*.The **sampling distribution** of a statistic is the distribution of that staitstic, considered as a random variable.

We will study two statistic: sample mean and sample variance

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

## 2.1 Distribution of the sample mean

Let $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} F(x)$ with $E(X_1) = \mu$ and $\text{Var}(X_1) = \sigma^2$. For any $F(x)$, we have

$$E(\bar{X}) = \mu \qquad \text{and} \qquad \text{Var}(\bar{X}) = \sigma^2/n$$

If $F(x)$ is the cdf of the normal distribution $N(\mu, \sigma^2)$, then

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

## 2.2 Distribution of the sample variance

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

For any $F(x)$, the sample variance is an ubiased estimator of the variance, i.e.,

$$E(S^2) = \sigma^2$$

If $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, then

$$S^2 \perp \bar{X} \quad \text{and} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

# 3 The $t$ and $F$ Distribution

Let $Y \sim \chi^2(n)$, $Z \sim N(0,1)$, and $Y \perp Z$. The distribution of

$$T = \frac{Z}{\sqrt{Y/n}}.$$

is called the $t$-distribution with $n$ degrees of freedom and is written as $T \sim t(n)$.

**Therorem 2.2** (Density of  distribution). The density of $T \sim t(n)$ is given by

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, -\infty < t < \infty.$$

**Proof**:

$$F(x) = Pr(T \leqslant x) = Pr\left(\frac{Z}{\sqrt{Y/n}} \leqslant x\right) = \int Pr()$$

Use to estimate the mean of normal sample. For example, let $X_1, ..., X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, using sample mean and sample variance, we have

$$T = \frac{Z}{\sqrt{Y/n}} = \frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{\sqrt{\frac{(n-1)\sigma^2}{\sigma^2}/(n-1)}} = \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1)$$

Let $U \sim \chi^2(m)$,$V \sim \chi^2(n)$ and $U \perp V$. The distribution of the r.v.

$$W = \frac{U/m}{V/n}$$

is said to have an $F$ distribution with $m$ and $n$ degrees of freedom. We write $W \sim F(m,n)$.

Theorem 2.3 (Density of the $F$ distribution). The density of $W \sim F(m,n)$ is given by

$$f(w) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1}\left(1+\frac{m}{n}w\right)^{-\frac{m+n}{2}}, w > 0.$$

**Thero 2.4** (Ration of two normal sample variances). If $S_1^2$ and $S_2^2$ are the sample variance of independent random samples of size $n_1$ and $n_2$ from normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, then

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

If $X \sim F(m,n)$, then $Y = 1/X \sim F(n,m)$.

# 4 Order Statistics

Here are random sample: $X_1, ....X_n \sim^{\text{iid}} F(\cdot)$, and $f(\cdot)$ is the pdf. Let $X_{(1)} = \min\,(X_1, X_{2,\cdots,X_n})$ be the smallest of $X_1, X_2, \cdots, X_n$; $X_{(2)}$ be the second samllest of $X_1, X_2, \cdots, X_n$. $X_{(n)} = \max\,(X_1, \cdots, X_n)$ be the largest of all random sample. $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ are called the ***order statistics*** and $X_{(r)}$ is called the $r-$th *orther statistic* for $r = 1, \cdots, n$.

$x_{(1)}, x_{(2)}, \cdots x_{(n)}$ is the realization of $X_{(1)}, X_{(2)}, \cdots X_{(n)}$.

The cdf of the largest order statistic $X_{(n)}$ is

$$\begin{aligned} G_n(x) &= \Pr\{\max\,(X_1, ..., X_n) \leqslant x\} \\ &= \Pr(X_1 \leqslant x, X_2 \leqslant x, ..., X_n \leqslant x) \\ &= F^n(x)\ \text{（为了保证}X_{(n)}\text{是最大的）} \end{aligned}$$

The pdf of $X_{(n)}$ is

$$g_{(n)}(x) = n f(x) F^{n-1}(x-1)$$

Similarly, the smallest oder statistic(反其道而行之)

$$G_{(1)}(x) = \Pr(X_{(1)} \leqslant x) = 1 - \Pr(\min(X_1, \cdots\cdots X_n) > x) = 1 - \{1 - F(x\}^n$$

The pdf of $X_{(1)}$ is

$$g_{(1)}(x) = n f(x)\{1 - F(x)\}^{n-1}$$

The cdf of the $r$-th oder statistic

$$\begin{aligned} G_{(r)}(x) &= \sum_{i=r}^{n} \binom{n}{i} F(x)^i (1 - F(x))^{n-i} \\ &= \frac{1}{B(r, n-r+1)} \int_0^p t^{r-1} (1-t)^{n-r} dt \end{aligned}$$

the pdf of $X_{(r)}$ is

$$\begin{aligned} g_{(r)}(x) &= \frac{d}{dx} G_r(x) \\ &= \frac{1}{B(r, n-r+1)} \cdot \frac{d}{dx} \int_0^{F(x)} t^{r-1} (1-t)^{n-r} dt \\ &= \frac{n!}{(r-1)!(n-r)!} f(x) F^{r-1}(x) \{1 - F(x)\}^{n-r} \end{aligned}$$

The joint pdf of $X_{(1)}, \ldots, X_{(n)}$ is, for $x_1 \leqslant \cdots \leqslant x_n$,

$$g_{(1), \ldots (n)}(x_{(1)}, \ldots, x_{(n)}) = n! f_X(x_{(1)}) \cdots f_X(x_{(n)})$$

# 5 Limit Theorems

Definition (Convergence in distribution) Given a sequence of r.v.'s $\{X_n\}_{n=1}^{\infty}$. Let $F_n(x)$ be the cdf of $X_n$, if there exists an r.v. $X$ with cdf $F(x)$ such that

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for all points $x$ at which $F(x)$ is continuous, then we say that $\{X_n\}_{n=1}^{\infty}$ converges in *distribution* or *in law* to $X$ and write $X_n \xrightarrow{\mathrm{D}} X$ or $X_L \xrightarrow{L} X$.

Definition (weak convergence). A sequence of r.v.'s $\{X_n\}_{n=1}^{\infty}$ is said to *weakly* converge *in probability* to an r.v. $X$, denoted by $X_n \xrightarrow{P} X$, if for any $\varepsilon > 0$,

$$\lim_{n \to \infty} \Pr(|X_n - X| \geqslant \varepsilon) = 0.$$

Definiation (Strong convergence). $X_n \xrightarrow{a.s.} X$, if

$$\Pr\left( \lim_{n \to \infty} X_n = X \right) = 1$$

Defination (Convergence in mean square). $X_n \xrightarrow{m.s.} X$, if

$$\lim_{n \to \infty} E(X_n - X)^2 = 0$$

The realationship of four classes of convergency is

$$\begin{matrix} X_n \xrightarrow{a.s.} X \\ X_n \xrightarrow{m.s.} X \end{matrix} \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{L} X$$

### 5.1 Central Limit Therorem

**Therorem 2.9** (Central limit theorem). Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of i.i.d. random variables with common mean $\mu$ and common variance $\sigma^2 > 0$. Let $\bar{X} = \sum_{i=1}^{n} X_i / n$ and $Y_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$, then $Y_n \xrightarrow{L} Z$ as $n \longrightarrow \infty$, where $Z \sim N(0,1)$.

Proof:(To be continue)

$$M(t;n) = E(e^{\mathrm{tY}_n})$$

#### 5.1.1 Some Challenging Questions

## 6 Reference

Wikipedia, note from SUSTech